

–Supplementary material for the paper –

Bio-Object, a stochastic simulator for post-transcriptional regulation

Nobukazu Ohki¹, Masatoshi Hagiwara^{1,2,*}

¹Department of Functional Genomics, Medical Research Institute, ²Laboratory of Gene Expression, School of Biomedical Science, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113-0034. Japan

1. Definitions and simulation process in Bio-Object

1.1 Definition of a virtual cell

In this model, a square with a side 20 μm long on the two dimensional plane is provided for a virtual cell space. The length of the side is adopted because it is the average size of animal cells. A virtual nucleus space is placed inside of the cell space as a circle with a 12 μm diameter (the average size of animal nuclei). The region that is inside of the square but outside of the circle is defined as the virtual cytoplasm space. The simulation process takes place by using this spatial information. The defined spaces described above and the locations of molecules moving around in this virtual cell are represented by a (x,y) coordinate system. The origin of the coordinate is defined as the upper left corner of the square and the range of both (x,y) coordinates is limited within 0.0 - 20.0 μm .

The virtual cell space is divided into 10,000 square arrays in order to decide whether or not chemical reaction occurs in the arrays. (See Fig. 1A in the paper) Only molecules in the same *Cell* (0.2 μm by 0.2 μm) are close enough to react or interact with each other. Therefore, the simulation model decides whether or not a reaction takes place among molecules for every one of these *Cells*.

1.2 Definition of molecular characteristics

In this simulation algorithm, each molecule is represented as an Object class, which contains five kinds of information about the molecule:

- 1) A molecular identification code to distinguish each molecule in the simulation;
- 2) A molecular type code to show the type of molecule (e.g. PER protein, *per* mRNA, ribosome and *per* transcription site);
- 3) A pair of (x,y) coordinates to represent the location of molecule in the (x,y) coordinate system;
- 4) A localization code to restrict the movement of the molecule (e.g. whole cell, within the nucleus or within the cytoplasm);
- 5) A lapse to delete the molecule over its lifetime;

1.3 Translation site

There are four transcription sites, *per*, *tim*, *dClk* and *cyc*. Each one of them is defined as a fixed molecule object inside of the nucleus. The locations of (x, y) coordinates are subjectively defined as (8.0, 8.0) for *per*, (12.0, 8.0) for *tim*, (8.0, 12.0) for *dClk* and (12.0, 12.0) for *cyc* because there is no experimental data about the location of the transcription site. Each transcription site contains information about a transcription rate,

which is regulated by transcriptional regulators. The mRNAs are produced from these fixed transcription sites.

1.4 Intracellular events

For the treatment in this simulation, biological events are classified into four processes: transcription, translation, complex formation, and degradation.

1.5 Transcription

Two significant parameters are defined for the transcription process in this simulation algorithm. One is a probability of successful transcription and the other is a number of transcription attempts in one unit of time at one transcription site. The probability of successful transcription is defined to represent a stochastic gene expression faithfully and is subjectively set as a 0.01 molecule / transcription attempt in this model. Thus, if 1,000 transcription attempts / minute are introduced, the average transcription rate will be 10 molecules / minute as the product of these two parameters.

1.5.1 Transcription regulation

In order to emulate a stochastic gene regulation process, three parameters are defined in the regulation of the transcription process. They are (1) a probability of successful transcription regulation, (2) probability of successful transcription deregulation and (3) number of transcription attempts under the regulated state. The probability of successful transcription regulation reflects the interaction between a regulating factor and a regulated promoter. On the contrary, the probability of successful transcription deregulation provides information on the frequency of clearance of the regulating factor on the promoter. The number of transcription attempts is set as large as a normal state and as small as a normal state or zero under an activating regulator and a repressing regulator, respectively.

1.5.2 Transcription process

If the transcription process is successful, a novel mRNA is produced and starts to move around in the cell (the localization of all mRNAs is not limited within the cell). The type of mRNA depends on the transcription site where they are produced. (*per* mRNA, *tim* mRNA, *dClk* mRNA and *cyc* mRNA arise from *per* transcription site, *tim* transcription site, *dClk* transcription site and *cyc* transcription site, respectively.) The initial lifetime of a newly produced mRNA is defined randomly, using a random number in the normal distribution. (See the following section 1.9. *the initial lifetime setting*) The identification

numbers of mRNAs are obtained systematically so that they do not overlap.

