\section{Introduction}

This is the modeling for project ATM allergen testing system of Team Tsinghua-A in iGEM2018.

\subsection{System processing}

\begin{figure}[H]

\centering

\includegraphics[scale=0.5]{process.png}

\caption{System Processing}

\label{fig:my\_label}

\end{figure}

\section{Reaction in system}

This part of model focuses on the reaction of histamine and receptors in the testing system, and figures out the relation between the amount of histamine and ligand-receptor combinations.

\begin{table}[!h]

\centering

\begin{tabular}{|c|c|c|}

\hline

Name & Description & Unit \\

\hline

$R$ & The inactivated histamine receptor & 1 \\

$RL$ & The activated histamine receptor in the system & 1\\

$RL\_p$ & The activated histamine receptor in per yeast & 1 \\

$L\_t$ &- The ligand histamine & 1\\

\hline

\end{tabular}

\caption{Definitions}

\label{tab:my\_label}

\end{table}

\begin{table}[!h]

\centering

\begin{tabular}{|c|c|c|c|}

\hline

Name & Description & value& Unit \\

\hline

$K$ & The affinity of receptor & 10 & nM\\

$V$ & The volume of the reaction system & $10^{-3}$ & L\\

$N$ & The number of yeasts in the system &$10^7$& 1\\

$R\_p$ & The receptor number in one yeast &$10^4$& 1\\

$N\_A$ & Avogadro's number &$6.02\times 10^{23} $ & 1\\

\hline

\end{tabular}

\caption{Parameters}

\label{tab:my\_label}

\end{table}

\subsection{Model Description}

The main reaction is described by the following function

$$

K=\frac{[R][L]}{[RL]}=\frac{(NR\_p-[RL])(L\_t-[RL])}{[RL]}=\frac{\frac{(NR\_p-RL)(L\_t-RL)}{{(VN\_A)}^2}}{{\frac{RL}{VN\_A}}}

$$

Replacing $RL$ as $x$, we can have follow function

$$

x^2-(NR\_p +L\_t +KVN\_A)x+NR\_pL\_t=0

$$

The solve of this quadratic function are

$$

x\_{1,2}=\frac{NR\_p+L\_t+KVN\_A\pm \sqrt{{(NR\_p+L\_t+KVN\_A)}^2-4L\_tNR\_p}}{2}

$$

According to the limitation $RL<R$,we remain this solve to calculate the concentration of ligand-receptor compound in the testing system.

$$

RL\_t=x\_1=\frac{NR\_p+L\_t+KVN\_A- \sqrt{{(NR\_p+L\_t+KVN\_A)}^2-4L\_tNR\_p}}{2}

$$

Now we simply use the expectation of the activated histamine receptor in each cell,which is defined as $ERL=\frac{RL\_t}{N}$ to approximate the reality.This figure below demonstrates the relation between concentrations of histamine and the activated histamine receptor.

$$RL\_t=f(L\_t)$$

\begin{figure}[H]

\centering

\includegraphics[scale=0.6]{RL.png}

\caption{Histamine and receptors reaction}

\label{fig:my\_label}

\end{figure}

\section{Signal pathway in a yeast}

This part o model displays the modified pheromone pathway in the yeast.The input signal is the concentration of activated receptor in a yeast.The system is contructed by GPCR,MAPK and the expression of EGFP gene.And the output is the concentration of EGFP.In this part,we use Simbiology in MATLAB to construct and simulate our model.\citep{Szallasi2006System}

\subsection{Compartment Diagram}

\begin{figure}[H]

\centering

\includegraphics[scale=0.8]{CDiag.png}

\caption{Compartment Diagram}

\label{fig:my\_label}

\end{figure}

\subsection{Model details}

the species, parameters and functions of the model can be found in appendix 1. \citep{Kofahl2004Modelling}

\subsection{Model Analysis}

First, we took a glance at the input time series (activated histamine receptor) and the transcription fact (ste12active) which is the output of MAPK cascade.

\begin{figure}[H]

\centering

\includegraphics[scale=0.6]{pulsebroadening.png}

\caption{Pulse Broadening}

\label{fig:my\_label}

\end{figure}

In this figure, the degradation of activated histamine receptor (RL) and the output of signal transmission (ste12active) is demonstrated.\citep{Bardwell1996Signaling} We are surprising that the system -- GPCR and MAPK in series connection has a function of pulse broadening. This function is because the amplification of MAPK cascade, which ensure the output transcription factor is stable while the input is really low. Thus, this part of the cascade ensures the stability of the output even if the input is a narrow pulse by pulse broadening.

