ANNEXURE 1

We had performed some computational studies to try and predict the enzyme kinetics of our reaction. These results are not conclusive but our findings suggest that it most likely follows the ordered bisubstrate reaction of enzyme kinetics. However, we have modelled for both ordered bisubstrate and random bisubstrate enzyme kinetics in our mathematical model. We undertook the following steps in our computational studies.

a. To predict the structure of our enzyme SulX, the I-TASSER server was used. The predicted structure was aligned with HsaA[6] as this was the protein used to homology model our enzyme in the literature[1][2][3] (PDB ID: 3AFF).
b. Protein-ligand docking for the same protein was performed for sulfonamide, FMNH2, degradation product (4-aminophenol), and FMN using Autodock VINA[4]. The pose (orientation) with the highest affinity and closest to the site predicted in the literature was used.
c. Molecular Dynamics simulations were performed for the SulX sulfonamide complex.

RESULTS AND FINDINGS

Computational methods were used to predict the structure of the protein (SulX) which has not been solved before. The protein structure that was predicted using the I-TASSER[8][9][10] server was observed to have a good overlap with HsaA with RMSD = 1.4. The corresponding substrate binding sites in the sequence were used to determine the grid box for docking studies and the predicted structure was used for molecular dynamics (MD) simulations.

a) Docking Studies

Docking was done using Autodock VINA[11]. The pose with the highest affinity and closest to the site predicted in the literature was used.

TABLE 1: Affinity value for the substrates and the the degradation product of the drug towards SulX

Substrate	Pose	Affinity (kcal/mol)
Sulfamethazine	1	-7.4
4-aminophenol	1	-4.6
FMNH2	4	-7.7
FMN	4	-6.8

TABLE 2: SASA (Solvent Accessible Surface Area) values for the two substrate-binding sites of SulX

Substrate	SASA value
FMNH2	135.298 $Å^2$
Sulfonamide	85.035 $Å^2$

b) Molecular Dynamics

MD was performed using DESMOND[12] for ligand and protein for 1.16ns where sulfonamide was taken as ligand and SulX was taken as protein. The position of sulfonamide used as predicted by the docking studies.

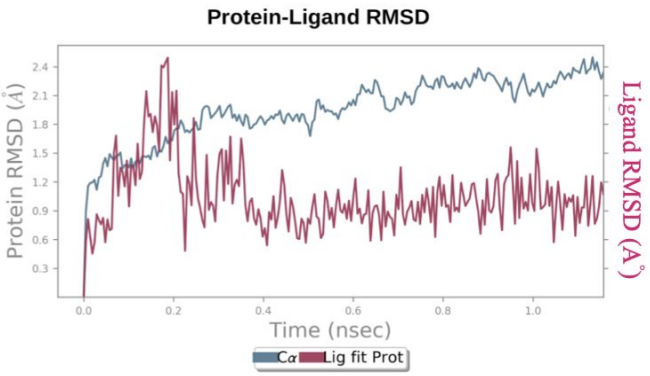


FIGURE 3: protein-ligand rmsd plot.

Docking and molecular dynamics studies have shown that there is a clear decrease in the affinity of the active site of the sulfonamide for the degradation product 4-aminophenol (Table 1). This suggests that the degradation product is not held as tightly as the drug itself. The SASA values obtained were found to be significantly higher for $FMNH_2$ than sulfonamide (Table 2). We hypothesize that the bi-substrate mechanism which SulX follows is most likely an ordered bi-substrate mechanism as $FMNH_2$ has more access to its active site than the drug. However, further studies will be performed to confirm the reaction mechanism. The ligand RMSD values from the simulation (Figure 3) are significantly lesser than the protein RMSD values for most of the simulation indicating that the ligand has not diffused away from the active site.

1. Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., & Zhang, Y. (2015). The. I-TASSER Suite: protein structure and function prediction. *Nature methods*, *12*(1), 7-8.
2. Roy, A., Kucukural, A., & Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nature protocols*, *5*(4), 725-738.
3. Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC bioinformatics*, *9*(1), 40.
4. Trott, O., & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of computational chemistry, 31(2), 455-461.
5. Kevin J. Bowers, Edmond Chow, Huafeng Xu, Ron O. Dror, Michael P. Eastwood, Brent A. Gregersen, John L. Klepeis, István Kolossváry, Mark A. Moraes, Federico D. Sacerdoti, John K. Salmon, Yibing Shan, and David E. Shaw, "Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters," Proceedings of the ACM/IEEE Conference on Supercomputing (SC06), Tampa, Florida, November 11–17, 2006
6. Kim, Dae-Wi, et al. (2019). "A novel sulfonamide resistance mechanism by two-component flavin-dependent monooxygenase system in sulfonamide-degrading actinobacteria." *Environment International* 127: 206-215.

ANNEXURE 3

Implementation of Coli Kaze

In this section we discuss some supplementary information for the proposed implementation of our project.

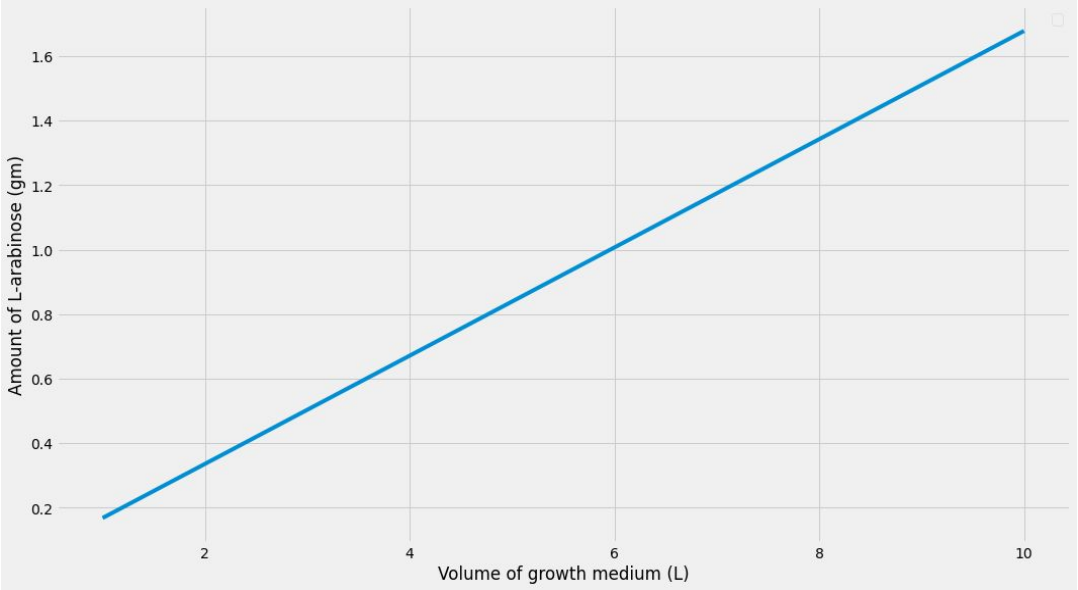
From survey[1]:
Dose: sulfamethazine (SMZ) given to each swine per day in finishing phase 0.069gm/day-swine
Assuming 90% recovery of SMZ as upper limit 0.063gm of SMZ is excreted by a swine in a day.
Assuming a swine flushes out 200 gm of solid excreta per day the, 0.315mg of sulfamethazine is recovered from 1 gram of swine excreta.For X gm of Swine excreta, estimated recovery of sulfamethazine is 0.315X mg (S')

From module 1: Antibiotic degradation
Sulfonamide monooxygenase is assumed to degrade 99.95% of its substrate.
After degradation 99.95% of sulfamethazine, the final conc. to be released, i.e. 0.05% of S' is 0.0001575mg (S)

