Biochemical and phenomenological models completed for the shinethanogen system

October 28, 2020

Contents

1 Introduction and design

Without access to a lab, this section of *in silico* molecular modelling effectively served as our experimental workbench, finetuning our gene circuit and testing our approach to experimental design.

1.1 Gene circuit

The overall gene circuit comprises two main parts: a shinorine-producing part called the shinogen – 'shinorine-producing' – and a killswitch called the 'thanogen'

– the 'death-producer', preventing the escape of the bacterium outside the environment of the skin. Together, they make the Shinescreen system. Hereafter, we name our genetically modified EcN *ShinE. coli*.

1.1.1 The thanogen circuit

We learnt from our discussion with Prof Nicola Allison, biogeochemist and coral expert the importance of a quarantining mechanism to prevent release of our micro-organism. Consequently, through many trials, we perfected a theoretical killswitch which accomplished both long-term bistability - the ability of the system to exist in two separate states, permissive and non-permissive - and one which minimised our biosafety concerns of HGT and nonfunctionalisation of the gene circuit caused by selection and drift.

Properties of the thanogen

- The thanogen is divided across two plasmids, severally carrying the first and last two shinezymes. Both plasmids contain toxins and antitoxin systems which neutralise the contralateral toxin i.e. the toxin on the other plasmid.
- Both plasmids must be present for the cell to enter a potentially permissive state. If either plasmid in the cassette were absent, the cell cannot be rescued from stasis or death.
- The thanogen incorporates two environmental inputs with different weighting: glucose concentration and light intensity. A steady-state permissive condition requires both glucose and light intensity of a specific wavelength to be arbitrarily high ($2200umol/m^2/s$). Glucose concentration primarily controls the fate of the cell and thus, in its absence, accelerates non-permissive conditions. Conversely, light intensity is an *attenuator*, maintaining viability when exposed to UV while prolonging cell life for a period after the signal is absent.
- The thanogen integrates both transcriptional control of gene expression and post-translational degradation to control the output of the system. This allows for three-dimensional regulation of the system at multiple check-points while forcing the cell into one distinct state (see engineering for more information about our previous killswitch designs).
- The UV-response element itself is an elegant ccaR-ccaS module extracted from Synechocystis (fig 1). Functioning as a green-light* receptor kinase, ccaS is a transmembrane protein which mediates the signal transduction of a green light signal into a phosphorylation event of ccaR, the secondary messenger protein. Once phosphorylated, ccaR dimerises and serves as a transcriptional activator, binding to a conserved operator sequence, csiR1. In the endogenous species, this cascade would upregulate expression of flagellar proteins needed for negative phototaxy. In our design, however, the ccaR/s module shall serve as an input to the killswitch by binding

upstream at the promoter of an antidote which rescues the cell from toxin-mediated death.

*Green light is used as a proxy for UV light after determining that the paralogous UirR/S is strongly inhibited in the presence of red light
Toxin-antitoxin modules are prevalent in most families of bacteria. [] to confer an adaptive advantage under environmentally demanding conditions. These TA systems comprise an antitoxin, either constitutively expressed or at a higher level than the toxin, which is stabler than its antitoxin. When antitoxin production is stopped due to loss of the right environmental input signal or plasmid segregation the antitoxin is degraded rapidly, ceasing to counteract the toxin. We chose ccdA-ccdB (controls cell death) system because it is endogenous to our chassis E. coli and the extreme versatility of ccdAB binding and its operon.

