Combating Wide range Cadmium poisoning in Agriculture using Genetically Engineered Yeast -A Mathematical Modelling Study

Nivedaa Dhandapani
19MS023

ABSTRACT

1)The mathematical model for the production of phytochelatin synthetase in the presence of Cadmium using **SpPCS1 gene,** a genetic model involving a **promoter, RBS, SpPCS1 genetic sequence and a terminator** is presented. Cadmium here serves as an activator to the promoter, needed to activate the **SpPCS1 gene** which produces phytochelatin synthetase.We attempt to show this is a clear case of a bistable mechanism(the ON/OFF switch)..

2) The mathematical model for the genetic circuit in the presence of Cadmium only is presented varying the concentrations of Cadmium and visualising the working of this circuit over a time period, documenting the change in concentration with respect to time.

INTRODUCTION

Bi-stability (fig. 1)in biochemical signalling model occurs when the system contains two possible steady states which are stable. The two steady states are called low activity and high activity state. This is visualised in a mathematical model as a **bifurcation,** where the system qualitatively changes its behaviour. Most of biological mechanisms are monostable, but there are several processes in nature such as cell division, fertilisation, apoptosis etc. where a graded response is insufficient to capture the nature of the processes which require persistence. This is where a bistable model is a more accurate representation of such phenomena.

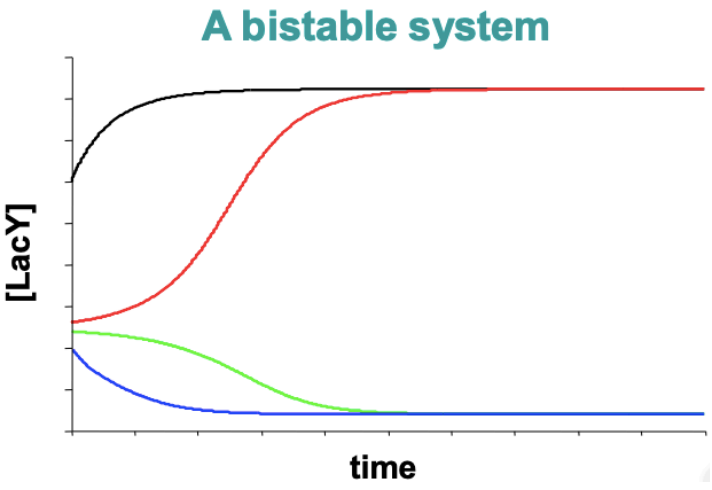


FIG .1:CONC. VS TIME PLOT FOR A LACY GENE IN THE LAC OPERON MODEL

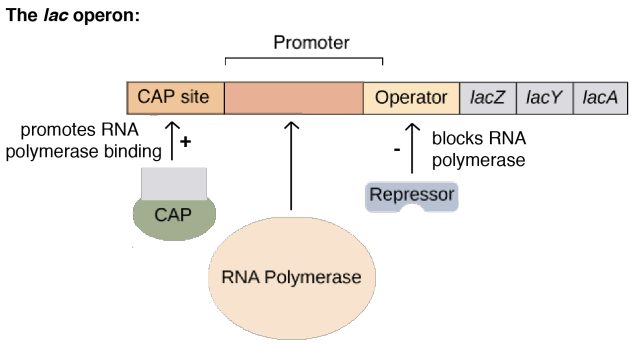


FIG. 2:SCHEMATIC REPRESENTATION OF A LAC OPERON

An ***Operon*** (fig. 2)is a functioning unit of DNA containing a cluster of genes under the control of a single promoter. The three basic DNA components that make up the operon are:

PROMOTER- a nucleotide sequence that enables genes to be transcribed. In case of an inhibited promoter, its activation is vital to facilitate transcription.

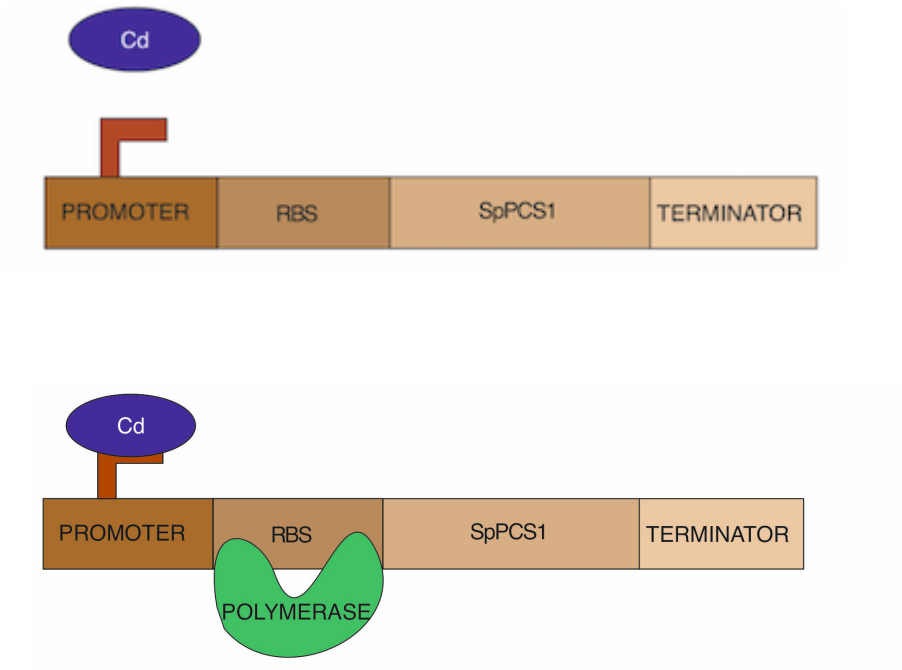
OPERATOR- the segment of DNA to which the repressor binds, preventing transcription by DNA/RNA polymerase.

STRUCTURAL GENES- the genes that are co-regulated by the operon that are needed to be transcribed to mRNA.

MODELLING SpPCS1 GENE CIRCUIT

1) GENETIC CIRCUIT: PHYTOCHELATIN SYNTHETASE PRODUCTION

This model is designed to understand the production of phytochelatin synthetase (hereafter referred to as PC-s) in the presence of the heavy metal cadmium (Cd). The cadmium acts as an activator to the SpPCS1 gene by binding to its promoter region. We see this process as an analogue to the well established Lac-operon model which uses allolactose to activate gene function by having it bind to the repressor. The key difference between the SpPCS1 gene and the Lac operon is that the binding happens at different sites of the gene, but the end result of the processes remain the same i.e., gene is activated and the required enzyme is produced. This is a classic example of a bistable system because the cadmium here switches the gene on initiating PC-s production upon binding and the gene remains off otherwise.



CADMIUM ACTIVATES THE PROMOTER AS A RESULT OF WHICH POLYMERASE BINDS TO THE RBS.

ASSUMPTIONS:

1. n molecules of cadmium are required to bind to the promoter(this is a reversible reaction) to activate the genetic circuit.
2. phytochelatin synthetase concentration is zero if the cadmium concentration in the system is below threshold value as we assume that without cadmium activation there cannot be PC-s in the system.
3. The initial cadmium concentration taken is either much higher or much lower than the threshold value of cadmium required to activate the circuit.

DERIVATION:

$$P + nA \leftrightarrow PA_n$$

The above equation is from the assumption that n molecules of metal A (cadmium in our case) binds to the promoter P. This is an equilibrium reaction with a constant $k_1 = \frac{PA_n}{P \cdot A^n}$.—(1)

$PA_n = k_1 \cdot P \cdot A^n$ from (1)

$P_{total} = P + PA_n$

$= P + k_1 \cdot P \cdot A^n = P \cdot (1 + k_1 A^n)$

Fraction of promoters bound by metal

$$= \frac{PA_n}{P_{total}}$$

$$= \frac{k_1 \cdot P \cdot A^n}{P \cdot (1 + k_1 A^n)} = \frac{k_1 \cdot A^n}{(1 + k_1 A^n)}$$ (*this is a classic example of a **Hill function***)

This is used in the differential equations to account for the rate of mRNA production due to its direct correlation to the cadmium promoter binding process.