From module 3: Kill Switch and DNA Degradation

Graph (1) : Implementation Phase II kill switch induction

Amount of L-arabinose added to media Vs Volume of Growth Medium



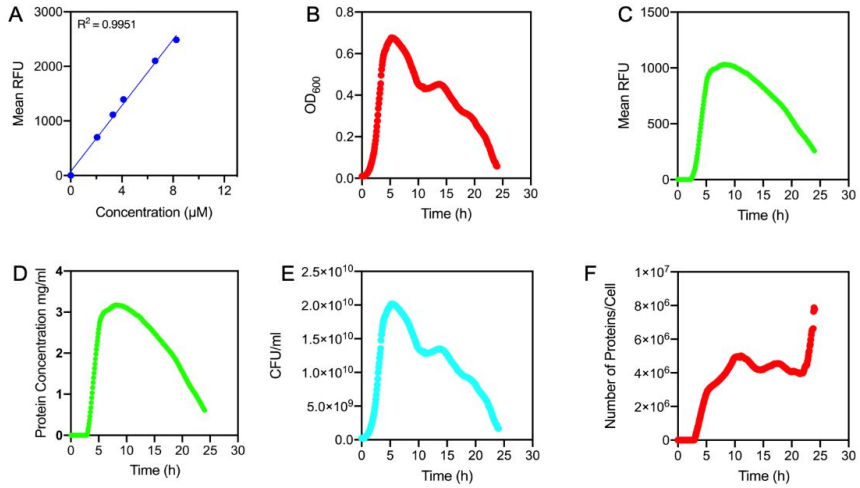
Equation(1):

$$A(gm) = 0.16783.V'(L)$$

A= amount of L arabinose added to growth medium for Pbad induction (in grams)
V'= volume of Coli Kaze growth medium (in Litres)

From Module 1 model :
Time required for enzyme: sulfonamide monooxygenase to degrade all sulfonamide in the slurry.
Values of K_m, $K_{cat.}$, α_1 & α_2 taken from module 1 model.
For V_m the total enzyme needed has to estimated from:

Characterisation of the strength of the constitutive promoter BBa_J23100[2][3]



The graph D shows that the maximum protein concentration is approximately 3mg/ml after 10 hours. The protein concentration (mg/ml) showed a similar pattern as observed in the fluorescence level (graph C). Subsequently they calculate the protein concentration in mg/CFU which is a better unit than mg/ml.

1D = 1.66054 X 10^{-21}mg, Molecular weight of eGFP = 28.179kDa = 28179Da = 4.677 x 10^{-17}mgmg/CFU = (mg/ml)/(CFU/ml)

Conclusion

Each cell has roughly 5 x 10^6 protein molecules after 10 hours (We should consider this). But it has 8 x 10^6 protein molecules after a period of 24 hours (This data is not reliable due to high uncertainties and errors). Considering an OD of 0.5 (= 4 x 10^8 cells/ml) at that point, total protein would be 2 x 10^{15}/ml. This is roughly 9.35 x 10^{-2}mg/ml. This means that a 1L culture of 0.5 OD would give around 93.5mg of eGFP.Hence, amount of any protein X (0.5 OD, 1L) = [93.5mg x 28.179/ Molecular weight of protein X (in kDa)] mg.For SulX, if we use this strong promoter, we would get around 57.34mg of protein (0.5 OD, 1L) in *E. coli*.

For SulX, if we use the medium strength promoter which has a really low metabolic burden value, we would get around 32.11mg of protein in *E. coli*.

Hence V_m=$K_{cat.}$(Total Enzyme)

= 20min^{-1}(0.683uM)