1.1.2 Steps in killswitch activation

- This bifunctional design using a type II toxin-antitoxin system relies therefore on three different parts across the two plasmids - the environmental input, an adaptor and an output. The environmental inputs, glucose and UV, respectively control ccdB and ccdA production. Depending on the proportion of expression of ccdB and ccdA, the adaptor mf-lon protease, a protein derived from *Mesoplasma florum*, is up- or down-regulated.
- When [ccdA] >> [ccdB] or [ccdA] = [ccdB], ccdA and ccdAB binding of the protease promoter is energetically favourable, thereby concealing the Pribnow sequence and repressing transcription. However, when [ccdB] ¿ [ccdA] because glucose levels or UV intensity are low, ccdB binds an allosteric site on the ccdA2 portion of ccdAB, fully de-repressing the promoter and exposing the promoter for RNA polymerase. Why not select the native *E. coli* lon protease which targets a conserved oligopeptide ssrA tag? We wanted to avoid background noise caused by lon protease already within the cell which could interfere with the switch-like function of the circuit or other essential functions within the cell. mf-lon protease recognises a different oligopeptide sequence, AANDENYALAA, and therefore does not overlap with any cellular targets.
- In turn, ccdA is tagged with this degradation tag at the N-terminal domain which ensures negative feedback once the protease is expressed to a critical concentration. Including this regulatory step, it minimises the chance of stochastic repression and de-repression of the protease leading to unpredictable life/death. Downstream of the protease, lacI-ssrA is constitutively expressed under P_{const} and binds a synthetic lacP' operon lacking the CAP binding site. It is the adaptor's function to transduce the environmental signal into degradation of lacI.

Glucose	UV	Adaptor	Repressor	Output toxin
1	1	0	1	0
1	0	$1 > x > 0$	$1 > x > 0$	> 0
0	1	1	0	1
0	0	1	0	1

Under high glucose, i.e. in the glucose-rich sunscreen medium, CAP-cAMP activity is dramatically lowered.

Nevertheless, an experimental and modelling challenge shall be the fine-tuning of this system for sufficient Shinorine production and controlling the effective range of UV intensities ($\mu M/l/s$) over which the gene circuit can protect. Furthermore, the *A. variabilis* shinogenic cluster contains four prerequisite enzymes which direct the flux of Shinorine from sedoheptulose-7-phosphate, an intermediate of the pentose phosphate pathway and precursor in glyceraldehyde-3-phosphate synthesis for glycolysis. The gene cluster and biochemical pathway are illustrated below in fig 3. Of these, prior modelling and literature indicate that the lattermost enzyme is the major rate-limiting enzyme or adopts the greatest control over total flux. Consequently, our experimental plan discusses approaches to its upregulation, notwithstanding altering the promoter stoichiometry and Shine-Dalgarno (ribosome-binding) site strength (fig 4).

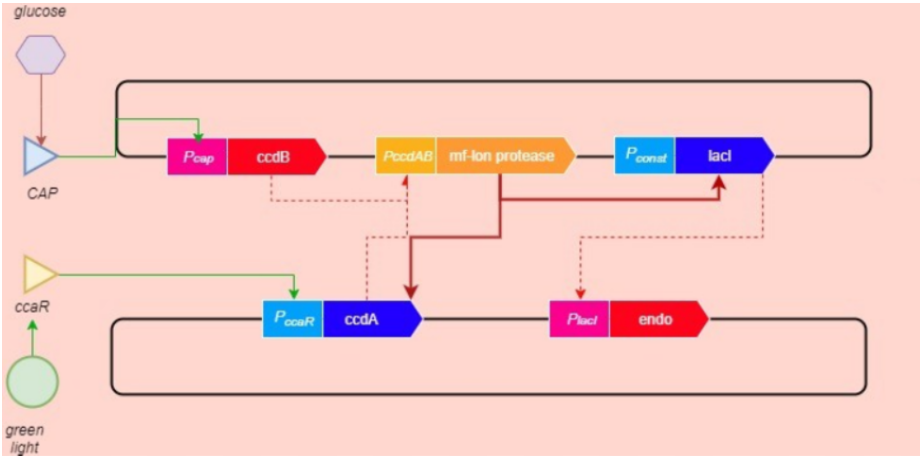


Figure 1: Thanogen gene circuit, showing the two plasmid back-bones carrying contralateral TA modules and the protease integrator, ccdA and endonuclease; ccdB and lacI

The thanogen circuit

Therefore, we sought to examine the behaviour of the two branches of our gene circuit: the shinogen, for *shinorine* production and the thanogen, our $AND_{Glucose,UV}$-gated killswitch. While our molecular models cannot precisely quantify expected concentrations of target compounds including shinorine, we have by exploring parameter space, estimated the predicted behaviour of the systems.