And we run some parameter scan to analysis how the model works under various parameter sets. And we focus on the affinity of histamine receptor and the degradation rate of the activated histamine, which are both reckoned as highly sensitive parameters before.

The the affinity of histamine receptor and degradation rate of activate receptor scan results are showed below.

\begin{figure}[H]

\begin{minipage}[t]{0.5 \linewidth}

\centering

\includegraphics[scale=0.4]{bingding\_rate.png}

\caption\*{Parameter Scan of the Affinity}

\label{fig:13}

\end{minipage}

\begin{minipage}[t]{0.5 \linewidth}

\centering

\includegraphics[scale=0.4]{RL\_degration\_rate.png}

\caption\*{Parameter Scan of receptor degradation rate}

\label{fig:13}

\end{minipage}

\end{figure}

From the figures, we figure that the peak of the concentration of EGFP increases and the decline of it is delayed as the affinity goes higher or the degradation rate of receptor gets lower, which implied us that we can artificially tune these two parameters to adjust the output. Besides,it reminds us that the randomness of these should be further considered in the stochastic model.

Then, we did an input scan and get a figure of output time series of various input, which is demonstrated below. As to the system, the input is the activated histamine receptor in a cell and the output is the concentration of EGFP.

\begin{figure}[H]

\begin{minipage}[t]{0.5 \linewidth}

\centering

\includegraphics[scale=0.4]{3dTimeSeries.png}

\caption\*{TimeSeries 3D}

\label{fig:my\_label}

\end{minipage}

\begin{minipage}[t]{0.5 \linewidth}

\centering

\includegraphics[scale=0.4]{3dTimeSeries2.png}

\caption\*{Time Series 3D in time-side}

\label{fig:my\_label}

\end{minipage}

\caption{Time Series}

\end{figure}

From this figure, it is obvious that the concentration of EGFP soars at the first time and reach a peak,then gradually decreases, since the activated histamine receptor degrade as the time goes by. In addition, the concentration of EGFP monotonously changes as that of RL (input) goes up , which is an evidence that it is possible to test the concentration of activated receptor by that of EGFP.

\section{Diagnosis Model}

Diagnosis Model is a model constructed to design,analyze and optimize the diagnostic criteria of histamine testing we developed.In this testing, our purpose is to test whether the histamine concentration in a sample reach a certain level.If it does, we consider this kind of result as positive result, which means allergy is possibly happened.However,if the histamine concentration maintains in a background level, we consider it as a negative result.So the diagnosis model is indeed a classification model which focuses on the strategy to determine how we announce the testing result and how much we can believe in it. And, to be emphasized, this model is on the basis of the models we constructed before -- 'Reaction in system' and 'Signal pathway in a yeast'\citep{Isermann2005Model}

\subsection{Diagnostic criteria design}

Now we begin with the EGFP concentration in a yeast when certain amount of the histamine receptor activated, which is the output in the signal pathway model. From the figure 14, it is clear that the output is dynamic. However, using the whole time series as an output result is complicated in calculation and analysis. Thus, we decide to choose the output at a certain time as the output.Now the dilemma is how to choose the proper time that makes the testing is most accurate and effective.