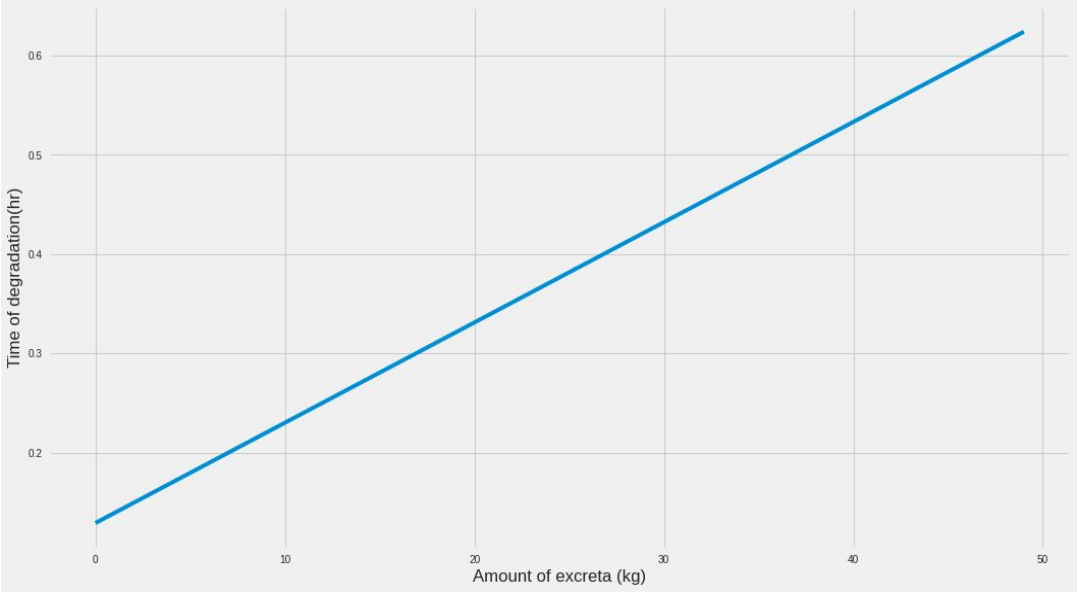
=13.66 uM/min

Equation:

$$t = \frac{(S' - S) + K_m(\alpha_1 + 1)\ln(\frac{S'}{S})}{\alpha_2 V_m}$$

Graphs (2) : Implementation Phase III Incubation

Time for degradation of SMZ Vs Amount of excreta being processed



Equation 2:

$$t(hrs) = 0.0005X(Kg.) + 0.124$$

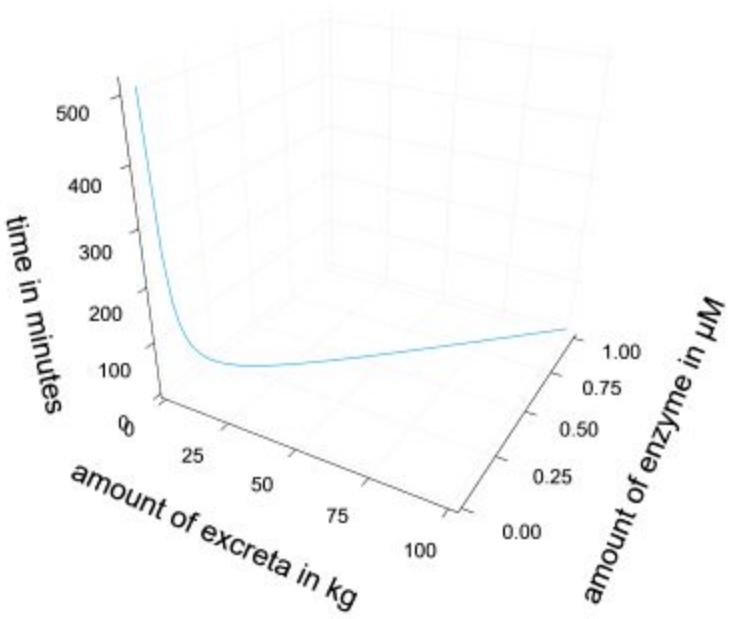
S'= 1015 uM
S= 0.50 uM
K_m= 16.83uM
V_m=13.66 uM/min
α_1= 0.8
α_2= 0.7

Modified Equation(2)

$$t = \frac{(S'-S)+K_m(\alpha_1+1)\ln(\frac{S'}{S})}{\alpha_2 K_{cat}.E}$$

For graph (3): Implementation Phase III Incubation

To estimate time of degradation Vs Amount of enzyme Vs the amount of excreta:



Equation:

$$t(hrs) = 0.02036\left[\frac{X(Kg))}{E(\mu M)}\right] + 5.085$$

Here E = total enzyme (uM)

For
S'= 1015 uM
S= 0.50 uM
K_m= 16.83uM
$K_{cat.}$= 20 min^{-1}
α_1= 0.8
α_2= 0.7

References:

[1] Krishnasamy, V., Otte, J. and Silbergeld, E., 2015. Antimicrobial use in Chinese swine and broiler poultry production. *Antimicrobial Resistance and Infection Control*, 4(1).

