Gene Circuit Model

Hyeon-Jae Seo and Erin Kim

June 2017

1 Overview

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. *csgG* and *csgE* will be treated as a combined *csgGE* complex in a 1:1 ratio, as the reported role of *CsgE* is 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 *csgGE* 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.

2 Transcription

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

2.1 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.

2.2 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.

3 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.

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

$$\mathrm{mRNA_{csg}} \xrightarrow{\mathrm{k_C}} \mathrm{mRNA_{csg}} + \mathrm{csgC}$$

$$\mathrm{mRNA_{csg}} \xrightarrow{\mathrm{k_E}} \mathrm{mRNA_{csg}} + \mathrm{csgE}$$

$$\mathrm{mRNA_{csg}} \xrightarrow{\mathrm{k_G}} \mathrm{mRNA_{csg}} + \mathrm{csgG}$$

However, since we will assume the optimal ratio of $csgG : csgE$ is $1 : 1$, we will henceforth refer to a combined $csgGE$ complex in place of the two individual proteins for the purposes of the model:

$$\text{mRNA}_{\text{csg}} \xrightarrow{\text{k}_{\text{A}}} \text{mRNA}_{\text{csg}} + \text{csgA}$$

$$\text{mRNA}_{\text{csg}} \xrightarrow{\text{k}_{\text{C}}} \text{mRNA}_{\text{csg}} + \text{csgC}$$

$$\text{mRNA}_{\text{csg}} \xrightarrow{\text{k}_{\text{EG}}} \text{mRNA}_{\text{csg}} + \text{csgEG}$$

3.1 RBS Effects

We will run a simulation with various RBS strengths.

3.2 Transcript Polarity

TBD

3.3 Protein Stability

Degradation

$$\text{CsgA} \xrightarrow{\text{k}_{-\text{A}}} \varnothing$$

$$\text{CsgC} \xrightarrow{\text{k}_{-\text{C}}} \varnothing$$

$$\text{CsgEG} \xrightarrow{\text{k}_{-\text{EG}}} \varnothing$$

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.

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

4 Protein Secretion

Due to *e. coli*'s double membrane, all the *csg* proteins must first pass through through the SecYEG protein complex in the inner membrane via the Type II secretion system into the periplasm, followed

by the secretion of $csgA$ through $csgG$ to reach the cell's extracellular space. This is described in the following reactions:

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

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

$$\mathrm{csgEG} \xrightarrow{\mathrm{k_{IM}}} \mathrm{csgEG_{per}}$$

$$\mathrm{csgA_{per}} \xrightarrow{\mathrm{k_{OM}}} \mathrm{csgA_1}$$

5 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:

$$\mathrm{CsgA_1 + CsgA_1} \underset{\mathrm{k_{nu,-1}}}{\overset{\mathrm{k_{nu,1}}}{\rightleftharpoons}} \mathrm{CsgA_2}$$

$$\mathrm{CsgA}_i + \mathrm{CsgA_1} \underset{\mathrm{k_{nu,-i}}}{\overset{\mathrm{k_{nu,i}}}{\rightleftharpoons}} \mathrm{CsgA}_{i+1}$$

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

$CsgA_1$ 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)

Ignore these next three equations, it needs to be looked into further.

$$\frac{dCsgA_1}{dt} = k_{OM}[CsgA_{per}] - k_{nu,-1}[CsgA_n]$$

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

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

6 Differential Equations

6.1 mRNA

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

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

6.2 Proteins in Cytoplasm

Next, we look at the rate at which protein is accumulated in the cytoplasm:

$$\frac{d[csgA]}{dt} = k_A[mRNA_{csg}] - k_{-A}[csgA] - k_{IM}[csgA] - k_{nu,1}[csgA]^2 + k_{nu,-1}[csgA_2]$$

$$\frac{d[csgA_2]}{dt} = k_{nu,1}[csgA]^2 - k_{nu,-1}[csgA_2]$$

$$\frac{d[csgC]}{dt} = k_C[mRNA_{csg}] - k_{-C}[csgC] - k_{IM}[csgC]$$

$$\frac{d[csgEG]}{dt} = k_{EG}[mRNA_{csg}] - k_{-EG}[csgEG] - k_{IM}[csgEG]$$

6.3 Proteins in the Periplasm

Once the proteins are produced, they are secreted into the periplasm via the Type II secretion system through the SecYEG protein complex.

