# Reverse engineering of gene regulatory networks for quorum sensing in the food borne pathogen "CRONOBACTER SAKAZAKII"

Report submitted in fulfilment of the degree of

**Bachelor of Technology**

**In**

**Bioinformatics**

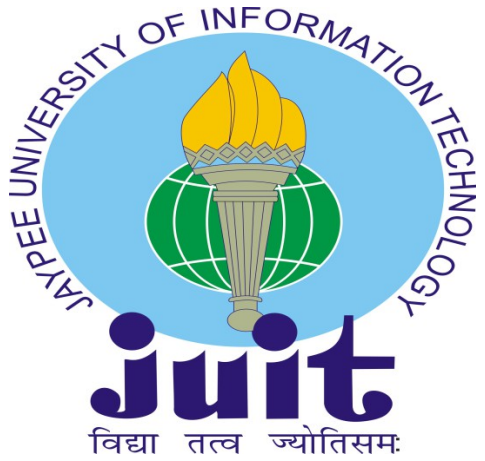
Under the Supervision of

**Dr. Tiratha Raj Singh**

By

**Himanshi Arora (111508)**

To




**Department of Biotechnology/Bioinformatics**
**Jaypee University of Information Technology**
**Waknaghat, Solan – 173234, Himachal Pradesh**

## TABLE OF CONTENTS

## Figures and Tables

# CERTIFICATE

This is to certify that the work titled "**Reverse engineering of gene regulatory networks for quorum sensing in the food borne pathogen "CRONOBACTER SAKAZAKII"** " submitted by "**HIMANSHI ARORA**" in fulfilment for the award of degree of B. Tech in Bioinformatics of Jaypee University of Information Technology, Waknaghat has been carried out under my supervision. This work has not been submitted partially or wholly to any other University or Institute for the award of this or any other degree or diploma.

Signature of Supervisor ……..
Name of Supervisor ……..
Designation ……..
Date ……..

# ACKNOWLEDGEMENT

I owe a great thanks to Ashwani sir who have been helping and supporting me during this project. Our deepest thanks to Dr. **Tirtha Raj Singh**, the Guide of the project for guiding and correcting us at every step of our work with attention and care. He has taken pain to go through the project and make necessary correction as and when needed. Thanks and appreciation to the helpful people at college, for their support. We would also thank our Institution and my faculty members without whom this project

would have been a distant reality. We also extend our heartfelt thanks to our family and well wishers.

Signature of the student ……..
Name of Student ……..
Date ……..

## SUMMARY

Project focus on reverse engineering of gene regulatory networks for quorum sensing in the food borne pathogen. Modelling gene regulatory networks (GRNs) is one of the most important issues in systems biology research. It uses time series gene profiles to characterize the phenotypic behaviour of a target system, and reverse engineering has been advocated to construct networks in an automated way. In the process of inferring gene networks, many computational models and methods have been proposed.

Data is collected from various papers and databases. Collected gene interaction is studied on the basis on STRING DB output. Biofilm formation pathway and Virulence pathway was studied using KEGG Database. Pathway was reconstructed using various tools. A model was build in J designer for comparison purpose. Model

was also built in Snoopy to look into the transitions. Transition level analysis is performed on the basis of interaction results in Snoopy.
Simulation is performed using Cell Designer. We are looking for putative genes/proteins involved in the regulation of the pathway. More parameters will be added to simulate the model and accordingly model will be modified. Predicted entities will provide biologically meaningful information for the regulation of quorum sensing pathway and finally in biofilm formation. Results of simulation graphs and Snoopy transition interaction predicted 3 genes which will be validated in wet lab.

## 1 INTRODUCTION

Reverse engineering is taking apart an object to see how it works in order to duplicate or enhance the object. The practice, taken from older industries, is now frequently used on computer hardware and software. Complex networks of genes, proteins, and small molecules interact to determine cellular function. The reverse engineering or inference of gene regulatory networks using DNA sequence data, protein–DNA binding data, and observed molecular abundance data has become a major interest of the biological community. Modelling gene regulatory networks (GRNs) is one of the most important issues in systems biology research. It uses time series gene profiles to characterize the phenotypic behaviour of a target system, and reverse engineering has been advocated to construct networks in an automated way. In the process of inferring gene networks, many computational models and methods have been proposed. The choices mainly depend on the biological levels to be studied and the computational complexities needed to solve the corresponding problems.

Quorum sensing is the regulation of gene expression in response to fluctuations in cell-population density. Quorum sensing bacteria produce and release chemical signal molecules called auto inducers that increase in concentration as a function of cell density.Gram-positive and Gram-negative bacteria use quorum sensing communication circuits to regulate a diverse array of physiological activities.

These processes include symbiosis, virulence, competence, conjugation, antibiotic production, motility, sporulation, and biofilm formation.

Cronobacter is the officially recognised bacterial genus name for the organism which before 2007 was named Enterobacter sakazakii. Enterobacter sakazakii has been reported to form biofilms, but environmental conditions affecting attachment to and biofilm formation on abiotic surfaces have not been described. We did a study to determine the effects of temperature and nutrient availability on attachment and biofilm formation by E. sakazakii on stainless steel and enteral feeding tubes. Cronobacter is a group of gram-negative bacteria that exists in the environment and which can survive in very dry conditions. The natural habitat for Cronobacter is not known. It has been found in a variety of dry foods, including powdered infant formula, skimmed milk powder, herbal teas, and starches. It has also been found in wastewater. Cronobacter illnesses are rare, but they are frequently lethal for infants and can be serious among people with immunocompromising conditions and the elderly. *Enterobacter sakazakii* (ES) is an emerging pathogen associated with the ingestion of contaminated reconstituted formula that causes necrotizing enterocolitis, sepsis, and meningitis in low-birth-weight preterm neonatal infants.

The major goals of this project will be separated into six main parts:

### 1.1 **Data Collection**

The first part of this project is to collect data for model development/reconstruction. For collection of genes playing important role in biofilm formation and virulence of *Cronobacter Sakazakii* and finding biological role and other features of these genes, two types of search was done.

#### 1.1.1 **Literature Search**

The relevant literature was also collected manually. Important genes playing role in biofilm formation and virulence were collected. After we obtained the literature, we read through the full text of each article to identify one or more genes involved in process. Also, we have collected information about the disease caused and cases in India.

#### **1.1.2 Database search**

Several databases were extensively searched to find out properties, taxonomy and biological function. Few databases were even used to find out interaction between the genes.

**Table1 Tools and Databases used to collect relevant information.**

| S. No. | Online Tools / Databases | Purpose / Usage |
|---|---|---|
| 1. | NCBI | Gene Information |
| 2. | KEGG | Gene Information and Pathway |
| 3. | Gene Card | Gene Information |
| 4. | UNIPROT | Complete Protein Information |
| 5. | Interpro | Gene Ontology |
| 6. | BioCyc | Pathway Information |
| 7. | STRING | Interaction |
| 8 | MINT and BIND | Interaction |

The above collected information has been introduced in the reference table. Once a non-redundant list of genes was extracted from the literature further important information regarding the gene was derived and has been integrated into other tables.

### 1.2 Reverse Engineering of gene regulatory network

A systematic methodology for analyzing the design of an existing device or system, either as an approach to study the design or as a prerequisite for re-designs. An important problem in systems biology is the inference of biochemical pathways and regulatory networks from post genomic data. Understanding the complex regulatory networks underlying development and evolution of multi-cellular organisms is a major problem in biology. Computational models can be used as tools to extract the regulatory structure and dynamics of such networks from gene expression data. This approach is called reverse engineering.

1.2 **Study of Pathway**

For this two different pathways were studied for Cronobacter Sakazakii, for the reconstruction of the model of these pathways and then further generating simulation graphs for the analysis. On the bases of analysis done some hypothesis will be purposed.

Following are the two pathways studied:-

#### 1.3.1 Biofilm Formation Pathway

Biofilm formation pathway was studied using KEGG Database.

Biofilms can be defined as communities of microorganisms attached to a surface. It is clear that microorganisms undergo profound changes during their transition from planktonic (free-swimming) organisms to cells that are part of a complex, surface-

attached community. These changes are reflected in the new phenotypic characteristics developed by biofilm bacteria and occur in response to a variety of environmental signals. In nature, bacteria often exist as biofilms. Here, we discuss the environmental signals and regulatory proteins that affect both the initiation of bacterial biofilm formation and the nature of the mature biofilm structure. Nutrient availability regulates the depth of the biofilm in such a way that the maximal number of cells in a biofilm appears to occur at suboptimal nutrient concentrations. Similarly, quorum-sensing control of the formation of channels and pillar-like structures may ensure efficient nutrient delivery to cells in a biofilm.

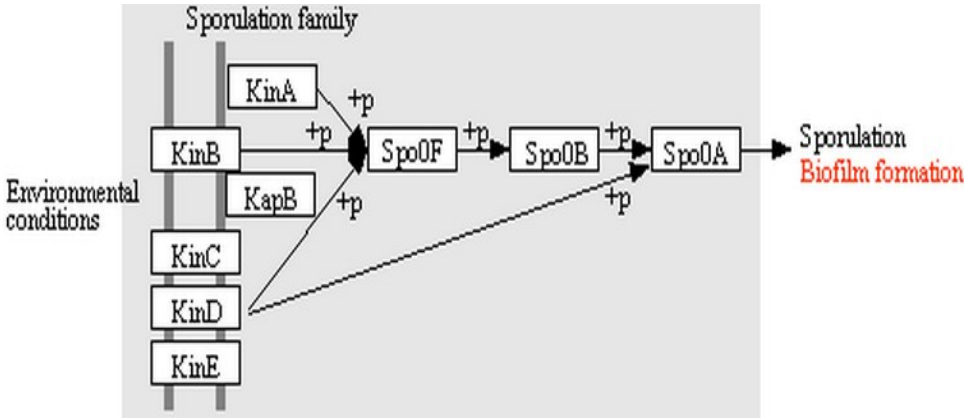



**Figure1. Role of kinase and sporulate genes in Biofilm formation.**

**1.3.2 Virulence Pathway**

*Cronobacter* spp. has come to prominence because of their association with severe neonatal infections, which can be fatal. Current knowledge of the virulence and epidemiology of this organism is limited. However, because neonates are frequently fed reconstituted PIF, this product has been the focus of attention for reducing infection risk to neonates because the number of exposure routes is limited. *Cronobacter* spp[]. has been shown to invade human intestinal cells, replicate in macrophages, and invade the blood–brain barrier. [13][ 14].

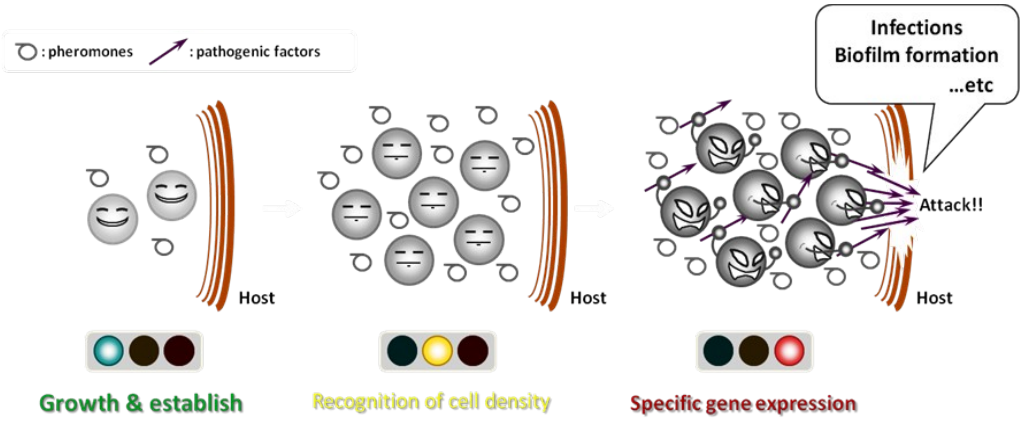



**Figure 2 Virulence process**

### 1.4 Reconstruction of Pathway Model

Using the information collected from above pathways, extensive literature search and database search, the model of the pathway was reconstructed using different pathway reconstruction tools. The important genes collected were used to reconstruct the model on the basis of their relationships and interaction with each other. Two different compartments were made for Virulence and Biofilm formation respectively. After the reconstruction of pathway using different tools the results were generated and compared.

Following three tools were used in this project for the reconstruction of model:-

- Cell Designer
- J Designer
- Snoopy

### 1.5 Simulations

Once the model was reconstructed using above three tools, simulations were done to analyze the behaviour of important genes used for the model reconstruction. Analysis was made on the basis of gene expression. It was observed that how the expression of different genes were regulated. Above mentioned three tools used different methods for simulations and results. They even use different mathematical equations to generate results.

- **Cell Designer**

In CellDesigner modeling tool, I used Runge-Kutta method for iterative analysis which is used for temporal descritization for approximate solution of ordinary differential equations (ODE). I used irreversible Michaelis-Menten equation to analyze the dynamic behavior of genes and protein involved in quorum sensing.