[2] 2019.igem.org. 2020. *Team:Newcastle/Results/Bronze characterisation - 2019.Igem.Org*. [online] Available at: <https://2019.igem.org/Team:Newcastle/Results/bronzecharacterisation> [Accessed 26 October 2020].

[3] Parts.igem.org. 2020. *Part:Bba J23100 - Parts.Igem.Org*. [online] Available at: <http://parts.igem.org/Part:BBa_J23100> [Accessed 26 October 2020].

ANNEXURE 2

1*

To determine a_1, the transcription rate per gene, we know the following:-

Average rate of m-RNA elongation= 39 nucleotides per second [1]
Average distance between adjacent RNA polymerases= 135 nucleotides [2]
Size of *araC* mRNA= 873 nucleotides.

Number of RNA polymerases per mRNA of this size
= (size of mRNA/ dist bw adjacent RNA polymerases)
= 6.467 arbitrary units

a_1 = (number of RNA pol on mRNA*elongation rate)/size of gene

Hence we calculate the number of mRNAs produced per gene copy per unit time to be 0.289 mRNAs per gene per second.

2*

T0 determine a_2, we used the salislab software [3], by entering our mRNA sequence with the RBS. we got the following:-

Translation initiation units (TIU) = 5.45 arbitrary units
Protein elongation rate = 22.91nuc/s
Size of *araC* mRNA= 873 nucleotides.

a_2 = (TIU*protein elongation rate)/size of mRNA

Hence we calculate the number of AraC proteins produced per mRNA per unit time to be 0.143 molecules per mRNA per second

3*

To determine the d_1, the degradation rate of mRNA in the cell, we use:-

Total number of mRNAs present in a cell = 10000 copies [4]
Total number of mRNAs degraded in a cell = 0.003 copies per second [4]

d_1= total number of mRNAs degraded/total number of mRNAs present in cell

Hence we get the degradation rate of 1mRNA to be 0.0000003 per second

4*

To determine the d_2, the degradation rate of protein in the cell, we use:-

Total number of proteins present in a cell = 2320000 copies [4]
Total number of proteins degraded in a cell = 0.000028 copies per second [4]

d_2= total number of proteins degraded/total number of proteins present in cell

Hence we get the degradation rate of 1mRNA to be 0.000000000012 per second

5*

To determine the k_{elong} of *DNASE1* gene, we use:-

Average elongation rate = 39 nucleotides per second [1]
Size of gene= 843 nucleotides

k_{elong} = average elongation rate/ size of gene

Hence we get the elongation rate as 0.0462 mRNAs per second

6*

To determine the η_1, we have [7]:-

η_1= [RP]/ total gene copies

$R+\theta_2 \rightarrow [R\theta_2]^* \rightarrow RP$

$[RP]= \theta_2(1-e^{\wedge}(-k_{obs}t))$

$k_{obs}= (1/k_2)+(k_d/k_2[R])$

Where k_d = 0.3 M = 260000000 molecules per cell**, k_2=0.167 second inverse [6]
[R]= average number of RNA polymerase molecules in a cell = 4486.39 molecules per cell** [7]

7*

To determine b_2, we use the following:-

Translation initiation units (TIU) = 25.4 arbitrary units
Protein elongation rate = 0.02 nuc/s
Size of *araC* mRNA= 843 nucleotides

b_2 = (TIU*protein elongation rate)/size of mRNA

Hence we calculate the number of DNASE1 proteins produced per mRNA per unit time to be 0.0006026 molecules per mRNA per second.