1.6 Translation

The mRNA, which has moved into the cytoplasm and meets a ribosome in the same Cell, has the chance to produce new proteins (translation attempt). However, this opportunity is also under a stochastic control. The probability of successful translation is subjectively defined as 0.2 molecule / translation attempt. If translation succeeds, a set of numbers (defined as translation rates) for a new protein is created and the new protein starts to move around in the cytoplasm (the localization of PER proteins, TIM proteins, dCLK proteins and CYC proteins are limited within the cytoplasm). The type of protein is dependent on the mRNA from the protein that is produced. (PER proteins, TIM protein, dCLK protein and CYC protein arise from *per* mRNA, *tim* mRNA, *dClk* mRNA and *cyc* mRNA, respectively). The initial lifetime of a newly produced protein is defined randomly, using a random number in the normal distribution. The identification number of each protein is obtained systematically so that it does not overlap.

1.7 Complex formation

When the monomers of PER, TIM, dCLK and CYC proteins moving around in the cytoplasm meet their partner protein in the same Cell, complexes of hetero-dimer are able to form. This step advances stochastically because only the molecules that have gained activation energy can form a complex. This activation energy is different for each type of reaction, thus each type of reaction has its own reaction probability. Therefore, we define the parameters of reaction probability for each reaction. A reaction probability is also defined for PER-TIM-dCLK-CYC hetero-tetramer complex formation by a PER-TIM dimer complex and a dCLK-CYC dimer complex in the same way as the reaction of hetero dimer formation (see Table 2 in the paper).

The newly produced dimer and the tetramer complexes take over one of the reactant's (x-y) coordinates as the initial location. (e. g. a PER-TIM complex takes over the PER (x-y) coordinates, a dCLK-CYC complex takes over the dCLK (x-y) coordinates and a PER-TIM-dCLK-CYC complex takes over the PER-TIM (x-y) coordinates.) The moving area of these complexes is not limited, so they can locate in the nuclei to regulate a gene expression as a transcription regulator. The initial lifetime of a newly produced complex is redefined randomly, using a random number in the normal distribution. The identification numbers of the complex is obtained systematically so that they do not overlap.

1.8 Degradation

The lifetime of a molecule decreases by two minutes per each calculation step. Eventually, when the lifetime becomes zero, the molecule is degraded and deleted from the virtual cell space.

1.9 Initial lifetime setting

The random number that is used to set an initial lifetime of the molecule significantly affects the simulated time course of a number of molecules. As shown in Fig. 7A, the number of molecules decreases in a linear manner when the uniform random number (average: 1.00, variance: 0.34) is applied to set the initial lifetime. However, the actual time course of the degradation process of mRNAs (proteins) mediated by RNAase (protease) as enzymes shows a sigmoid curve. By using a random number (average: 1.00, variance: 0.04) in the normal distribution, the decrease pattern in the sigmoid curve is emulated (Fig. 7B). Thus, our model uses random numbers in the normal distribution to set an initial lifetime of molecules.

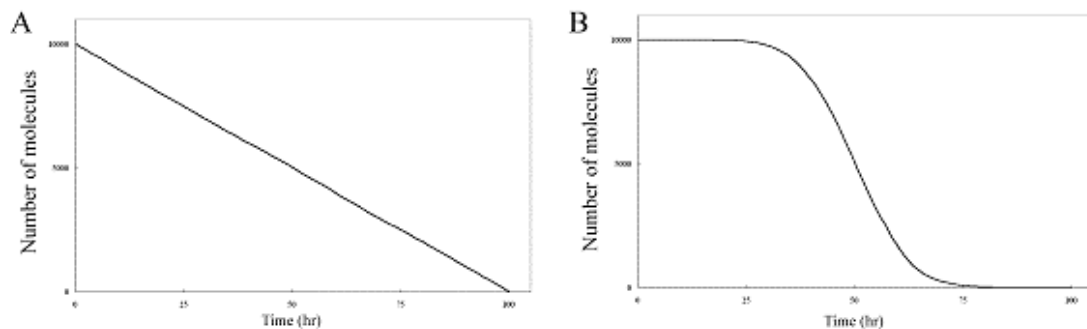


Fig. 7. Evaluation of our simulation model in degradation. To evaluate the adequacy of the model and algorithm, we applied the Bio-Object simulation method to simple models of degradation. The initial lifetime determines the degradation pattern by using a uniform random number (A) and a random number in the normal distribution (B).