DIFFERENTIAL EQUATIONS OF RATE OF CHANGE OF CONCENTRATIONS OF PARAMETERS:

1)Rate of change of mRNA concentration

$$\frac{d[mRNA]}{dt} = \frac{\alpha_m k_1 (e^{-\mu\tau_m}[Cd])^n}{1 + k_1 (e^{-\mu\tau_m}[Cd])^n} + \Gamma_0 - (\gamma + \mu)[mRNA]$$

The above equation accounts for the delay in mRNA transcription by including the $e^{-\mu\tau_m}$ term(**growth dependent dilution of Cd**) where τ_m is the time taken for the transcription of SpPCS1 gene by RNA polymerase.
$(\gamma + \mu)[mRNA]$ is the loss term which accounts for degradation($(\gamma)[mRNA]$) and loss due to dilution $(\mu)[mRNA]$. Γ_0 is the spontaneous rate of mRNA production(which is zero for our model)

2)Rate of change of PC-s concentration

$$\frac{d[PC-s]}{dt} = \alpha_b (e^{-\mu\tau_b})[mRNA]_{\tau_b} - (\gamma_b + \mu)[PC-s]$$

Here τ_b accounts for the time taken for mRNA translation and $e^{-\mu\tau_b}$ accounts for mRNA dilution due to cell growth.$[mRNA]_{\tau_b}$ is the concentration of mRNA at a time τ_b before the current time. $(\gamma_b + \mu)[PC-s]$ is the loss term which accounts for degradation ($\gamma[PC-s]$) and loss due to dilution of PC-s.($\mu[PC-s]$)

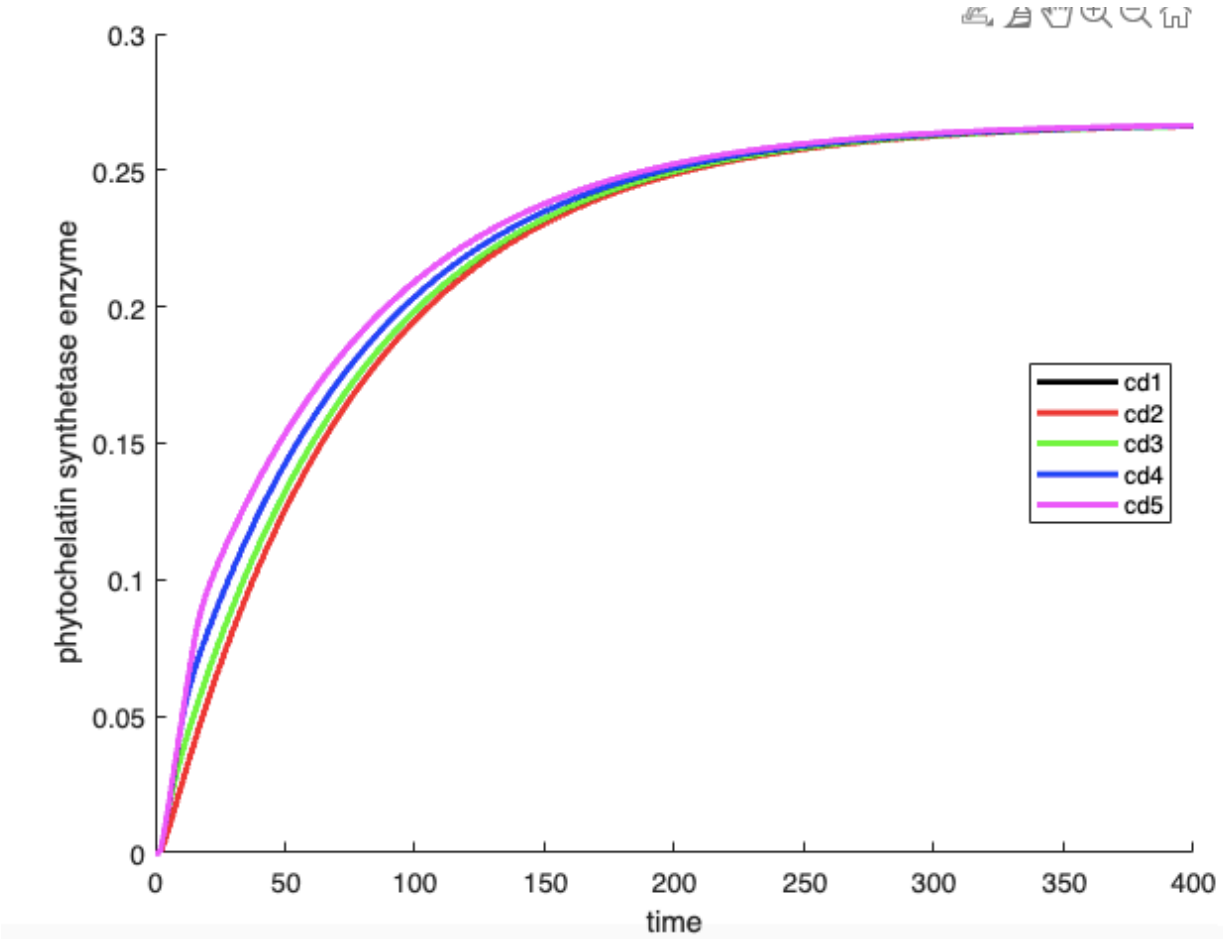
3)Rate of change of Cd

$$\frac{d[Cd]}{dt} = \alpha_c - (\gamma_c + \mu)[Cd]$$

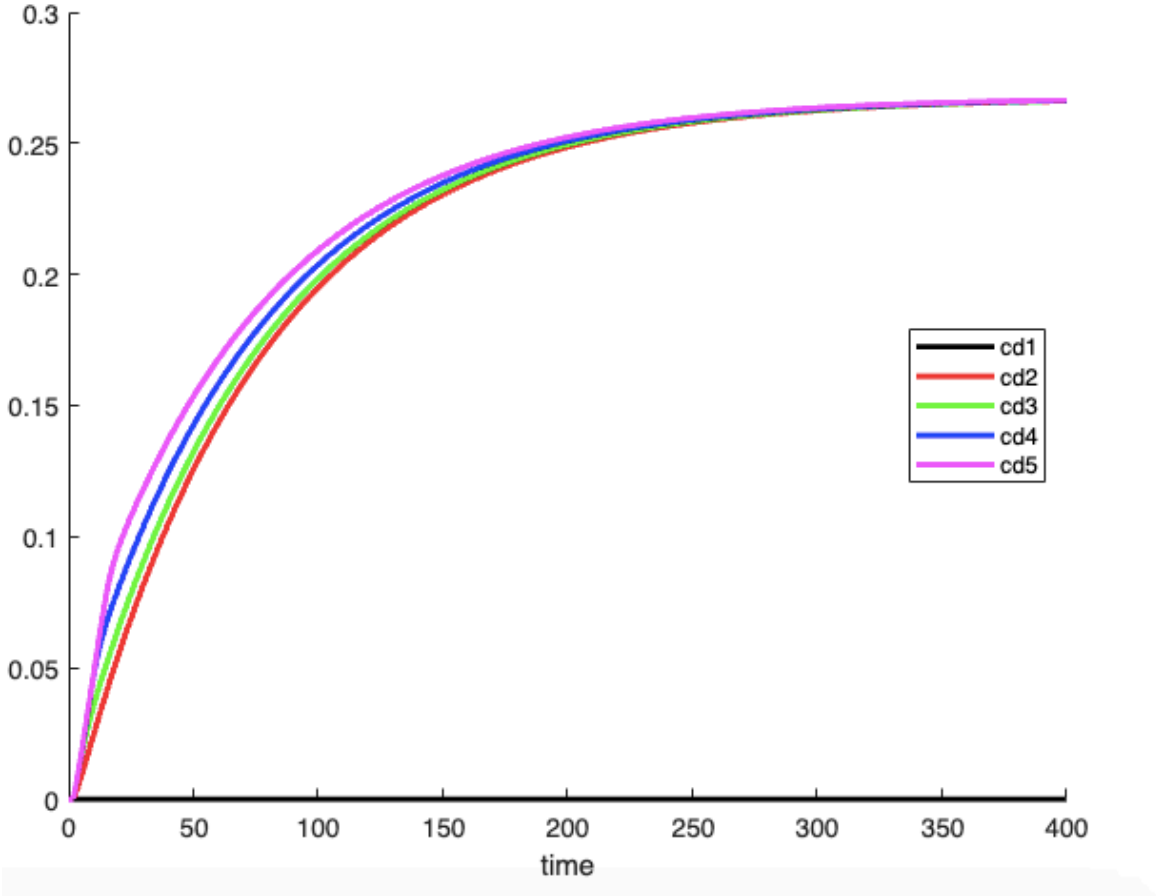
Here α_c is the constant rate of binding of cadmium to the promoter and $(\gamma_c + \mu)[Cd]$ is the loss term with degradation ($(\gamma_c)[Cd]$)and loss due to dilution ($(\mu)[Cd]$).

