Part 1 Model objective

1.1 Introduction

The nepetalactol synthesis pathway is divided into two parts and delegated to *E. coli* and yeast separately to decrease the genetic-engineering per organism and take advantages of the unique properties of each organism. (Link to Project>>Design>>Co-culture) For our project, *E. coli* is in charge of producing the metabolic intermediate geraniol; then geraniol is translocated into yeast to be converted into nepetalactone.

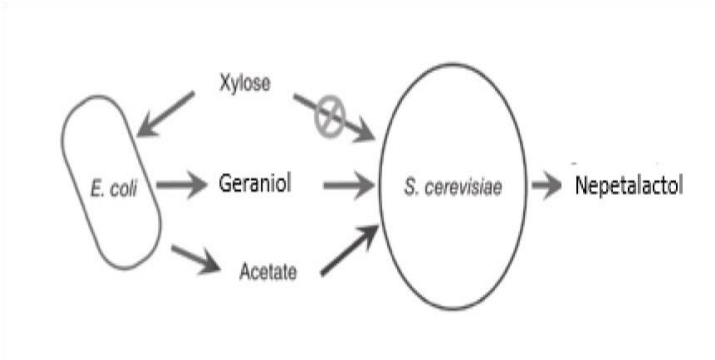

Figure 1. The co-culture system

Co-culture is achieved through using xylose as the sole carbon source, which can only be used by *E. coli* but not yeast. Upon growing on xylose, *E. coli* would produce acetate which serves as yeast's carbon source. Then yeast would consume acetate inhibitory to *E. coli* growth. Thus, the relationship is mutualistic: yeast relies on *E. coli* to convert xylose to acetate for growth; without yeast, *E. coli* growth would eventually halt with the accumulation of acetate.

Aside from the mutualistic relationship, the model would also feature the inhibition effect of the products, geraniol and nepetalactol. There is usually a trade-off between product formation and microbial growth, because increased level of production means additional toxicity to the producing hosts — the two microbial producers are self-preventing growth by production.

1.2 Exploration outline and aim

The co-culture model is consist of the growth functions of the two population, yeast and *E. coli;* the specific growth rate equations taking into account the effect of limited substrate and product inhibitions on growth rate; and four kinetics expressions for the substrate and products, nepetalactol, acetate, geraniol and xylose.

With the equations wrote out, a sensitivity analysis is conducted to determine how each parameter impact the outcome of the production. By interpreting the result and finding the production-determined factor, the system can be understood and modified.

Overall, we hope the unstructured model can model the co-culture system and help us determine the inoculation times and concentrations of *E. coli* and yeast and the optimal production rates of geraniol and nepetalctol.

2 General assumptions and parameters

To best analyze the growth of cells, we use unsegregated model, in which average cellular properties and balanced growth are assumed, each cell is deemed identical, the population is treated as one-component solute and cells of the same quality are produced without any variation of composition — at various stage of the co-culture, new cells are assumed to be similar with old cells.

The model is also unstructured, in the sense that the intermediate reactions are omitted, as well as the toxic effect of the metabolites from these reactions.

Parameters: For ease of reading, substances are represented by their capital initials:

Concentration of a substance (g/L)	**Symbol**
E. coli	E
S. cerevisiae, or yeast	Y
Geraniol	G
Nepetalactol	N
Acetate	A
Xylose	X

Table 1. Abbreviations for key substance concentration

Part 2 Co-culture model

2.1 Using logistic function for the basic population growth

By assuming *E. coli* and yeast follow logistic growth, we can obtain:

$$\frac{dE}{dt} = (\mu_E - d_E)E\left(1 - \frac{E+Y}{K}\right).$$
$$\frac{dY}{dt} = (\mu_Y - d_Y)Y\left(1 - \frac{E+Y}{K}\right).$$

Since the bioreactor would be accommodating two populations, *E. coli* and *S. cerevisiae's* total concentration combining together should not exceed the carrying capacity of the bioreactor, K, which is assumed to be non-time-varying. d_E and d_Y represent the specific cell death rate of *E. coli* and yeast cell.

2.2 Deriving specific growth rate based on Monod kinetics

Specific growth rate, μ, affected by the limited substrate and the toxicity of the product, is expressed using Monod kinetics.

a. Specific growth rate with limited substrate

By assuming the total substrate uptake follows Michaelis-Menten kinetics and the substrate is used for non-growth associated metabolic maintenance during stationary phase and microbial growth, the substrate uptake rate r_S is expressed as:

$$r_S = K_{max\,S/X} \cdot \frac{S}{K_{x-S}+S} \cdot X = \mu_{X/S} \cdot Y_{X/S} \cdot X + m \cdot X$$

In which $K_{max\ S/X}$ is the maximum substrate uptake rate by cells. S represents substrate concentration, X represents cell concentrate and Kx-s is the substrate affinity. With constant maintenance requirement assumed, m is the maintenance constant, $Y_{X/S}$ is the yield coefficient of the biomass to the substrate consumed.

Therefore, the specific growth rate is

$$\mu_{X/S} = \frac{1}{Y_{X/S}} \cdot \left(\frac{K_{max\,S/X} \cdot S}{K_{x-S} + S} - m\right)$$

And assuming S >> K_{x-S} the maximum growth rate should be:

$$\mu_{X/Smax} = \frac{1}{Y_{X/S}} \cdot (K_{max\,S/X} - m)$$

And the substrate affinity can be expressed as:

$$K_{max\ S/X} = Y_{X/S} \cdot \mu_{X/Smax} + m$$

Thus, the specific growth rate can be reformulated into:

$$\mu_{X/S} = \frac{1}{Y_{X/S}} \cdot \left(\frac{(Y_{X/S} \cdot \mu_{X/Smax} + m) \cdot S}{K_{x-S} + S} - m\right) = \left(\mu_{X/Smax} + \frac{m}{Y_{X/S}}\right) \cdot \frac{S}{K_{x-S} + S} - \frac{m}{Y_{X/S}}$$

The substrate consumption rate becomes:

$$r_S = K_{max\ S/X} \cdot \frac{S}{K_{x-S} + S} \cdot X = (Y_{X/S} \cdot \mu_{X/Smax} + m) \cdot \frac{S}{K_{x-S} + S} \cdot X$$

b. Product inhibition

In addition, nepetalactol and geraniol can both inhibit microbial growth, and acetate is inhibitory to *E. coli* growth. Therefore, they are taken into account of the specific growth rate expression, again using an extended Monod expression that accounts for multiple product inhibition effects.