8*

Initial amount of AraC molecules in cell is calculated by:-

Average number of proteins produced per gene copy = 47
Number of gene copies of *araC* = 225

Initial number of AraC molecules
= average number of proteins produced per gene copy*number of gene copies
= 47*225 = 10575 molecules of protein per cell

9*

The maximum number of c-AMP molecules per bacterial cell can be 89727.8 molecules per cell ** [9]
This can happen at any minimal concentration of external glucose because,
at t=3600s (graph not shown), the external concentration of glucose corresponding to this value of c-AMP is way more than normal, 480000 moles per liter (which is impractical), and any concentration of glucose below this, produces more c-AMP than the threshold amount (89727.8), so we assume that c-AMP in the cell is constant(i.e. 89727.8 molecules) in the presence of even very small amounts of glucose.

10*

The total number of gene copies of both *araC* and *DNASE1* is taken to be 225 because each set of these genes are present in both the plasmids we are introducing in the bacteria. The average copy number of plasmids of type 1 is 200 and that of type 2 is 25, hence a total average of 225 is assumed.

11*

The total number of phosphodiester bonds present in a bacterial cell is given by:-
(number of base pairs-1)*2 in case of linear strands and (number of base pairs*2) for plasmids.
Our bacteria is assumed to have a single long linear chromosome of length 4600000 base pairs and plasmid 1 (200 copies) of length 6893 base pairs and plasmid 2 (25 copies) of length 6586 base pairs.

Hence the total number of phosphodiester bonds in a bacterial cell will be= 12286498 numbers per cell

12**

But by the time this happens, the genes coding for AraC and DNASE1 protein also would have degraded. As the time taken for degradation is very less, the degradation of the genes would hardly matter. But to be accurate and develop a rigorous mathematical model, we do the following,

According to the [10]:-

The mass distribution of DNA fragments at a given interval depending upon the amount and time of UV radiation it is exposed is given by:-

$$F(M)dM = \left(\frac{\mu}{M0}[2+\mu(1+\frac{M}{M0})]^{-\mu\frac{M}{M0}} + 2^{-\mu*\delta(M0-M)}\right) dM$$

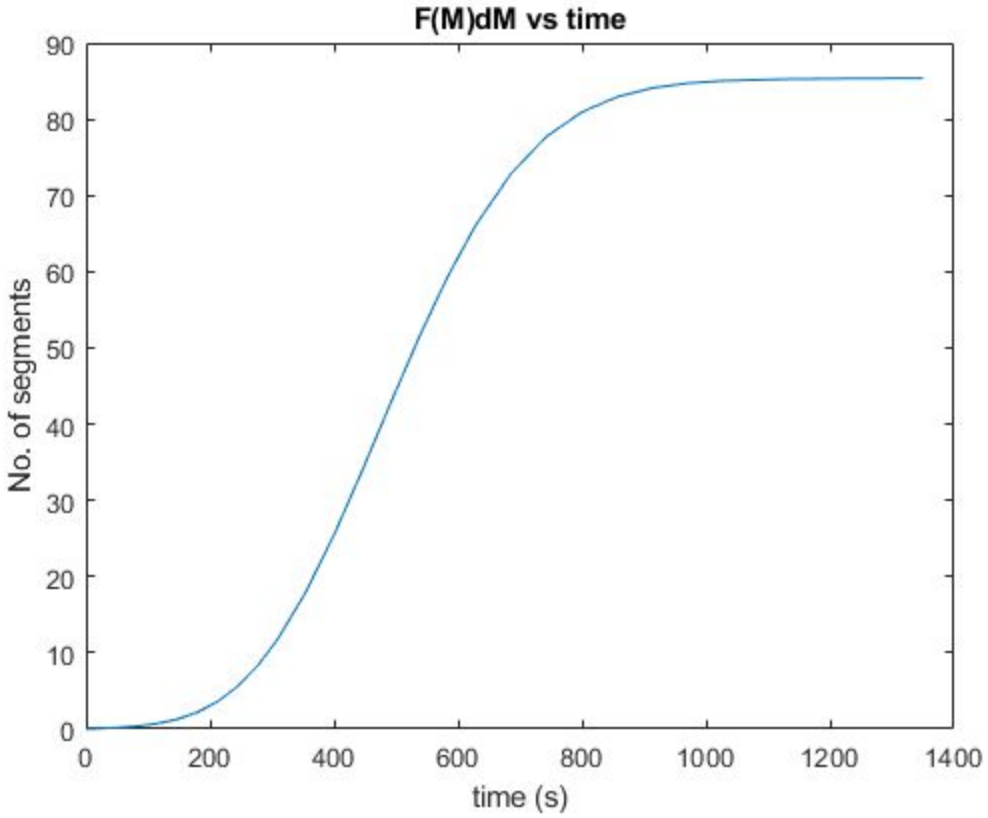
$$F(M) = number\ of\ fragments\ with\ mass\ ranging\ from\ M\ to\ M+dM$$
$$\mu = r\ (actually\ =\ no.\ of\ breaks = M0 * dose\ of\ UV * yield)$$
$$M0 = initial\ mass\ of\ fragments$$
$$\delta(M-M_0) = dirac\ function = \lim_{a\to 0} \frac{e^{-(\frac{M-M_0}{a})^2}}{a\sqrt{\pi}}$$

