CHAPTER 1.

INTRODUCTION
The continuous discovery of new vaccines and therapeutics products has concomitantly created the need for the development of efficient systems for the production of these pharmaceutical proteins. The choice of an appropriate expression system and suitable production conditions is crucial for the overall design of such a system especially with regard to downstream processing efficiencies and the final product yield. Among all hosts *Escherichia coli* still remains the most favoured and frequently used microbe for the industrial production of recombinant proteins with therapeutic applications, since it offers several advantages. These include growth on inexpensive carbon sources, rapid biomass accumulation, amenability to high cell density fermentations and simpler process scale-up. Because of its long history as a model system, its genetics is very well characterized and many tools have been developed for chromosome engineering and to facilitate gene cloning and expression (involving combination of expression vectors, recombinant fusion partners, mutant strains etc).

A comprehensive process design for the development and production of a therapeutic recombinant protein involves three major steps: a genetic design; a bioprocess design; and finally a downstream processing design. A major consideration in genetic design is whether the gene product should be produced intracellularly in the cytoplasm or a 'secretion system' should be preferred.

### 1.1 Genetic strategies

Different genetic design strategies together with the inherent properties of the target protein determine the most suitable expression route. Production of proteins in different compartments of *E. coli* offers multiple options in terms of the overall process design with each compartment presenting a few advantages as well as a few disadvantages.

Expression in the cytoplasm can be in the form of a soluble protein or as misfolded protein aggregates usually called 'inclusion bodies'. Soluble cytoplasmic expression is favored in cases where the expressed proteins are small (<100 residues) single domain proteins with fast folding kinetics (e.g.
thioredoxin), are non-toxic to host cells, and the proteins are devoid of disulphide bonds. Accumulation of over-expressed proteins (in their native form) after a threshold level can interfere with the host cell machinery/metabolism, thus affecting the cellular health. Also, simple overcrowding in the cytoplasmic milieu can result in increased rates of protein aggregation as the ‘foldases’ (the folding modulators) get titrated out. This often results in the lowering of the final product yield. Alternatively, failure to rapidly reach a native conformation or to interact with the folding modulators in a timely fashion results in the expression of the recombinant proteins as inclusion bodies (IB’s – which are insoluble aggregates of non-native proteins) in the cytoplasm. Expression as IB’s may help increase the levels of recombinant protein accumulation in the cytoplasm, as these are relatively ‘inert’ molecules. They usually do not interfere with the host cell machinery and additionally are resistant to the action of the host proteases. Though IB isolation procedures from cells are facile, refolding can be a testing and rate-limiting step in downstream processing. Soluble expression in the cytoplasm on the other hand means tedious downstream processing especially in the purification part primarily due to the presence of native cytoplasmic proteins and ‘endotoxins’.

Periplasmic expression is necessary in cases, where disulphide bond formation is critical for protein function or there are multiple disulphides in the protein. It is also important in cases where refolding of the expressed protein is a tedious or problematic procedure. It should however be noted that the efficiency of protein transport across the membrane and the periplasmic space is a serious limitation for high-level recombinant protein expression in comparison to cytoplasmic expression. Thus in general lower expression levels are achieved. However periplasmic expression has a number of advantages over cytoplasmic expression, such as: the expressed protein is obtained with an authentic N-terminus which is critical for therapeutic applications (since the signal sequence gets processed in transport), it facilitates disulphide bond formation, there are fewer proteases present in comparison to the cytoplasm which means reduced levels of proteolytic degradation, fewer contaminating proteins are present in the
Introduction

Periplasm and selective release by osmotic shock facilitates downstream purification. Although, in a majority of expression studies there exists a general tendency to achieve maximum production levels within the cytoplasm, targeting of the recombinant proteins to the extracellular compartment may offer an interesting alternative, especially when cytoplasmic expression leads to cellular toxicity or improper folding. In addition to the advantages associated with periplasmic expression, targeting of the recombinant protein to the extracellular compartment presents other advantages like – the lowest levels of proteolysis as the extracellular compartment is protease deficient, no space limitation for the accumulation of the target protein as is the case with periplasmic expression, the cells need not be lysed or processed for extracting the desired protein, all of which tremendously simplifies the downstream processing steps. However, given that *E. coli* is not known for secreting many proteins into the extracellular medium, manipulation of the various transport pathways to facilitate extracellular secretion of proteins remains a formidable task. Depending upon the expression or the production strategy, the genetic design to develop an optimized expression vector system becomes a critical task (many options are available such as: fusion of the target protein to various partners in an attempt to increase/decrease the solubility; enhancing protein stability from degradation by proteases; selection of the best promoter system/induction system; selection of the most efficient leader sequence for export to periplasm/extracellular space; co-expression of facilitator proteins; selection of an appropriate host). All these factors are decisive for the success of the designed strategy.