- **J Designer**

I used irreversible mass action method to calculate rate equation for genes and proteins.

$$A + B \rightarrow C + D$$

Applying the law of mass for the forward reaction and backward reactions

$$\text{Rate of forward reaction} \propto [A]\,[B] = kf\,[A]\,[B]$$

Kf is the rate constant for the forward reaction. [A] and [B] are the molar concentrations .The rate of backward reaction, similarly, would be:

$$\text{Rate of backward} \propto [C]\,[D] = Kb\,[C]\,[D]$$

- **Snoopy**

A Petri net is a mathematical graph with a strictly defined syntax. A Petri net is directly executable, if it is represented with the help of an appropriate tool like Snoopy. Snoopy translates automatically and thus reproducibly the graphical scheme into a set of equations used by the program to run simulations. In other words, a graphical representation of a Petri net drawn in Snoopy can be executed, i.e. simulations can be run with a mouse click; no special additional encoding is required.

### 1.6 Proposal of Hypothesis

Since Reverse Engineering is the back bone of the project, after the generation of results and our analysis on the basis of simulation we will purpose some hypothesis or some important genes what are playing important role in the pathway. On the basis of hypothesis or genes purposed some wet lab experiments will be conducted for the validation process.

## REVIEW OF LITRATURE

### 2.1 *Cronobacter Sakazakii*

Cronobacter is the officially recognised bacterial genus name for the organism which before 2007 was named *Enterobacter sakazakii. Cronobacter sakazakii* is a bacterium that causes a rare but often fatal infection of the bloodstream and central nervous system. Infants with weakened immune systems, particularly premature infants, are

most likely to contract and Cronobacter infection, although the bacteria have caused illnesses in all age groups.Named in honor of Japanese microbiologist **Riichi Sakazakii.**

### 2.1.1 Properties

- Formerly Enterobacter sakazakii (till 2007), belongs to Enterobacteriaceae.
- Gram-negative, non-sporulating straight rod.
- Size of cells: 0.5 – 1.5 x 2 – 3 μm.
- Ferments glucose and other sugars.
- Difference from related Enterobacter species consists in its inability to ferment D-sorbitol and ability to produce extracellular deoxyribonuclease.
- Cronobacter species is opportunistic food-borne pathogens that can cause severe and sometimes lethal infections in neonates. In some outbreaks, the sources of infection were traced to contaminated powdered infant formula (LACTOGEN 1,2,3 in India) or contaminated utensils used for PIF reconstitution. We investigated biofilm formation in Cronobacter sakazakii strain ES5.

### 2.1.2 Natural Habitat and Period of Incubation

- Cronobacter sakazakii is an opportunistic pathogen, affecting mainly newborns, and causes neonatal necrotizing enterocolitis and neonatal meningitis. In older infants, children and adults, it can cause sepsis and/or respiratory illness. The microbe has been found in many countries around the world, and has been recovered from Mexican fruit flies and the larvae of stable flies. It is found in food production facilities and households, and has been detected at low levels in powdered infant formulas in a number of countries.

- *Cronobacter sakazakii* infections of newborns have been traced to the use of reconstituted powdered infant formula, especially in hospital settings where formula is prepared in bulk and stored for several hours under refrigeration until needed. In neonatal intensive care units (NICU), feedings can take several hours, during which time the reconstituted formula remains at room temperature – a recipe for microbial multiplication. The incubation period for a Cronobacter sakazakii infection may be as short as one day or as long as three weeks; typically, infections show themselves within one week.

### 2.1.3 Cronobacter Sakazakii and Biofilm Formation

Cronobacter species are opportunistic food-borne pathogens that can cause severe and sometimes lethal infections in neonates. In some outbreaks, the sources of infection were traced to contaminated powdered infant formula (PIF) or contaminated utensils used for PIF reconstitution. Genetic characterization of the mutants led to identification of genes that are associated with cellulose biosynthesis and flagellar structure and biosynthesis and genes involved in basic cellular processes and virulence, as well as several genes whose functions are currently unknown. In two of the mutants, hypothetical proteins ESA_00281 and ESA_00282 had a strong impact on flow cell biofilm architecture, and their contribution to biofilm formation was confirmed by genetic complementation. In addition, adhesion of selected biofilm formation mutants to Caco-2 intestinal epithelial cells was investigated. Our findings suggest that flagella and hypothetical proteins ESA_00281 and ESA_00282, but not cellulose, contribute to adhesion of Cronobacter to this biotic surface.

The presence of biofilms is a relevant risk factors in the food industry due to the potential contamination of food products with pathogenic and spoilage microorganisms. The majority of bacteria are able to adhere and to form biofilms, where they can persist and survive for days to weeks or even longer, depending on the microorganism and the environmental conditions. The biological cycle of biofilms includes several developmental phases such as: initial attachment, maturation, maintenance, and dispersal. Bacteria in biofilms are generally well protected against environmental stress, consequently, extremely difficult to eradicate and detect in food industry. In the present manuscript, some techniques and compounds used to control and to prevent the biofilm formation are presented and discussed. Better knowledge on the architecture, physiology and molecular signalling in biofilms can contribute for preventing and controlling food-related spoilage and pathogenic bacteria. The present study highlights basic and applied concepts important for understanding the role of biofilms in bacterial survival, persistence and dissemination in food processing environments.

### 2.1.4 Cronobacter Sakazakii in Infant Formula

Most cases of *Cronobacter sakazakii* come from powdered infant formula contaminated with the bacterium. However, this type of infection is still very rare.

High temperatures reached in preparing the formula usually kill the bacteria, but they are known to survive even after preparation.

Powdered infant formula is most likely contaminated after production, since the pasteurization process is normally adequate to kill *Cronobacter sakazakii* bacteria. However, if the powder is produced using the dry blending process, and not heated, Cronobacter bacteria can survive in the formula.

**2.2 Disease Caused by *Cronobacter Sakazakii***

*Enterobacter sakazakii* (ES) is an emerging pathogen associated with the ingestion of contaminated reconstituted formula that causes necrotizing enterocolitis, sepsis, and meningitis in low-birth-weight preterm neonatal infants. Necrotizing enterocolitis remains the most common gastrointestinal surgical emergency in these infants.[15]

**C. Sakazakii is associated with neonatal infections**:

- **Meningitis**

Meningitis is an acute inflammation of the protective membranes covering thebrain and spinal cord, known collectively as the meninges. The inflammation may be caused by infection with viruses, bacteria, or other microorganisms, and less commonly by certain drugs. Meningitis can be life-threatening because of the inflammation's proximity to the brain and spinal cord; therefore, the condition is classified as a medical emergency.

- **Necrotizing enterocolitis**

Necrotizing enterocolitis (NEC), which typically occurs in the second to third week of life in premature, formula-fed infants, is characterized by variable damage to the intestinal tract, ranging from mucosal injury to full-thickness necrosis**.**

- **Sepsis**

Sepsis is a potentially life-threatening complication of an infection. Sepsis occurs when chemicals released into the bloodstream to fight the infection trigger inflammatory responses throughout the body.

### 2.2.1 Symptoms

**Meningitis**

The most common symptoms of meningitis are headache and neck stiffness associated with fever, confusion or altered consciousness, vomiting, and an inability to tolerate light (photophobia) or loud noises (phonophobia). Children often

exhibit only nonspecific symptoms, such as irritability and drowsiness. If a rash is present, it may indicate a particular cause of meningitis; for instance, meningitis caused by meningococcal bacteria may be accompanied by a characteristic rash.

- **Necrotizing enterocolitis**

Symptoms include feeding intolerance, increased gastric residuals, abdominal distension and bloody stools. Symptoms may progress rapidly to abdominal discoloration with intestinal perforation and peritonitis and systemic hypotension requiring intensive medical support.

- **Sepsis**

To be diagnosed with sepsis, you must exhibit at least two of the following symptoms:

- Body temperature above 101 F (38.3 C) or below 96.8 F (36 C)
- Heart rate higher than 90 beats a minute
- Respiratory rate higher than 20 breaths a minute
- Probable or confirmed infection

### 2.2.2 Treatment

A *Cronobacter sakazakii* infection is usually treatable with antibiotics, although some antibiotic-resistant strains have recently been discovered. If a newborn exhibits any of the above symptoms, consult a doctor to see whether the infant might need treatment.

#### 2.2.2.1 Prevention

**Clean utensils**

- Wash hands, forearms and fingernails thoroughly before handling any feeding materials or preparing formula.
- All bottles, nipples, caps and rings should be washed in hot, soapy water with thorough rinsing.

**Preparing formula**

- Before use, powdered formula should be kept dry in an airtight container with a firm cap or lid and stored in a cool, dark area. Make sure the expiration date has not passed.
- During formula preparation, bring water to a bubbling boil for two minutes and allow the water to cool before mixing
- Do not use a microwave oven to warm the formula.

**Storing formula**

- Formula should be prepared in small amounts immediately before feeding time to minimize the need for storing reconstituted formula.
- Reconstituted formula should not be stored at room temperature for more than one hour or more than four hours in the refrigerator after preparation.
- Throw out any formula left in a bottle after feeding.

### 2.2.3 Cases in India

**Table2. Cases found in India**

| Reference (date of report) | Gender | Weight at time of birth | Age of illness | Symptoms | Outcome | Powdered infant | Death | Source and Comments |
|---|---|---|---|---|---|---|---|---|
| Ray, et al. 2007 | Female | 1.4kg | 5d | Meningitis | Died | yes | 1(1) | Case from 1992; Source assumed to be IF as infant on tube feeding - however no analyses performed |
| Above | Female | Unknown | 60d | Bacteraemia | Recovered | No | 1 | Case from 2006. Breast feeding - nosocomial infection assumed. |

### 2.3 Types of Cronobacter Species

**Genus *Cronobacter* is composed of following species:**

- *C. Sakazakii*
- *C. Malonaticus*
- *C. Turicensis*
- C. Muytjensii
- C. Dublinensis
- C. Condimenti
- C. Universalies
- C. Helveticus
- C. Zurichenesis
- C. Pilveris
- C. Collectics

**The first three species from above are opportunistic human pathogens.**

## 2.4 Biofilm Formation

Biofilms are colonies of living micro-organisms (e.g., bacteria, fungi, algae) growing on any surface (e.g. metals, plastics, tissue, soil particles, teeth, and so forth).

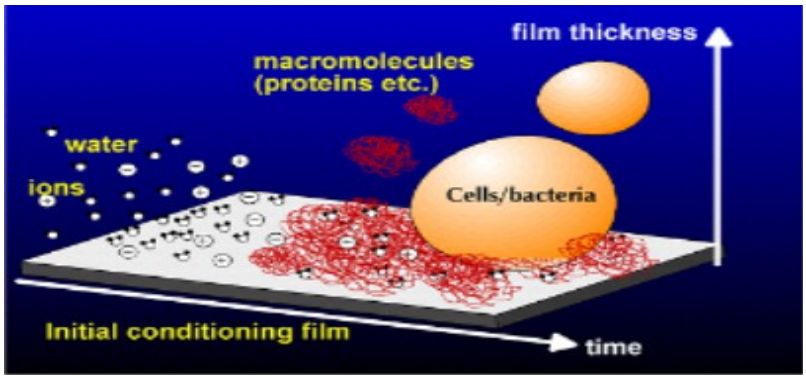



**Figure 3 General process of Biofilm formation in gram negative bacteria.**

Formation of Biofilms

- Form in places with access to water
- Attach to a solid surface using several means:
    - Flagella
    - Hydrophobic Cell Walls
    - Sticky Polymers

**Microbial Mats = Biofilms**

### 2.4.1 Factors Responsible For Biofilm Formation

PIGMENTATION

- To determine the effect of pigmentation colorless mutant (crtE, crtX, crtY) were compared to the yellow wild type.
- It was found that colorless mutant grew at significantly increased rates under osmotic stress compared to that of the yellow wild type.
- Moreover pigments play a role in the survival of bacteria in harmful environments and have been found to increase the virulence of pathogen.

FLAGELLA

- It was found that flagella from *C. sakazakii* are involved in biofilm formation.
- Flagella seem to aid cronobacter cell to adhere to mammalian cells.

## 2.5 Quorum Sensing

Quorum sensing is a system of stimulate and response correlated to population density. Many species of bacteria use quorum sensing to coordinate gene expression according to the density of their local population. In similar fashion, some social insects use quorum sensing to determine where to nest. In addition to its function in biological systems, quorum sensing has several useful applications for computing and robotics.

Quorum sensing can function as a decision-making process in any decentralized system, as long as individual components have:

(a) A means of assessing the number of other components they interact with and

(b) A standard response once a threshold number of components is detected.

Some of the best-known examples of quorum sensing come from studies of bacteria. Bacteria use quorum sensing to coordinate certain behaviours such as biofilm formation, virulence, and antibiotic resistance, based on the local density of the bacterial population. Quorum sensing can occur within a single bacterial species as well as between diverse species, and can regulate a host of different processes, in essence, serving as a simple indicator of population density or the diffusion rate of the cell's immediate environment.

