CHAPTER 2

LITERATURE REVIEW

2.1 RECOMBINANT PROTEIN PRODUCTION IN ESCHERICHIA COLI

With the advent of recombinant DNA technology, it was discovered that recombinant therapeutic proteins could be produced by *E. coli* in a robust and economic manner (Huang et al 2012). In early 1980s, the FDA approved first *E. coli*-produced recombinant insulin for diabetes treatment. Currently, one third of approved therapeutic proteins have been produced in *E. coli*. The current technology used for commercial production of recombinant therapeutics in *E. coli* and the recent advances that can potentially expand the use of *E. coli* system towards more sophisticated therapeutic proteins have been summarized by Huang et al (2012).

The rising demands for the higher expression levels of recombinant protein necessitate the development of High Cell Density Cultivation (HCDC) techniques in *E. coli* as this may guarantee high product levels. The expression level of foreign proteins in recombinant *E. coli* depends on many factors. They can be classified into genetic factors and operating or environmental factors (Hellmuth et al 1994, Sawers and Jarsch 1996, Saraswat et al 1999, Gill et al 2000 and Grabherr et al 2002). Extensive research work has been carried out to analyze the challenges involved in growing *E. coli* to HCDC using growth strategies, medium optimization and the molecular biology methods (Jung et al 1988, Yoon et al 1994, Lee 1996, Shiloach and Fass 2005 and Shojaosadati et al 2008).
One of the most popular methods to achieve high cell density is fed-batch culture by controlling the nutrient feeding. The feeding strategy in fed-batch culture is critical for its success, as it not only affects the cell’s maximum attainable concentration, but also its productivity. Fed-batch culture is normally carried out under nutrient limited conditions. The effects of post induction nutrient strategies on the production of recombinant proteins have been analyzed by many researchers.

In order to avoid the accumulation of substrate in fed-batch culture broth a new feeding strategy of combining exponential feeding with pH-stat has been suggested by Kim et al (2004). It has been concluded that, to achieve high cell density with reducing formation of acetate, glucose concentration has to be kept low by controlling the specific growth rate at a low value.

Wong et al (1998) investigated the effect of post induction nutrient feeding strategies on the production of bioadhesive protein using an IPTG inducible expression system in *Escherichia coli*. From the experimental work it was observed that linearly changing post induction feeding rate with a suitable slope allowed production of bioadhesive protein up to 5.3 g/L, which was higher than that obtained by the other post induction feeding strategies.

The pilot production of recombinant streptokinase by fed-batch cultivation of *E. coli* has been established by Zhang et al (1999) and Goyal et al (2009). The effects of post induction nutrient feed rates, on recombinant streptokinase production in fed-batch processes, have been investigated by Ramalingam et al (2007) using various feed profiles. A maximum activity per unit biomass has been achieved for exponentially decreasing feed and linearly increasing feed. The decrease in specific growth rate during the post induction
phase was also less pronounced in these cases in comparison to other fed-batch experiments.

2.2 MODELING OF FERMENTATION PROCESS

The mathematical model with kinetic equations explaining the mechanism is required to explain the behavior of the microbial process. A process model is essential for numerical optimization, process scheduling and model-based control of a process to achieve a particular goal, generally achieving higher volumetric productivity of the target protein in recombinant fed-batch processes. The bioprocess model is organized according to five levels of complexity: the molecular level model, the single cell model, the population model, the bioreactor model and the bioplant model (Bailey et al 1983). The first two levels are microscopic process kinetics, and the next two levels are macroscopic process kinetics; a combination of all the levels gives the overall bioprocess kinetics. The model for a dynamic unsteady state process like fed-batch fermentation will be a set of differential equations coupled with nonlinear algebraic equations representing the kinetics of the process. They are generally classified into structured and unstructured models. Since the focus of this research work is to model the fed-batch fermentation process, the following review section is dedicated for structured and unstructured models.

2.2.1 Structured and Unstructured Models

Various structured models have been formulated for the cell level analysis of recombinant microorganism. Lee et al (1985) has formulated structured model for product formation by unstable recombinant microorganism in batch and continuous flow reactors. The cell population is characterized by three different genotypes. The cell level model has been
elaborated to reactor level. The developed model has been used to determine the reactor performance and sensitivity of reactor performance to certain process and host-vector design parameters. Kim and Schuler (1990) have developed a structured, segregated model by asynchronously growing population of genetically modified $E. coli$ cells. The model has been used to predict plasmid instability and distribution of plasmid- originated properties in a population without priori determination of growth rates. Nielson and Villadsen (1992) have developed a structured model describing microbial kinetics by means of selected cell components rather than by the undifferentiated biomass. The substrate consumption rate and the formation of metabolic product have been described by the energy level of the cell such as intracellular ATP and ADP concentration.