$$\frac{d[csgA_{per}]}{dt} = k_{IM}[csgA] - k_{-A}[csgA_{per}] - k_{OM}[csgA_{per}] - k_{n,1}[csgA_{per}]^2 + k_{n,-1}[csgA_{per,2}]$$

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

$$\frac{d[csgEG_{per}]}{dt} = k_{IM}[csgEG] - k_{-EG}[csgEG_{per}]$$

6.4 Secreted *csgA*

Finally, we look at the rate at which csgA is exported from the periplasm into the extracellular matrix through csgG.

$$\frac{d[csgA_1]}{dt} = k_{OM}[csgA_{per}] - k_{-A}[csgA] - k_{nu,i}[csgA_i][csgA_1] + k_{nu,-i}[csgA_i]$$

7 Optimizing Stoichiometric Ratios

Solutions to the system of differential equations above are functions of the rates of translation of the proteins. These rates are in turn dependent upon the relative strengths of the RBS's associated with each protein. Of particular importance is the ratio of strengths between the RBS's associated with *csgA* and *csgG*. By finding the ratio that maximizes $\frac{d[F]}{dt}$, we can determine the optimal relative RBS strengths to assign to each respective protein.

8 Symbols

8.1 Species Symbols

Symbol	Definition
g_{csg}	Operon encoding csg proteins
$mRNA_{csg}$	Transcript for csg proteins
$csgEG$	"Black box" referring to the csgE+csgG complex
$csgA - G$	csg proteins in the cytoplasm
$csgA - G_{per}$	csg proteins in the periplasm
$csgA_1$	csgA monomer in the extracellular matrix
$csgA_i$	csgA oligomer of length i
F	Fibril

8.2 Rate Constants

Symbol	Definition	Value	Units
k_t	Transcription Rate	TBD	TBD
k_{-t}	Transcript Degradation Rate	TBD	TBD
k_A	csgA translation rate	TBD	TBD
k_{-A}	csgA degradation rate	TBD	TBD
k_C	csgC translation rate	TBD	TBD
k_{-C}	csgC degradation rate	TBD	TBD
k_E	csgE translation rate	TBD	TBD
k_{-E}	csgE degradation rate	TBD	TBD
k_G	csgG translation rate	TBD	TBD
k_{-G}	csgG degradation rate	TBD	TBD
k_{EG}	csgEG translation Rate	TBD	TBD
k_{-EG}	csgEG degradation rate	TBD	TBD
k_{IM}	Rate of secretion through SecYEG	TBD	TBD
$k_{nu,1}$	Rate of nucleation b/w 2 csgA monomers	TBD	TBD
$k_{nu,-1}$	Rate of dissociation of csgA dimer	TBD	TBD
$k_{nu,i}$	Rate of addition of csgA monomer to oligomer of length i	TBD	TBD
$k_{nu,-i}$	Rate of dissociation of csgA monomer from oligomer of length i	TBD	TBD
$k_{fb,i}$	Rate of elongation of a fiber	TBD	TBD
$k_{fb,-i}$	Rate of dissociation of a monomer from a fiber	TBD	TBD

9 References

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

Selinger DW, Saxena RM, Cheung KJ, Church GM, Rosenow C. Global RNA Half-Life Analysis in Escherichia coli Reveals Positional Patterns of Transcript Degradation. Genome Research. 2003;13(2):216-223. doi:10.1101/gr.912603.

10 Updated Basic Framework

The basic idea is that the rate at which A is made should be equal to the rate at which A is exported from the cell.

The rate at which A is made is:

[((mRNA - $mRNA_{1/2}$)(translation rate) - $A_{1/2}$)][rate of sec complex secretion] = concentration of A in the periplasm.

The rate of A polymerization and depolymerization was not included in the model because the enucleation of A (especially without the facilitation of B and F) is a very slow rate. Therefore we make the assumption that there will be relatively little enucleation in the cell itself, especially if the cell is optimally exporting A.

The rate of exportation by GE should also match the rate of exportation activity by GE:

[((mRNA - $mRNA_{1/2}$)(translation rate) - $GE_{1/2}$)][rate of sec complex secretion][GE activity constant] = secretion activity of GE in the periplasm.

The GE activity constant is how many "things" is exported given a unit of time, specifically A.

Now:

A_{made} - $A_{GEsecreted}$ = 0

is the optimized model.