### 2.5.1 Process in different Species

Gram negative and Gram positive both show quorum sensing. Gram-positive and Gram-negative bacteria use different types of QS systems. Quorum sensing is achieved through the coordinated production of special kind of particles called autoinducers.

**GRAM NEGATIVE**

Gram-negative bacteria communicate using small molecules as AIs. These are either acyl-homoserine lactones (AHLs) or other molecules whose production depends on *S*-adenosylmethionine (SAM) as a substrate. AIs are produced in the cell and freely diffuse across the inner and outer membranes.

Example: - Escherichia coli

**GRAM POSITIVE**

QS in Gram-positive bacteria relies on principles common to all QS circuits: production, detection, and response to AIs. In many Gram-positive bacteria, the AIs are oligopeptide AIPs that are detected by membrane-bound two-component signal transduction systems.

Example: - *Bacillus cereus*

### 2.5.2 Quorum sensing and biofilm formation

Bacteria form multicellular biofilm communities on most surfaces. Genetic analysis of biofilm formation has led to the proposal that extracellular signals and quorum-sensing regulatory systems are essential for differentiated biofilms. Although such a model fits the concept of density-driven cell-cell communication and appears to describe biofilm development in several bacterial species and conditions, biofilm formation is multifactorial and complex. Hydrodynamics, nutrient load and intracellular carbon flux have major impacts, presumably by altering the expression of cellular traits essential for bacterial adaptation during the different stages of biofilm formation. Hence, differentiated biofilms may also be the net result of many independent interactions, rather than being determined by a particular global quorum sensing system.

## METHODOLOGY

### 3.1 Reverse Engineering

A systematic methodology for analyzing the design of an existing device or system, either as an approach to study the design or as a prerequisite for re-designs. An important problem in systems biology is the inference of biochemical pathways and regulatory networks from post genomic data. Modern experimental biology is moving away from analyses of single elements to whole-organism measurements. Such measured time-course data contain a wealth of information about the structure and dynamic of the pathway or network. The dynamic modelling of the whole systems is formulated as a reverse problem that requires a well-suited mathematical model and a very efficient computational method to identify the model structure and parameters.

### 3.1.1 Reverse engineering and Gene Regulatory Network

Understanding the complex regulatory networks underlying development and evolution of multi-cellular organisms is a major problem in biology. Computational models can be used as tools to extract the regulatory structure and dynamics of such networks from gene expression data. This approach is called reverse engineering. It

has been successfully applied to many gene networks in various biological systems. The major goal of computational biology is to derive regulatory interactions between genes from large-scale gene expression data and other biological sources.

There have been many attempt s to reach this goal, but the field needs more research before we can claim that we have reached a complete understanding of reverse engineering of regulatory networks. When modelling regulatory networks, genes are often considered as “black boxes”, where gene expression level is an input signal and changed level of expression is the output. We need to shed light on reverse engineering of regulatory networks by modelling the gene “boxes” at a more detailed level of information, e.g., by using regulatory elements as input to gene boxes as a complement to expression levels. Another problem in the context of inferring regulatory networks is the difficulty of validating inferred interactions because it is practically impossible to test and experimentally confirm hundreds to thousands of predicted interactions. Therefore, we need to develop an artificial network to evaluate the developed method for reverse engineering.

Following figure explains the process of reverse engineering.

**Figure 4 Pipeline- Reverse engineering for model building and verification of from biological data.**

### 3.2 Tools

Following were three tools used for the reconstruction of model followed by simulations, generation of graphs and their analysis. On the basis of the analysis done some hypotheses were purposed.

#### 3.2.1 STRING

In molecular biology, STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) is a biological database and web resource of known and predicted protein.

The STRING database contains information from numerous sources, including experimental data, computational prediction methods and public text collections. It is freely accessible and it is regularly updated. The latest version 9.0 contains information on about 5.2 millions proteins from 1133 species. STRING has been developed by a consortium of academic institutions including CPR, EMBL, KU, SIB, TUD and UZH.

#### 3.2.1.1 Usage

Protein-protein interaction networks are an important ingredient for the system-level understanding of cellular processes. Such networks can be used for filtering and assessing functional genomics data and for providing an intuitive platform for annotating structural, functional and evolutionary properties of proteins. Exploring the predicted interaction networks can suggest new directions for future experimental research and provide cross-species predictions for efficient interaction mapping.

#### 3.2.2.2 Features

The data is weighted and integrated and a confidence score is calculated for all protein interactions. Results of the various computational predictions can be inspected from different designated views. There are two modes of STRING: Protein-mode and COG-mode. Predicted interactions are propagated to proteins in other organisms for which interaction has been described by inference of orthology. A web interface is available to access the data and to give a fast overview of the proteins and their interactions. A plug-in for Ccytoscape to use STRING data is available. Another possibility to access data STRING is to use the application programming interface (API) by constructing a URL that contains the request.

### 3.2.2 Cell Designer

CellDesigner is used for modeling and simulation purpose. CellDesigner is a modeling tool for biochemical and gene-regulatory network. CellDesigner was selected for simulation purposes against other contemporary methods as it supports parameter search, estimation, and simulations through sophisticated graphical user interface. For quantitative methods and analysis CellDesigner is comparable to other contemporary tools. Based on a fundamental hypothesis, existing global biological knowledge is collected and processed to build up an integrative conceptual model of a biological system. This conceptual model is then extended with more information to

get the corresponding analytical model. For the analytical model, constraints have to be considered in the buildup model while, the parameters that are missing have to be guessed and diverse assumptions have to be made to get a usable model. Computer simulations were performed for the developed model. The results obtained from the simulation were compared to the data from the biological knowledge of explicit experiments. The once constructed model has to be revised continuously in an iterative process and adapted dynamically to a validated model. This validated model can then be used for diagnostic purposes. [2 ] [12]

#### 3.2.2.1 Model Refinement

We have applied the same approach where we started our model as phenomenological to a comprehensive one while progressing and refining it step-by-step. Starting from a simpler model, we achieved a complex model after few refinements.

### 3.2.3 J Designer

J Designer is a Win32 application which allows one to draw a biochemical network and export the network in the form of SBML. The Designer has an SBW interface that allows it to be called from other SBW compliant modules, for example Python.

In addition, J Designer has the ability to use Jarnac as a simulation server (via SBW) thus allowing models to be run from within J Designer. In this mode J Designer is both a network design tool and simulator.

### 3.2.4 Snoopy

- Snoopy is a tool to design and animate hierarchical graphs, among others Petri nets. Thus, Snoopy allows constructing of Petri nets, as well as to animate and simulate the resulting token-flow.
- The tool has been developed - and is still under development - at the Dep. of Computer Science, Data Structures and Software Dependability, University of Technology in Cottbus.
- The tool is in use for the verification of technical systems, especially software-based systems, as well as for the validation of natural systems, i.e., biochemical networks as metabolic, signal transduction, gene regulatory networks.

The three outstanding main characteristics of Snoopy are:

- **It is extensible:** it's generic design facilitates the implementation of new Petri net classes,
- **It is adaptive:** several models can be used simultaneously, the graphical user interface adapts dynamically to the network class in the active window,
- **It is platform independent**: it is executable on all popular operating systems (Linux, Mac, Windows).

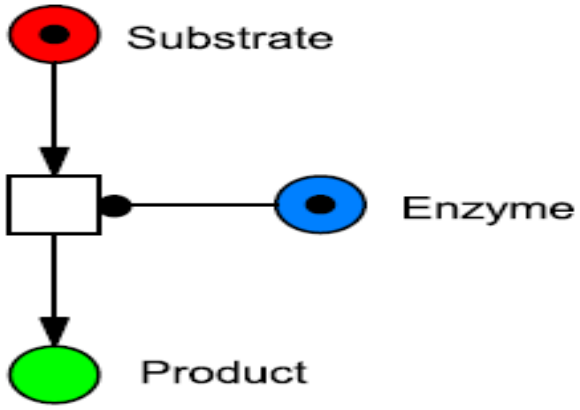



**Figure 5 Snoopy tool diagram for general chemical equation for species (genes and proteins) involved enzymes.**

### 3.3 Process

Following figures explain the step by step methodology of the project in brief.

**Figure 6(a) Flowchart of model building, simulation and verification analysis of biological data.**

**Figure 5(b) Procedure for model building and simulation.**

## WORK DONE

### 4.1 Literature Review

Literature was extensively review and genes playing important role were collected in tabular form.

#### 4.1.1 Important Genes Involved

**Table 3(a) Description and characteristics of important mutants involved in serum tolerance (QS). [1]**

| MUTANT | ANNOTATION/PHENOTYPE | LOCUS TAG CLOSEST TO BLASTX | PROTEIN NAME | DESCRIPTION |
|---|---|---|---|---|
| **67.1a** | Reduced serum resistance | ESA_04343/Cronobacter sakazakii BAA-894 | Putative uncharacterized protein (100%) | Putative membrane protein IgaA homolog (C. turicensis z3032) |
| **BF4b** | Reduced serum resistance | ESA_04103/Cronobacter sakazakii BAA-894 | Putative uncharacterized protein (100%) | Hypothetical protein, conserved domain:Wzy C superfamily O-antigene ligase |
| **51_C4c** | Reduced serum resistance | ESA_03258/Cronobacter sakazakiiBAA-894 | DNA binding transcriptional regulator FruR | Fructose repressor |
| **51_C6c** | Reduced serum resistance | CSE899_07155/Cronobacter sakazakii E899 | Hypothetical protein (100%) | FadR, GNTR family of transcriptional regulator, winged helix-turn helix DNA binding domain. |
| **69_F1c** | Reduced serum resistance | ESA_01368 Cronobacter sakazakii BAA-894 | Hypothetical protein (98%) | DnaJ domain protein |
| **1_E1c** | Increased serum resistance | CSE899_13864 Cronobacter sakazakii E899 | Copper homeostasis protein CutC | Uncharacterized protein involved in copper resistance |
| **4_G12c** | Increased serum resistance | ESA_03283 Cronobacter sakazakii ATCC BAA-894 | Hypothetical protein (99%) | DjlA |
| **21_G1c** | Increased serum resistance | ESA_02809/Cronobacter sakazakii BAA-894 | Hypothetical protein (99%) | Hha toxicity attenuator, YbaJ "biofilm formation regulator" |
| **24_H4c** | Increased serum resistance | ESA_03832/Cronobacter sakazakii BAA-894 | Hypothetical protein (100%) | ribonuclease activity regulator protein RraA |
| **37_A11c** | Increased serum resistance | Ctu_3p00270/ Cronobacter turicensis z3032 | Hypothetical protein (99%) | On Plasmid pCtu3 of C. turicensis z3032 – no annotation available |

**Table 3(b)Description and characteristics of important mutants involved in biofilm formation (QS). [4]**

| Functional group | Mutant | Locus tag(gene ID no.) | Gene | Description(Organism) | Accession no. | No. of +ve amino acid/total no.(%) | BioFilm in CV assay(% of wild type, mean +- SD) | CR agar | Flagella (EM) |
|---|---|---|---|---|---|---|---|---|---|

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Cellulose Biosynthesis** | BF3, BF10 | ESA_04207 (5550384) | *bcsC(yhjL)* | Cellulose synthase subunit BcsC(Cronobacter sakazakii BAA-894)^d | YP_001440 233.1 | 1167/1167(100) | 33.6+25.4 | saw^h | ND^k |
| | BF6 | ESA_04199(5550446) | *bcsG(yhjU)* | Conserved synthase protein(Cronobacter sakazakii)^d | CAM32322.1 | 557/561(99) | 20.5+-8.9 | saw | ND |
| | BF12 | ESA_04204(5550385) | *bcsA(yhjO)* | Cellulose synthase catalytic subunit(Cronobacter sakazakii BAA-894) | YP_001440 220.1 | 872/872(100) | 15.5+-5.9 | saw | ND |
| **Flagella structure or biosynthesis** | BF2 | ESA_01356(5549172) | *flhE* | Flagellar protein FlhE(Chronobacter Turicensis) | CBA31743.1 | 118/119(99) | 19.7+-6.3 | pdar | like wild type |
| | BF5 | ESA_01287(5549196) | *fliD* | Flagellar hook-associated 2 domain protein2(Chronobacter Turicensis) | CBA31883.1 | 471/476(98) | 9.3+-5.1 | pdar | Shorter brittle |
| | BF13 | ESA_02266(5551730) | *flgJ* | Peptidoglycan hydrolase (Cronobacter sakazakii) | YP_001438 351.1 | 321/321(100) | 77.9+-6.4 | pdar | Absent |
| **Basic cellular Processes cell Division** | BF1 | ESA_02449(5549447) | *ftsK* | DNA translocase FtsK(Cronobacter turiensis) | CBA29614.1 | 1264/1364(92) | 8.3+-2.6 | pdar | ND |
| **Energy production** | BF8 | ESA_02873(5548784) | *cyoD* | Cytochrono bacter abiquinol oxidase | CBA28615.1 | 109/109(100) | 19.3+-6.1 | pdar | ND |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | protein CyoD(Chronobacter Turicensis) | | | | | |
| | BF9 | ESA_00861(5550827) | *alsS* | Acetolactate synthase(Chronobacter sakazakii | YP_001436 968.1 | 559/559(100) | 16.5+-3.4 | pdar | ND |
| **Possibly invoked in Virulance** | BF17 | ESA_pESA3p05536 | *mgiB* | Magnesium transporting ATPase, P-type 1(Cranobacter Turicensis) | CBA34129.1 | 853/904(94) | 22.5+-9.1 | pdar | ND |
| | BF4 | ESA_04103(5551795) | *NA* | HYper protein Ent63_0111(Enterobector sp. 638) | ABP58801.1 | 361/425(84) | 66.9+-20.1 | pdar | Like wild type |
| **Unknown Function** | BF11, BF15 | ESA_00282(5548918) | *NA* | Hyponetical Protein(Cranobacter Turicensis) | CBA33797.1 | 164/166(98) | 72.8+-21.3 | pdarse | Like wild type |
| | BF14 | ESA_00281(5549972) | *NA* | Hyponetical Protein(Cranobacter Turicensis) | CBA33799.1.1 | 132/133(99) | 60.5+-36.9 | pdarse | Like wild type |
| | BF16 | NA | *NA* | Similar to Escherichia coli plasmids pMG828-5 p6148, and pE2348^f | DQ995355.1, EU580136.1, FM180570.1 | NA | 34.3+-24.0 | pdar | ND |

**4.2 Interaction of Genes**

Initially interaction of genes found out using STRING. The STRING code for Cronobacter Sakazakii species is “esa” and “csk. The interaction shown below is found using gene entry ESA_00281.