1) The Ierusalimsky equation:

$$\mu = \mu_{max\ P/S} \cdot S' \cdot \frac{K_I}{K_I + I}$$

Where K_I a constant that describe the inhibition effect of the inhibitor, commonly measured as the concentration of the inhibitor under which half the population growth is halted. I is the concentration of the inhibitor.

2) Zeng and Deckwer 's multiple product inhibition equation:

$$\mu = \mu_{max\ P/S} \cdot S' \cdot \left[\prod\left(1 - \left(\frac{I}{I^*}\right)^n\right)\right]$$

S' represents the substrate uptake kinetics term.

With the presence of multiple inhibitors, K_I would be hard to measure since the effect of multiple inhibitors cannot be separated. In comparison,* , the critical inhibitor concentration with the presence of other inhibitors, is obtainable. Also, in Ierusaliminsky's equation, microbial growth would not stop until I, the inhibitor concentration reaches infinity, which is contradictory to the general observation that when a certain level of inhibitor concentration is reached, cell growth would stop.

Based on Zeng and Deckwer's equation, specific growth rates can be expressed with the multiplicative form of the inhibition terms K:

$$\mu_E = [(\mu_{E/Xmax} + \frac{m_{E-X}}{Y_{E/X}}) \cdot \frac{X}{K_{E-X} + X} - \frac{m_{E-X}}{Y_{E/X}}] \cdot I_{G-E} \cdot I_{N-E} \cdot I_{A-E} \cdot$$

$$\mu_Y = [\left(\frac{A}{G+A}\left(\mu_{Y/Amax} + \frac{m_{Y-A}}{Y_{Y/A}}\right) \cdot \frac{A}{K_{Y-A} + A} - \frac{m_{Y-A}}{Y_{Y/A}}\right) + \left(\frac{G}{G+A}\left(\mu_{Y/Gmax} + \frac{m_{Y-G}}{Y_{Y/G}}\right) \cdot \frac{G}{K_{Y-G} + G} - \frac{m_{Y-G}}{Y_{Y/G}}\right)] \cdot I_{G-Y} \cdot I_{N-Y}$$

In which geraniol's inhibition to *E. coli* is:

$$I_{G-E} = \begin{cases} (1 - \left(\frac{G}{I^*_{G-E}}\right))^{n_G}, & when\ G \leq I^*_{G-E}. \\ 0, & when\ G > I^*_{G-E} \end{cases}$$

The same are written for nepetalactol inhibition to *E. coli* (I_{N-E}) and to yeast (I_{N-Y}), geraniol's inhibition to yeast (I_{G-Y})and acetate's inhibition to *E. coli*

n is an empirically determined exponent.

2.3 Product formation

Considering product formation may be growth-associated and non-growth associated (produced at the stationary phase), the product formation is depicted by the Leudeking-Piert equation as:

$$\frac{dP}{dt} = \alpha_P \cdot \frac{dX}{dt} + \beta_P \cdot X$$

P represents product concentration, X is the cell concentration. S represents substrate concentration; α_P and β_P are the growth-associated and non-growth associated yield coefficients for the product.

2.4 Xylose, acetate, geraniol and nepetalactol concentration

Considering substrate utilization, product formation, in which nepetalactol, geraniol and acetate productions take place during both growth and the stationary phase, and degradation kinetics:

Process	**Explanation**	**Equations**
Xylose		
Substrate utilization	Xylose is utilized as substrate by *E. coli*	$(\mu_{E/X} \cdot Y_{E/X} \cdot E + m_{E-X} \cdot E)$

Degradation		$-D_X \cdot X$
Acetate		
Substrate utilization	Acetate is utilized as substrate by yeast and it is assumed to not be utilized by *E. coli*	$-(\mu_{Y/A} \cdot Y_{Y/A} \cdot Y + m_{Y-A} \cdot Y)$
Product formation	Acetate is produced by *E. coli*	$+\alpha_A \cdot \frac{dE}{dt} + \beta_A \cdot E$
Degradation		$-D_A \cdot A$
Geraniol		
Substrate utilization	Geraniol is converted to nepetalactol and metabolize into side products when during yeast's detoxification while *E. coli*' s capability of converting of geraniol to other metabolites is assumed to be minimal and thus not accounted for	$-(\mu_{Y/G} \cdot Y_{Y/G} \cdot Y + m_{Y-G} \cdot Y)$
Product formation	Geraniol is produced by *E. coli*	$+\alpha_G \cdot \frac{dE}{dt} + \beta_G \cdot E$
Degradation		$-D_G \cdot G$
Nepetalactol		
Product formation	Nepetalactol is produced by yeast	$+\alpha_N \cdot \frac{dY}{dt} + \beta_N \cdot Y$
Degradation		$-D_N \cdot N$

Table 2. Explanations for dXdt, dAdt, dGdt, and dNdt.

We derive:

$$\frac{dX}{dt} = -\ (\mu_{E/X} \cdot Y_{E/X} \cdot E + m_{E-X} \cdot E)\ - D_X \cdot X$$

$$\frac{dA}{dt} = -\ (\mu_{Y/A} \cdot Y_{Y/A} \cdot Y + m_{Y-A} \cdot Y)\ + \alpha_A \cdot \frac{dE}{dt} + \beta_A \cdot E\ - D_A \cdot A$$

$$\frac{dG}{dt} = -\ (\mu_{Y/G} \cdot Y_{Y/G} \cdot Y + m_{Y-G} \cdot Y)\ + \alpha_G \cdot \frac{dE}{dt} + \beta_G \cdot E\ - D_G \cdot G$$

$$\frac{dN}{dt} = \alpha_N \cdot \frac{dY}{dt} + \beta_N \cdot Y \quad - D_N \cdot N$$

Part 3 Model analysis

Parameters	Range	Notes	Matlab representation	Unite
K	500	The capacity of the bioreactor (Constant)	K	g/L
$\mu_{maxX/S}$: Maximum growth rate of X on substrate S				
$\mu_{max\ E/X}$	0.1-2.0		rE_X	h^{-1}
$\mu_{max\ Y/G}$	0.087-.0.200		rY_G	
$\mu_{max\ Y/A}$	0.1-2.0		rY_A	
K_{X-S} : Substrate affinity of X's growth on substrate S				
K_{E-X}	0.01-3		KE_X	g/L
K_{Y-A}	0.01-3		KY_A	
K_{Y-G}	0.01-3		KY_G	
I^*_{P-X} : Critical inhibition concentration of product P on X				
I^*_{G-E}	10-100		IG_E	g/L
I^*_{N-E}	10-100		IN_E	
I^*_{A-E}	10-100		IA_E	
I^*_{G-Y}	10-100		IG_Y	
I^*_{N-Y}	10-100		IN_Y	
$Y_{X/S}$ Yield coefficient of X on substrate S				