A three-compartment structured model has been proposed by Tamerler et al (2001) for simulating batch cultivation of growth, substrate utilization and intracellular $EcoRI$ endonuclease production by $E. coli$, an overproducing recombinant strain containing plasmid pPG430. Model development and verification was based on the dynamic changes in biomass formation and production of intracellular components such as total protein, total RNA, plasmid DNA and $EcoRI$ endonuclease in supplemented minimal media with glucose as the carbon source.

Wong et al 1997 and Santillan and Mackey (2004) have formulated mathematical model of $lac$ operon in $E. coli$ which includes all of unknown regulatory mechanisms, including external glucose dependent catabolite repression and inducer inclusion. They have investigated the influence of external glucose, by means of catabolite repression and regulation of lactose uptake, on the bistable behavior of this system.
Nielsen et al (1991) developed a simple structured model to describe the fermentation process for the production of recombinant protein in *Escherichia coli* where the different biomass components are lumped together in a few intracellular variables. The model was used to describe the dynamic changes in plasmid copy number, e.g., runaway plasmid replication. Mathematical models have been used in systems biology to understand complex biological systems, and to modify and redirect metabolic and regulatory systems based on quantitative predictions (Ishii et al 2004, Goryanin et al 2006 and Usudaa et al 2010).

Unstructured models are the simplest; considering the cell mass as a uniform quantity without internal dynamics whose reaction rates depend only upon the conditions in the liquid phase of the reactor. Thus the unstructured models developed by many researchers contain only the kinetics of growth, substrate uptake and product formation. In contrast to unstructured model, structured models provide information about the physiological state of the cell, i.e., changes in the composition and regulatory adaptation to the environmental changes (Rehm et al 1991). Segregated, unstructured model describing the kinetics of growth and foreign protein production in *E. coli* has been developed by Baheri et al (1997). The developed model has been used to predict the biomass and enzyme concentrations as well as the optimal induction time for higher concentration of substrate.

Cockshott and Bogle (1999) have been developed a simple unstructured model to describe the recombinant protein production in *E. coli*. The have designed model based optimization to predict the optimal glucose feed rate to maximize productivity. Birol et al (2002) have developed simulation software based on detailed unstructured model for penicillin production in fed-batch fermentor. The simulation package was used for monitoring and fault diagnosis of the fermentation process. Ajbar and
Fakeeha (2002) have developed unstructured model of continuous bioreactor for analysis of static and dynamic behavior of the process. From the analysis results they have illustrated the fundamental weakness of unstructured model in predicting the transient behavior in continuous cultures. The effect of kinetic and operating parameters on the stability characteristics of the model also been investigated.

A family of 10 competing, unstructured models has been developed to model cell growth, substrate consumption, and product formation of the pyruvate producing *E. coli* strain in fed-batch processes (Zelic et al 2004). The most suitable model has been identified that reflected the pyruvate and biomass curves adequately by considering a pyruvate inhibited growth and pyruvate inhibited product formation.

Esener et al (1983) have discussed Microbial kinetics and energetics in connection with the formulation of unstructured growth models. The development of microbial energetics and the use of macroscopic methods in the study of microbial growth have been briefly evaluated. A simple unstructured model based on Monod kinetics and the linear relation for substrate consumption is evaluated with reference to extensive experimental and simulation data obtained in batch, fed-batch, and continuous cultivation modes. Influence of an environmental factor, the temperature, on the unstructured model parameters is also quantitatively described.

Kinetic model that considers cell segregation to optimize hEGF expression in fed-batch cultures of recombinant *E. coli* has been proposed by Zheng et al (2005). The effect of cell segregation on the kinetics of growth and foreign protein production has been investigated with the fed-batch
fermentations. The optimal induction strategy has been predicted using the kinetic model for maximum expression of recombinant protein.

Nadri et al (2006) have developed a mathematical model to describe the response of *E. coli* for the production of foreign protein β-galactosidase. A Nonlinear observer has been designed to estimate the substrate consumption during the fermentation process.

