Modelling of Bioprocess for Streptokinase Production Using Mechanistic and Neural Network Approaches

Pavan KUMAR
Department of Biotechnology, Indian Institute of Technology Roorkee
Roorkee, Uttarakhand 247667, India

Sanjoy GHOSH
Department of Biotechnology, Indian Institute of Technology Roorkee
Roorkee, Uttarakhand 247667, India

ABSTRACT

Streptokinase is a vital fibrinolytic drug produced by β-hemolytic streptococci often used to treat myocardial infarction and pulmonary embolism. While growing recombinant strain of E.coli to produce streptokinase in a bioreactor, we are dealing with a controlled environment experiencing the role several indispensable factors that are associated to structured and unstructured aspects of the system. Since a cell itself can be assumed to be an entire structured system that shows the accountability of various parameters. On the other hand, unstructured factors are found influencing the active existence of cell in the fermentation media environment. A model has been established on the basis of both structured and unstructured constraints that has key role in dealing with the plasmid stability. Our effort is to configure a composite model which represents the over all dynamics in a well defined algorithm that depicts the behaviour of the microbial population in the entire bioreactor operational environment. The simulation of the process has clearly shown the role of each and every parameter viz. metabolite concentration, dilution rate etc, to swot up competitive dynamics and segregational instability.

Keywords: Growth dynamics, fibrinolytic drug, competitive theme, bioreactor environment, segregational instability, microbial population, structured and unstructured constrains.

1. INTRODUCTION

Streptokinase is a 47 kDa multi-domain protein constituted of 440 amino acid residues in which about 36 percent are non-polar. The enzyme streptokinase is listed in ‘Model of essential medicines, 2005’ by WHO. It is a vital fibrinolytic drug produced by β-hemolytic streptococci often used to treat myocardial infarction and pulmonary embolism. Being a life saving drug it has high commercial demand. There requires an improvement in technology to increase its production from recombinant strain.

Bioprocess modelling strategies are now playing a key role in the production of high value products like streptokinase. Structured and unstructured constraints are modelled in a mechanistic way that has key role in dealing with the plasmid stability. Our effort is to configure a composite hybrid model which represents the overall dynamics in a well defined algorithm that depicts the behaviour of the microbial population in the entire bioreactor operational environment. The estimators like segregational instability, plasmid copy number and metabolite concentration are utilized to estimate some of the aspects of fermentation process which may lead to improve supervision and control. The simulation of the process has been done using numerical simulation to evaluate and predict the process behaviour using neural network approach. It is observed that predicted output obtained from neural network had close statistical resemblance to the actual input values. The idea is to utilize the approach of input-output mapping to optimize the output.

There should be a kinetic model to analyze the decrease of plasmid copy number. Also there is a need to develop fast determination of copy number during the process operation. Plasmid number in plasmid containing cells decreases rapidly while in the process and hence fast determination of the copy number which is very much essential. The existing method, e.g. CsCl gradient centrifugation, HPLC etc is offline methods and time consuming. One of the important tasks is to measure the copy number in-process in order to take corrective measure to minimize its loss. In order to maintain a high copy number and high fraction of recombinant cells in the culture medium strategy would be to optimize key parameter, dilution rate in continuous culture production of streptokinase. It is also used to govern product formation by regulating the sustenance of fraction of recombinant cells in the culture medium.

Understanding the growth dynamics it become easier to recognize the pivotal factors that has influence in formation of the desired product hence modelling such a system in future will be of great help to enhance the production adjusting those key parameters. Production of an enzyme or product within a cell is associated to various pathways. So, several components and sub-components together have their role in an enzyme production dynamics. Using the theme of structured model we may assume the subcellular level dynamics and its sustainability. Several subcomponents involve in the production dynamics used to contribute their some vital role and hence control the product formation. Although there is a decline in the product formations in respect to the considered time frame in general.

While evaluating the performance of a unit carrying out production process consideration is to be made on associated prime pathways with their subcomponents. The ultimate component is giving the product after the entire interaction of all subunits substantially in a defined manner. The sequential interaction of subunits, their failure and repair in the considered time span is quite interesting. A probabilistic framework can show a better implication of the approach to
design a model in order to evaluate the performance of such dynamic units. The feasibility of this approach is to be made on the basis of reliability assessment of the production unit in a mechanistic way. It is rather more interesting to ensure the applicability of stochastic process in running any subunit in its working state or getting repaired at any instance of time, the sole network will be expressed in the form of a unitary system, which is responsible for the production of a particular product which lies on the retaining of its productive state with varying level of product formation which is very likely to be time dependent. The release of metabolites etc may have its impact to negatively affect the subunits involve in pathways, results into cease in production and hence decline in the reliability of such units taken into account. Mortality is often considered to lower the extent of inter specific competition and thereby promote the coexistence of competing species [1].