2 Thanogen

The thanogen (etyma from Greek jointly meaning 'death producer') is a carefully devised construct whose function responds to the biosafety hazards discussed in the *Experimental design section*. Essentially, in the absence of permissive conditions, the default state of the engineered bacterium should be bacteriostasis - prolonged non-growth - and death. Furthermore, due to its theoretical proximity to native flora of the microbiome, the killswitch circuit was designed to incorporate genes expressing toxins endogenous to *E. coli* alone.

In this section, we present three parallel models: a Boolean, deterministic and stochastic model. Further, we detail the biochemical model of our precedent circuit involving tlpA.

Modelling characteristics of the thanogen:

- The killswitch takes two environmental inputs, glucose concentration and UV intensity and gives three conditional outputs (fig 2).
- The output dependence is weighted towards glucose insofar as short-term absence of UV results in bacteriostasis (and in the long term, death) while in glucose, short-term cell death.
- In permissive conditions, the viability of *E. coli* is unaffected.

Assumptions

- The concentration of RNA polymerase (RNAp) is saturating and thus can be considered 0 order with respect to the reaction kinetics

- Transcription and translation rates are directly proportional to promoter and Shine-Delgarno binding strength; minimal 1D scanning occurs (This may be problematic for genes integrated into the genome due to non-specific interactions and scanning longer stretches of DNA)

Rate equations

Reaction kinetics with transcription factors, L, LacI repressor and g a proxy for CAP. [g] and CAP activation are inversely proportional; thus, when [g] is high, the level of activated CAP is low and gene expression is leaky. The principal advantage of this approach overcomes the spatial dependence of adenylate cyclase responsible for cAMP synthesis under low glucose concentration. Further, evidence from the lac operon suggest that the activation by CAP at the operator region obeys a switch-like behaviour in the presence/absence of glucose by binding upstream of the -35 box and enhancing RNAP CTD binding. Nevertheless, a consequence may be a quicker response time to glucose than *in vivo*.

Figure 2 shows the plots of the key molecular species controlling cell life/death behaviour. A2B2 is the 1:1 complex formed between ccdA and ccdBl, A2B4, 1:2 complex with ccdB in excess of ccdA. This complex emerges when relative expression of ccdB outweighs ccdA and can be reversed by increased ccdA synthesis which preferentially binds one dimer of ccdB. Importantly, A2B2 is a repressor of the mf-lon protease integrator. When A2B4 ¿ A2B2, derepression of the protease occurs, initiating a cascade of ccdA and lacI degradation. Without lacI, the endonuclease is over-expressed beyond leaky level, committing the cell to death (fig 3).

Transcriptional rates (1)

$$\mathrm{C} \longrightarrow \mathrm{C} + \mathrm{B}_2 \tag{2}$$

$$\mathrm{U} \longrightarrow \mathrm{U} + \mathrm{A}_2 \tag{3}$$

$$\mathrm{L} \longrightarrow \mathrm{L} + \mathrm{E} \tag{4}$$

$$\mathrm{A_2B_2} \longrightarrow \mathrm{A_2B_2} + \mathrm{P} \tag{5}$$

$$\mathrm{A}_2 \longrightarrow \mathrm{A}_2 + \mathrm{P} \tag{6}$$

$$\mathrm{A}_2 + \mathrm{B}_2 \rightleftharpoons \mathrm{A_2B_2} \tag{7}$$

$$\mathrm{A_2B_2} + \mathrm{B}_2 \rightleftharpoons \mathrm{A_2B_4} \tag{8}$$

$$\mathrm{nL} + \mathrm{P_1}' \xrightarrow{\mathrm{k}_a} \mathrm{P} \tag{9}$$

$$\mathrm{P_1}' \xrightarrow{\mathrm{k}_b} \mathrm{P_1}' + \mathrm{B}_2 + \mathrm{L} \tag{10}$$

$$\mathrm{P_2P_2} + \mathrm{g} \xrightarrow{\mathrm{k}_c} \mathrm{P}_2 \tag{11}$$