The core concept is the greedy method here. As the result we observed in the figure 14, the EGFP outputs at certain time are different from the input, which means we can define a gap at certain time as follow

$$

gap(t)=max\{EGFP(t,RL\_{max})\}-min\{EGFP(t,RL\_{min})\}

$$

The $RL\_{max}$ and $RL\_{min}$ in the definition mean the $RL$ values that let the EGFP in time $t$ has the maximum or minimum value respectively.Generally, $RL\_{max}=RL\_p$ and $RL\_{min}=0$.The gap can be calculated and shown in the follow chart.

\begin{figure}[H]

\centering

\includegraphics[scale=0.6]{gap.png}

\caption{Gaps through the time}

\label{fig:my\_label}

\end{figure}

Then,the time that has biggest gap $t\_{max}$ can be easily found by algorithms, for example, the simulated annealing we used.In this set of parameter, we find that

$t\_{max}=145$

Next, we focus on the relation between the concentration of activated histamine receptor (input) and that of EGFP (output) in the time $t\_{max}$ (145), which is demonstrated in the follow line charts.

\begin{figure}[H]

\centering

\includegraphics[scale=0.6]{plot\_in\_tmax.png}

\caption{Plot in $t\_{max}$}

\label{fig:my\_label}

\end{figure}

As the chart shows, the concentration of EGFP soars when that of activated histamine receptor is less than approximate 250nM, but the growth rate decreases and maintains stable after this.And, by interpolation, we figure out the function of relation between the concentration of activated histamine receptor and that of EGFP in one yeast.

Then, we assume that the fluorescence intensity ($Flo\_p$) and the concentration of EGFP in one yeast (EGFP) follow this simple linear relation.

$$

Flo\_p=kEGFP=kf(RL\_p)

$$

Recalling the number of yeasts in the testing system $N$, we can derive the relation between the fluorescence intensity ($Flo\_t$) and the expectation of concentration of EGFP in one yeast.

$$

Flo\_t=k\times N \times EGFP=k N f(RL\_p)

$$

The $k$ is scale coefficient and $N$ is the number of yeast in the system.To further improve, we combine the model 'Reaction in system' that provides the relation of histamine in the system and the expectation activated histamine in one cell. Thus, we have a full model from the histamine in the cell to the fluorescence intensity.And we consider it as a known function

$$

Flo\_t=Testing(L\_t)

$$

Taking $k=1000,N={10}^7$, the plot of this function is showed below.

\begin{figure}[H]

\centering

\includegraphics[scale=0.6]{systemplot.png}

\caption{Testing System}

\label{fig:my\_label}

\end{figure}

In this testing, we can flexibly control the input concentration of histamine by diluting or concentrating. However, we know from other researchers and clinical doctors that the expectation histamine concentration of positive result ($EH\_p$) is often ten times of that of negative one ($EH\_n$).

$$

log(EH\_p)=log(EH\_n)+1

$$

Thus, with the above rule, we have to find a proper negative input, which ensures the testing system has the best discrimination. In other words, we need to decide the expectation of histamine input of negative result, which makes the negative result and positive result is the furthest. We can define the difference between the negative and positive result as follow as we know that $EH\_p=10EH\_n$

$$

dif=Testing(EH\_p)-Testing(EH\_n)=Testing(10EH\_n)-Testing(EH\_n)

$$

The $dif$ is a function of $EH\_n$, so we can find a $EH\_n$ that maximize the $dif$.The figure of $dif $ is shown below.

\begin{figure}[H]

\centering

\includegraphics[scale=0.6]{dif.png}

\caption{Difference between negative and positive result}

\label{fig:my\_label}

\end{figure}

In this parameter set we choose, the proper expectation histamine input of negative result that makes the difference biggest is $EN\_n=3.16nM$, and the corresponding input of positive result is $EN\_p = 10EN\_n =31.65nM$. We can directly point these in the plot of the system function as below.

\begin{figure}[H]

\centering

\includegraphics[scale=0.4]{ouput.png}

\caption{System function with results in expectation}

\label{fig:my\_label}

\end{figure}

In summary, after analyzing the testing system, we figure out that letting the negative histamine input be $EN\_n=3.16nM$(With this parameter set) and testing fluorescence in time $t\_{max}=145$(With this parameter set) should be the most efficient and accurate criteria.