### 2.2.2 Neural Network Based Modeling of Fermentation Process

Although first principle models have the advantage of being valid over a wide range of process operations, building such models often requires physical insight into the batch processes and a large amount of time and resources (Xiong and Zhang 2005). To address this issue, empirical models based on process operational data can be utilized. Empirical models can generally be developed very quickly without requiring detailed insight into the processes.

2.2.3 Neural Network based Hybrid Models

Even though the neural network approaches are powerful, the lack of dependence upon physical relationships and poor capacity for extrapolation limits its application in bioprocess applications. Hybrid neural network approaches which combine mechanistic and neural network models have received considerable attention. In this combination the neural network part accounts for the unknown and nonlinear part of the mechanistic model and results in more efficient and accurate prediction of the process dynamics in industries (Lee et al 2002). Psichogios and Ungar (1992) combined a partial first principles model with a neural network. Estimation of process parameters using hybrid model approach has been found to be better than Kalman filtering and NLP optimization methods.

HNN and genetic algorithms has been used to optimize and control a fed-batch fermentation system to cultivate *Bacillus thuringiensis* for thuringiensin production (Zuo and Wu 2000). Semi-real time optimization has been carried out using the HNN model and the measured state variables to re-optimize the system to enhance the productivity.

A full scale coke-plant wastewater treatment process was modeled based on hybrid concept by Lee et al (2002). The Neural network model incorporated into the mechanistic model in parallel configuration, resulted in more accurate prediction of the process with good extrapolation properties, even when the process is upset by shock load of toxic compounds.

Laursen et al (2007) have developed a dynamic hybrid neural network model of an industrial fed-batch fermentation process to produce foreign protein. The fed-batch process has proven difficult to control due to the complex behavior of the bacteria after induction. A gray box modeling approach of parameter function neural networks has been used to capture
dynamic systems and it has been very well used to predict the response of fermentation process.

A neural network based modeling scheme of parameter function modeling that captures the inherently nonlinear dynamics of fed-batch bioreactor systems has been developed by Tholudur and Ramirez (1996), Tholudur and Ramirez (1998) and Tholudur et al (2000). A neural network model in conjunction with basic material and energy balances provides a better representation of the system dynamics than a purely empirical model. It was proved that the combined balance relation and neural network parameter function model can be used with dynamic programming optimization to generate optimal operational policies. The hybrid modeling approach for two fed-batch bioreactor systems were presented and shown to agree well with exact optimal strategies.

Rani and Rao (1999) have listed the following advantages of hybrid techniques in connection with modeling the fed-batch fermentation process. It results in more generalization and extrapolation over standard neural networks and requires less data for training. For process parameters which are rapidly time varying and which are not easily described by parameterized model the hybrid model outperform other estimation techniques such as Extended Kalman filter.

2.2.4 State Estimation

Most of the bioprocess facilities are limited by the unavailability of low-cost online sensors which give reliable measurements reflecting the physiological state of the culture. Model-based state estimation techniques are required which relate the measurable variables with the model variables. The use of the mathematical technique in conjunction with the measurements can
enable the estimation of parameters or process variables that cannot be directly measured.

The power of on-line state estimators which provide estimation of process variables and parameters deliver for better monitoring and control of bioreactor has been demonstrated by Venkateswarlu (2004). Various developments in state and parameters estimation of bioreactors were also discussed.

Adilson and Filho (2000) have presented the state of the art of state estimator techniques. Special attention has been given to filtering techniques, namely extended Kalman filter, adaptive observers, and artificial neural networks (ANN). It has been shown that software based state estimation can be successfully used to enhance the performance of biological systems through monitoring, control and on-line optimization. The multi objective optimization approach to estimate the kinetic parameter of batch and fed batch fermentation processes for ethanol production has been discussed by Wang and Sheu (2000).

Dochain (2003) has presented an overview of state and parameter approaches for chemical and biochemical processes. The basic design concepts of state observers for extended Luenberger and Kalman followed by nonlinear observer and online parameter estimation has been discussed in detail. Komives and Parker (2003) have focused on methods to estimate parameters and process variables that cannot be measured directly and consequent use of inferred measurements for the control of bioprocesses. The great potential of ANN based black box model and metabolic flux model has been discussed for dynamic modeling.