Leaky production when the promoter is inactive

$$\cdot \xrightarrow{\mathrm{k}_l \mathrm{k\text{-}P_1'}} \mathrm{B}_2 + \mathrm{L} \tag{12}$$

$$\cdot \longrightarrow \mathrm{A}_2 \tag{13}$$

$$\mathrm{P}_2 \xrightarrow{\mathrm{k}_e} \mathrm{A}_2 \tag{14}$$

$$\mathrm{A}_2 \longrightarrow \cdot \tag{15}$$

$$\mathrm{B}_2 \longrightarrow \cdot \tag{16}$$

$$\mathrm{L} \longrightarrow \cdot \tag{17}$$

$$\mathrm{L} + \mathrm{P'}_2 \longrightarrow \mathrm{P}_2 \tag{18}$$

$$\mathrm{B}_2 \longrightarrow \cdot \tag{19}$$

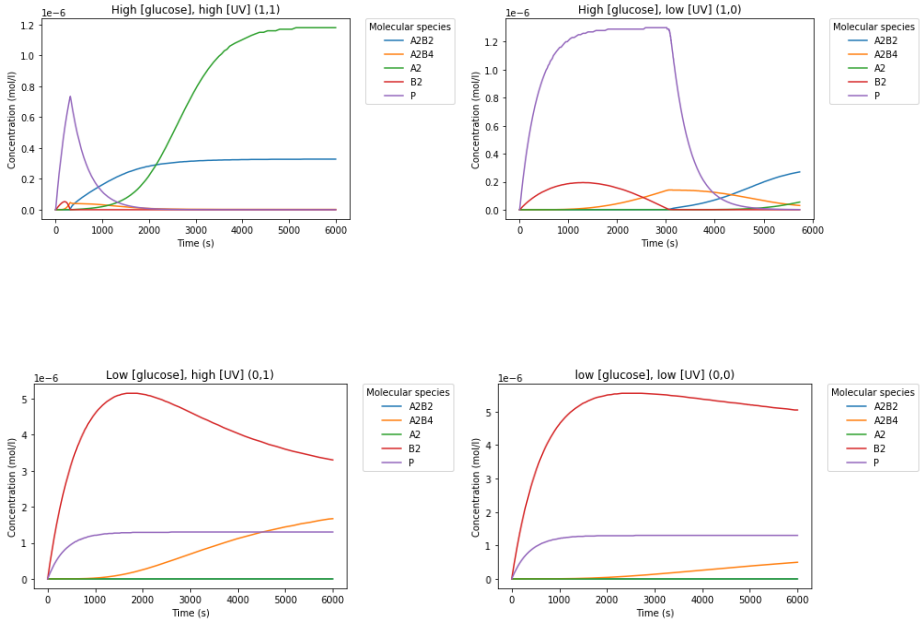


Figure 2: Boolean behaviour of the thanogen when the environmental inputs, glucose and UV (with green light as a correlated proxy for UV intensity) are changed from arbitrarily high to low concentrations. In high glucose and high UV ensure dominant production of $ccdA_2$ (A2) after 1500s. In the top right, note that the amplitude and phase of mf-lon protease (P) is greater, skewing the state of the cell towards death in the absence of UV. In the bottom left and right, absence of glucose controls expression of ccdB (B2), drives DNA gyrase poisoning and derepression of the protease integrator, thus indirectly inhibiting lacI and raising endonuclease concentration (fig 3)

2.1 Steady states and transcription factor dynamics

Solved and plotted using the odeint package in Python From this, the probabilities of cell death under conditions of green light and glucose were extrapolated. For green light, since ccdB:ccdA interactions are the main mediator of bacteriostatic cell death, a Michaelis-Menten type equation was derived for ccdB binding its target DNA gyrase to estimate probability of cell death. For glucose, since a toxic threshold of endonuclease could not be confirmed from the literature, concentration of endonuclease was assumed to be proportional with $P(death)$.