2. METHODOLOGY AND MODEL DEVELOPMENT

Streptokinase is biochemically a well known fibrinolytic activator that enhances the conversion of plasminogen to the fibrinolytic enzyme plasmin. Workers have over-expressed the streptokinase in E.coli. It is 47 kDa protein encoded by skc gene of Streptococcus equisimilis which is a gram positive bacterium. Though skc gene is known to have many rare codons in its composition, overexpression was achieved instead of having some negative effect. Earlier the constructs were transformed into BL21 (DE3) and grown in LB medium till 0.6 OD at A600. Then the cultures were induced with 1 mM IPTG at 37deg C for 3 h. The expression profile of the streptokinase samples were analyzed [2]. The analysis of relative codon frequency of skc gene in E.coli reveals the presence few position specific rare codons that affect the heterologous protein expression to an extent. In the way of growing E.coli BL21 recombinant strain in a bioreactor, we are dealing with a controlled environment experiencing the role several indispensable factors that are associated to structured and unstructured aspects of the system. Since a cell itself can be assumed to be an entire structured system that shows the accountability of various parameters. On the other hand, unstructured factors are found influencing the active existence of cell in the fermentation media environment. With the progress of fermentation process the initial population of recombinant cells tends to divide into two different cell populations, including those that are having plasmid (as vector pRSET-B) and other that are devoid of any plasmid copy number.

Structured Model Outline

The dynamics and model framework are confined to the concept of reliability has its relevance in assessing the durability of the functionality of any mechanistic unit for a defined time frame. Over all idea is to support the associated pathway networks in view of conducting possible estimation of the enzyme streptokinase production taking into consideration as an instance. Often time dependent interaction of few prominent parameters among the ample of parameters in natural cellular system has been given a desirable weight age in the developing model. The outline of general model in terms of configured subunits and dynamic pathways with an optimal outcome has been expressed on the basis of reliability in numerical terms, as shown in Fig. 1.

The fundamentals governing topology includes, the fact that any production unit has initially maximum reliability and it decreases on basis of interaction of several subunits in a specific topological fashion with variable dynamics. The varying consequence of transformation in topology can be taken into account to depict the observed level of productivity in any terms.

![Figure 1. Model framework showing processing details in structured model form](image)

Considering all the possibilities together, the running original part with general decline in reliability, repair of the partially failed component at any time instant and replacement of the completely failed component at the instant; the putative model can be designed. Although overall process is Stochastic. Supportive algorithms to adapt the property of existing regulatory networks are now in the current trend. Stochastic assumptions are made to generate random decision to regulate the variable pathways by the predefined optimal criteria. Time dependent variables play crucial role in governing the multidimensional dynamics thus preparing the model for showing close proximity to the real-time process dynamics.

Regulatory control and computational intelligence techniques are of major help in artificial framing the dynamics associated to bioprocess [4]. Cellular intelligence is now a days a most excellent tool to amplify the information from cellular level activities in order to utilize it for higher level computational applications in prospects of getting automated [3]. Neural network and hybrid Neural Network modelling are being applied for control of fed-batch like process and to evaluate the bioprocess performance under non-ideal conditions [6]. Also Signal oriented modelling is remarkably an approach for utilizing the intra and inter cellular level signals to emphasize over the fundamentals of structured model. The use of regulatory control enable the models based on cells to utilize information gained from experience and thereby respond intelligently to external stimulus or conditions.

Generally quite complex mechanistic models are required to describe adequately the metabolic dynamics of multi-cellular systems especially under non-ideal conditions. It is evident that cells have internal regulatory control to govern all biochemical pathways in a legitimate manner. Hence it coordinates and directs the adaptive machinery to cope up with external or extra-cellular variations maintaining the supportive mechanism that operates to serve simultaneously. The utility of the devised algorithm would be proved to a greater extent only by ensuring its usefulness under realistic conditions. This instrumental work can only be made feasible employing computational and stochastic approaches which easily support to constitute the variable pathway types.
Functional unit’s identification within the network of pathways is a vital task to complete for giving a minimal framework. An optimal and reliable representation of the overall model using integrated approach for bioprocess design [7] is a mandatory task in this direction. All the model parameters of plasmid replication control can be obtained independently and no adjustable parameters are needed for the plasmid model [8]. Online estimators for biomass and recombinant protein concentration were constructed using information available online by the application of Neural Network [9]. The accuracy of simulations allows a description of the complex interaction between substrate, microorganism and metabolite (acetate) which is a product and also a carbon source for the microbes [10]. In continuous culture the increase in dilution rate to certain point seemed to induce a rapid decrease in plasmid copy numbers [11]. A rapid quantification of plasmid copy number is important process variable that can be used in process control [12].