1.10 Simple gene regulation model

We also constructed a simple gene expression regulation model to confirm whether or not the simulation process produces reasonable results for the transcription regulation. In Fig. 8A, the transcription activator is included in the system to regulate the expression of the gene, while in Fig. 8B, the transcription repressor is included. These transcriptional regulators are gradually degrading in the simulation. Concurrently, the activating transcription regulation level becomes low and eventually it completely stops the transcription. On the other hand, when the amount of the repressor becomes low, the transcription increases to the maximum level. This simulation results in

satisfactorily emulating the actual gene regulation that sequentially and gradually changes.

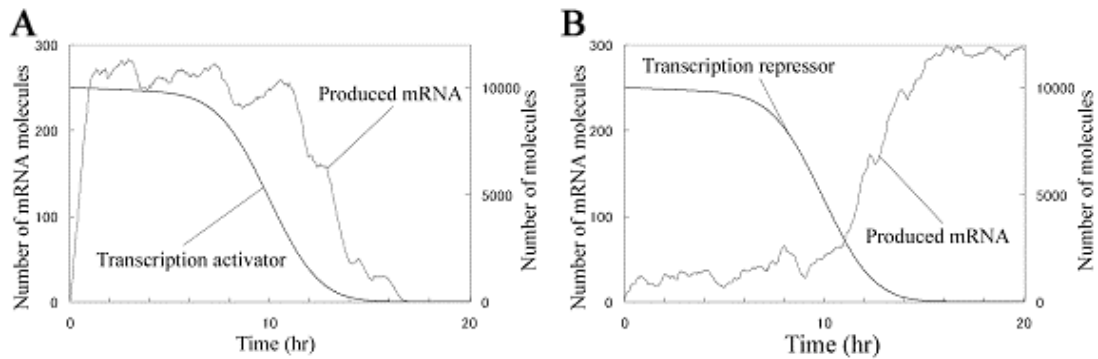


Fig. 8. Evaluation of our simulation model in gene regulations. Each plot shows temporal change in the number of various molecules of simple models of gene regulations for the transcriptional activator (A) and for the transcriptional repressor (B).

2. *in silico* Knockout study of *Drosophila* circadian rhythm

We performed virtual knockout studies of genes which regulate *Drosophila* circadian rhythm using Bio-object. The representative result is shown as Fig.9.

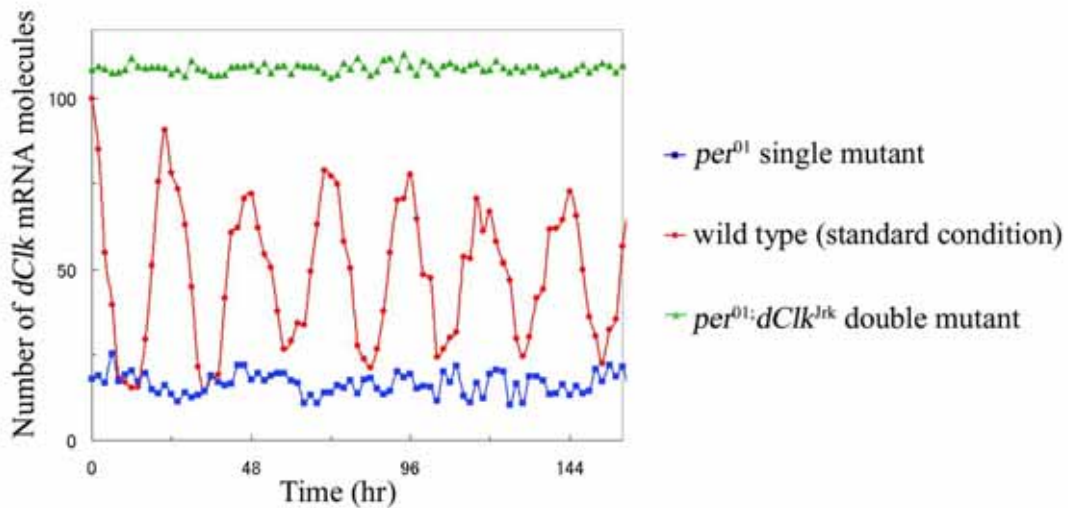


Fig. 9 Simulation of single and double mutants. Average relative levels of *dClk* mRNA at each time point (every 2 hours) which are obtained from 10 independent simulations are shown in *per*⁰¹ single mutant (blue), wild type (red) and *per*⁰¹;*dClk*^{Jrk} double mutant (green). The peak level of *dClk* mRNA in the wild type is set to 100%. In *per*⁰¹ mutant, sustaining oscillation is abolished and *dClk* mRNA level is decreased to 15% of the peak level in the wild type. In *per*⁰¹;*dClk*^{Jrk} mutant, sustaining oscillations are also abolished and the *dClk* mRNA level is restored to 115% of the peak level in the wild type. Parameter values are as in Table 1, expect for “translation rate of PER protein” = 0

molecule for per^{01} , “translation rate of PER protein” = “translation rate of dCLK protein” = 0 for $per^{01};dClk^{Jrk}$.