### 4.2.1 STRINGS

In molecular biology, STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) is a biological database and web resource of known and predicted protein.

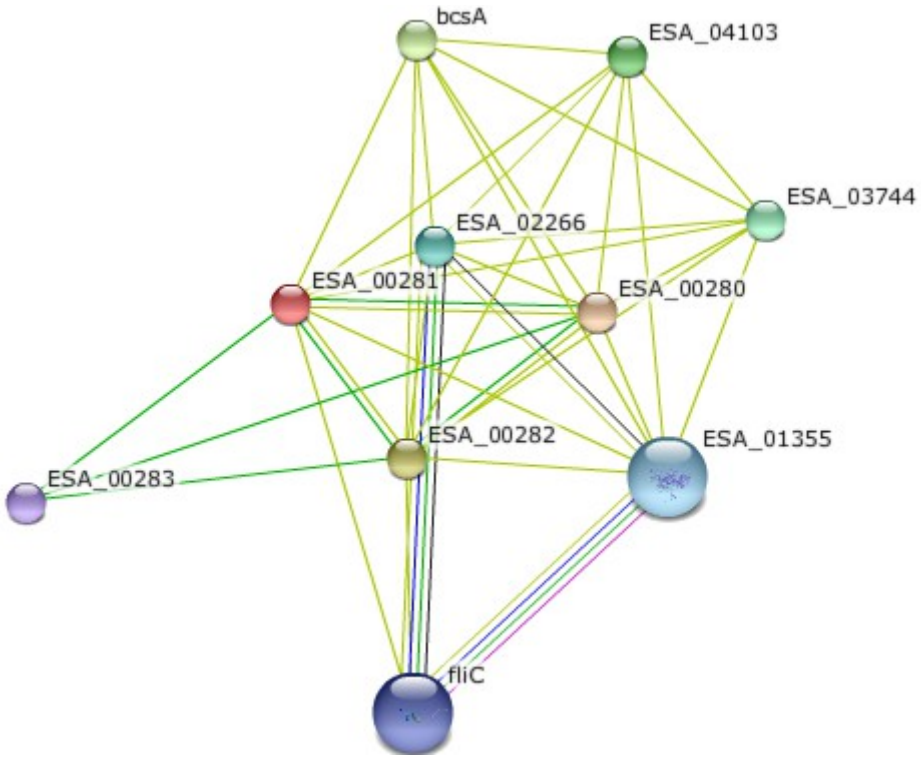



**Figure 7 Protein interactions in STRING database**

### 4.3 Model Reconstruction Using Different Tools

Quorum Sensing (virulence and Biofilm formation) pathway model for Cronobacter Sakazakii was reconstructed using following tools.

#### 4.3.1 STRINGS and CellDesigner

Initially a very basic model was constructed in CellDesigner on the basis of interactions found using STRING.

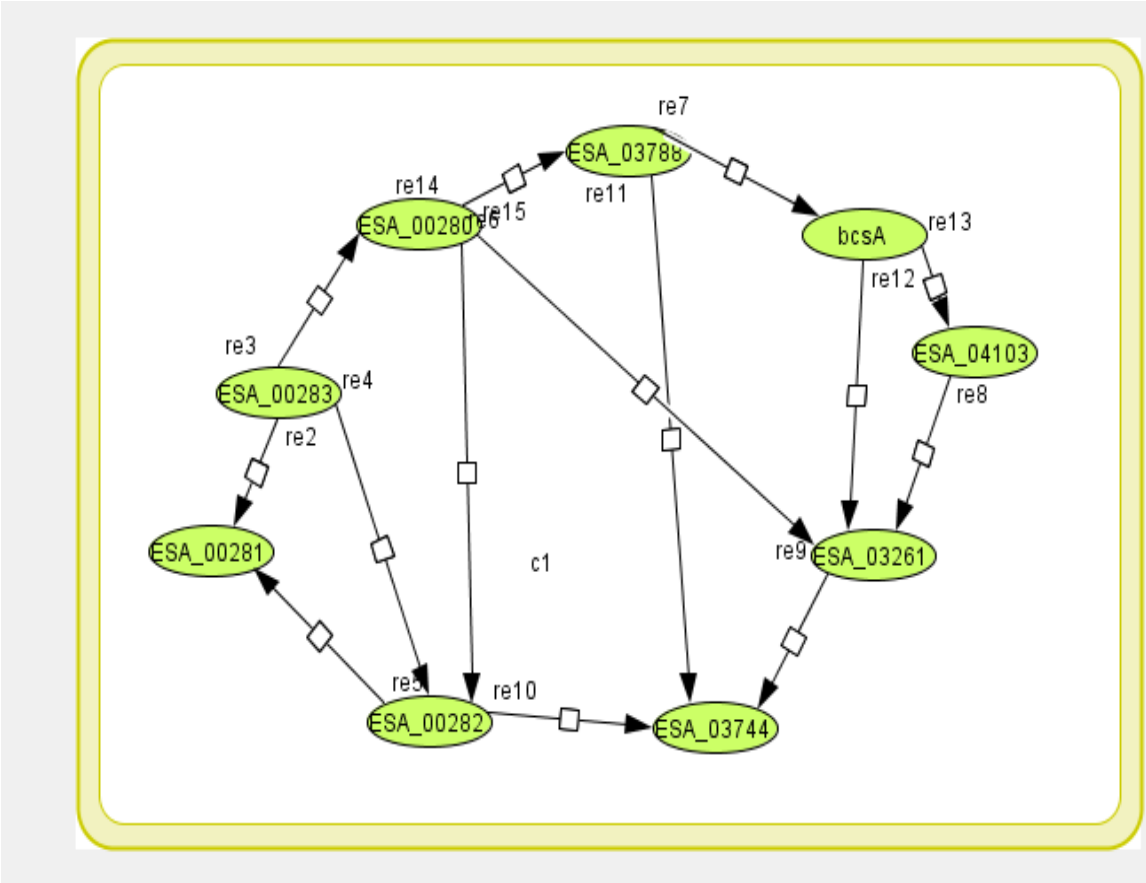



**Figure 8(a) Elementary model for genes in CellDesigner tool.**

### 4.3.2 Cell Designer

Then more literature was studied and searched to find out more genes which play an important role. On the basis of it a new model was build.

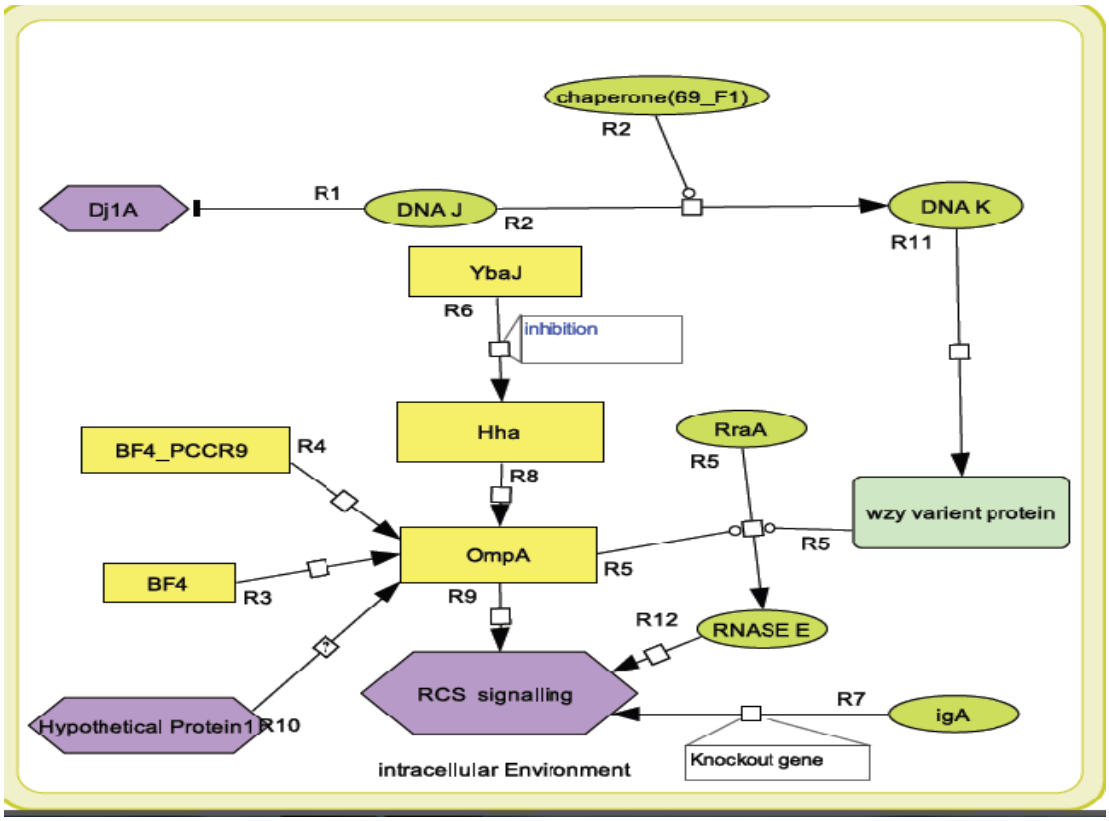



**Figure 8(b) Model reconstruction for important genes and protein involved in virulence (QS) in CellDesigner canvas tool.**

**Few Facts Related To the Above Model**

- Omp A significantly contributes to the survival of bacteria in human blood.
- Concentration Range: - -2 to 10.
- RcsCDB is two component system involved in both negative and positive regulation of target gene involved in biofilm formation and pathogencity.
- WZY protein is responsible for formation of lipopolysaccaride in outer membrane of gram negative bacteria and is quite variable in nature.
- BF4(ESA_04103) and its mutant (BF4_pccr9) are quite important.
- igA knockout in RCS signaling which is responsible for virulence.

#### 4.3.2.1 Model Refinement

Now for model refinement more literature was studied and we extensively searched out important and suitable genes for Virulence and Biofilm formation in Cronobacter

Sakazakii. Even some existing pathways were consulted. On the basis of information collected a new and final model was reconstructed.