ANALYSIS OF MODEL EQUATIONS:

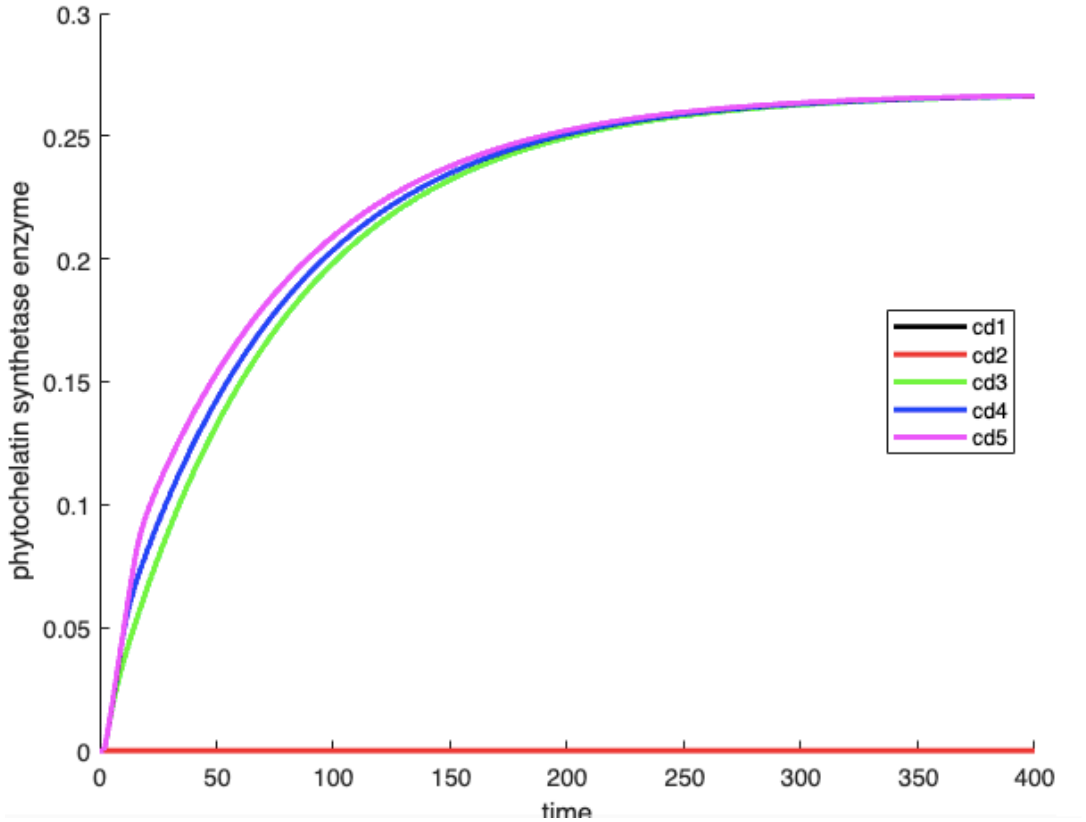
We now plot the above mentioned equations in **MATLAB R2020b.** The program was run assuming arbitrary values for the constant parameters in the ODEs since we haven't performed laboratory experiments to conclude accurate values.



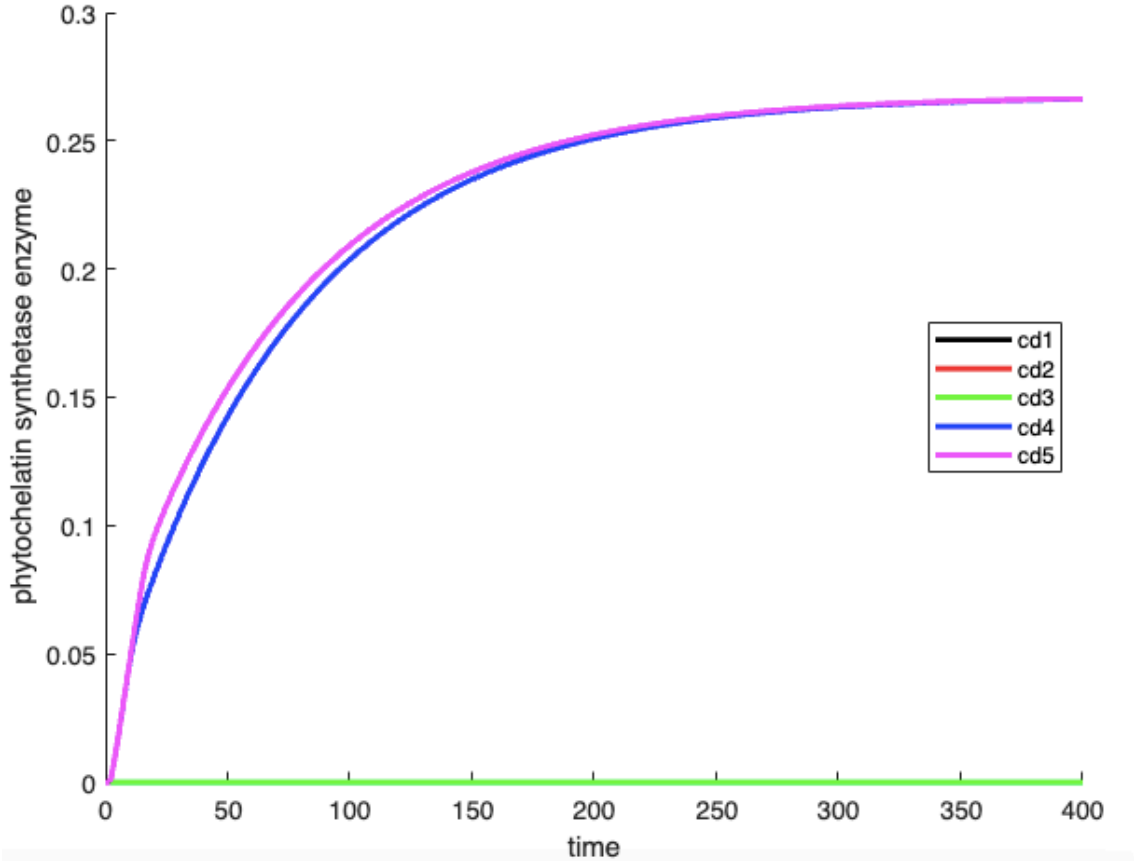
ALL CADMIUM CONC. ARE GREATER THAN THRESHOLD VALUE



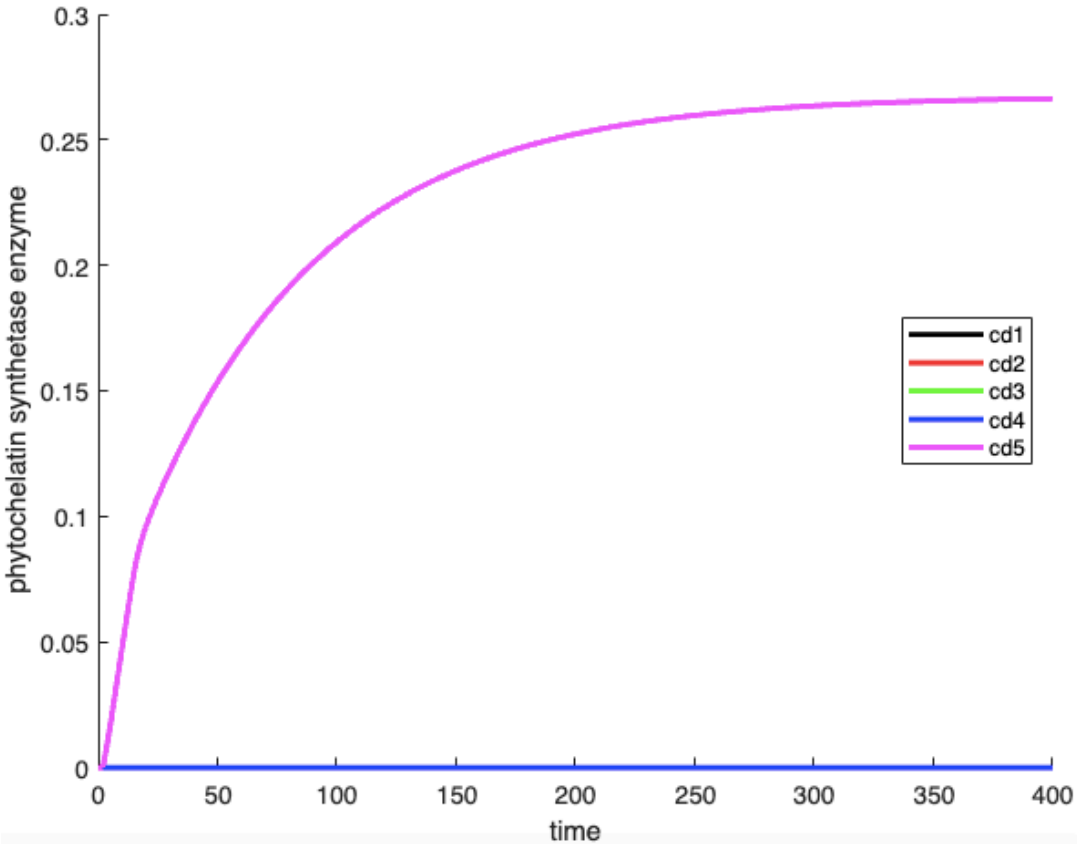
CD_1 CONC. IS LESSER THAN THE THRESHOLD VALUE



CD_1 AND CD_2 CONC. ARE BELOW THE THRESHOLD VALUE



CD_1 ,CD_2 AND CD_3 CONC. ARE BELOW THE THRESHOLD VALUE



CD_1 ,CD_2 , CD_3 AND CD_4 CONC. ARE BELOW THE THRESHOLD VALUE

1) The above five plots represent the **change in concentration of PC-s with respect to time** for different Cd thresholds. We observe that for different initial concentrations of Cd, the concentration of enzyme reaches the **same** steady state value after a certain instant of time. This outcome is in accordance with the hypothesis that production of enzyme stabilises and reaches a steady state value after a well defined time. From this instant on one can consider the concentration of the enzyme to be constant.
2) We notice that the enzyme concentration is zero initially (according to our assumption).When Cd in system is lesser than the threshold value the enzyme concentration is zero.
3) The trend of enzyme concentration for Cd5 shown by the pink line in the plot has a sharp increase before it rises gradually. This is because Cd5 concentration is very high (10^5 times greater than Cd1), implying that more Cd binds to the promoter resulting a sharp increase in enzyme production in a lesser amount of time.
4) On close observation, we see that for a small amount of time after t=0, there is no significant increase in enzyme concentration. This is because mRNA translation takes place in that time delay.
5) Initial Cd concentration has no bearing on the final steady steady state value of the enzyme.