Classical interpretation has always been appreciated to support and realize the emergent role of any component. A highly developed model presents designing of composite architecture that optimally blends cellular intelligence, artificial intelligence and mechanistic models [3]. Indeed constituting a framework revealing some link between intra-cellular and extra-cellular processes is reasonably of high utility to generate a robust model with an ample of indispensable dimensions.

Unstructured Model Development

On the other hand approach to utilize the unstructured model to quantify the performance in terms of magnitude is rather equally important. Since the production of the streptokinase is directly associated to stability of plasmid in the recombinant cells thus time dependent loss of plasmid [5] can be found associated accordingly by the given Eq. (1),

\[
\frac{dq}{dt} = \left(1 - e^{-mf(\frac{X_p}{X_i + X_o})}\right)\mu_2 
\]

\[\text{Eq. (1)}\]

\[X_i, X_o \text{ Concentration of two cell populations (g/l), } X_p \text{-plasmid bearing, } X_o \text{-plasmid lacking cells, } M- \text{Metabolite concentration (g/l), } m_\text{ Collective probability factor, } \mu_2 \text{ Specific gr. rate of plasmid free cells (1/h) and } t \text{ Time (h).}\]

Overall the probability of plasmid loss in process operation,

\[q = f(t, X_1, X_2, M, \mu 2) \]

\[\text{Eq. (2)}\]

Above parameters are in turn depends on several other variables being the function of more other parameters. Plasmid loss probability (p) is directly found to govern the number of plasmid bearing (pb) and lacking (pl) cells in the preceding and subsequent generations. Computing the magnitude of cells count, from Eq. (1) – (6) we can estimate the system dynamics by simply performing the computer simulation in Matlab 8.0.

\[n_i = pb + pl \]

\[\text{Eq. (3)}\]

\[pl_i = (pb_{i-1} + p + pl_{i-1})^2 \]

\[\text{Eq. (4)}\]

\[pb_{i} = (pb_{i-1} - pb_{i-1}^* + p)^2 \]

\[\text{Eq. (5)}\]

\[pb_{i+1} = n_i - pl_i \]

\[\text{Eq. (6)}\]

In the expressions n is the total number of cells in the bioreactor system at a time. The index ‘i’ stands for the number of generations. The dynamics dealing with count of two types of cells viz. plasmid bearing and lacking cells has its direct association with the production of product like streptokinase. The byproducts like acetic acid etc has its adverse impact over these variety of cells variably due to different constraints associated with them. The unstructured model variables are made to be influenced by the structured model dynamics with time dependent paradigm. So the impact of intra-cellular factors including genetic etc has its profound impact in governing the overall model performance.

Classical Population Dynamics Models

An earliest basic model for population dynamics was proposed earlier by Lotka and Volterra [13] which interprets the population interaction for a prey-predator like system. The microbial ecosystem was considered as a functional entity characterized by certain macroscopic measurements such as the total quantity of biomass or the total number of cells in the medium. It is possible to work with rapid growth of species in well controlled environments, such as “chemostat” [14].

During the culture and production of microorganisms, the control of the bioprocess sometimes depends on the microorganism concentration or the biomass density and conditions provided for that time period, for instance, in some aerobic microorganisms the Dissolved Oxygen (DO) is a key factor to their growth [15].

While performing fermentation process in the bioreactor, recombinant cells are found to lose their plasmid. After some time as the fermentation proceeds two types of cell populations are developed. Since substrate is a growth-limiting nutrient (or factor) so there starts a competition between two populations. The organisms carrying the plasmid or plasmid bearing cells are likely to be weaker competitors than one without because of the added load on its metabolic machinery [16]. There could be a number of factors that regulates the dynamics of plasmid carrying cells within the reactor. One such major factor is probability of plasmid loss due to segregation during cell division that can be described by segregative instability.