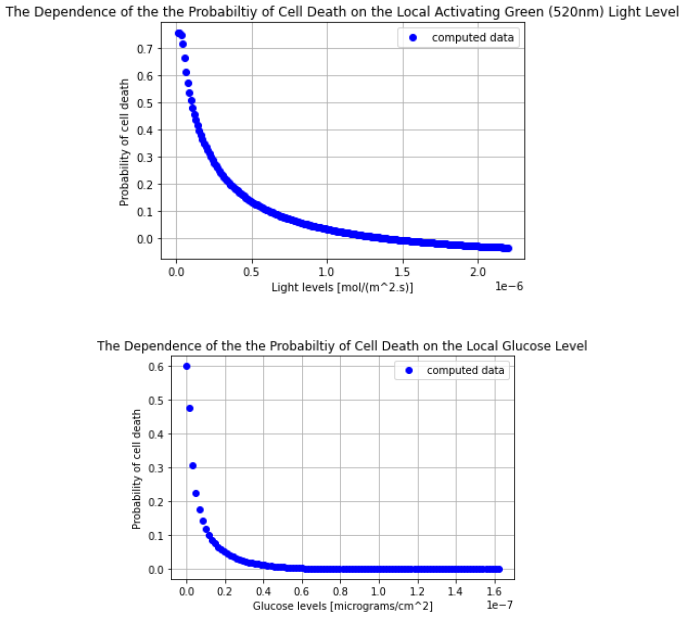


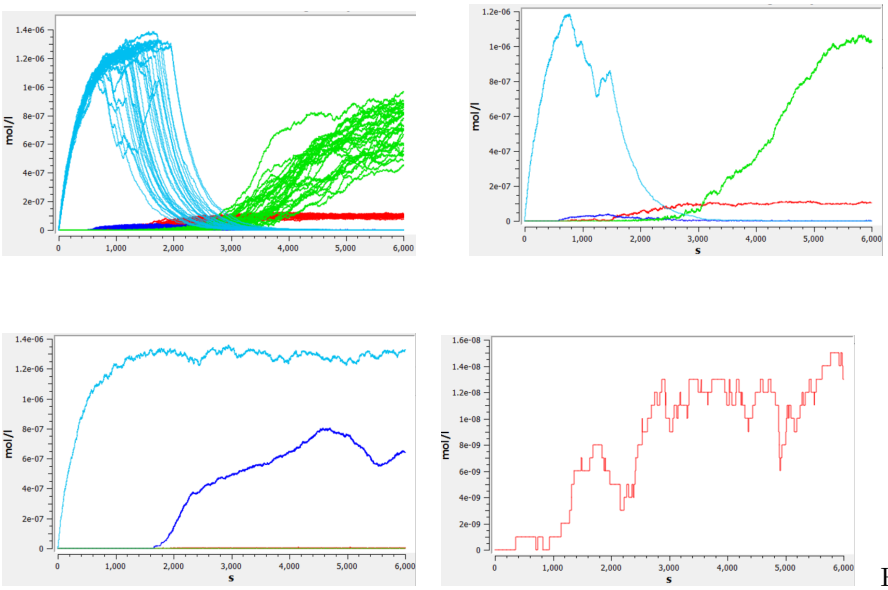
Figure 3: Interpolated probabilities of cell death calculated for green light and glucose dependence based on A2B4 (2*ccdB:ccdA) and endonuclease concentrations at a fixed timepoint, respectively. Behaviour

2.1.1 Estimating parameters

COPASI is a biomolecular pathway modelling software with built-in optimisation and parameter estimation tools. Using parameters assumed from the literature - protein degradation rates and basal transcription rates, unknown parameters were modelled using the software's pairwise genetic algorithm for bistable steady state concentrations of toxins and antitoxin.