PLOTTING NULL-CLINES AND BIFURCATION:

After phytochelatin synthetase reaches steady state concentration, as per the definition there is no more change in the amount of the enzyme in the system.

$$\Rightarrow \frac{d[PC-s]}{dt} = 0$$

This happens because the circuit is no longer activated and no new mRNA is produced in the system,

$$\Rightarrow \frac{d[mRNA]}{dt} = 0$$

The circuit is not activated because there is no influx of cadmium into the system,

$$\Rightarrow \frac{d[Cd]}{dt} = 0$$

Solving,

$$\frac{d[mRNA]}{dt} = \frac{\alpha_m k_1 (e^{-\mu\tau_m}[Cd])^n}{1 + k_1 (e^{-\mu\tau_m}[Cd])^n} + \Gamma_0 - (\gamma + \mu)[mRNA]$$

$$\frac{d[PC-s]}{dt} = \alpha_b (e^{-\mu\tau_b})[mRNA]_{\tau_b} - (\gamma_b + \mu)[PC-s]$$

and $\frac{d[Cd]}{dt} = \alpha_c - (\gamma_c + \mu)[Cd]$ after substituting $\frac{d[mRNA]}{dt} = 0, \frac{d[Cd]}{dt} = 0, \frac{d[PC-s]}{dt} = 0$, we get

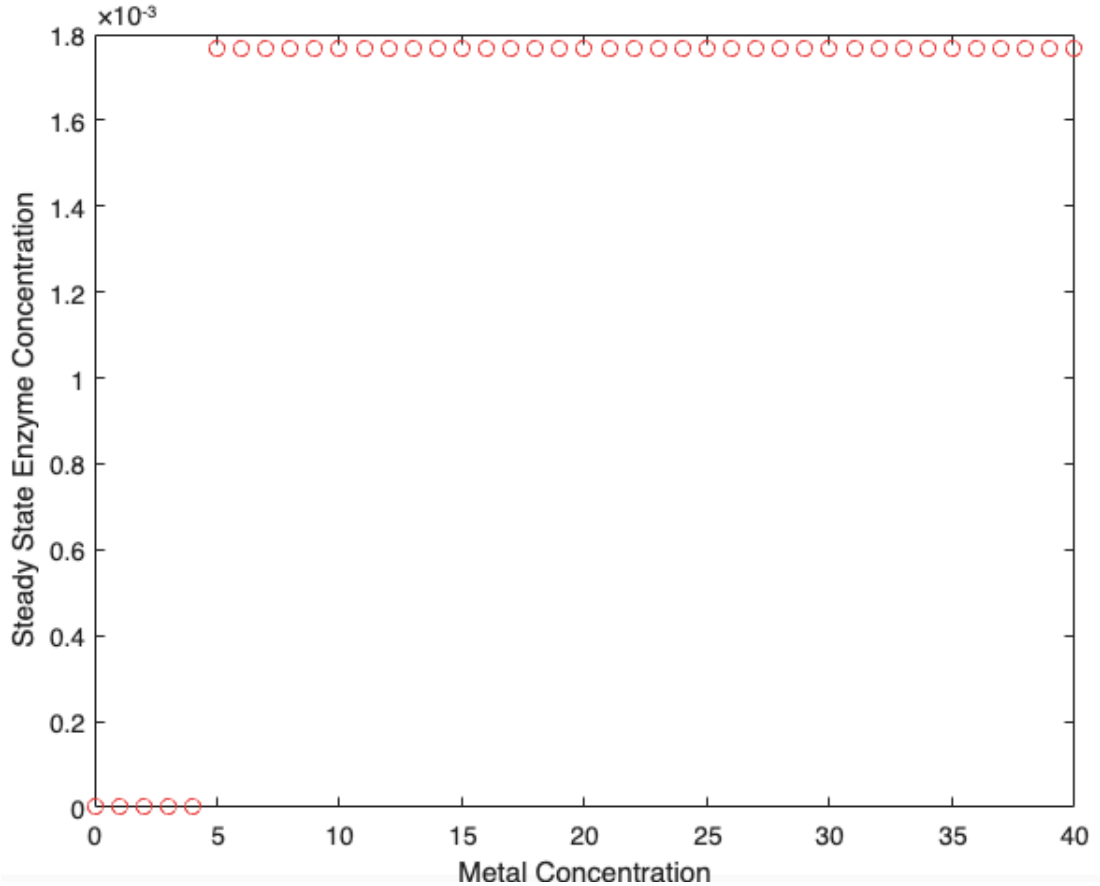
$$[PC-s] = \frac{\alpha_b (e^{-\mu\tau_b})(\alpha_m k_1 (e^{-\mu\tau_m}[Cd])^n)}{(\gamma_b + \mu)(\gamma + \mu)(1 + k_1 (e^{-\mu\tau_m}[Cd])^n)}$$

Substitute $[Cd] = \frac{\alpha_c}{\gamma_c + \mu}$ in the above equation. We get

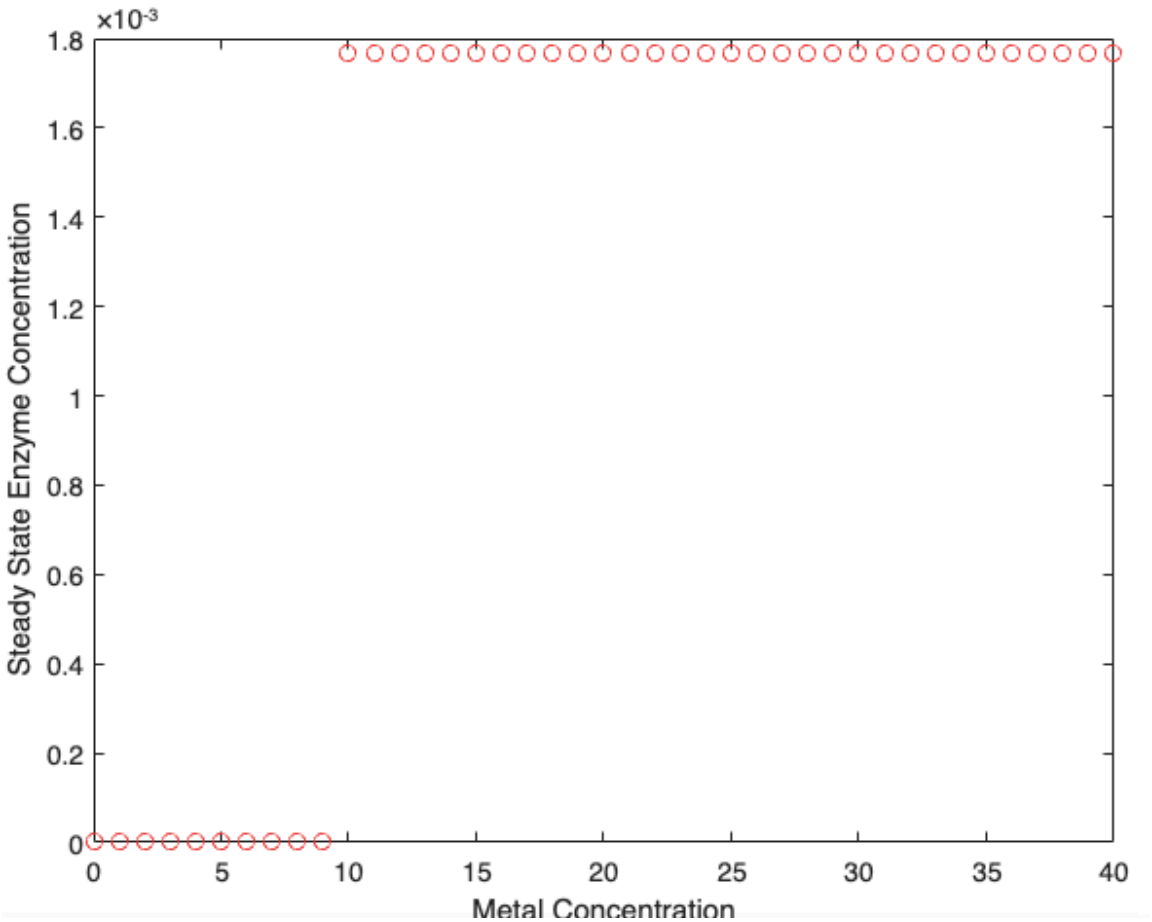
$$[PC-s] = \frac{\alpha_b (e^{-\mu\tau_b})(\alpha_m k_1 (e^{-\mu\tau_m}\frac{\alpha_c}{\gamma_c + \mu})^n)}{(\gamma_b + \mu)(\gamma + \mu)(1 + k_1 (e^{-\mu\tau_m}\frac{\alpha_c}{\gamma_c + \mu})^n)}$$

This above equation satisfies when Cd concentration is greater than the threshold value.