The probability of plasmid loss in selective medium and difference in specific growth rate between recombinant and plasmid free cells [17] are affected by factors like genetic make-up of the host cells and the reactor operating parameters such as temperature, pH and growth medium composition. Moreover the selective pressure of selection medium is less effective due to the leakage of gene product [18] which is responsible for selective mechanism. A chemostat model of competition can be established between plasmid bearing and plasmid free organism for a single nutrient where plasmid bearing organism can produce toxin against plasmid free organism at same cost to its reproductive abilities [19]. [20] studied the effect of an inhibitor on two populations. They considered the degree of inhibition in presence of inhibitor or toxicant on growth rate. It is studied that estimation of immeasurable biological variables is important in fermentation process, directly influencing the optimal control performance of the fermentation system as well as quality and yield of the targeted recombinant product, applying some novel strategy for state estimation of fed-batch fermentation [21].

The input substrate concentration and dilution rate serve as operating parameters and they are to be controlled by the experimenter. The study for the cases where nutrient supplied at constant rate and time dependent manner were performed earlier and a delay in the growth response of organism to...
nutrient uptake is obtained. Varying feed profiles were employed in the post-induction phase of recombinant streptokinase protein expression, including constant feed rates, linearly increasing feed rate and exponentially varying feed rates [22] to evaluate the requirement of variable feed strategy.

**Our Model Constraints and Framework**

Considering the case of selective media, a relevant factor selection stress coefficient can be taken into account which is favourable for plasmid bearing cells [18] while it doesn’t favour the plasmid free cells. Since it favours the population of plasmid bearing cells so this parameter would be considered to resist the phenomenon of plasmid loss. Therefore probability of plasmid loss in selective medium is smaller than that in non-selective medium. During the process of fermentation metabolites formation [23] occur as by-products of metabolism which are toxic [24] and inhibit the growth of both types of cells to different extent. The plasmid bearing cells are likely to lose their plasmid because of the permeability of such metabolites into the cells from the environment while other type that is free of plasmid do not have that much extent of harm. In other words we can say that toxicity of metabolite equally harm both the population simultaneously but since plasmid bearing cells are liable to lose their plasmid in response, so plasmid free cell population is strengthening in the same time. So plasmid free cell population or its specific gravity would have an increasing trend despite of decreasing due to this event.

It is observed that probability of plasmid loss is not constant throughout the fermentation process due to the formation of metabolites which shows its presumed toxic effect after certain threshold concentration, so it needs to be taken as variable parameter in the model with respect to time. A criterion of threshold policy can be implemented to evaluate collective probability factor which influence the time dependent variation in probability. Also mortality or formation of dormant cells is possible which is due to toxification developed on account of toxic metabolite or by-product formation, like acetic acid formation in case of fermentation process carried for streptokinase production. Considering the above assumptions we may write a part of the dynamic model [5] as given below:

\[
\begin{align*}
\mu_1 & = \mu_{1_{\max}} S/(K_1 + S) \\
\mu_2 & = \mu_{2_{\max}} S/(K_2 + S) \\
f_1 & = m_0 - r + z, \text{ while, } f_2 = z - \tau \\
\end{align*}
\]

where, \( m_f = \begin{cases} 
  f_1 & \text{if } M > M_{th} \\
  f_2 & \text{if } M < M_{th} 
\end{cases} \)

Here, \( m_1 > m_2 \) also, \( 0 < f_1, f_2 < 1 \)

Since metabolic toxicity has an influence over plasmid bearing cell population to a larger amount.

\( m_1 \) & \( m_2 = 0, \text{ if } M < M_{th} \)

\( M_{th} \) required to be evaluated for different recombinant strain of micro-organisms and media composition under varying set of operational conditions and it depends upon experimental setup with presumed parameters for a bioprocess. Likewise, values of constants \( m_0, m_1, m_2 \) and \( r_1 \) could be assessed for a defined set of conditions. The Nomenclature used for the above unstructured model and its plot are depicted here.

\( S_1, S_2 \) Concentration of two cell populations (g/l), \( N_{1}\)-plasmid bearing, \( N_{2}\)-plasmid lacking cells, \( Y\)-Yield i.e., ratio of gram of cells formed and gram of substrate consumed, \( Y_0 \) Yield in terms of metabolite, \( M_0 \) Threshold concentration of metabolite that can effect population growth (g/l), \( m_3 \) Metabolite toxicity coefficient for plasmid bearing cells, \( m_2 \) Metabolite toxicity coefficient for plasmid free cells, \( p \) Probability of plasmid loss, \( K_1 \) Monod coefficient for plasmid bearing cells (g/l), \( K_2 \) Monod coefficient for plasmid lacking cells(g/l), \( k_0 \) Decay constant for plasmid bearing cells (1/h), \( k_2 \) Decay constant for product (1/h), \( D \) Dilution rate or washout rate (1/h), \( q \) Probability of plasmid loss, \( S_i \) Initial substrate concentration (g/l), \( \mu_1 \) Specific growth rate of plasmid bearing cells (1/h), \( \mu_2 \) Specific growth rate of plasmid free cells (1/h), \( \mu_{1_{\max}} \) Maximum specific growth rate of plasmid bearing cells (1/h) and \( \mu_{2_{\max}} \) Maximum specific growth rate of plasmid free cells (1/h).