2.2 Stochastic model of toxin-antitoxin production

While the assumptions of the deterministic equations may hold true for reactions on the molar scale (in systems with 6.6×10^{23} molecules) and give a good prediction of the dynamics of the system, transient fluctuations in the number of discrete reacting particles may alter the behaviour of the system at time-points where the state is unstable. Particularly in introduced plasmid-borne genes where gene expression may generate proteins on the order of 10000s.
In the thanogen, the life-death decision is mediated by a turning point in ccdA which both neutralises DNA gyrase-induced ccdB toxicity and acts as a co-repressor to mf-lon protease.



For an interval of time Δt, the change in $P(s,t)$ equals the probability of entering state s from another state s' minus the probability of leaving s.

Figure 4: Behaviour of thanogen under high glucose and UV (green light), plotted with COPASI. In the top left, 50 stochastic trials of medium UV-high glucose conditions were run, showing the same characteristic behaviour as the differential model. The turqoise trials are protease, green, A2, red, A2B2 and dark blue, A2B4. In the top right, a single trial is shown. Bottom left is a single trial for low UV, low glucose showing protease, turqoise, and A2B4, dark blue in excess of the A2, committing the cell towards cell stasis and death. In the bottom right is a trial for endonuclease production under low glucose, low uV.

At transient ccdA concentrations, random noise in the cell could perturb the robustness of the killswitch or generate unpredictable outputs.

If instead of using the law of mass-action, we consider a probability of the system being in a given state. Calling that vector, x as a master list of all the species involved in the system and S, the vector containing the number of molecules of individual chemical species, S_i $P(x,t)$. Hence, the probability that the thanogen is in state s at time t is $P(x,t|x_0,t)$ where x_0 is the given set of starting concentrations. The probability of being in a given state must be determined by the net flux through that given state. Species may enter and exit states via a vector of reactions R $(R_1...R_T)$ where T is the total number of reactions in the system (see equations above). These reactions include degradations, syntheses and binding. Therefore,

$$\Delta \mathrm{P}(\text{system being in s, at t}) = [\mathrm{P}(\text{entering } s \text{ from } s') - \mathrm{P}(\text{leaving } s \text{ to } s')]\Delta \mathrm{t} \tag{20}$$

2.2.1 Assumptions

- The volume, V of the cell is fixed $(1.66 \times 10^{-15} L)$

- The system is spatially homogeneous and thus the probability of finding a given species is uniform throughout the cell

Following from these assumptions, COPASIUI and Python were used to run the Tau-Leaping method, an approximation method for solving stochastic systems.

3 Shinogen

This section of the modelling attempted to achieve the following aims:

- Justify aspects of the experimental design, i.e. whether chromosomal integration could achieve the required expression levels of the shinezymes
- Determine any bottlenecks in the metabolic pathway to be addressed in the wetlab flux optimisation. In the dearth of experimental data, this was approximated using the translation rate proportionality assumption
-

3.1 Biochemical model for constitutive shinogen production

The behaviour of the genetic circuit was mapped out in kinetic reactions describing RNA polymerase binding, dissociation, protein synthesis and degradation. Initially, we postulated that the pool of available RNA polymerase would influence the net expression quota of our gene cluster. Conversely, it appeared that the population of promoters were small enough (within the copy number range of 1-120 [P]) to consider RNAp 0 order and saturating. Graph of RNAp use

Shinezyme parameters: P denotes available promoters, P', occupied promoters, D_1 DHQS, D_2, O-MT, D_3, ATPG and D_4, NRPS.