Jenzsch et al (2006 b) has devised a method based on ANN, to estimate biomass concentrations in E. coli fermentation process for the
production of recombinant protein. Veloso et al (2009) have demonstrated the effectiveness of soft sensor for monitoring fed-batch *E. coli* fermentations. They have proved that compared to extended Kalman observer, soft sensor based on asymptotic observer resulted in slightly better performance in monitoring the fermentation process.

### 2.3 OPTIMIZATION AND CONTROL OF FERMENTATION PROCESS

Optimization and control of fermentation process has become an active area of research mainly due to the fact that it is extremely difficult to control. Considerable emphasis on the control of fed-batch fermentation process has been placed because of prevalence in industries (Henson and Seborg 1992).

The determination of optimal feeding rate of a fed-batch bioreactor is an attractive control problem since small improvements in the performance can be significantly cost effective. Several reports have been reported on the optimal operation of a fed-batch fermentor with respect to various objective functions and control variables (Morari and Zafiriou 1997 and Johnston et al 2002). The most common measure of the performance of these systems is the productivity of the culture. It is usually optimized by manipulating the nutrient flow rate according to a predetermined program derived using a fermentation model and an optimizing principle Agrawal et al (1987). Weigand (1981) utilized Pontryagin’s maximum principle to derive optimal feeding strategy to maximize the biomass productivity in a repeated fed-batch fermentor. Menawat et al (1987), Modak and Lim (1992), Seo et al (1992) and Fujioka and Shimizu (1994) have obtained the optimal open-loop time profiles of the feed rate to enhance the productivity of the process.
A systematic approach based on the application of a stochastic optimization method using Genetic Algorithms is widely used for optimizing the bioprocess (Garlapati et al. 2010). Sarkar and Modak (2004) have developed an optimization procedure based on genetic algorithm for the determination of substrate feed profile of fed batch bioreactor. The efficiency of the algorithm has been demonstrated with two fermentation processes secreted protein and yeast cell mass production. Model based optimization of viral capsid protein production in fed-batch culture of recombinant *E. coli* has been proposed by Levisauskas et al (2003).

Optimization of fermentation process is generally performed based on deterministic mechanistic model. The major challenges in developing accurate model have been overcome by development of powerful tools such as artificial intelligence, fuzzy logic and neural network (Tholudur and Ramirez 1999). Application of these tools to optimize the fed-batch fermentation process has been used in the bioprocess industry. Optimization of a fed-batch bioreactor using a cascade neural network model and modified genetic algorithm has been studied by Chen et al (2004 b). It was proved that, the final biomass quantity that yields from the optimal feed rate profile based on neural network model has been greater than the mechanistic model.

Fermentation processes involve a complex interaction between the cell and its environment. Feedback control is necessary because, in real processes, there are variations in the quality of the inoculum and raw materials as well as unanticipated process disturbances which will lead to variations in performance. Control strategies are designed to manipulate the environment based on macroscopically observable variable and not on intracellular parameters which directly affect regulation of metabolism. Skillful design of process control requires a good model for the control law(s)
and the selection of control strategies which manipulate the appropriate process variables to reach the appropriate goal at the right time (Coonor et al 1992). Lee et al (1999) have reviewed various methods of optimization and control of fed-batch fermentation process.

Rani and Rao (1999) have reviewed the developments in the control of batch, fed-batch and continuous fermenters. They have discussed starting from basic conventional controller to advance neural network based controllers for bioprocess industries. They have also explored several tools used to capture the inherent nonlinear and time varying characteristics of fermentation processes. Schugerl (2001) has elaborated on progress in monitoring, modeling and control of bioprocess in his review article.

Extensive developments in the area of process control have recently begun to impact bioprocess development, but much work remains to be done to couple model-based control methods to biochemical reactor technology (Komives and Parker 2003 and Chen et al 2011). Smet et al (2002) has presented an overview of optimal adaptive control of biochemical reactor. Maximum principle of pontryagin the derivation of optimal control sequence for fed-batch fermentation process is briefly revisited.

The specific growth rate ($\mu$) is an essential process variable above all because it is characteristic of the physiological state of microorganisms and related to the biosynthesis of many products of interest. Specific growth rate has an influence on biomass production, product quantity, cellular state, and product quality (Schuler and Marison 2012). In recombinant protein production, the protein synthesizing ability of cells also depending on $\mu$. The metabolic burden associated with recombinant protein production can decrease the growth rate significantly (Panda et al 1999 and
Pinsach et al 2006). High specific growth rates and the presence of high glucose concentrations can lead to overflow metabolite such as acetate which leads to the inhibition of both growth and protein formation. Hence to maximize product yield it is important to maintain low specific growth rates during protein expression phase, (Panda et al 1999, Jenzsch et al 2006 a and Babaeipour et al 2008).