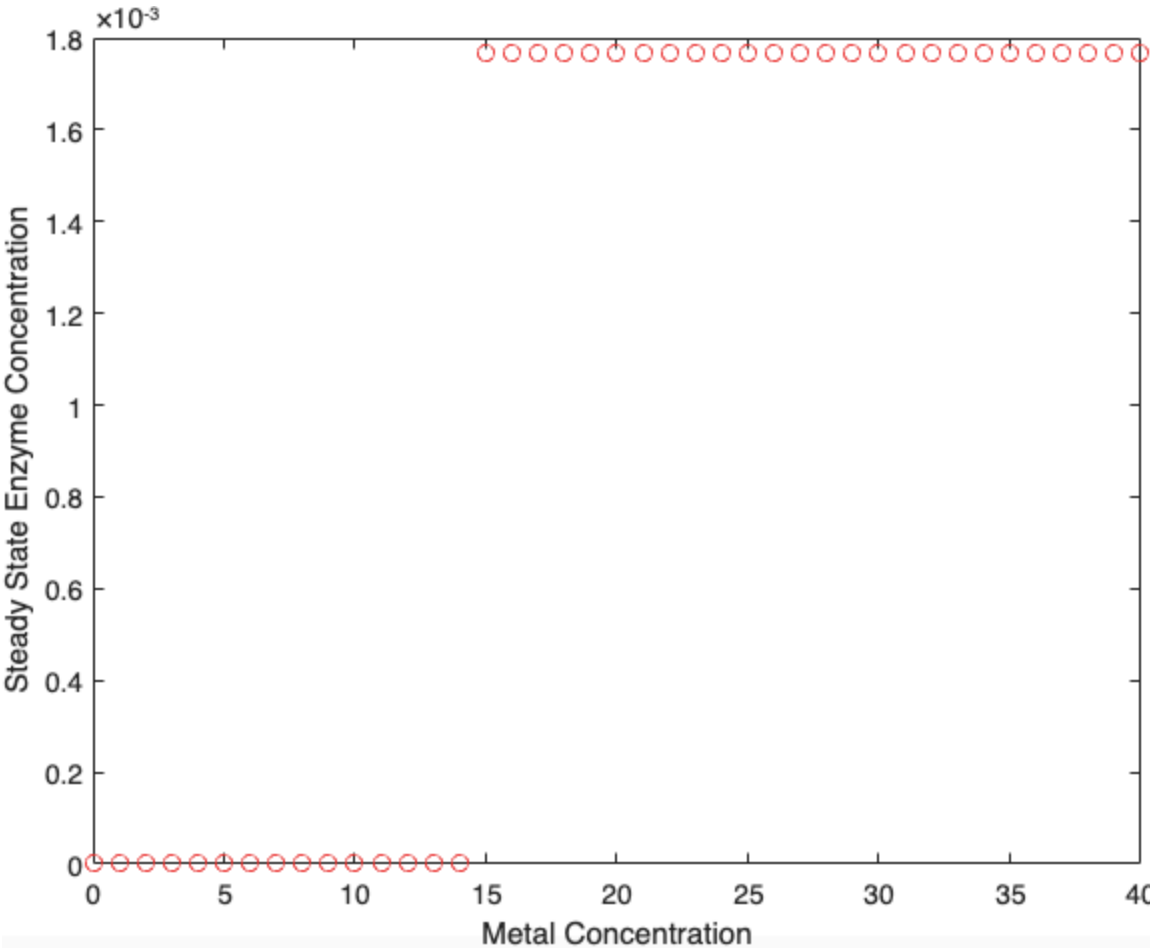
Else, the circuit won't be activated and hence the $[PC-s]$ concentration at steady state is zero.