Determining steady-state concentrations of the shinezymes Considering that shinorine and its shinezymes remain contained within the cell cytoplasm, the species were assumed to be homogeneous within the solution; model took into account N(t).

$$\mathrm{P} \underset{\mathrm{k_{-1}}}{\overset{\mathrm{k_1}}{\rightleftharpoons}} \mathrm{P'} \tag{3.21}$$

$$\mathrm{P'} \xrightarrow{\mathrm{k_2}} \mathrm{P} + \mathrm{D_1} \tag{3.22}$$

$$\mathrm{P'} \xrightarrow{\mathrm{k_3}} \mathrm{P} + \mathrm{D_2} \tag{3.23}$$

$$\mathrm{P'} \xrightarrow{\mathrm{k_4}} \mathrm{P} + \mathrm{D_3} \tag{3.24}$$

$$\mathrm{P'} \xrightarrow{\mathrm{k_5}} \mathrm{P} + \mathrm{D_4} \tag{3.25}$$

Degradation terms

$$\mathrm{D_1} \xrightarrow{\mathrm{d_{D_1}}} \cdot \tag{3.26}$$

$$\mathrm{D_2} \xrightarrow{\mathrm{d_{D_2}}} \cdot \tag{3.27}$$

$$\mathrm{D_3} \xrightarrow{\mathrm{d_{D_3}}} \cdot \tag{3.28}$$

$$\mathrm{D_4} \xrightarrow{\mathrm{d_{D_4}}} \cdot \tag{3.29}$$

3.2 Derivatives

$$\frac{\partial P'}{\partial t} = k_1 P - K P' \tag{3.30}$$

$$\frac{\partial P}{\partial t} = -k_1 P + K P' \tag{3.31}$$

$$\frac{\partial D_1}{\partial t} = k_2 P' - \alpha D_1 \tag{3.32}$$

$$\frac{\partial D_2}{\partial t} = k_2 P' - \beta D_2 \tag{3.33}$$

$$\frac{\partial D_3}{\partial t} = k_3 P' - \gamma D_3 \tag{3.34}$$

$$\frac{\partial D_4}{\partial t} = k_4 P' - \epsilon D_4 \tag{3.35}$$

$$\frac{dS}{dt} = Nv \tag{3.36}$$

Where $K = k\_1 + k_2 + k_3 + k_4 + k_5$ and S denotes the set of all species

3.3 Matrix of reaction terms for N and v

N is a stoichiometric matrix of the species corresponding to the individual rates of reaction. -1 defines a reactant, 1, a product and 0, an unaffected species.

$S^T = (P, P', D_1, D_2, D_3, D_4)$

$v^T = (k_1 P, k_{-1} P', k_2 P', k_3 P', k_4 P', k_5 p', \alpha D_1, \beta D_2, \gamma D_3, \epsilon D_4)$

$$N = \begin{bmatrix} -1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & -1 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \end{bmatrix}$$

3.4 Solving for steady states

On analysis of the matrix, one detects a redundancy between P' and P. The availability of promoters, P, is dependent upon the concentration of P'.

Ergo, $P_T = P' + P$

To calculate steady-state concentrations of the enzymes When $\frac{dS}{dt} = 0$, simple algebraic manipulation yields the following equations for [Protein] at steady state.