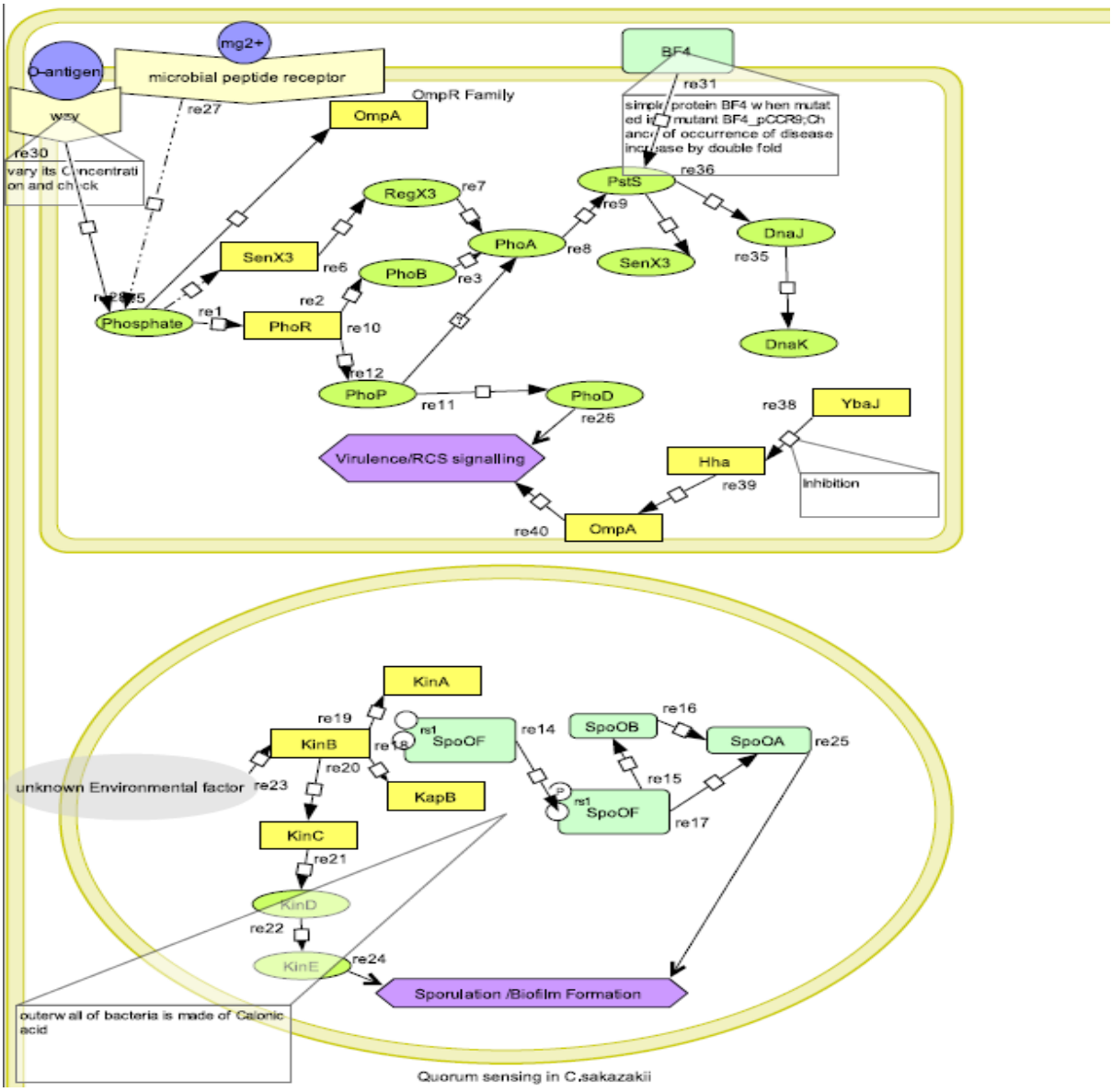



**Figure 8 (c) Refined model for quorum sensing (virulence and biofilm) of Cronobacter sakazakii.**

### 4.3.3 J Designer

After construction of model in cell designer the similar model was reconstructed using J Designer another tool for pathway reconstruction. This was done to make a suitable comparison between the two models. The xml file of the model build in cell designer was extracted and was exported to J Designer.

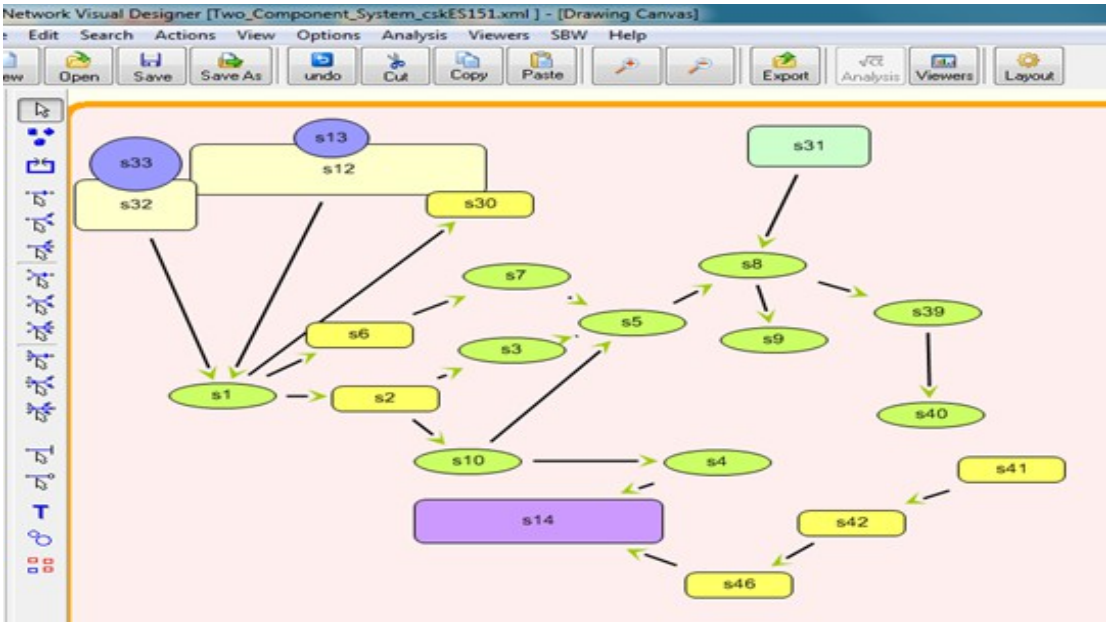



**Figure 9(a) Virulence process representation in J-Designer tool.**

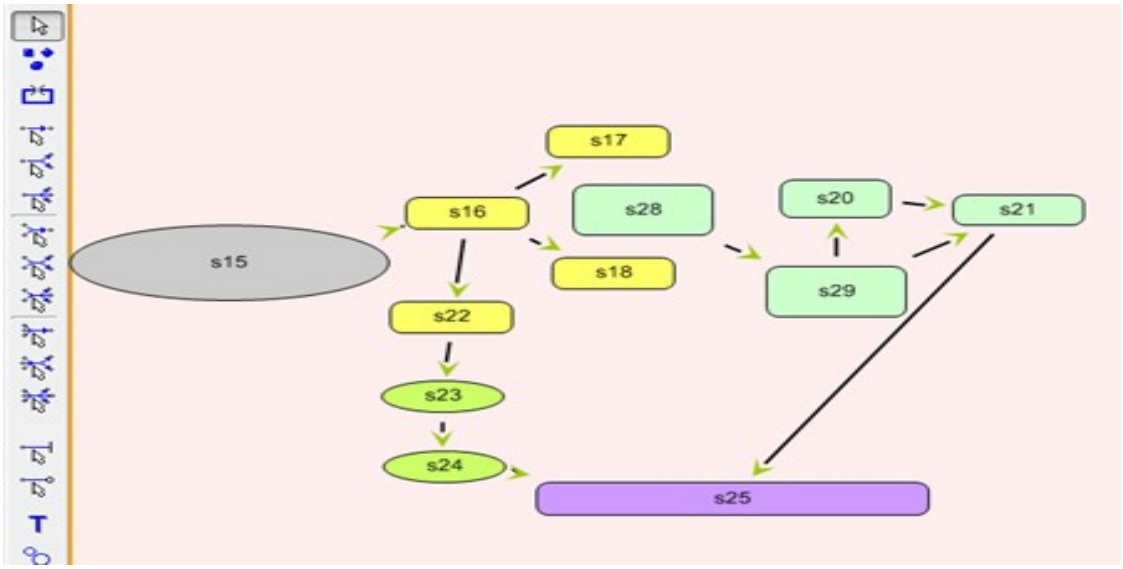



**Figure 9(b) Biofilm formation process representation in J-Designer.**

### 4.3.4 Snoopy

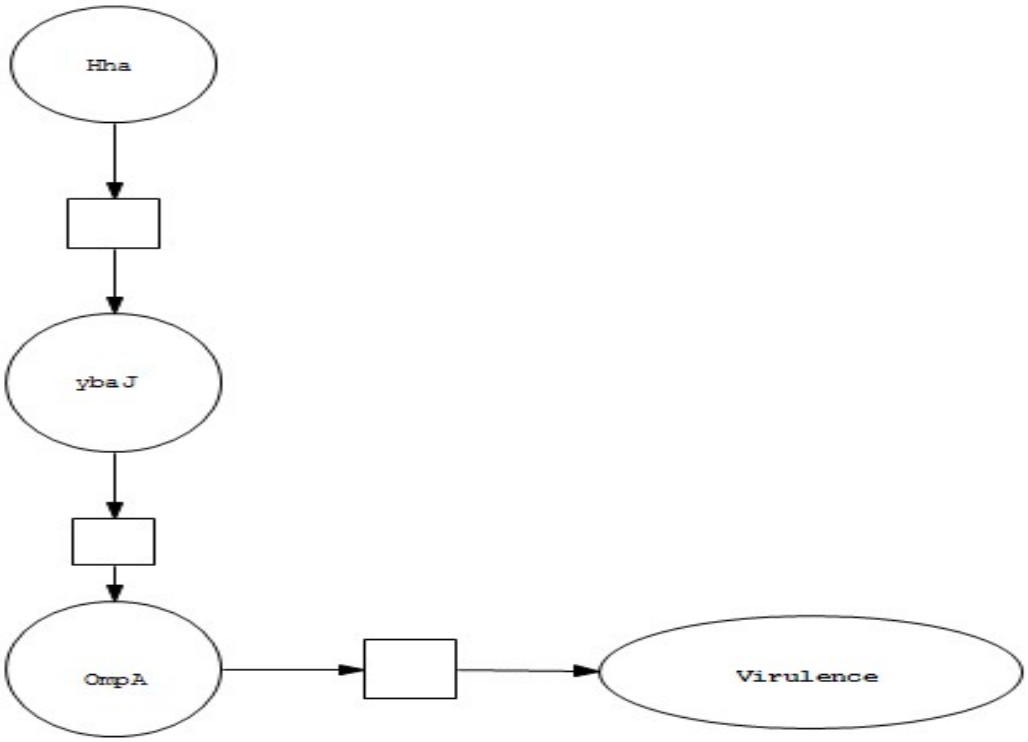



**Figure 10 Snoopy diagram of three important genes involved in virulence pathway.**

**4.4 Model Dynamics for Cell Designer (Basic)**

**Table 4 Dynamical behaviour analysis of species involved in quorum sensing.**

At Time t= 5 milliseconds

| Species1 | Species2 | Irreversible Michialis Menten Method | | Rate Equation |
|---|---|---|---|---|
| **Phosphate(s1)** | PhoR(s2) | V=8 | K=5 | (v1*s1)/(k1+s1) |
| **PhoB(s3)** | PhoA(s5) | V=6 | K=3 | (v1*s3)/(k1+s3) |
| **SpoOB(s20)** | SpoOA(s21) | V=5 | K=1 | (v1*s20)/(k1+s20) |
| **SpoOA(s21)** | Biofilm(s25) | V=9 | K=4 | (v1*s21)/(k1+s21) |
| **KinD(s23)** | KinE(s24) | V=4 | K=3 | (v1*s23)/(k1+s23) |
| **YbaJ(s41)** | Hha(s42) | V=3 | K=-1 | (v1*s41)/(k1+s41) |
| **Virulence(s14)** | OmpA(s46) | V=6 | K=2 | (v1*s46)/(k1+s46) |

**4.5 Simulations**

After reconstructing the model using above tools simulations were done by varying different parameters. Different results were obtained which are pasted in *Results and Discussions* and then compared.

4.5.1 Cell Designer

4.5.2 J designer

4.5.3 Snoopy

## RESULTS and DISCUSSION

### 6.1 Interpretation

Graphs were generated in CellDesigner and J Designer and animation of transitions in Snoopy for interpretation of models.