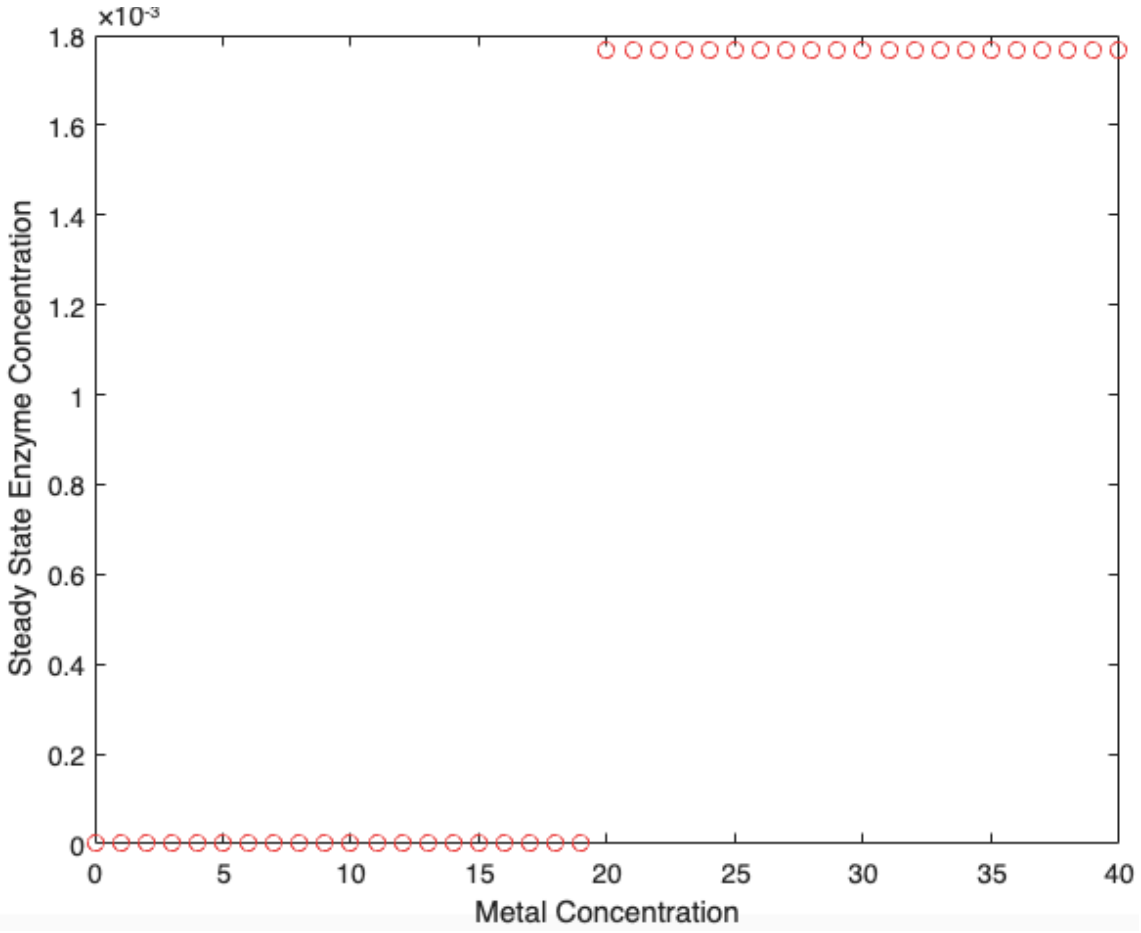
BIFURCATION PLOT FO R CADMIUM THRESHOLD CONC. 1



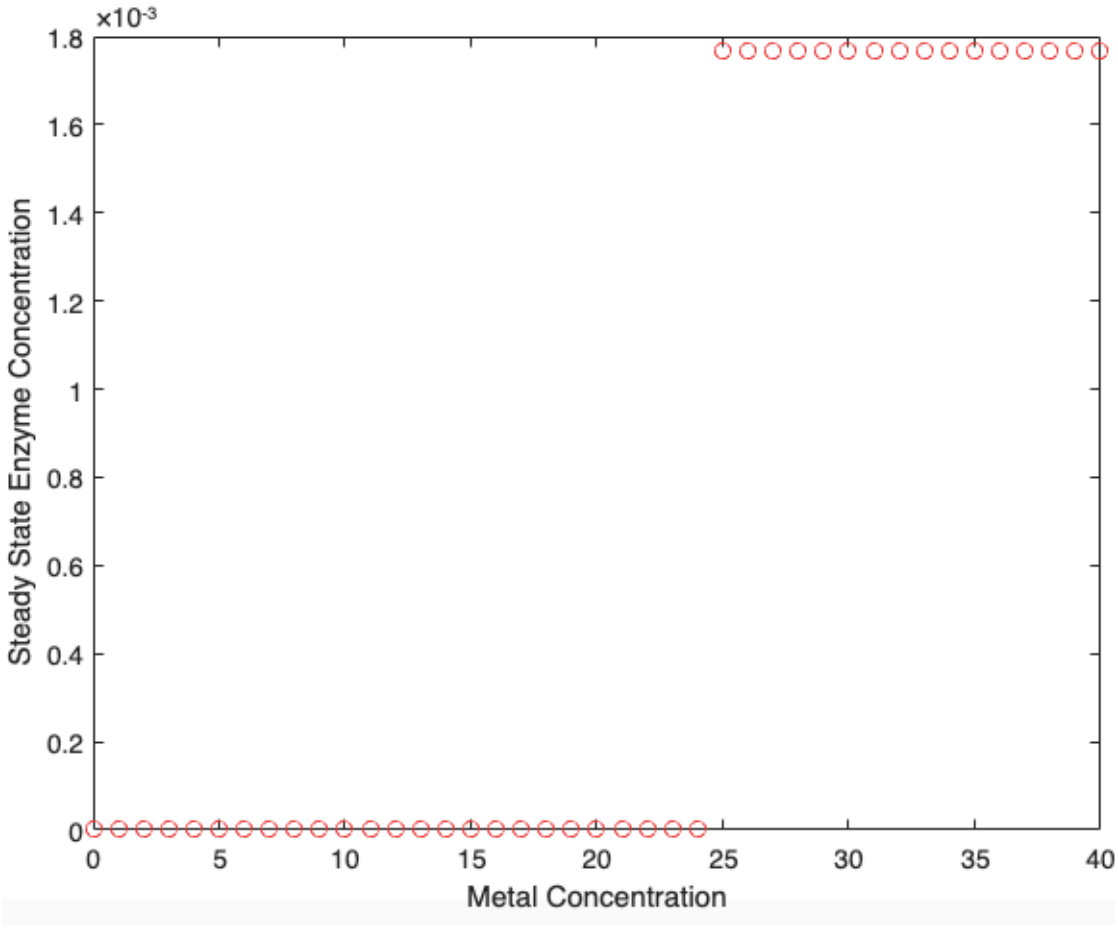
BIFURCATION PLOT FO R CADMIUM THRESHOLD CONC. 2



BIFURCATION PLOT FO R CADMIUM THRESHOLD CONC. 3



BIFURCATION PLOT FO R CADMIUM THRESHOLD CONC. 4



BIFURCATION PLOT FO R CADMIUM THRESHOLD CONC. 5

ANALYSIS

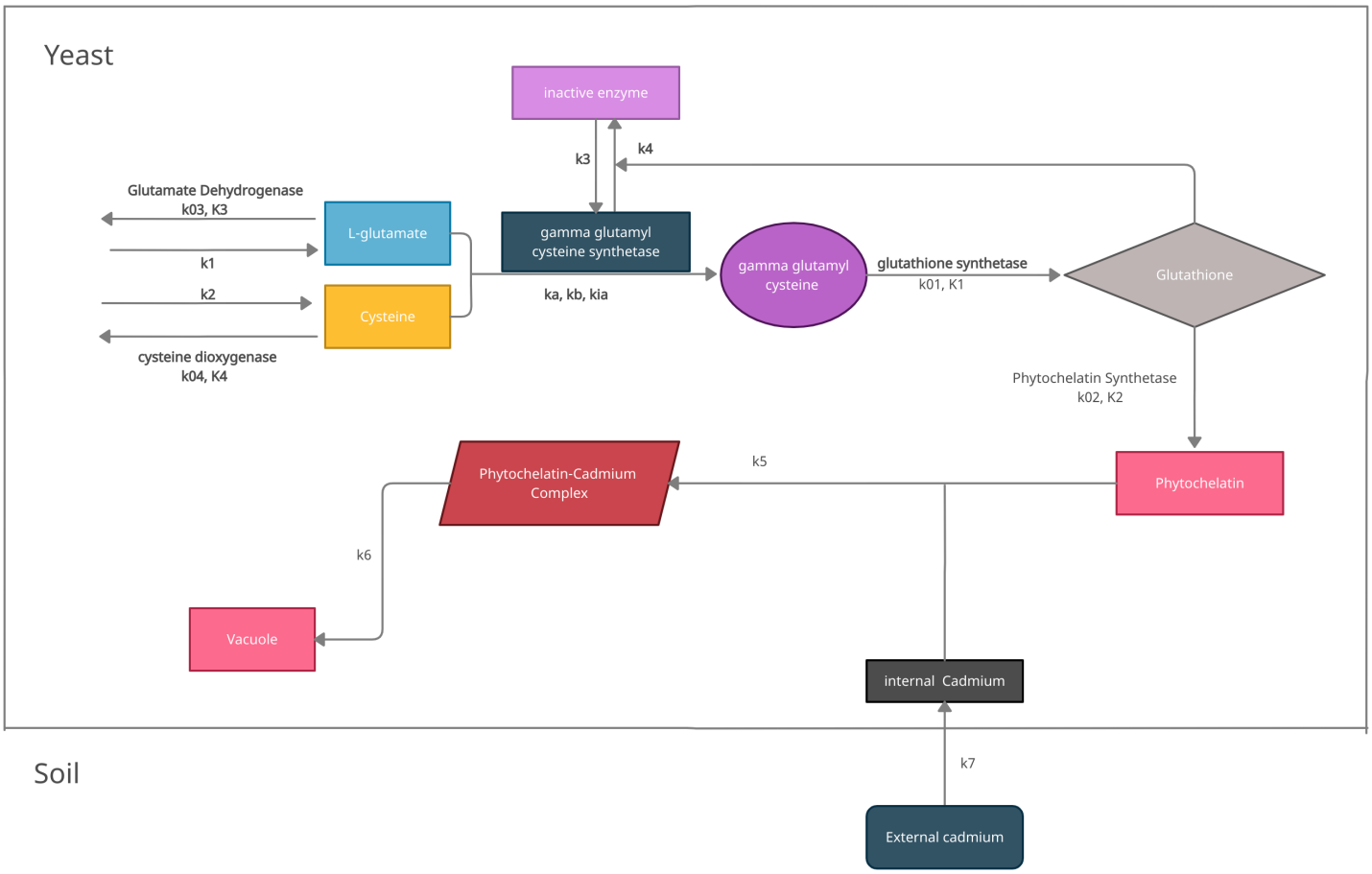
The above graphs represent steady state enzyme concentration **with respect to Cd concentration.** We see a visible change in the qualitative behaviour of the graph beyond a particular concentration of Cd, which is a **bifurcation.** This proves that our circuit is **bistable.**

When Cd concentration is lesser than the threshold value the steady state enzyme concentration is zero (as shown in the graph).

Cd conc. threshold 1<2<3<4<5, which clearly makes a difference regarding at which Cd concentration the discontinuity occurs in the graph behaviour. Greater is the threshold value, greater is the concentration of Cd at discontinuity which is in line with our theoretical predictions. Plotting experimental values of steady state concentrations of PC-s versus metal concentration will accurately determine the Cd threshold value which is none other than the Cd concentration at the discontinuity.

We observe this sudden change in the qualitative behaviour of the plot because the initial cadmium concentrations considered are either much greater or much lesser than the threshold value required to activate the genetic circuit.

MODELLING THE LIGATION OF CADMIUM



Using mathematical modelling we try to understand how the chelation of cadmium by yeast affects the concentration of cadmium both inside and outside the soil. In order to formulate differential equations explaining the same we make the following assumptions:

ASSUMPTIONS MADE IN THE MODEL:

1) Michaelis Menten kinetics is used to determine rate of change of concentrations of substrates in the cycle.
2) The rate of intake of cadmium from the soil by the yeast cell transporters is assumed to be directly proportional to the concentration of cadmium in the soil.
3) The chelated cadmium is removed from the soil at a constant rate.
4) There is no arsenic in the system (this activates another genetic circuit triggering a different response which will interfere with the processes of the cadmium chelation cycle)
5)Phytochelatin synthetase and phytochelatin in the system is absent when the cadmium in the system is below the threshold.
6)The temperature of the system is constant, ensuring no change in rate constant values.

DIFFERENTIAL EQUATIONS FOR THE CADMIUM CHELATION CYCLE:

Using Michaelis Menten kinetics for enzyme substrate reactions, basic rate equations governed by equilibrium laws and law of conservation of mass, we arrive at the following equations:

$$\frac{d[L-glu]}{dt} = k_1 - \frac{k_a k_b [L-glu][cys][\gamma-glu-cys-syn]}{k_{ia}k_a + k_b[L-glu] + k_a[cys] + [L-glu][cys]} - \frac{k_{03}[glutamate-dehydrogenase][L-glu]}{K_3 + [L-glu]}$$

$$\frac{d[cys]}{dt} = k_2 - \frac{k_a k_b [L-glu][cys][\gamma-glu-cys-syn]}{k_{ia}k_a + k_b[L-glu] + k_a[cys] + [L-glu][cys]} - \frac{k_{04}[cysteine-dioxygenase][cys]}{K_4 + [cys]}$$

$$\frac{d[\gamma-glu-cys]}{dt} = \frac{k_a k_b [L-glu][cys][\gamma-glu-cys-syn]}{k_{ia}k_a + k_b[L-glu] + k_a[cys] + [L-glu][cys]} - \frac{k_{01}[\gamma-glu-cys][glutathione-synthetase]}{K_1 + [\gamma-glu-cys]}$$

$$\frac{d[gsh]}{dt} = \frac{k_{01}[\gamma-glu-cys][glutathione-synthetase]}{K_1 + [\gamma-glu-cys]} - k_4[\gamma-glu-cys-syn][gsh] - \frac{k_{02}[PC-s][gsh]}{K_2 + [gsh]}$$

$$\frac{d[phytochelatin]}{dt} = \frac{k_{02}[gsh][PC-s]}{K_2 + [gsh]} - k_5[phytochelatin][Cd]$$

$$\frac{d[Pn-Cd-complex]}{dt} = k_5[phytochelatin][Cd] - k_6$$

$$\frac{d[Cd]}{dt} = k_7[Cd]_{out} - k_5[phytochelatin][Cd]$$

$$\frac{d([Cd]_{out})}{dt} = -k_7[Cd]_{out}$$

$$\frac{d[\gamma-glu-cys-syn]}{dt} = k_3([\gamma-glu-cys-syn]_{total} - [\gamma-glu-cys-syn]) - k_4[\gamma-glu-cys-syn][gsh]$$