The dynamics related to plasmid bearing and lacking cell populations together with effects impart due to metabolite toxicity and selective stress. Above Monod expressions are to compute specific growth rates. For the collective probability factor \( m_0 \) two variants are used as \( f_1 \) and \( f_2 \), in expression which has got variable net probability tested with the level of threshold metabolite concentration to assume its value. Since initially plasmid lacking cell population and metabolite concentration is absent in the medium, both can be taken to zero. In the very start of the process all cells present are plasmid bearing, so the probability of plasmid loss is also to be taken as zero. The magnitude of recombinant cell population and substrate has a pivotal role in governing the dynamics.

### 3. RESULTS AND DISCUSSION

Process simulation has been done using Matlab 8.0, in order to numerically evaluate the role of each and every parameter. The initial values [25] taken for different parameters at time zero. The simulation is done using most of the standard values taken from a previous model data meant for streptokinase [26] together with various other model constraints with smaller magnitude assumed for different set parameters, which has been used earlier in our work [5], taken on the basis of their apparent role in this dynamical system. The probability of plasmid loss is zero at time \( t_0 \) since at the beginning of process all recombinants cells have plasmid machinery. Different dilution rates are considered which proved to be the most relevant factor for continuous operation. Dilution rate was started at a very low value and increased to high values to evaluate the sensitivity of response at different levels.

Plotting all the three variable parameters in Fig. (2), together consequently justifies the correlation among different component variables simultaneously. The dynamics of plasmid copy number is shown in Fig. (3). It is quite reasonable to investigate that low dilution rate allows the fermentation to last longer generates higher magnitude of concentration of plasmid bearing cells and favour the consumption of substrate. The numerical simulation of the model equations shows plasmid loss tends to occur after a certain limit of metabolite level, which depicts that the increasing concentration of metabolite tends to support the population of plasmid lacking cells that witnesses the trend in loss of plasmid from recombinant cells.
Figure 2. Showing dynamics in figure at dilution rate of \( D = 0.21 \)

Figure 3. Streptokinase production and cell growth

Figure 4. Impact of dilution rate on plasmid containing cells with time duration of bioprocess

Role of Dilution Rate and Metabolite Concentration

The influence of dilution rate \( D \) on concentration of plasmid bearing and plasmid lacking cells with respect to time duration of continuous process is represented in two plots, Fig. (4) and (5). The plots in respect to dilution rate have their importance in deciphering the behavior of biased inter-population cell dynamics in this case. The results emphasizes that delayed plasmid loss occurs at lower dilution rates. Numerical simulation of the continuous fermentation process could be rather helpful to enhance the performance adjusting dilution rate to achieve product in amplified amount.

The plasmid bearing cell has shown an abrupt declining trend at initial hours of the culture, while increasing the dilution rate, as shown in figure (4). There is a very high probability of plasmid loss even at low metabolite concentration on moving towards a high dilution rate. Particularly the probability has shown a rapid hike on increasing dilution rate beyond \( D = 0.45 \) in figure (5).

Figure 5. Showing probability variation of plasmid loss with respect to concentration of metabolite at different rate of dilutions

4. CONCLUSIONS

In our approach structured and unstructured model together taken into an amalgamated form, we can have a more sophisticated mathematical representation that is required to be experimentally validated. An approach of designing sophisticated model put forward a higher level of challenge accommodates all sorts of noise viz. noises at intra-cellular level: genetic noise and interaction noise as well as environmental noise at extra-cellular level. These noise associated properties often incorporates more real time properties to such flexible models.

So, population dynamics model for plasmid bearing and plasmid lacking cells in bioreactor has been made more robust to develop an insilico dynamical system which has the characteristics of a chemostat that we used to employ for streptokinase production. The representation of the dynamical system through modelling has its relevance in predicting the behaviour of the system on disturbances or changes made in initial conditions. In the proposed model new factors are taken into account like selection stress coefficient and metabolite toxicity coefficient that have resolved the simultaneous variation in other parameters and their interaction criteria. A relatively composite model ensures a higher degree of flexibility since it has a number of adjustable parameters and properties.

5. REFERENCES