$$P_s = P_T - P'_s \tag{3.37}$$

$$D_{1s} = \frac{1}{\alpha} k_2 P'_s \tag{3.38}$$

$$D_{2s} = \frac{1}{\beta} k_3 P'_s \tag{3.39}$$

$$D_{3s} = \frac{1}{\gamma} k_4 P'_s \tag{3.40}$$

$$D_{4s} = \frac{1}{\epsilon} k_5 P'_s \tag{3.41}$$

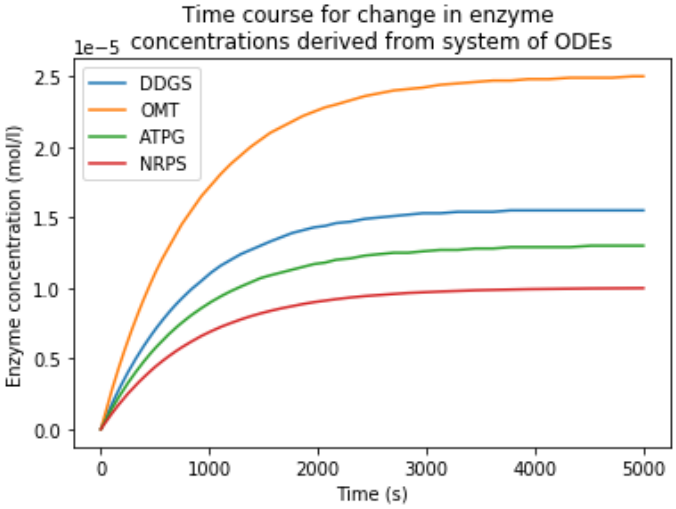


Figure 5: Time course for shinorine-producing enzymes. The non-ribosomal peptide synthetase (NRPS) appears most rate-limiting.

Inputting values determined below using the approximate rates derived from nucleotide sequence length and assuming a rate of transcription/translation of 60nt/s, the succeeding rates were computed for $k_1 - k_5$. Employing the Avogadro relationship, the steady state number of the four shinezymes were determined to be 42, 60, 34 and 19 molecules, respectively.

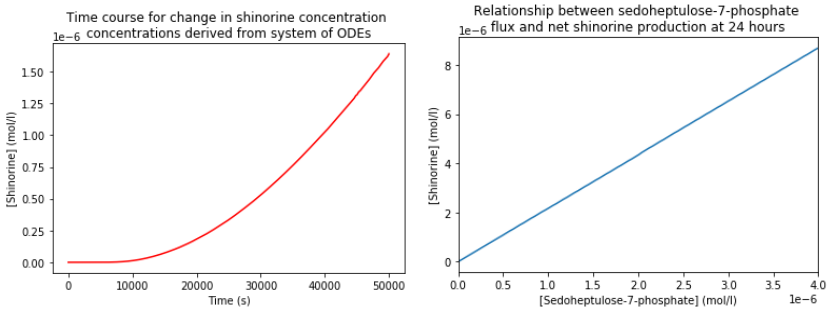


Figure 6: Production of shinorine under wild-type flux of sedoheptulose-7-phosphate (SH7P) left, and changing flux of SH7P on shinorine concentration at t=24 hours.

This demonstrates that the final non-ribosomal peptide synthetase (NRPS) is indeed a likely rate-limiting step. This influenced our design of the composite part for NRPS synthesis; strong promoter/RBS were incorporated to gain better stoichiometry among the four enzymes. Total shinorine production across 14 hours is shown in fig 6, left. Within the model, shinorine does not reach a steady state insofar as it is assumed that (1) the cell has no mechanism for degradation of shinorine and (2) the effects of diffusion are not applicable since shinorine can absorb and dissipate incident UV at any point within the cytoplasm of the bacterium.

Additionally, flux through the pentose phosphate pathway was determined to directly influence shinorine production (fig 6, left). By increasing sedoheptulose-7-phosphate, the net shinorine at 24 hours post-induction increases linearly.

3.5 Parameters

Parameter	Value	Unit
V_{cell}	1.66×10^{-15}	L
P_T	10^{-9}	M
k_1	10^5	s^{-1}
k_-1	0.01	s^{-1}
k_2	0.049	s^{-1}
k_3	0.07	s^{-1}
k_4	0.042	s^{-1}
k_5	0.022	s^{-1}
C_N	1-30	au
k_d	0.00116	s^{-1}

4 Bibliography

The bibliography can be located at the bottom of the modelling page at https://2020.igem.org/Team:St$_Andrews/Model$