#### 6.1.1 CellDesigner

Three models were build in Cell Designer, starting from a simpler model build on the basis of interactions found out using STRING and ending up to a complex model after model refinements

##### 6.1.1.1Model1

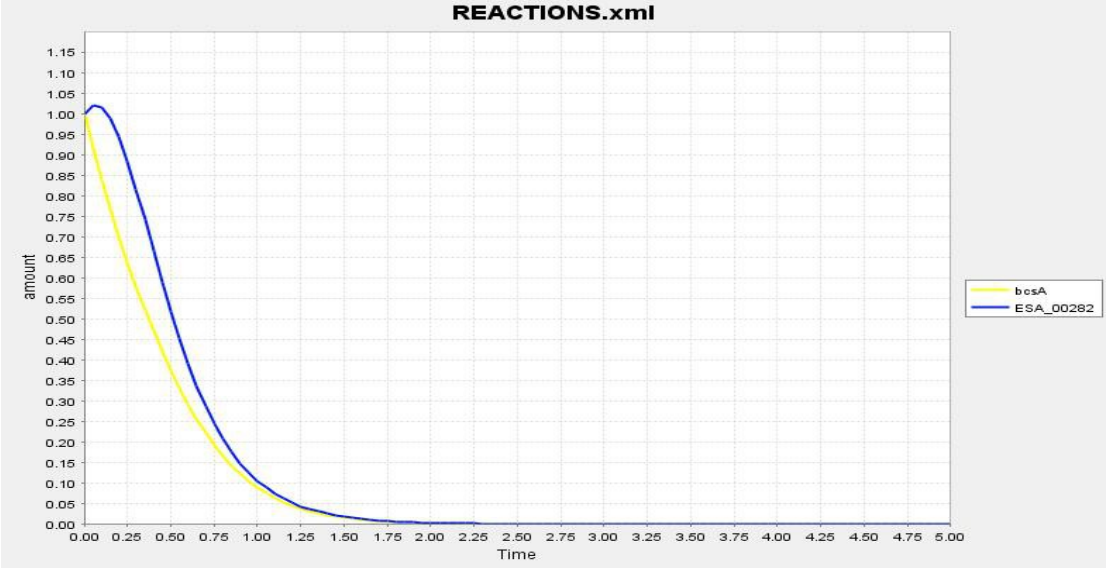



**Figure 11(a) Simulation graph for elementary model.**

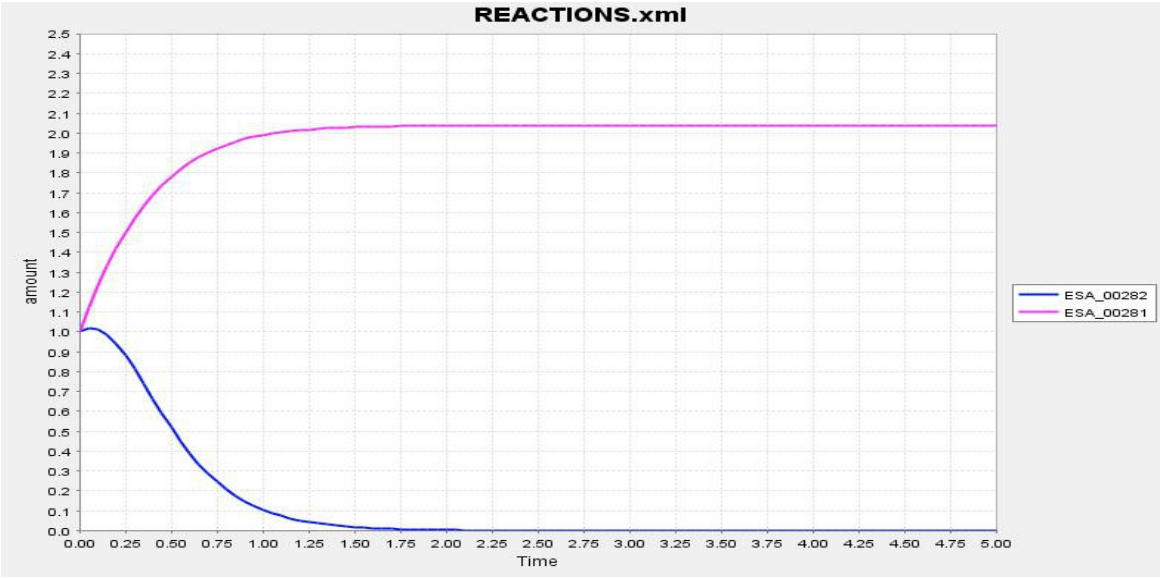



**Figure 11(b) Simulation graph for elementary model.**

### 5.1.1.2 Model 2

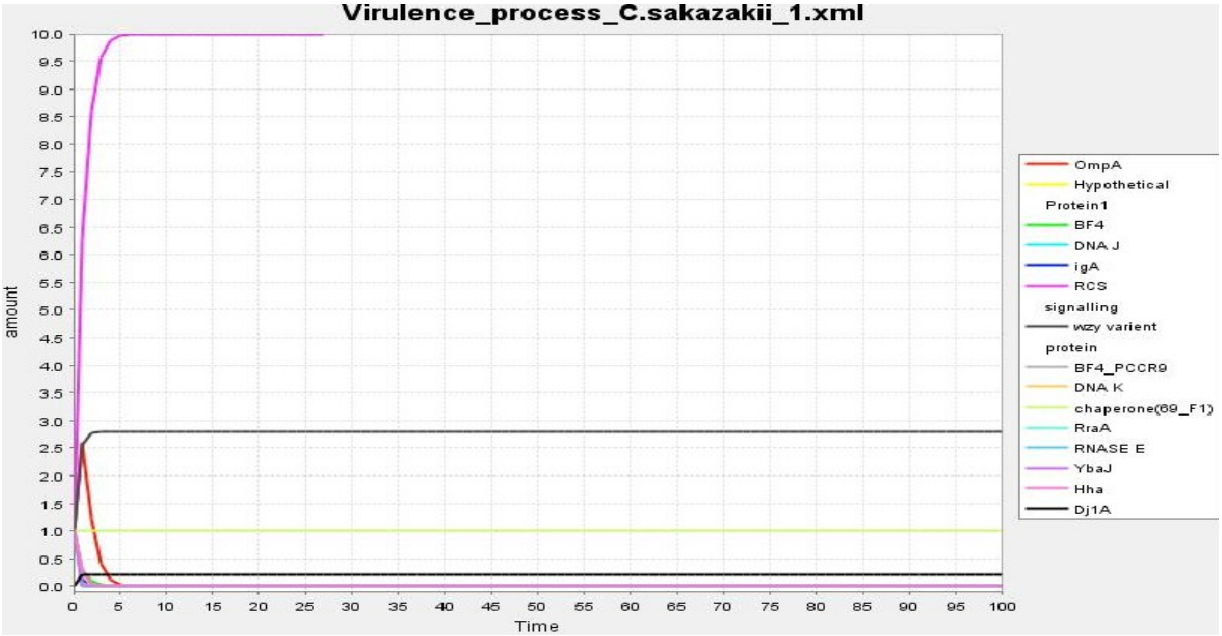



**Figure 12(a) At end time 1 milli second. Simulation graph of different species (gene and protein) at varied time and concentration.**

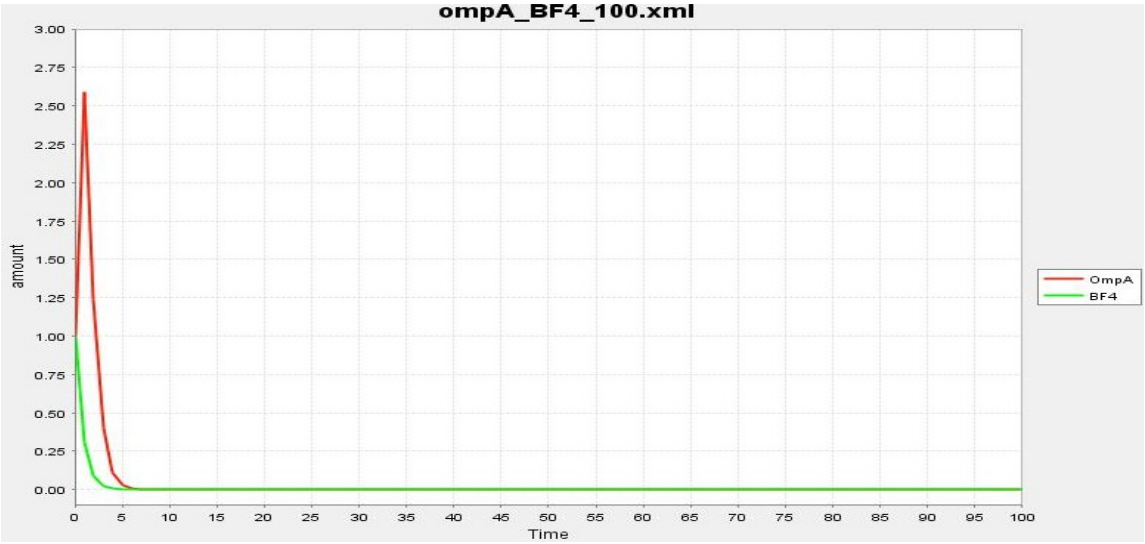



**Figure 12(b) At end time 100 milli second. Simulation graph at time 2 milli second and concentration 2.6 µmole show peak activity for OmpA and for Bf4 the curve shows retrogressive growth which is starting at 1 µ mole and stabilizing at 0.1 µmole . Equilibration point of two species is almost same which is at 6 milli second concentration 0.1 µmole.**

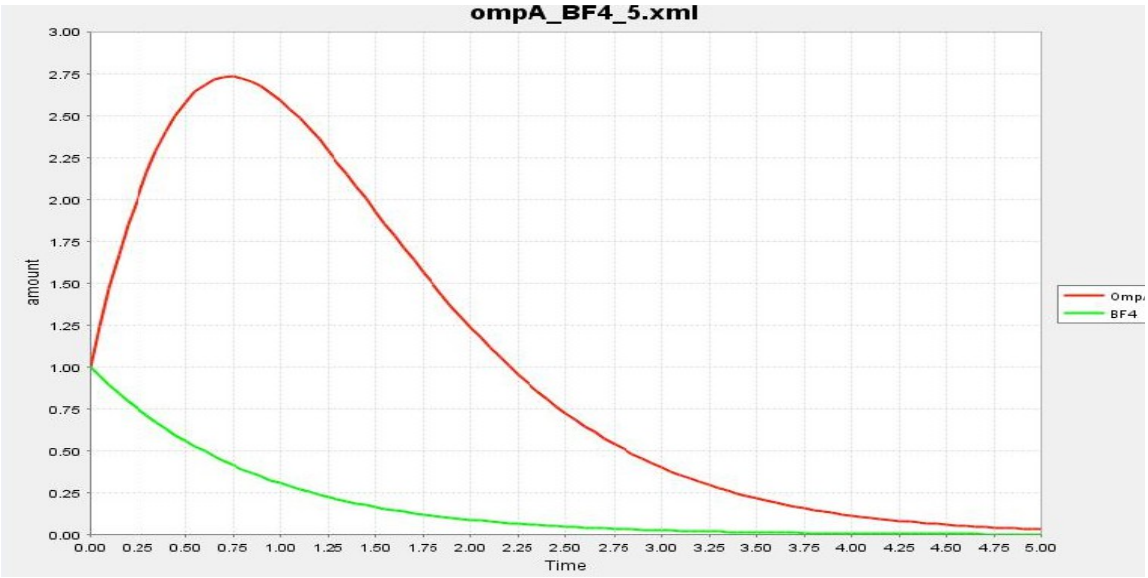



**Figure 12(c) At end time 5 milli second. Simulation graph at time 0.75 milli second and concentration 2.75 µmole for OmpA is showing max activity and time 2.75 milli second and concentration 0.01 µmole for Bf4. Equilibration point of two species is varied and for ompA it is coming near 4.6 milli second and for Bf4 it is coming near 3 milli second.**

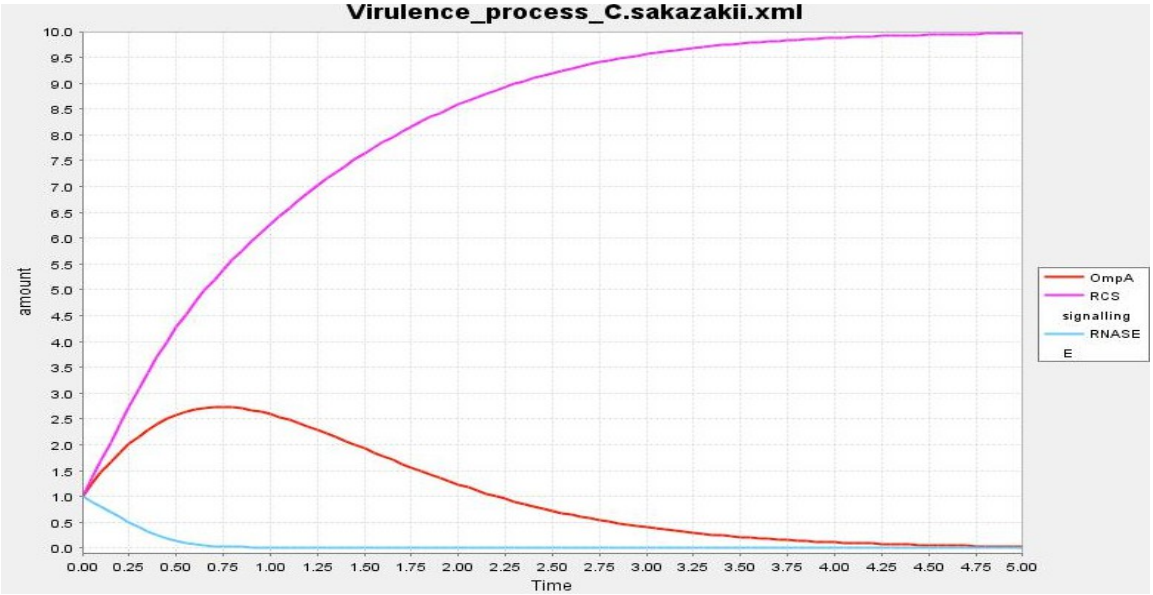



**Figure 12(d) Simulation graph between OmpA , RCS Signalling , RNASE E.**


**Figure 12(f) Simulation graph at time 1.5 mili second and concentration 2.6 µ mole for OmpA and time 0.5 mili second and concentration 1 µ mole for Bf4. Equilibration point of two species are not much varied and is coming near at time 5 mili second.**