We take M to be 5074 (number of base pairs on *araC and DNASE1* gene together) and r is dS/dt. And obtain:-

$$F(M)dM = \left(\frac{r}{9200000}[2+r(1.00028)].e^{-0.00028r} + 2.e^{-r*\delta(9197417)}\right) dM$$
$$+ 200\left(\frac{r}{13784}[2+r(1.1874)].e^{-0.1874r} + 2.e^{-r*\delta(11210)}\right)dM$$
$$+ 25\left(\frac{r}{13170}[2+r(1.1961)].e^{-0.1961r} + 2.e^{-r*\delta(10587)}\right)dM$$

Each term for 1 long chromosomal DNA, 200 copies of plasmid 1 and 25 copies of plasmid 2.

We plot F(M)dM vs time:-



Inference :- maximum concentration of segments of length 5074 is attained at 1025 seconds. So we assume that only after this point of time, our *araC* and *DNASE1* genes start to degrade.

REFERENCES:-

** We take the volume of bacterial cell to be 1.49*10^-15 L [5] for conversion from molarity to molecules per cell.

[1] https://bionumbers.hms.harvard.edu/files/Growth-rate-dependent%20parameters.pdf
[2] https://bionumbers.hms.harvard.edu/file kos/Parameters%20pertaining%20to%20the%20macromolecular%20synthesis%20rates%20in%20exponentially%20growing%20E.%20coli%20Br%20as%20a%20function%20of%20growth%20rate%20at%2037%20degrees%20celsius.pdf
[3] https://salislab.net/
[4] https://bionumbers.hms.harvard.edu/files/Total%20number%20of%20transcripts%20and%20proteins%20and%20transcript%20and%20protein%20degradation%20rates.pdf

[5] https://bionumbers.hms.harvard.edu/bionumber.aspx?id=114924&ver=0&trm=volume+of+ecoli+cell&org=
[6] https://bionumbers.hms.harvard.edu/files/Kinetic%20parameters%20of%20open-complex%20formation.pdf
[7] Wong, P., Gladney, S., & Keasling, J. D. (1997). Mathematical model of the lac operon: inducer exclusion, catabolite repression, and diauxic growth on glucose and lactose. *Biotechnology progress*, *13*(2), 132-143.

[8] https://bionumbers.hms.harvard.edu/files/Summary%20of%20translational%20parameters.pdf
[9] Notley-McRobb, L., Death, A., & Ferenci, T. (1997). The relationship between external glucose concentration and cAMP levels inside Escherichia coli: implications for models of phosphotransferase-mediated regulation of adenylate cyclase. *Microbiology*, *143*(6), 1909-1918.
[10] Belli, M., Campa, A., Dini, V., & Esposito, G. (2000). A mathematical model of DNA fragmentation induced by gamma-rays. *Mathematics and Computers in Modern Sciences. World Scientific and Engineering Society Press, Singapore*, 183-186.