[1] Simultaneous utilization of glucose, xylose and arabinose in the presence of acetate by a consortium of *Escherichia coli* strains | Microbial Cell Factories | Full Text https://microbialcellfactories.biomedcentral.com/articles/10.1186/1475-2859-11-

[2] Engineering the biocatalytic selectivity of iridoid production in *Saccharomyces cerevisiae* John M. Billingsleya, Anthony B. DeNicolaa, Joyann S. Barberb, Man-Cheng Tanga, Joe Horeckac,d, Angela Chuc,d, Neil K. Gargb, Yi Tanga,b,

$Y_{E/X}$	0.001-0.030		YE_X	g-substrate/ g-cells or gproducts
$Y_{Y/A}$	0.001-0.030		YY_A	
$Y_{Y/G}$	0.001-0.030		YY_G	
m_{X-S}: Maintenance coefficient for X on substrate S				
m_{Y-G}	0-0.2		mE_X	
m_{E-X}	0-0.2		mY_A	(g-substrate/ g-cells)/ h^{-1}
m_{Y-A}	0-0.2		mY_G	
α_P: growth associated yield coefficient of product P, β_P non growth associated yield coefficient of product P.				
α_A	0-1.5		aA	$g(g)^{-1}$
β_A	0-1.5		bA	$g\ (gh)^{-1}$
α_G	0-1.5		aG	$g(g)^{-1}$
β_G	0-1.5		bG	$g\ (gh)^{-1}$
α_N	0-1.5		aN	$g(g)^{-1}$
β_N	0-1.5		bN	$g\ (gh)^{-1}$
D_X: Degradation r ate of substance X				
D_G	0.001	Constant	DG	h_{-1}
D_N	0.001	Constant	DN	
D_X	0.001	Constant	DX	
D_A	0.001	Constant	DA	
k_X: Specific death rate of X				
k_E	0.001-0.200		kE	h^{-}
k_Y	0.001-0.200		kY	

Table 3. Parameters and their ranges

Our team also perform a global sensitivity analysis to identify critical parameters affecting and controlling our co-culture behavior. We numerically integrated our ODE with 1,000 sets of parameter values sampled from uniform distribution, with the adaptation of Latin hypercube selection. For each parameter, partial rank correlation coefficients (PRCC) and p-value (Significance) were calculated with six output metrics, dAdt, dEdt, dGdt, dNdt, dXdt and dYdt, whose integration derives acetate concentration, *E. coli* concentration, geraniol concentration, nepetalactol concentration, xylose concentration and yeast concentration. The most significant parameters (|PRCC|>0.4 and p-value<0.05) are highlighted and recognized as the most influential factors in our co-culture system.

Latin Hypercube Sampling (LHS) technique is first used to create a matrix (1000x25) with 1000 sets (samples) of 25 parameters whose values are randomly generated in the ranges we set. The LHS procedure is implemented by dividing the range of values for each parameter into equally probable intervals. The LHS scheme is a so-called stratified scheme whereby probability distributions are assigned to parameters, the intervals in the distribution are divided into equally probable regions, and these intervals are then each sampled without replacement. The LHS method assumes the sampling is performed independently for each parameter and the sampling is done by randomly selecting values from each probability distribution function.

Parameter values are sampled from either a normal distribution or uniform distribution. For a normal distribution:

$$a_{i,j} = F^{-1}(p_{i,j} ; \mu_j ,(c_{v,j}*\mu_j))$$

where $a_{i,j}$ is the value of parameter j in LHS sample i, $F^{-1}(p_{i,j} ; \mu_j ,(c_{v,j}*\mu_j))$ is the normal inverse cumulative distribution function, $p_{i,j}$ is the LHS probability for parameter j in sample i, μ_j is the mean value of parameter j, and $c_{v,j}$ is the coefficient of variation for parameter j (with $\sigma = c_{v,j} * \mu_j$).

For a uniform distribution:

$$a_{i,j} = a_{j,min} + p_{i,j} (a_{j,max} - a_{j,min})$$

where $a_{j,min}$ is the lower bound on parameter j, $a_{j,max}$ is the upper bound of parameter j, and other terms are as described above.

Since we want to investigate the effect of each parameter on the outputs at any time point in the simulation, our ODE Model was numerically integrated with each set of sampled parameter values, and the output metrics dAdt, dEdt, dGdt, dNdt, dXdt and dYdt are calculated. We also employs ode15s with time span from 0 to 50 (with interval of 1) to record the output values at any time point throughout the simulation. Six output metrics (50*1000) are generated, with each row representing the output values at a particular time point and each column representing the output values for a sample of parameters' values.

The LHS Matrix of parameters' values and the six output metrics are then used to carry out global sensitivity analysis. In biology, input factors are often very uncertain and therefore local sensitivity analysis techniques are not appropriate for a quantitative analysis, instead global sensitivity analysis techniques are needed. Specifically, we adopt partial rank correlation coefficient (PRCC) to characterize the relationship between input parameters and the outputs. PRCC is a robust sensitivity measure for nonlinear but monotonic relationships between parameters and outputs. By combining the uncertainty analyses with PRCC, we are able to reasonably assess the sensitivity of our outcome variable to parameter variation. To perform PRCC, we first calculate correlation coefficient. A correlation coefficient between x j and y is calculated as follows:

$$r_{xjy} = \frac{\mathrm{Cov}(xj; y)}{\sqrt{\mathrm{Var}(xj)\mathrm{Var}(y)}} = \frac{\sum_{i=1}^{N}(x_{ij}-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{N}(x_{ij}-\bar{x})^2(y_i-\bar{y})^2}} \qquad j=1,2,\ldots k.$$

Cov (x_j, y) represents the covariance between x_j and y, while Var(x_j) and Var(y) are respectively the variance of x_j and the variance of y ($\bar{x}$ and $\bar{y}$ are the respective sample means).