### 5.1.1.3 Model 3 (Final Model)

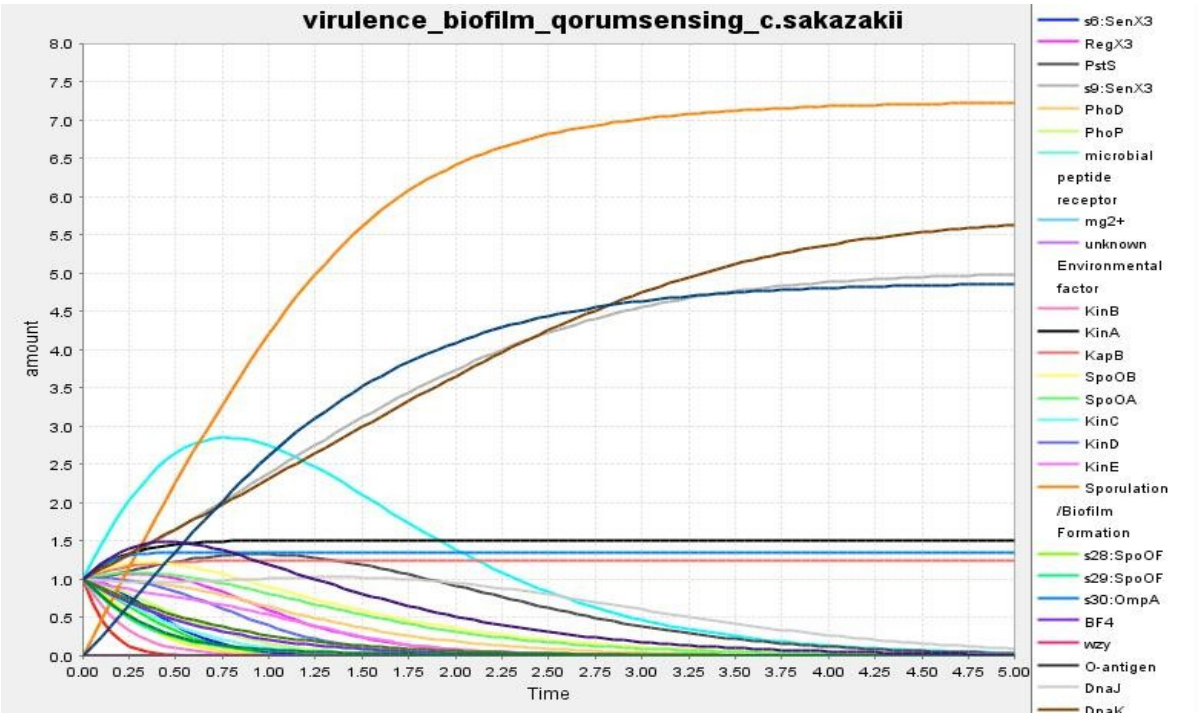



**Figure 13(a) Default simulation graph at end time 5 milliseconds.**

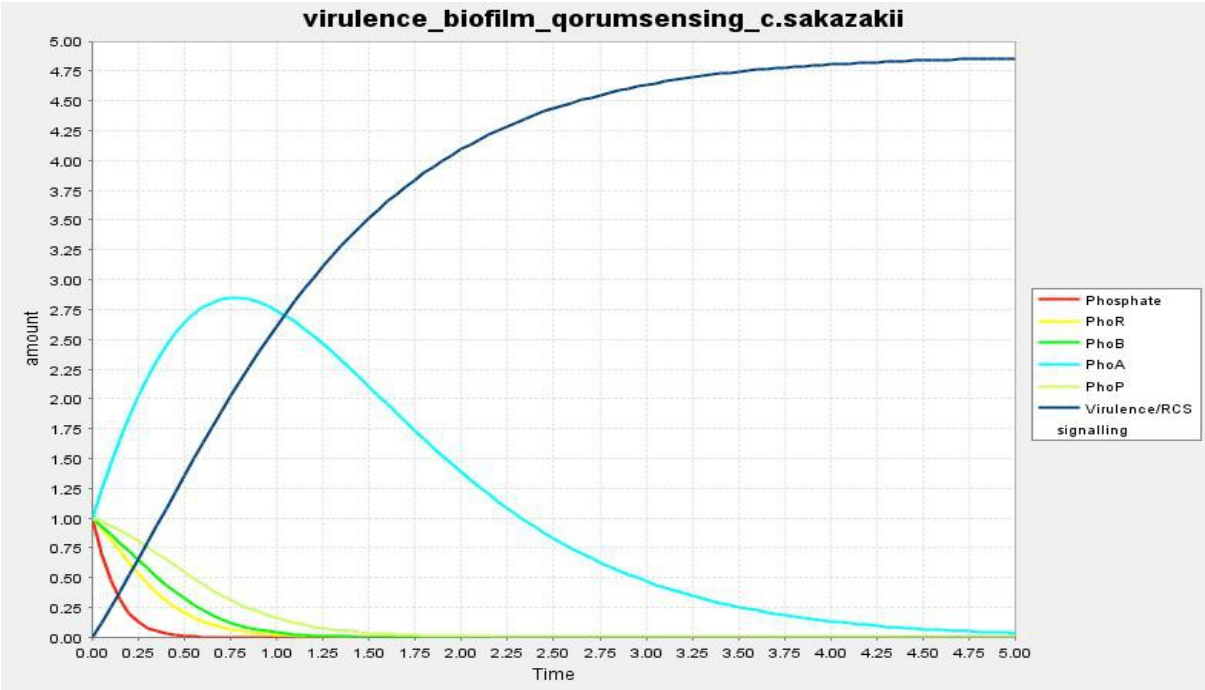



**Figure 13(b) Simulation graph of phosphatase sepecies(*R,B,A,P*) in virulence at end time 5 mili second and initial concentration 1 µ mole.**

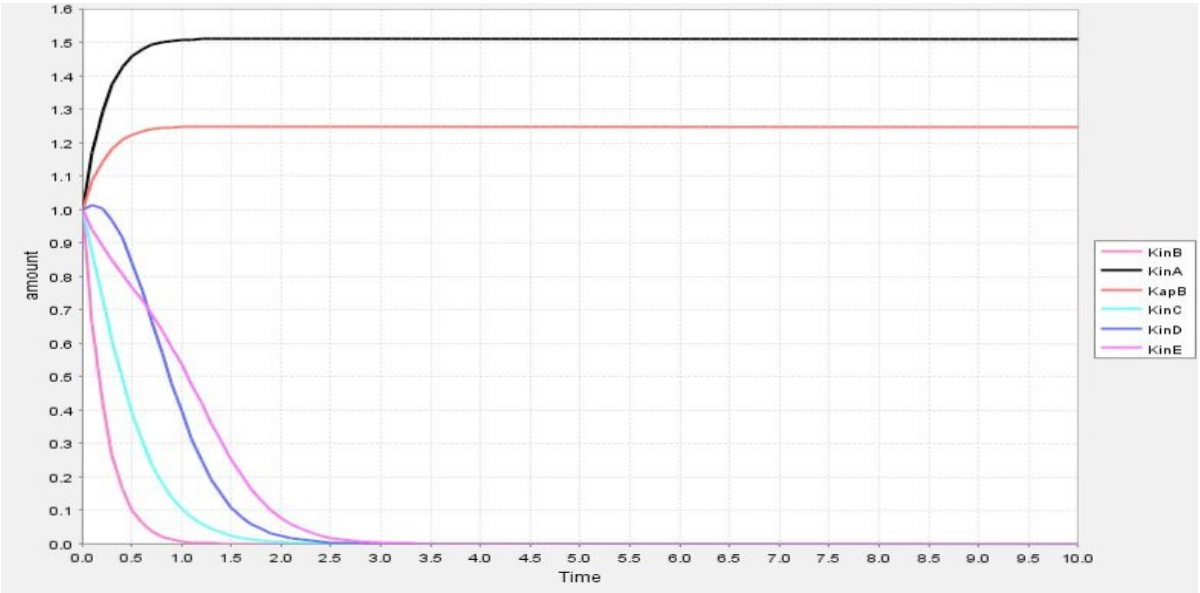



**Figure 13(c) Simulation graph of kinase species (*A,B,C,D,E*) in which species B,C,d,E attain stability at time 3 mili second and the initial concentration of these species is 1 µmole. Whereas kineseA and KapB have similar dynamic behaviour. These two species attain stability at 1 millisecond.**

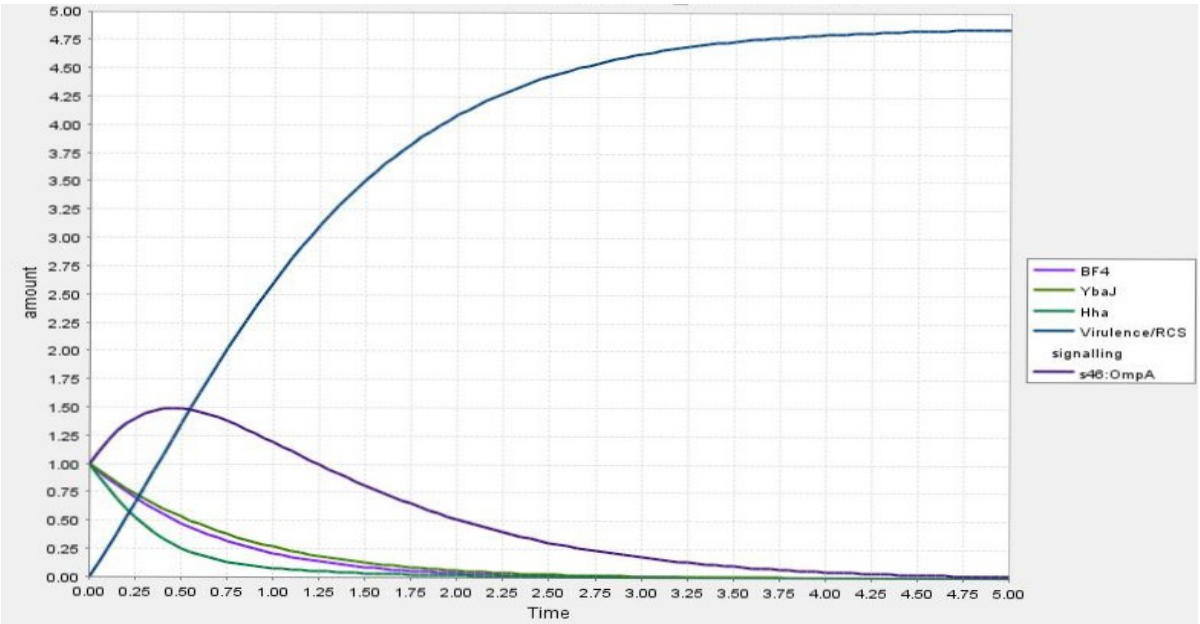



**Figure 13(d) Inhibitors of Bf4, Ybaj, Hha (similar characteristic) and ompA gene increase the phenotypic virulence activity.**

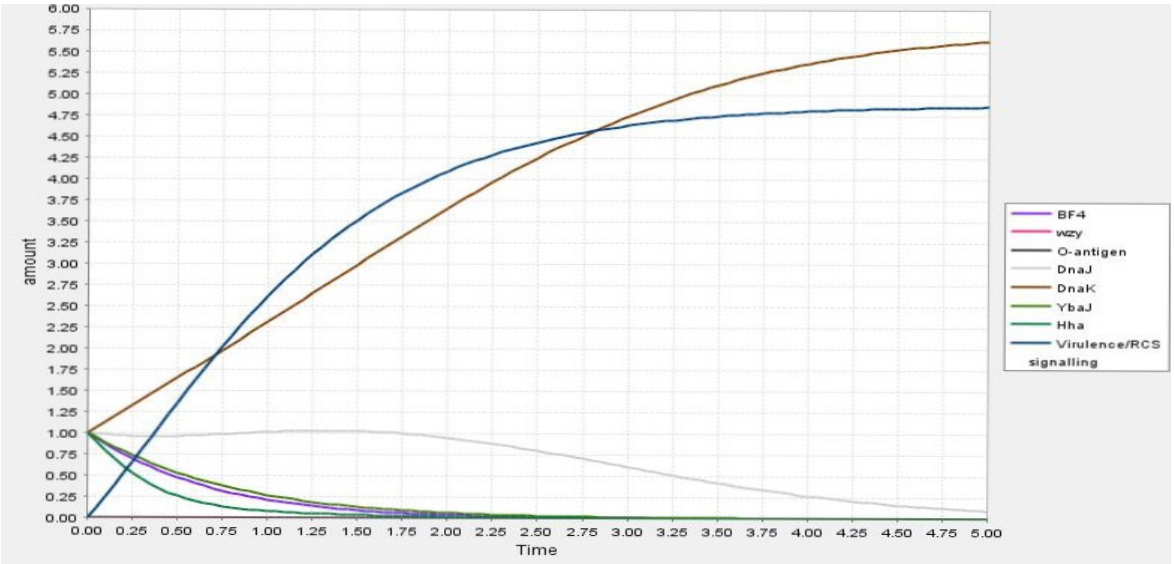



**Figure 13(e) This simulation graph depict effect of Dna (J,K) on virulence. At initial concentration 1 µmole and end time 0.75 and 2.8 are two check points in this graph.**