Here,

$[L-glu]$ ⇒ L-glutamate

$[cys]$ ⇒ cysteine

$[\gamma-glu-cys-syn]$⇒ γ-glutamylcysteine synthetase

$[\gamma-glu-cys]$ ⇒ γ-glutamylcysteine

$[gsh]$ ⇒ glutathione

$[PC-s]$ ⇒ Phytochelatin synthetase

$[Pn-Cd-complex]$ ⇒ Phytochelatin-Cadmium complex

$[Cd]$ ⇒ Cadmium

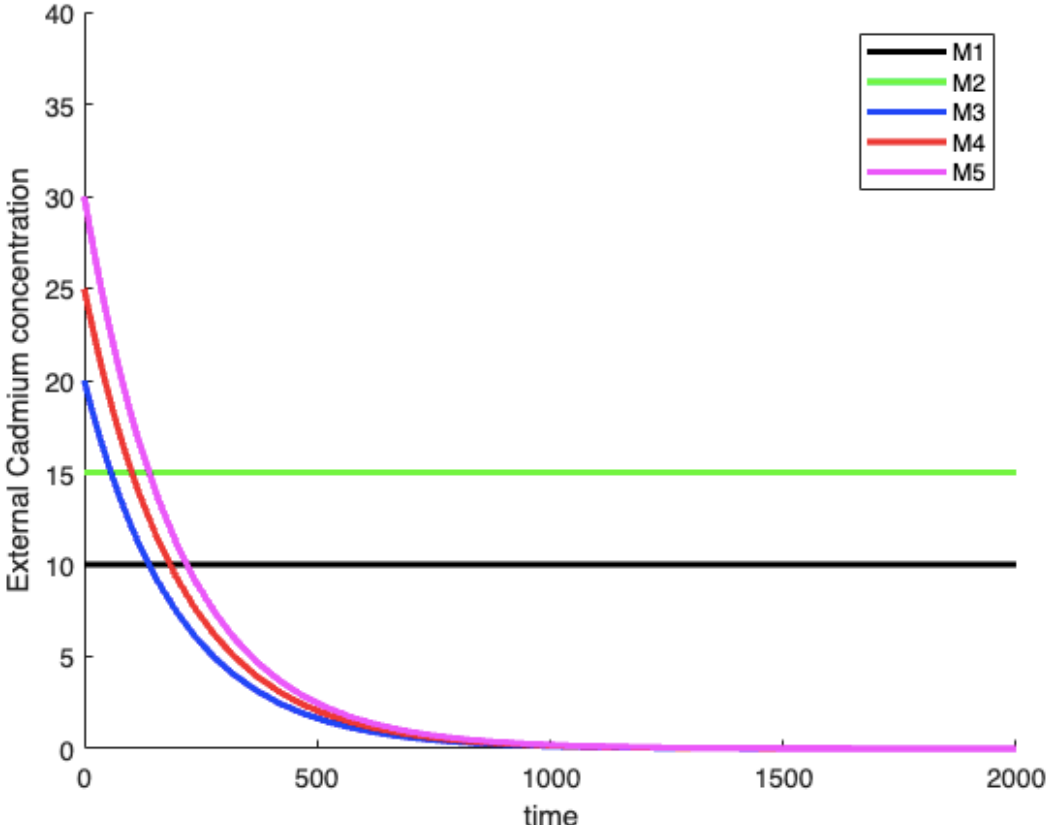
All the rate constants are assumed to not deviate significantly due to temperature fluctuations.

GRAPHS:

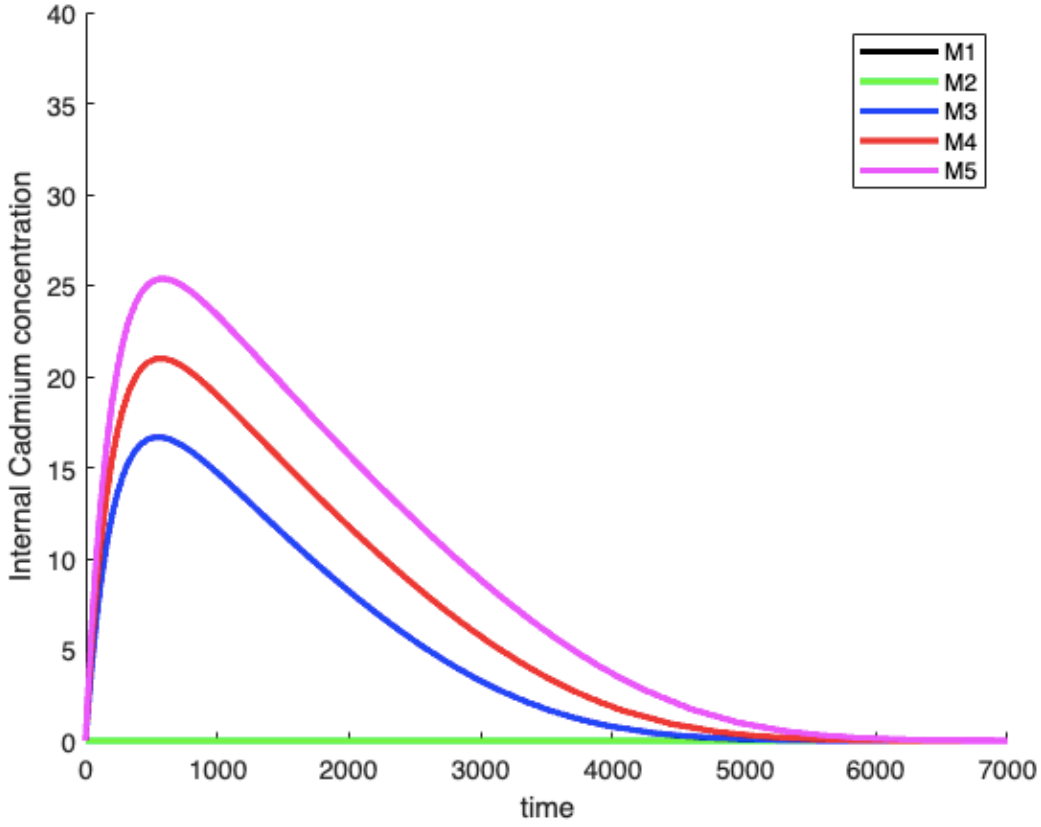
We plot the above differential equations using **MATLAB R2020b** to visualise the dynamic nature if the system at hand and to verify the working of the system. We assume arbitrary initial concentration for the above mentioned parameters and fix rate constants using trial and error since finding the accurate values of these constants require rigorous laboratory experiments.

Here we are plotting various initial concentrations of cadmium (both internal and external) versus time:

In the below graphs concentrations M1<M2<M3<M4<M5 and M1 and M2 concentrations are lesser than the threshold value required to activate the SpPCS-1 gene.



EXPONENTIAL DECAY IN THE CONCENTRATIONS OF CADMIUM IN THE SOIL WHOSE CONCENTRATIONS ARE ABOVE THE THRESHOLD.



INTERNAL CONCENTRATION OF CADMIUM WITH RESPECT TO TIME

ANALYSIS:

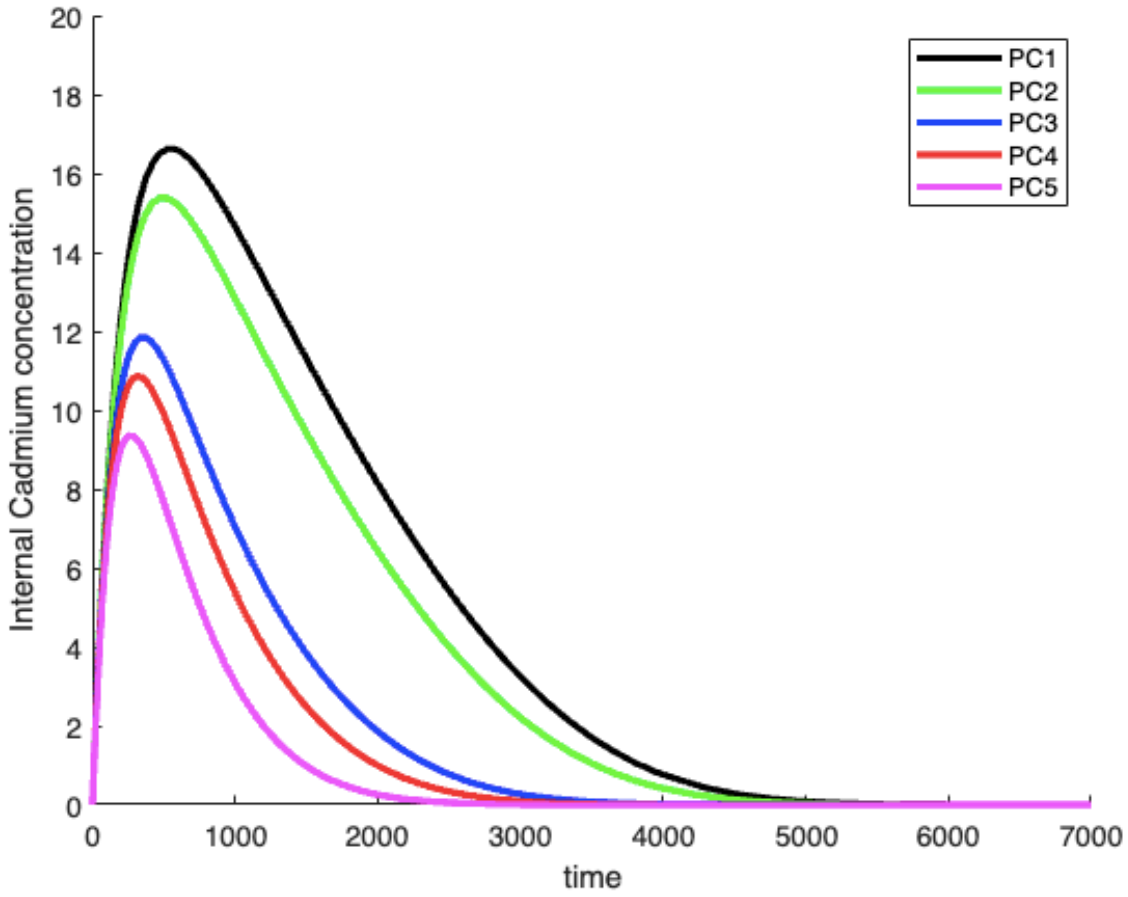
1) We clearly see that higher the external concentration of cadmium, greater the time it takes to chelate all of the residual cadmium in the soil and the concentration drops to zero.
2) We see that below the certain threshold concentration of the cadmium required to activate the circuit, the external concentrations remain constant and the internal cadmium concentration is zero.
3) In the internal concentration of cadmium graph we see an increase initially and then we see a gradual decrease. This is because the slope of the graph for some time after t=0s remains positive because in $\frac{d[Cd]}{dt} = k_7[Cd]_{out} - k_5[phytochelatin][Cd]$, the $[Cd]_{out}$ value is significantly high. Then with more Cadmium in the system, phytochelatin production in the system increases rapidly and reaches a steady state value which dominates over the $[Cd]_{out}$ term making the slope negative again.

We are successfully able to show that through effective chelation of the cadmium in the system, in an acceptable amount of time, the concentration of cadmium reaches zero affirming the working of the theoretical model.

Now we try to study the changes in our model by varying the initial concentration of phytochelatin (chelating agent) for the same initial cadmium concentration:

Internal initial concentration is plotted with respect to time.

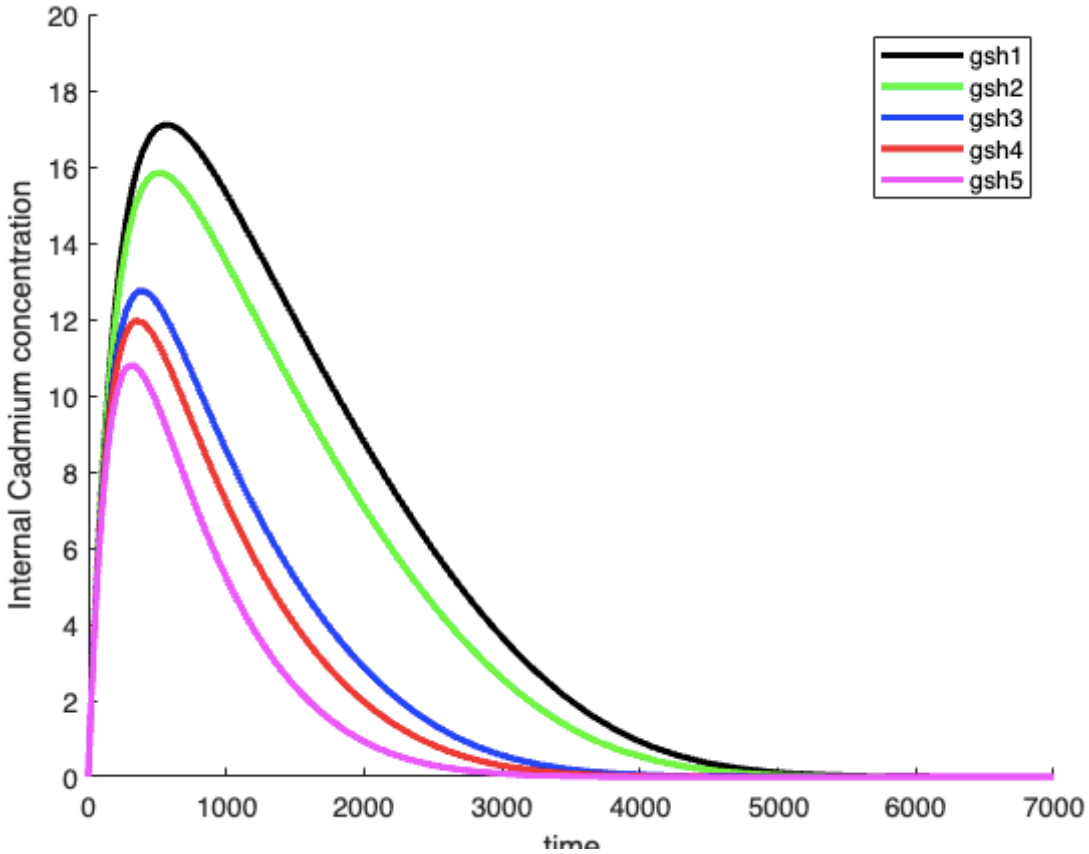
Here the concentration PC1<PC2<PC3<PC4<PC5 and the initial cadmium concentration taken is greater than the threshold concentration.



INTERNAL CONCENTRATION OF CADMIUM WITH RESPECT TO TIME

ANALYSIS:
We see that with increase in the concentration of Phytochelatin, the time taken for the cadmium concentration to reach zero decreases. Greater the amount of chelating agent present in the cell, more cadmium is chelated at a particular time and as a result of which the amount of cadmium in the system falls more drastically with respect to time.This also affects the maxima of the curve i.e., greater phytochelatin concentration implies that the slope $\frac{d[Cd]}{dt} = k_7[Cd]_{out} - k_5[phytochelatin][Cd]$ becomes negative at a lower time and for a lower concentration of internal cadmium.

Now we try to study the changes in our model by varying the initial concentration of glutathione (chelating agent) for the same initial cadmium concentration:
Internal initial concentration is plotted with respect to time.
Here the concentration gsh1<gsh2<gsh3<gsh4<gsh5 and the initial cadmium concentration taken is greater than the threshold concentration.



INTERNAL CONCENTRATION OF CADMIUM WITH RESPECT TO TIME

ANALYSIS:
We see that with increase in the concentration of glutathione, the time taken for the cadmium concentration to reach zero decreases. Glutathione combines with phytochelatin synthetase to form Phytochelatin (the trend of varying concentrations of phytochelatin is already observed) . Hence the way change of glutathione concentration affects the system will be similar to that of change in phytochelatin. But glutathione does not directly affect cadmium chelation as much as phytochelatin does. Greater the amount of chelating agent present in the cell, more cadmium is chelated at a particular time and as a result of which the amount of cadmium in the system falls more drastically with respect to time.This also affects the maxima of the curve i.e., greater phytochelatin concentration (as a result of greater glutathione concentration) implies that the slope $\frac{d[Cd]}{dt} = k_7[Cd]_{out} - k_5[phytochelatin][Cd]$ becomes negative at a lower time and for a lower concentration of internal cadmium.(just like the previous graph).

CONCLUSION:

- With the help of our bifurcation plots we were able to establish that the genetic circuit has two stable steady states i.e., inactive state at metal concentrations lower than the threshold value and active state at a concentration much higher than the threshold.
- We formulated differential equations which were simulated to visualise the chelation of cadmium in yeast cell in the form of concentration of cadmium (both inside and outside the cell) with respect to time. We observed the trends for different initial concentrations of cadmium.
- We have also varied the concentrations of glutathione and phytochelatin to see how it affects the internal concentration of cadmium in the cell with respect to time.

We have, therefore, verified the working of this theoretical model by establishing that the concentration of the cadmium after a practically feasible time reaches zero after undergoing chelation and losing its potency as a poison.

POSSIBLE IMPROVEMENTS TO THE MODEL:

From visualising and simulating the change in cadmium concentration in the soil with respect to time, we can also make predictions about the amount of yeast required per unit volume of soil to achieve maximum results.

Bifurcation plots an be used a way to experimentally measure the threshold concentration of cadmium required to Activate the gene circuit. *(This is achieved by plotting different steady state concentrations of cadmium for different initial metal concentrations and the concentration at which we see a sharp discontinuity is the threshold concentration).*