3. Changed half-lives of PER protein

We performed several theoretical experiments altering half-lives of molecules using Bio-Object. As a representative, we show the result of a theoretical experiment as for PER protein in Fig.10.

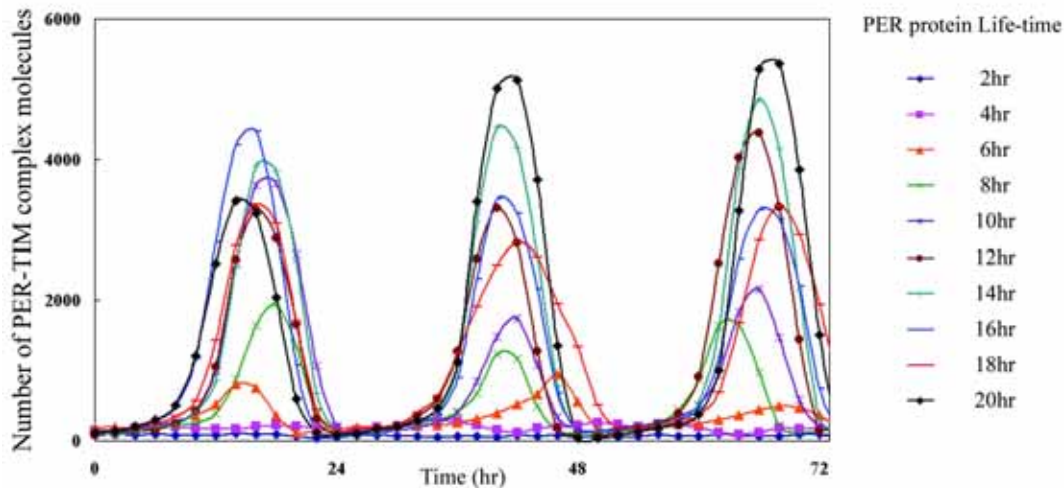


Fig. 10 Effects of changes in the parameter related to the PER protein life-time on circadian oscillations. The average life-time of PER protein changes from 2 hours to 20 hours with a 2-hour interval for an initial simulation condition. The time courses of the number of PER-TIM complex molecules at each time point (every 2 hours) are obtained from each simulation and plotted in the chart. The small number of simulations (only once for each condition) enhanced the fluctuation of the oscillation due to the stochastic property of the model. However, the results show the tendency that the stable PER protein which has a long life-time produces a stable oscillation with a large amplitude.

4. Summary of Simulation results

The duration and amplitude of circadian oscillations obtained from the simulation results shown in the Fig. 5 are summarized in Table 3. See the discussion in section 4 of the paper for further explanations.

Table 3. A summary of simulation results

Changed parameter	value	Oscillation	Duration	Amplitude
(standard condition)	-	stable oscillation	24.0 hr	286%
Average half-life of per mRNA	1 hr	completely dumped	-	-
Average half-life of per mRNA	6 hr	stable oscillation	23.4 hr	383%
Average half-life of dCLK mRNA	1 hr	stable oscillation	23.4 hr	459%
Average half-life of dCLK mRNA	6 hr	stable oscillation	30.0 hr	201%
Translation rate of PER protein	4 molecules/min.	stable oscillation	24.8 hr	284%
Translation rate of PER protein	1 molecule/min.	completely dumped	-	-
Translation rate of dCLK protein	4 molecules/min.	attenuated and damped	26.0 hr	205%
Translation rate of dCLK protein	1 molecule/min.	stable oscillation	23.2 hr	383%
Average half-life of PER	5 hr	completely dumped	-	-
Average half-life of PER	20 hr	stable oscillation	24.3 hr	372%
Average half-life of dCLK	5 hr	attenuated and damped	20.4 hr	249%
Average half-life of dCLK	20 hr	stable oscillation	32.4 hr	192%
Possibility of PER-TIM complex formation	0.005 molecules / attempt	stable oscillation	22.5 hr	369%
Possibility of dCLK-CYC complex formation	0.005 molecules / attempt	completely dumped	-	-
Possibility of PER-TIM-dCLK-CYC complex formation	0.1 molecules / attempt	completely dumped	-	-