Gene Circuit Model

Hyeon-Jae Seo and Erin Kim

June 2017

1 Overview

This week, I focused on finishing the gene circuit model for curli production. To do so, I will continue from where I left off in last week's writeup.

2 Mass Action Kinetics

Mass Action Kinetics state that the rate of a reaction is the product of a rate constant (k) and the mass of the substrate (S). Several assumptions will be made in the following model:

1. *csgC*, *csgG*, *csgE*, and *csgF* will be treated as a combined *csgCGEF* complex in a 1:1:1:1 ratio, as the reported roles of *csgC*, *CsgE*, and *csgF* are to assist *csgG* in the successful export of *csgA*, and no additional information can be gleaned by treating each as a separate species. Thus, the two main species of this model are the *csgCGEF* complex and *csgA*.
2. Transcription factor binding achieves equilibrium much faster than transcription, translation, and protein accumulation, so it can be considered to be at steady state on the time scale of proteins.
3. Spatial parameters, heat and diffusion gradients, etc. will not be considered to avoid partial differential equations.

The focus of this model will be on the transcription, translation, and secretion aspects of the curli pathway, since the accumulation of *csgA* in the extracellular space is the point of interest.

First, we will detail the relevant reactions and rate constants in each module of the pathway, then present the differential equations that represent the system.

3 Transcription

$$\mathrm{RNA_{pol}} + \mathrm{g_{csg}} \xrightarrow{\mathrm{k}_t} \mathrm{RNA_{pol}} + \mathrm{g_{csg}} + \mathrm{mRNA_{csg}}$$

3.1 Gene Regulation

The rate of transcription, k_t, is mainly governed by the promoter controlling the operon. Since all the relevant proteins in this model are encoded on the same operon, the transcript for all the proteins are generated at the same rate.

The rate of transcription, k_t, can be regulated by either activators or repressors. Thus, we must modify our differential equations to take this into account. We will define rate of transcription, α, to be

$$\alpha = Bf(x)$$

where B is the basal, or maximum, expression rate when the gene is "on", and $f(x)$ represents the probability of expression, a function of the concentration of x. This function $f(x)$ can be described by the Hill function model:

$$f(x)_{activation} = \frac{x^n}{x^n + K^n}$$

$$f(x)_{repression} = \frac{1}{1 + (\frac{x}{K})^n}$$

Here, x is the level of activator or repressor, n is the Hill coefficient, and the K is the value of x where the probability $f(x) = 0.5$. The Hill coefficient is context dependent and represents the cooperativity of the transcription factor. In other words, it describes the switch-like behavior of the transcription factor, increasing the nonlinearity of the function. The parameter K defines the functional concentration range of x.

3.2 Transcript Stability

$$\mathrm{mRNA_{csg}} \xrightarrow{\mathrm{k_{-t}}} \varnothing$$

The degradation rate of the mRNA transcript can be described as an exponential decay function of time:

$$k_{-t} = A_0 e^{rt}$$

Need to take barcodes and switches into account.

4 Translation

Since the genes coding for each *csg* protein is controlled by a different RBS and are translated independently of each other, each segment of the mRNA transcript will be treated as a separate gene

$$\mathrm{mRNA_A} \xrightarrow{\mathrm{k_A}} \mathrm{mRNA_A} + \mathrm{csgA}$$

$$\mathrm{mRNA_C} \xrightarrow{\mathrm{k_C}} \mathrm{mRNA_C} + \mathrm{csgC}$$

$$\mathrm{mRNA_E} \xrightarrow{\mathrm{k_E}} \mathrm{mRNA_E} + \mathrm{csgE}$$

$$\mathrm{mRNA_F} \xrightarrow{\mathrm{k_F}} \mathrm{mRNA_F} + \mathrm{csgF}$$

$$\mathrm{mRNA_G} \xrightarrow{\mathrm{k_G}} \mathrm{mRNA_G} + \mathrm{csgG}$$

4.1 RBS Effects

4.2 Transcript Polarity

4.3 Protein Stability

Degradation

$$\mathrm{CsgA} \xrightarrow{\mathrm{k_{-A}}} \varnothing$$

$$\mathrm{CsgE} \xrightarrow{\mathrm{k_{-E}}} \varnothing$$

$$\mathrm{CsgG} \xrightarrow{\mathrm{k_{-G}}} \varnothing$$

The same exponential decay function used to describe the rate of mRNA degradation earlier can be used to represent the rate of protein degradation, adjusting the rate of decay to account for the appropriate protein half-life.