Then, we perform the PRCC between x j and y, which is the partial correlation coefficient between the two residuals ($x_j - \hat{x}_j$) and ($y - \hat{y}_j$)) on rank-transformed data, where $\hat{x}$ and $\hat{y}$ are the following linear regression models:

$$\hat{x}_j = c_0 + \sum_{\substack{p=1 \\ p \neq j}}^{k} c_p x_p \quad \text{and} \quad \hat{y}_j = c_0 + \sum_{\substack{p=1 \\ p \neq j}}^{k} b_p x_p$$

In addition to PRCC, significance tests are performed to assess the level of uncertainty an LHS parameter contributes to the model. Significance tests also show if a PRCC is significantly different from zero (Even small correlations may be significant) and if two PRCC values are significantly different from each other. Thus, the significance of the PRCC value of a parameter indicates that parameter's contribution to the model's prediction imprecision. The parameters with large PRCC values (|PRCC|>0.4) and corresponding small p-values (< 0.05) are deemed the most influential in our model.

The results of our PRCC are shown in the diagrams below, with y axis demonstrating the time span of the whole simulation and x axis showing different parameters.

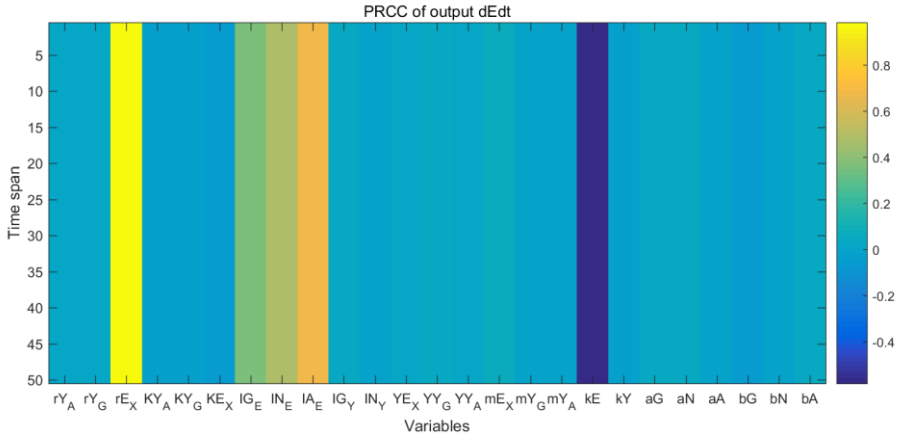

Figure 2. PRCC between 25 model parameters and the output of *E. coli* concentration at particular time points (Hours) throughout simulation.

In Figure 2, the categories of variables from left to right are maximum specific growth rates, substrate affinities, critical inhibitions, yield coefficients, maintenance coefficients, specific death rates and LP coefficients respectively. It is notable that the maximum specific growth rate of *E. coli* on xylose (rE_X) has a partial rank correlation coefficient of 0.983 and a p-value of zero, which reveal that the parameter has a strong effect on the *E. coli* concentration with a high degree of certainty. As revealed by the unchanged color intensity of every column, the sensitivity of each parameter remains almost constant throughout the whole simulation, measured at intervals of 1 hour from 0 hour to 50 hours. In addition to *E. coli* concentration, other outputs like xylose concentration (Figure 3) and nepetalactol concentration (Figure 4) demonstrate the same pattern throughout the simulation.

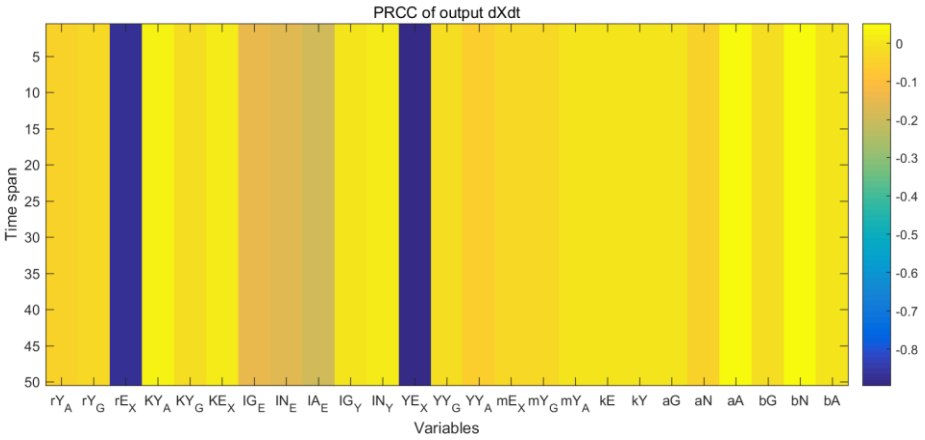

Figure 3. PRCC between 25 model parameters and the output of xylose concentration at particular time points (Hours) throughout simulation.

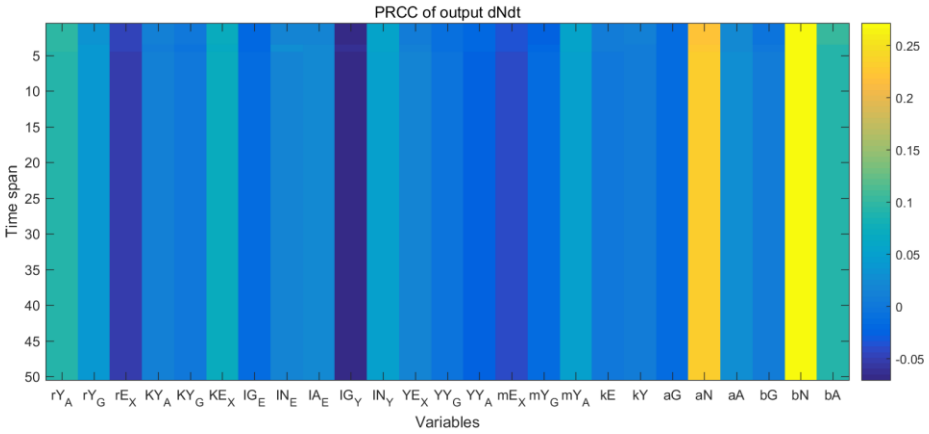

Figure 4. PRCC between 25 model parameters and the output of nepetalactol concentration at particular time points (Hours) throughout simulation.

In spite of the fact that we have varied the length of time span for the simulation, the sensitivities of the parameters remain unchanged. Therefore, we combine the PRCC between the 25 parameters and the 6 outputs into a single diagram which demonstrate the relationships between every input parameter and outputs throughout the simulation (Figure 5).