\subsection{Threshold Calculation}

After we designed the deterministic model of diagnosis, the error of the input and system should be considered. It is reasonable that there are errors in both negative and positive input, as the histamine in example is different and the manual operation also brings errors. Thus, we tend to assume that the negative and positive inputs ($H\_n$ and $H\_p$) are subject to normal distributions as below

$$

H\_n \sim N(EH\_n,\sigma\_n^2)

$$

$$

H\_p \sim N(EH\_p,\sigma\_p^2)

$$

Besides, the error of affinity of receptor and receptor amount in per yeast, as system errors, are also considered. From papers, we know that affinity $K$ is probably between $5nM$ to $15nM$, and the receptors in per yeast $R\_p$ is probably $10^4$,thus

$$

K \sim N(10,2.5)

$$

$$

R\_p \sim N(10^4,\sigma\_R^2)

$$

And we also consider an uniformity error of the system output ($Flo\_t$), which is caused by the error of fluorescence detection hardware.

$$

Flo\_t \sim N(E(Flo\_t),\sigma\_f^2)

$$

After defining deviations of the system, we simply take follow parameters to testing.

$$

\sigma\_n^2=5\%EH\_n \; \sigma\_p^2=5\%EH\_p \; \sigma\_R^2=500 \; \sigma\_f^2=1\%E(Flo\_t)

$$

Then, we can use Monte Carlo methods to simulate the negative and positive result. We choose 100000 iterations Gibs Sampling to simulate this system. And the outcome is demonstrated below.

\begin{figure}[H]

\centering

\includegraphics[scale=0.4]{distrib.png}

\caption{Distribution of outputs}

\label{fig:my\_label}

\end{figure}

And we can use the ROC to decide the proper threshold that makes the testing most accurate and efficient. The ROC plot is showed below

\begin{figure}[H]

\centering

\includegraphics[scale=0.4]{ROC.png}

\caption{ROC}

\label{fig:my\_label}

\end{figure}

From the ROC, it is clear that the threshold of the testing should be 8.822 and the AUC that measure the quality of testing is 0.929, which is adequately high.

\subsection{Risk Analysis}

At the final part of the modeling, we are interested in how the errors of input and system influence the testing accuracy. First, we continue to use the definition of error in the previous part,and more specific class the errors into two parts.

\begin{table}[H]

\centering

\begin{tabular}{|c|c|}

\hline

Relative system errors & \tabincell{c}{The relative error of the testing system, \\mainly because of the uncertainty of sensitive parameters \\and fluorescence detection error of the hardware.} \\

\hline

Relative input errors & \tabincell{c}{The relative error of input concentration of histamine,\\basically because of the operation errors\\ and the natural variance of samples}\\

\hline

\end{tabular}

\caption{Error definitions}

\label{tab:my\_label}

\end{table}

And we also use Monte Carlo method to simulate it and get a heat plot, which give us hints about how errors in system and input influence the testing quality. In this analysis, we use AUC to indicates the testing quality.

\begin{figure}[H]

\centering

\includegraphics[scale=0.5]{errorcomparer.png}

\caption{AUC under various level of errors}

\label{fig:my\_label}

\end{figure}

From the heat chart, it is clear that the relative error in system has much less effect on the accuracy of testing than that of input, which imply us that it is vital to control the deviation of input. Thus, we have to find a stable and precise way to deal with our histamine samples.The better the input error is controlled, the more accurate the testing is.

\section\*{Acknowledgement}

At the point of finishing this paper, I’d like to express my sincere thanks to all those teammates Li Yaqi Zuo Duyun etc. and advisors Mr. Wang Xiaobo Mr. Xie Zhen and Ms. Wang Jingjing who have lent me hands in the course of my writing this paper. Besides I would like to thank friends below who are not in our team but still give generous support to me.\\

\\

Qin Zihan \\

Renmin University of China \\

For statistic and questionnaire survey advise\\

Xu Haoyang \\

Tsinghua University\\

For algorithm and latex guidance \\

Liu Qiaochu \\

Tsinghua University\\

For chemistry counseling\\