$$k_{-t} = A_0 e^{rt}$$

5 Protein Secretion

Due to *e. coli*'s double membrane, all the *csg* proteins must first pass through the inner membrane through the SecYEG protein complex via the Type II secretion system, followed by the secretion of *csgA* through *csgG* to reach the cell's extracellular space. This is described in the following reactions:

$$\mathrm{csgA} \xrightarrow{\mathrm{k_{IM}}} \mathrm{csgA_{per}}$$

$$\mathrm{csgC} \xrightarrow{\mathrm{k_{IM}}} \mathrm{csgC_{per}}$$

$$\text{csgE} \xrightarrow{\text{k}_{\text{IM}}} \text{csgE}_{\text{per}}$$

$$\text{csgF} \xrightarrow{\text{k}_{\text{IM}}} \text{csgF}_{\text{per}}$$

$$\text{csgG} \xrightarrow{\text{k}_{\text{IM}}} \text{csgG}_{\text{per}}$$

$$\text{csgA}_{\text{per}} \xrightarrow{\text{k}_{\text{OM}}} \text{csgA}_{\text{sec}}$$

6 Differential Equations

From these, we derive a system of differential equations to describe the rate of change of mRNA and protein levels over time:

mRNA

$$\frac{d[mRNA_{csg}]}{dt} = k_t[g_{csg}] - k_{-t}[mRNA_{csg}]$$

Proteins in Cytoplasm

$$\frac{d[csgA]}{dt} = k_A[mRNA_{csg}] - k_{-A}[csgA] - k_{IM}[csgA]$$

$$\frac{d[csgE]}{dt} = k_E[mRNA_{csg}] - k_{-E}[csgE] - k_{IM}[csgE]$$

$$\frac{d[csgG]}{dt} = k_G[mRNA_{DEFG}] - k_{-G}[csgG] - k_{IM}[csgG]$$

Proteins in the Periplasm

$$\frac{d[csgA_{per}]}{dt} = k_{IM}[csgA] - k_{-A}[csgC_{per}] - kOM[csgA_{per}]$$

$$\frac{d[csgC_{per}]}{dt} = k_{IM}[csgC] - k_{-C}[csgC_{per}]$$

$$\frac{d[csgE_{per}]}{dt} = k_{IM}[csgE] - k_{-E}[csgE_{per}]$$

$$\frac{d[csgF_{per}]}{dt} = k_{IM}[csgF] - k_{-F}[csgF_{per}]$$

$$\frac{d[csgG_{per}]}{dt} = k_{IM}[csgG] - k_{-G}[csgG_{per}]$$

Secreted CsgA

$$\frac{d[CsgA_{OM}]}{dt} = k_OM[csgA_{per}] - k_{-A}[csgA]$$

7 Optimizing Stoichiometric Ratios

The optimal ratio of $csgA$, $csgC$, $csgE$, $csgF$, and $csgG$ is $k_{OM} : 1 : 1 : 1 : 1$.

8 Fiber Formation

The main reactions involved in the fibril formation is nucleation ($k_{n,i}$ and $k_{n,-i}$) and elongation of fibrils ($k_{fb,i}$ and $k_{fb,-i}$). These two steps proceed until the species involved reach an equilibrium. These are described by the following reactions:

$$\mathrm{CsgA}_1 + \mathrm{CsgA}_1 \underset{\mathrm{k}_{n,\ -1}}{\overset{\mathrm{k}_{n,\ 1}}{\rightleftharpoons}} \mathrm{CsgA}_2$$

$$\mathrm{CsgA}_n + \mathrm{CsgA}_1 \underset{\mathrm{k}_{n,\ -i}}{\overset{\mathrm{k}_{n,\ i}}{\rightleftharpoons}} \mathrm{CsgA}_{n+1}$$

$$\mathrm{F} + \mathrm{Csg}_n \underset{\mathrm{k_{fb,-i}}}{\overset{\mathrm{k_{fb,i}}}{\rightleftharpoons}} \mathrm{F}$$

$CsgA_1$ and $CsgA_n$ represent a monomer and an oligomer containing n monomers, respectively. Although fibril entanglement has been reported, the ends of the fibers are assumed to be the only active chemical sites, and thus fibers of all lengths are treated equally.

Nucleation is considered the rate limiting step in amyloid formation (Chapman 2008)

$$\frac{dCsgA_1}{dt} = k_{sec}[CsgA_{per}] - k_d[CsgA_n]$$

$$\frac{dCsgA_n}{dt} = k_{fb,i}[CsgA_{n-1}][CsgA_1] - k_{fb,-i}[CsgA_n]$$

$$\frac{dF}{dt} = k_n[CsgA_1]^2 + k_{fb,1}[F][CsgA_1] - k_{fb,-1}[F]$$

9 References

Csicsery, Nick and O'Laughlin, Ricky. (2013). *A Mathematical Model of a Synthetically Constructed Genetic Toggle Switch.*