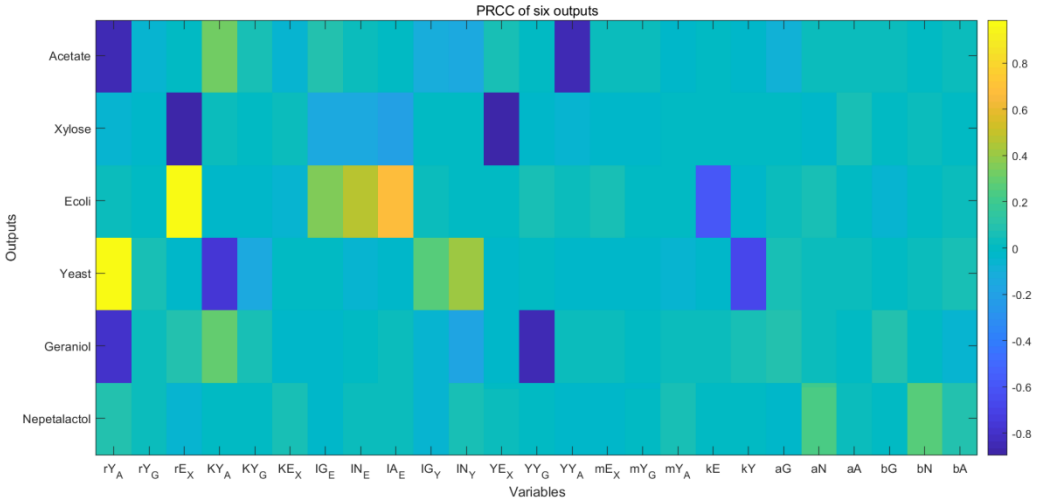

Figure 5. PRCC between 25 model parameters and the six outputs throughout simulation.

Several sensitive parameters are identified in Figure 5 and their significance is further considered. For acetate concentration, it is strongly negatively affected by the maximum specific growth rate of yeast on acetate and the yielding of yeast biomass on acetate, and is also positively affected by the substrate affinity between yeast and acetate, which has a PRCC of 0.294. For xylose concentration, no positive relationship between parameters and the output is identified. The maximum specific growth rate of *E. coli* on xylose and the yielding of *E. coli* biomass on xylose are the most sensitive parameters to xylose concentration, with PRCC of -0.882 and -0.893. For *E. coli* concentration, the maximum specific growth rate of *E. coli* on xylose (rE_X), inhibition of acetate, nepetalactol and geraniol on *E. coli* growth(IA_E, IN_E, IG_E) are the most positively significant parameters. It is notable that the maximum specific growth rate of *E. coli* on xylose has a very high partial rank correlation coefficient of 0.983 with a p-value of zero, which means the prediction of the high sensitivity of this parameter is very precise and concrete. Whereas the p-values of inhibition of acetate, nepetalactol and geraniol on *E. coli* growth are 0.392, 0.584 and 0.444 respectively, which are relatively high compared with the standard 0.05. The high p-values weaken the prediction of the strong relationship indicated by the PRCC between these three parameters and the *E. coli* concentration. For the negative relationship with *E. coli* concentration, the specific death rate of *E. coli* (kE) has a PRCC of -0.60 and a p-value of 2.33*e-16, which is also considered as significant. For yeast concentration, the inhibition of nepetalactol on yeast growth (IN_Y) has a relatively significant impact with the PRCC of 0.425. Since the p-value and the PRCC of the maximum specific growth rate of yeast on acetate (rY_A) are zero and 0.891 respectively, rY_A is considered to have the strongest positive relationship with the yeast concentration among all the parameters tested. Substrate affinity between yeast and acetate (KY_A) and the specific death rate of yeast (kY) has significant negative effect on the yeast concentration, with a PRCC of -0.762 and -0.71 respectively.

Since the production of geraniol and nepetalactol is very crucial and harsh, we consider the parameters with |PRCC|>0.2 and p-value<0.05 as being influential. For the concentration of Geraniol, p-values of rY_A, KY_A and YY_G are very close to zero with p-values of 7.89e-225, 2.07e-20 and 1.79e-282 and PRCC of -0.843, 0.289 and 0.857 respectively. This demonstrates a strong negative relationship between parameter rY_A and the concentration of Geraniol, and also a strong positive relationship between parameter YY_G and the concentration of Geraniol.

For the concentration of Nepetalactol, p-values of Luedeking-Piert coefficients of nepetalactol production (aN and bN) are 9.06e-20 and 1.89e-20 and the PRCC of aN and bN are 0.286 and 0.291 respectively, which reveal a significant positive relationship between the individual parameter and the concentration of Nepetalactol.

Part 4 Results and conclusion

Experimental validation of our co-culture model.

To study the ecology between the *E. coli* and yeast and verify our models, we experimentally characterized the co-culture of *E. coli* and yeast under xylose, measuring the proportion of between these two populations. The results we obtain match with our model simulations and provide insights into the values of our key parameters. It is reasonable that the population of *E. coli* outweighed the population of yeast as shown in figure 1, since the *E. coli* has shorter double time. We have

calculated the specific growth rate of *E. coli* based on the data we derive from our experiments and modify the relevant parameters in our model accordingly.

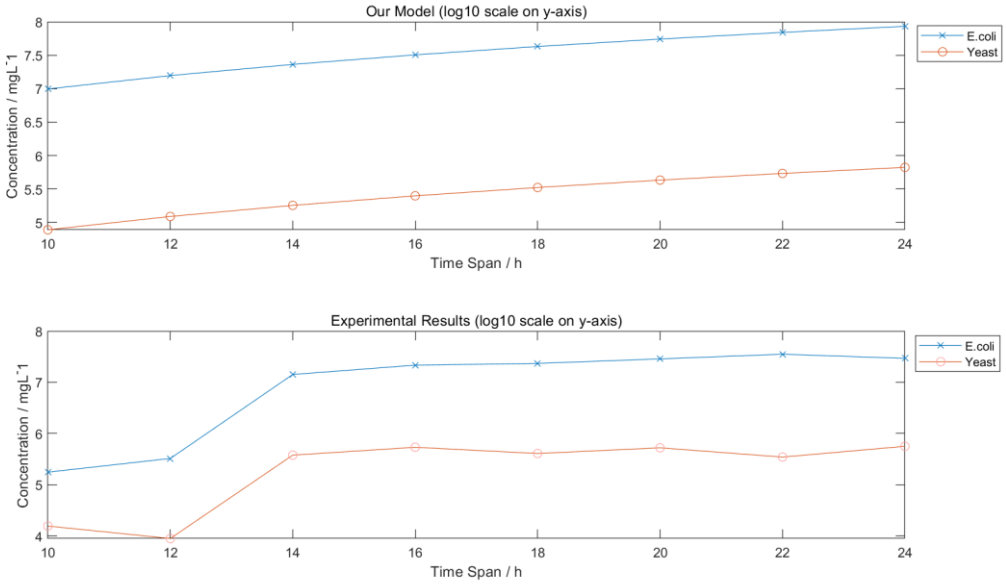

Figure 6. Growth curve for *E. coli* and yeast based on our model and co-culture experiments