Since the specific growth rate cannot be measured directly it is solely estimated from other variables. Hence, estimation of $\mu$ is dependent on reliable process measurements on the one hand and on accurate, yet simple and robust models. Real-time monitoring tools for quantifying unknown process variables used to estimate the specific growth rate in microbial cultures has been tabulated by Schuler and Marison (2012).

The concept of generic model control has been introduced by Lee and Sullivan (1988). The nonlinear models which can be directly embedded into the controller without resorting to linearization have been demonstrated in their work. The robust controller performance has been achieved for both process disturbance and plant mismatch.

Jenzsch et al (2006 a) demonstrated Generic model control (GMC) as a powerful tool for keeping a microbial cultivation process close to its predetermined control profile. GMC technique has been demonstrated for the green fluorescent protein expressed in genetically modified *Escherichia coli* host cells, found to be very effective in maintaining a predefined complex profile of the specific cell growth rate.

Many research works are still focused on the improvement of GMC controller performance for varies nonlinear processes (Zhoi et al 1992,
Lee et al 2002 and Lee and Samyudia 2003). The Predictive GMC (PGMC) method smooths the control signal by the use of control predictions. It reduces control variation by influencing the control at a particular time with predicted future control inputs (Istre 2004). Further, the robustness of GMC can be improved by changing the control law to incorporate parameter intervals and adapting the process model used in the controller to move smoothly within those intervals.

2.4 METABOLIC ENGINEERING AND FLUX ANALYSIS

The expression of foreign protein in a host organism often changes the metabolism of that organism and result in the imposition of a metabolic load. Glick (1995) has reviewed the causes of metabolic load, the resultant physiological changes to the cells and several strategies to overcome the problems associated with metabolic load.

With the recent advent of systems biology tools, mechanistic understanding of metabolic behavior of the *E. coli* and the underlying changes that are taking place after induction of foreign protein can be analyzed using metabolic flux analysis (Edwards and Palsson 2000, Weber et al 2005 and Taymaz-Nikerel et al 2010).

A metabolic model for cell growth and recombinant overproduction in *E. coli* which includes expression vector properties has been developed by Ozkan et al (2005). By performing metabolic flux analysis it has been observed that the transition in growth condition shift the usage of substrate from anabolic to catabolic pathways to increase their ability to produce energy for protein synthesis. It was also observed that the increase in catabolic fluxes has increased the acetate secretion rate.
The issue of bioprocess dynamical modeling has been addressed by Provost and Bastin (2004). A metabolic flux analysis has been performed to test the consistency of the metabolic network. The elementary flux modes are computed and translated into a set of macro-reactions connecting the extracellular substrates and products. The approach has been illustrated with the example of CHO cells cultivated in stirred flasks on a serum-free medium.

Heyland et al (2011) have investigated the metabolic consequences of recombinant protein synthesis in *E. coli*. The level of gene expression has been controlled by varying inducer concentrations. Changes in the intracellular flux distribution have been quantified from the $^{13}$C-labeling patterns of the DmpA amino acids. A correlation of ATP production and productivity is obtained using metabolic flux analysis to strengthen the hypothesis.

Edwards and Palsson (2000) have utilized Flux balance analysis to interpret and analyze the metabolic capabilities of *E. coli*. They have computationally mapped the metabolic capabilities of *E. coli* using FBA and examined the optimal utilization of the *E. coli* metabolic pathways as a function of environmental variables. An in-silico analysis has been used to identify seven gene products of central metabolism essential for aerobic growth of *E. coli*. They have proved that the computational models of *E. coli* metabolism based on physicochemical constraints can be used to interpret mutant behavior.

The estimation of intracellular fluxes through MFA not only enables the metabolic engineers to tailor the organism for enhanced performance but also can assist the process engineers for the better prediction of the response of the host for the production of recombinant proteins.
Carinhas et al (2011) has presented a hybrid frame work by combining classical metabolic flux analysis with projection to latent structures to further link estimated metabolic fluxes with measured productivities. It has been proved that the hybrid framework is an advantageous tool for metabolic identification and quantification of incomplete or ill-defined metabolic networks.