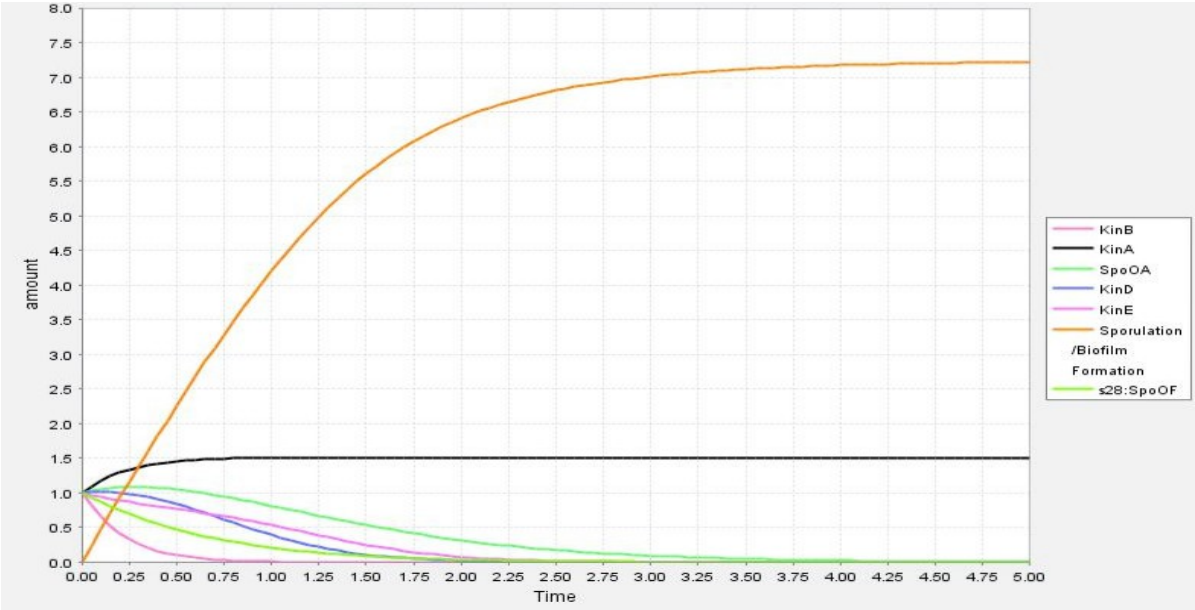



**Figure 13(f) Simulation graph of different genes from kinase family and genes from sporulate family at varied time and concentration.**

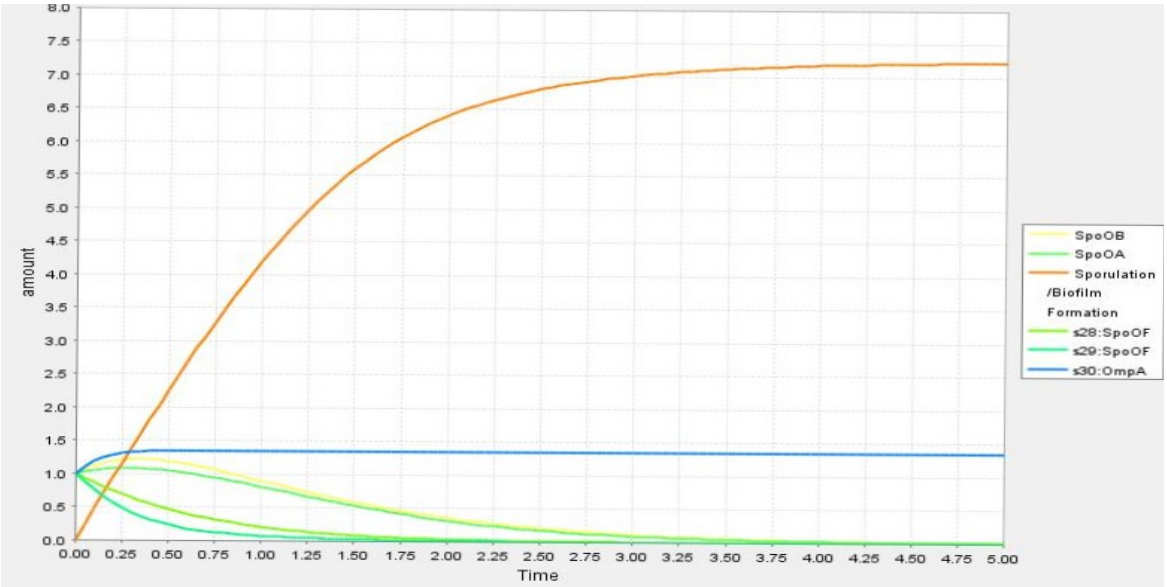



**Figure 13(g) Dynamic behaviour study of different species from sporulate family and OmpA on Biofilm formation process. At initial concentration 1 µmole for each species increase the phenotype but later at different concentration phenotype become stable.**

### 5.1.2 J Designer

A model was also build in J designer for comparison purpose.

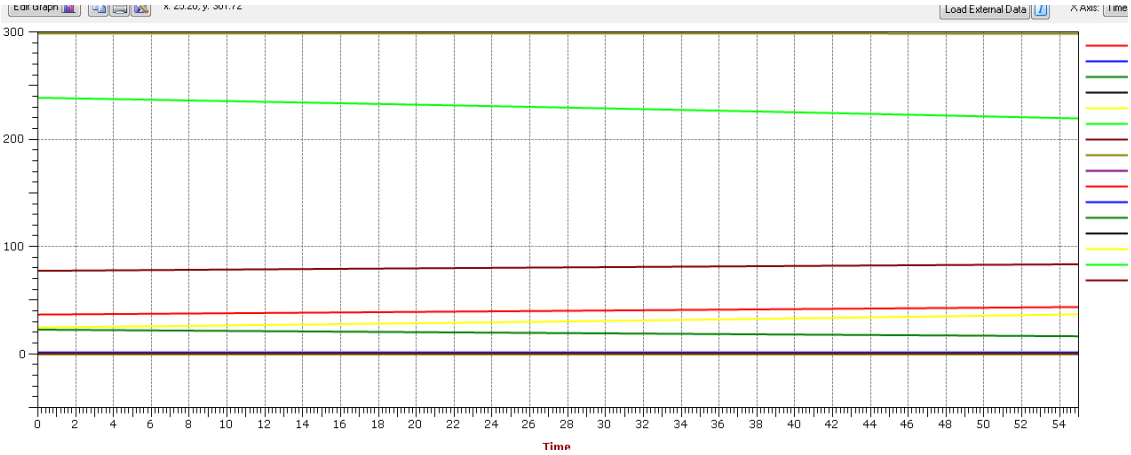



**Figure 14 Simulation graph for important genes and protein in J-designer**

### 5.1.3 Snoopy

Model was also built in snoopy to look into the transitions

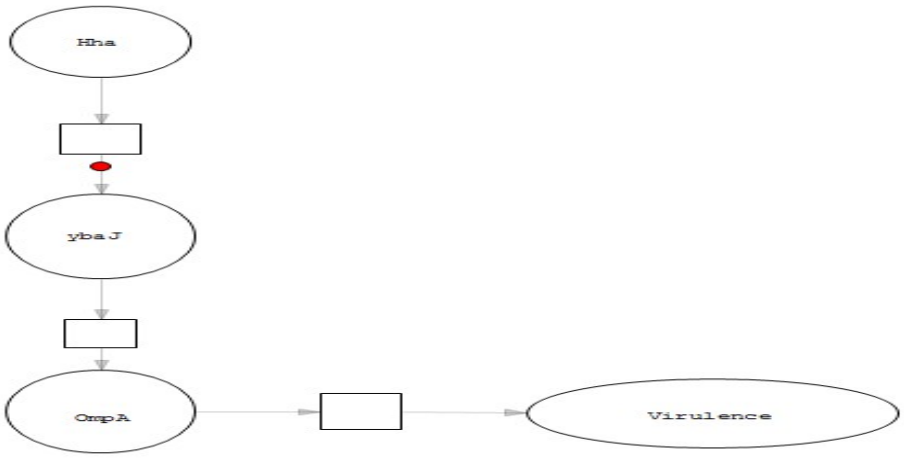



**Figure 15(a)**

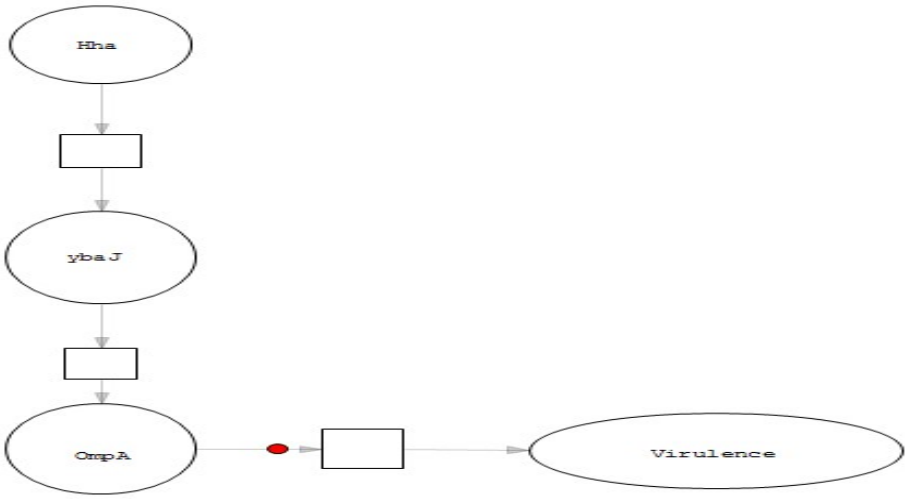



**Figure 15(b)**

**Both Figures A and B are showing transitions. Schematic representation by red mark shows directional flow of reaction in important species of virulence.**

## Discussion

Every aspect of the physical, biological, social, economic and engineering sciences is now critically dependent on insights gained from simulations of complex systems. Simulation is providing a fundamentally new approach to Science addressing problems beyond the scope of traditional Experimental or Theoretical Science. Thus Simulation Study is central in enabling everything from developing new materials and

new technologies for energy, information processing, and communication, to designing new medicines and treatments, and predicting how biological and social networks respond to different stimuli, from infections to financial crashes. The use and application of Simulation Study is critical to drive forward future innovations in the competitive environment of industry and scientific research. Understanding the behaviour of bacteria is very critical and Quorum sensing is the process which entirely responsible for all vital process in both gram positive as well as gram negative bacteria and my finding of function related to important genes and proteins related to QS is beneficial for experimentalist and computational biologist.

**5.2 Future Aspects**

- We are looking for putative genes/proteins involved in the regulation of the pathway.

- More parameters will be added to simulate the model and accordingly model will be modified.

- Predicted entities will provide biologically meaningful information for the regulation of quorum sensing pathway and finally in biofilm formation.

**CONCLUSION:-**

The uniqueness of our approach compare to the traditional differential equations is that it provides various interactive results that could be used for qualitatively studying the biological system. The results of this framework highlight HOW and WHY the system behaves in certain ways, whereas other models such as differential equation based models are able to capture solid states of the systems' behaviour. In this work, we have used quorum sensing as a complex biological system for evaluating our platform. In context of quorum sensing, this tools allow researchers to gain insight to aspects of the system never explored before. As an example, our model allows for looking at application of interaction rules and their frequency of application which

allows for gaining new insights to stability and robustness of the system. We have used a biological language that is able to simulate any biological system that can be expressed in terms of membranes.

In nutshell of our work, some important finding comes out for virulence and biofilm formation respectively. Some gene such as YbaJ, Hha, and OmpA are regulator for virulence process in studied bacteria and we tried to analyze the dynamic behaviour of these important genes in Snoopy tool. Similarly, for biofilm formation study in same bacteria we observed gene expression of mcp , Flagellar family genes by simulation study using Runge-Kutta method and applying Irreversible Michaelis-Menton action law.

Our work can be summarized simply as a tool for biologists to evaluate and monitor complex systems. In order to fill the gap between computer science and biology. With simple interface and pre-defined notations for rule implementation, this model provides a user friendly tool to biologists to develop complex models of different biological system. Proposed genes as regulator for various biological processes will definitely help the biologists, and biotechnologists to evaluate them at experimental level.