Conclusion

Maximum predicted production of nepetalactol

We numerically integrated our models and derived a set of parameters under which we gain the maximum production of nepetalactol.
The predicted maximum production of nepetalactol based on our models is approximately$10^4\mathrm{mg/L}$. It is notable that the population of yeast outweighs the population of *E. coli* when the maximum production of neptalactol is achieved as shown in figure 2. This conclusion sheds light on the proportion of two microbes that the consortia should reach in order to maximize the production of nepetalctol. In order to reach an inter-species balance and a yeast-dominate consortia, we plan to silent the gene *atpFH* in *E. coli* to increase the production of Acetate.

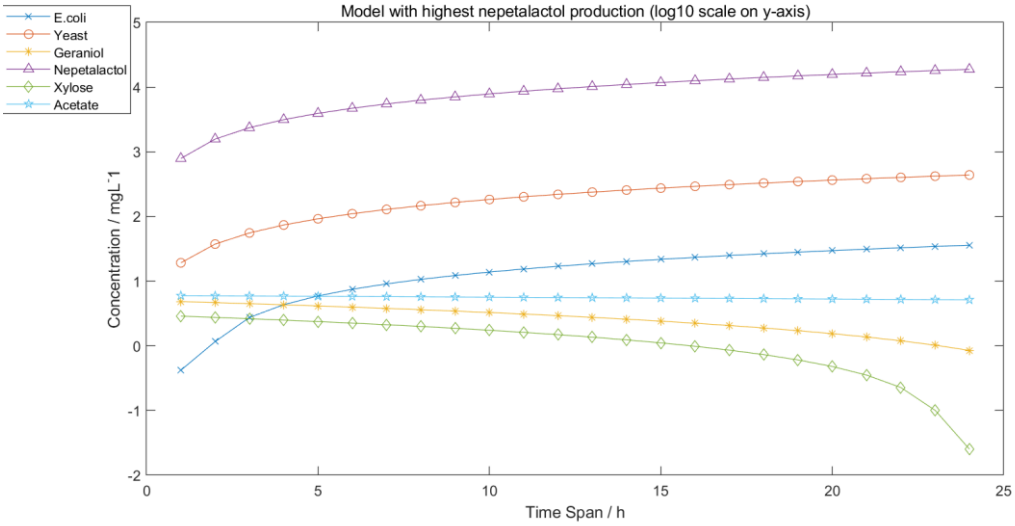

Figure 7. Simulation of *E. coli* and yeast consortia with highest neptalactol production.

Verification of our co-culture design and future optimization based on the models

The results we gain from global sensitivity test and PRCC provide key insights on future optimization of our bio-engineered consortia and prove the effectiveness of our co-culture design. According to the results of PRCC, the population of *E. coli* and yeast at lag phase during co-culture is highly related to the maximum specific rate of *E. coli* and yeast under this circumstances. Since the inserted genetic circuits will increase the burdens on host cells and cause competition among the shared intracellular resources in the cell, the division of synthetic routine into Yeast and *E. coli* will significantly decrease the burden caused by exogenous genetic circuits and increases the growth rate of microbes. The stability of bioengineered *E. coli* and yeast consortia can be further improved by developing and introducing a burden-based molecular feedbacks that robustly control the gene expression. The tolerance of *E. coli* and yeast to acetate and nepetalactol is also positively correlated to the concentration of these two microbes based on the PRCC we derive, which indicates that an increase in neptalactol and acetate tolerance may facilitate the growth of both microbes. In order to do so, we planned to express a Transporter, which can pump the Geraniol out of the *E. coli*.

Part 5 Supplements

Matlab

ODEs

```
  timespan = [0 3];  %Time span
  init = [0.33 1 0.02 0 0 1];  %Initial values of the functions
  %ODE settings
  odesettings = odeset('AbsTol', 1e-12, 'RelTol', 1e-6);
  %simulating the ODE
  [t,f] = ode15s(@coculture, timespan, init, odesettings);
function [out] = coculture(t,f)
```

```
%starting
Y = f(1);
E = f(2);
X = f(3);
N = f(4);
G = f(5);
A = f(6);
%constants
K = 500;
%maximum growth rates
rY_A = 0.35;
rY_G = 0.001;
rE_X = 0.81 ;
%substrate affinity
KY_A = 2.03;
KY_G = 0.001;
KE_X = 0.1;
%critical inhibition
IG_E = 5;
IN_E = 3;
IA_E = 8.9;
IG_Y = 0.3;
IN_Y = 0.5;
%yield coefficient
YE_X = 10;
YY_G = 0.01;
YY_A = 5;
%maintenance coefficient
mE_X = 1;
mY_G = 0.001;
mY_A = 1;
%specific death rate
kE = 0.00001;
kY = 0.00001;
%LP coefficients
aG = 0.01;
aN = 0.001;
aA =  1.43;
bG = 0.001;
bN = 0.0001;
```

```
  bA = 0.045;
  %Degradation
  DN = 0.00001;
  DG = 0.00001;
  DX = 0.00001;
  DA = 0.00001;
 %specific growth rates
  if (N >=IN_E)||(G >=IG_E)||(A >=IA_E)
     uE = 0;
  else
     uE = ((rE_X+mE_X/YE_X)*X/(KE_X+X)-mE_X/YE_X)*(1-G/IG_E)*(1-A/IA_E)*(1-N/IN_E);
  end
  if (G>=IG_Y)||(N >=IN_Y)
     uY = 0;
  else
     uY = (A/(A+G)*((rY_A+mY_A/YY_A)*Y/(KY_A+A)-
mY_A/YY_A)+G/(G+A)*((rY_G+mY_G/YY_G)*Y/(KY_G+G)-mY_G/YY_G))*(1-G/IG_Y)*(1-N/IN_Y);
  end
  %functions
  dYdt = (uY-kY)*Y*(1-(E+Y)/K);
  dEdt = (uE-kE)*E*(1-(E+Y)/K);
  dXdt = -uE*YE_X*E-mE_X-DX*X;
  dNdt = aN*dYdt-bN*Y-DN*N;
  dGdt = -uY*YY_G*Y-mY_G+aG*dEdt+bG*E-DG*G;
  dAdt = -uY*YY_A*Y-mY_A+aA*dEdt+bA*E-DA*A;

  out = [dYdt; dEdt; dXdt; dNdt; dGdt; dAdt];
End
```

Model of LHS

```
% Sample size N
runs=1000;

% LHS MATRIX
xsd=zeros(runs,25);
nsample=1000;
distrib='unif';
s_LHS=LatinHS_Call(xmin,xmean,xmax,xsd,nsample,distrib);

%%
```

```
for x=1:runs %Run each set of parameters
f=@func
 X
 aaa=s_LHS(x,:);
 [t,y]=ode15s(@(t,y)f(t,in,aaa),tspan,ynames);
 A=[t y]; % [time y]

 %% Save the outputs at ALL time points [tspan]
dYdt(:,x)=A(:,1+1);
 dEdt(:,x)=A(:,2+1);
dXdt(:,x)=A(:,3+1);
dNdt(:,x)=A(:,4+1);
dGdt(:,x)=A(:,5+1);
dAdt(:,x)=A(:,6+1);
End

%%
% CALCULATE PRCC
 dYdt(1,:)=[];
dEdt(1,:)=[];
dXdt(1,:)=[];
dNdt(1,:)=[];
dGdt(1,:)=[];
dAdt(1,:)=[];
[PRCCdYdt SigndYdt sign_label]=PRCC(s_LHS,dYdt,1:50,PRCC_var,1);
[PRCCdEdt SigndEdt sign_label]=PRCC(s_LHS,dEdt,1:50,PRCC_var,1);
[PRCCdXdt SigndXdt sign_label]=PRCC(s_LHS,dXdt,1:50,PRCC_var,1);
[PRCCdNdt SigndNdt sign_label]=PRCC(s_LHS,dNdt,1:50,PRCC_var,1);
[PRCCdGdt SigndGdt sign_label]=PRCC(s_LHS,dGdt,1:50,PRCC_var,1);
[PRCCdAdt SigndAdt sign_label]=PRCC(s_LHS,dAdt,1:50,PRCC_var,1);
%%
%Plot Diagram
imagesc(PRCCdAdt);
set(gca,'xtick',1:1:25)
set(gca,'xticklabel',PRCC_var)
title('PRCC of output dAdt')
xlabel('Variables')
ylabel('Time span')
Colorbar
%%
PRCCALL=[PRCCdAdt;PRCCdXdt;PRCCdEdt;PRCCdYdt;PRCCdGdt;PRCCdNdt];
```

```
imagesc(PRCCALL);
set(gca,'xtick',1:1:25)
set(gca,'xticklabel',PRCC_var)
set(gca,'ytick',25:50:325)
set(gca,'yticklabel',PRCC_out)
title('PRCC of six outputs')
xlabel('Variables')
ylabel('Outputs')
Colorbar
%% Import data
xylose=xlsread('C:\Users\mac\Desktop\Xylose Coculture of E. coli and Yeast.xlsx');
xylose(1:4,:)=[];
glucose=xlsread('C:\Users\mac\Desktop\Glucose Coculture of E. coli and yeast.xlsx');
glucose(1:4,:)=[];
%%
for c=1:100
        if dGdt(25,c)==max(dGdt(25,:))
       mark=c
        End
 End
 ycoculture=[dEdt([10 12 14 16 18 20 22 24],mark),dYdt([10 12 14 16 18 20 22 24],mark)];

 %% Plots
 Figure
 ax1 = subplot(2,1,1);
 ax2 = subplot(2,1,2);
 plot(ax1,xylose(:,1),log(ycoculture(:,1)),'-x',xylose(:,1),log(ycoculture(:,2)),'-O')
 title(ax1,'Our Model (y-axis with log10 scale)')
 xlabel(ax1,'Time Span / h')
 ylabel(ax1,'Concentration / gL^-1')

 plot(ax2,xylose(:,1),xy(:,2:3),'-x')
 xlabel(ax2,'Time Span / h')
 title(ax2,'Experiment Results (y-axis with log10 scale)')
 ylabel(ax2,'Concentration / gL^-1')

 plot(1:24,log10(dEdt(2:25,mark)),'-x',1:24,log10(dYdt(2:25,mark)),'-o',1:24,log10(dGdt(2:25,mark)),'-
*',...
  1:24,log10(dNdt(2:25,mark)),'-^',1:24,log10(dXdt(2:25,mark)),'-d',1:24,log10(dAdt(2:25,mark)),'-p')
 title('Model with highest nepetalactol production (log10 scale on y-axis)')
 ylabel('Concentration / gL^-1')
```

```
xlabel('Time Span / h')
% x:E. coli   o: Yeast   *: Geraniol   Triangle: Nepetalactol   Diamond: Xylose    Pentagram: Acetate
```

Function of LHS

```
function s=LatinHS_Call(xmin,xmean,xmax,xsd,nsample,distrib,threshold)

if nsample==1
   s=xmean;
   return
end
if nargin<7
   threshold=1e20;
end

[sample,nvar]=size(xmean);

if distrib == 'norm'   % you only need to specify xmean & xsd
   ran=rand(nsample,nvar);
   s=zeros(nsample,nvar);

   for j=1: nvar
      idx=randperm(nsample);
      P=(idx'-ran(:,j))/nsample;
      s(:,j) = xmean(j) + P; % this can be replaced by any inverse distribution function
   end
end


if distrib == 'unif'
   for q=1:nvar
      if xmin(q)==0
         xmin(q)=1e-300;
      end
   end
   nvar=length(xmin);
   ran=rand(nsample,nvar);
   s=zeros(nsample,nvar);
   for j=1: nvar
      idx=randperm(nsample);
      P =(idx'-ran(:,j))/nsample;
      xmax(j);
      xmin(j);
      xmax(j)/xmin(j);
      if (xmax(j)<1 & xmin(j)<1) || (xmax(j)>1 & xmin(j)>1)
         'SAME RANGE';
```

```
        if (xmax(j)/xmin(j))<threshold %% It uses the log scale if the order of magnitude of [xmax-xmin]
is bigger than threshold
            '<1e3: LINEAR SCALE';
            s(:,j) = xmin(j) + P.* (xmax(j)-xmin(j));
        else
            '>=1e3: LOG SCALE';
            s(:,j) = log(xmin(j)) + P.*abs(abs(log(xmax(j)))-abs(log(xmin(j))));
            s(:,j) = exp(s(:,j));
        end
    else
        'e- to e+';
        if (xmax(j)/xmin(j))<threshold %% It uses the log scale if the order of magnitude of [xmax-xmin]
is bigger than threshold
            '<1e3: LINEAR SCALE';
            s(:,j) = xmin(j) + P.* (xmax(j)-xmin(j));
        else
            '>=1e3: LOG SCALE';
            s(:,j) = log(xmin(j)) + P.*abs(log(xmax(j))-log(xmin(j)));
            s(:,j) = exp(s(:,j));
        end
    end
  end
end
```

Partial Rank Correlation Coefficient Function

```
function [prcc sign sign_label]=PRCC(LHSmatrix,Y,s,PRCC_var,alpha)
Y=Y(s,:)';
[a k]=size(LHSmatrix); % Define the size of LHS matrix
[b out]=size(Y);
for i=1:k
  c=['LHStemp=LHSmatrix;LHStemp(:,',num2str(i),')=[];Z',num2str(i),'=LHStemp;LHStemp=[];'];
  eval(c);
  % Loop to calculate PRCCs and significances
  c1=['[LHSmatrix(:,',num2str(i),'),Y];'];
  c2=['Z',num2str(i)];
  [rho,p]=partialcorr(eval(c1),eval(c2),'type','Spearman');
  for j=1:out
    c3=['prcc_',num2str(i),'(',num2str(j),')=rho(1,',num2str(j+1),');'];
    c4=['prcc_sign_',num2str(i),'(',num2str(j),')=p(1,',num2str(j+1),');'];
    eval(c3);
    eval(c4);
  end
  c5=['clear Z',num2str(i),';'];
  eval(c5);
end
```

```
prcc=[];
prcc_sign=[];
for i=1:k
   d1=['prcc=[prcc ; prcc_',num2str(i),'];'];
   eval(d1);
   d2=['prcc_sign=[prcc_sign ; prcc_sign_',num2str(i),'];'];
   eval(d2);
end
[length(s) k out];
PRCCs=prcc';
uncorrected_sign=prcc_sign';
prcc=PRCCs;
sign=uncorrected_sign;

sign_label_struct=struct;
sign_label_struct.uncorrected_sign=uncorrected_sign;
%sign_label_struct.value=prcc;

%figure
for r=1:length(s)
   c1=['PRCCs at time = ' num2str(s(r))];
   a=find(uncorrected_sign(r,:)<alpha);
   ['Significant PRCCs'];
   a;
   PRCC_var(a);
   prcc(r,a);
   b=num2str(prcc(r,a));
   sign_label_struct.index{r}=a;
   sign_label_struct.label{r}=PRCC_var(a);
   sign_label_struct.value{r}=b;
end
sign_label=sign_label_struct;
```

Equations

$$\frac{dE}{dt} = (\mu_E - d_E)E\left(1 - \frac{E+Y}{K}\right).$$
$$\frac{dY}{dt} = (\mu_Y - d_Y)Y\left(1 - \frac{E+Y}{K}\right).$$

$$\mu_E = [(\mu_{E/Xmax} + \frac{m_{E-X}}{Y_{E/X}}) \cdot \frac{X}{K_{E-X} + X} - \frac{m_{E-X}}{Y_{E/X}}] \cdot I_{G-E} \cdot I_{N-E} \cdot I_{A-E}$$

$$\mu_Y = [\left(\frac{A}{G+A}\left(\mu_{Y/Amax} + \frac{m_{Y-A}}{Y_{Y/A}}\right) \cdot \frac{A}{K_{Y-A} + A} - \frac{m_{Y-A}}{Y_{Y/A}}\right) + \left(\frac{G}{G+A}\left(\mu_{Y/Gmax} + \frac{m_{Y-G}}{Y_{Y/G}}\right) \cdot \frac{G}{K_{Y-G} + G} - \frac{m_{Y-G}}{Y_{Y/G}}\right)] \cdot I_{G-Y} \cdot I_{N-Y}$$

$$\frac{dX}{dt} = -\ (\mu_{E/X} \cdot Y_{E/X} \cdot E + m_{E-X} \cdot E)\ \ - D_X \cdot X$$

$$\frac{dA}{dt} = -\ (\mu_{Y/A} \cdot Y_{Y/A} \cdot Y + m_{Y-A} \cdot Y)\ \ + \alpha_A \cdot \frac{dE}{dt} + \beta_A \cdot E\ \ - D_A \cdot A$$

$$\frac{dG}{dt} = -\ (\mu_{Y/G} \cdot Y_{Y/G} \cdot Y + m_{Y-G} \cdot Y)\ \ + \alpha_G \cdot \frac{dE}{dt} + \beta_G \cdot E\ \ - D_G \cdot G$$

$$\frac{dN}{dt} = \alpha_N \cdot \frac{dY}{dt} + \beta_N \cdot Y\ \ - D_N \cdot N$$

Part 6 References

1, Gomero, Boloye, "Latin Hypercube Sampling and Partial Rank Correlation Coefficient Analysis Applied to an Optimal Control Problem." Master's Thesis, University of Tennessee, 2012.

2, Minty, J. J., Singer, M. E., Scholz, S. A., Bae, C.-H., Ahn, J.-H., Foster, C. E., … Lin, X. N. (2013). Design and characterization of synthetic fungal-bacterial consortia for direct production of isobutanol from cellulosic biomass. Proceedings of the National Academy of Sciences, 110(36), 14592 – 14597. doi:10.1073/pnas.1218447110

3, Marino, S., Hogue, I. B., Ray, C. J., & Kirschner, D. E. (2008). A methodology for performing global uncertainty and sensitivity analysis in systems biology. Journal of Theoretical Biology, 254(1), 178–196.

4, Xia, T., Eiteman, M. A., & Altman, E. (2012). Simultaneous utilization of glucose, xylose and arabinose in the presence of acetate by a consortium of Escherichia coli strains. Microbial Cell Factories, 11(1), 77-77.

5, Zhou, K., Qiao, K., Edgar, S., & Stephanopoulos, G. (2015). Distributing a metabolic pathway among a microbial consortium enhances production of natural products. Nature Biotechnology, 33(4), 377-383.