1.2 Bioprocess strategies

Bioprocess strategies involve: media optimization, selection of the most productive carbon source, optimization of the culture conditions (temperature, pH, aeration, dissolved oxygen concentration) and specific growth rate of the
culture. Any of these can be critical for developing an efficient process design. A high cell density is required to get high volumetric activities but this can only be achieved by fed-batch cultivation. However the final product yields are critically dependent on the specific productivity (i.e. the production capability of the individual cell). The specific productivity is a function of many parameters where both intracellular and extracellular factors play a role. Many intracellular bottlenecks to expression exist such as the availability of critical metabolites for transcription/translation of desired protein. The extracellular micro environment of the cell determines the exogenous availability of many of these metabolites. Macro parameters like specific growth rate are often useful indicators of this availability and hence determine the overall cellular health. The stress associated with recombinant protein expression can be partially overcome by appropriate bioprocess design which ensures the supply of critical metabolites. Thus the specific product formation rate and the specific growth rate have often been found to be strongly correlated to each other. High cell density has also been suggested to be an important signal for activation of general stress response in *E. coli*. Cell density effects are generally thought to be mediated by external signal molecules that affect gene expression when they reach a threshold concentration via cell to cell communication, a phenomenon called as 'Quorum sensing'. The fact that complex media is required for high growth rates also lead to build up of by products all of which affects productivity in many complex ways. One important technique for controlling the bioprocess has been the design of various feed profiles, both pre and post induction, thus the method of nutrient feeding is critical to high cell density cultivation (HCDC) as it not only affects the final biomass achieved but more importantly helps in maintaining optimum cellular health and hence maximizes expression capability. This feed profile can be designed and the point of induction determined so as to get the maximum expression.
1.3 Downstream strategies

Downstream processing considerations for an intracellular product are quite different from those of an extracellular or secreted product. Systems where the product is secreted from the cell into the extracellular space are gaining favor due to the tremendous simplification and reduction of the downstream purification steps. Product recovery is often the most difficult and expensive part and, for some recombinant DNA derived products, purification accounts for 80 - 90% of the total processing cost. Though downstream processing is equally important as the other two aspects mentioned earlier, it is not discussed in detail here, as it is not the major focus of the present work.

1.4 Asparaginase

The present work deals with the upstream development and preliminary scale-up studies of a therapeutically important protein namely asparaginase. Asparaginase is a chemotherapeutic agent used in treatment of acute lymphoblastic leukemia (ALL) for more than 35 years. ALL is the most common malignancy in children, it accounts for one fourth of all childhood cancers and approximately 75% of all cases of childhood leukemia. Approximately 2500 to 3000 children are diagnosed with acute leukemia each year in the United States; though exact figures citing the number of affected and new cases of ALL in India does not exist, the number is expected to be high as ALL is the most common cancer occurring in children with an annual rate of approximately 30 to 40 new cases per million.

E. coli L-asparaginase II is administered in combination with other drugs to induce a remission of the disease (induction therapy) and at a later stage continued application of asparaginase helps in preventing further outbreaks (maintenance therapy). Although long-term therapy results in some complicated side effects, E. coli asparaginase still remains the most widely used in disease remission and maintenance therapy. The cure rate has now improved to 80% in
children and 30 - 40% in adults by optimization of use of the existing drugs. The general medical approach to leukemic therapy is based upon the metabolic defect in L-asparaginase synthesis of some malignant cells, though the enzyme is also known to inhibit protein synthesis by L-asparagine hydrolysis. Apart from acute lymphoblastic leukemia (ALL) asparaginase is also used in treatment of Hodgkin disease, acute myelocytic leukemia, acute myelomonocytic leukemia, chronic lymphocytic leukemia, lymphosarcoma, reticlesarcoma and melanosarcoma.

Keeping in mind the above-mentioned considerations, the overall objective of this work was the design of a bioprocess strategy to obtain a high-level recombinant asparaginase production. This would include both genetic and bioreactor based strategies to optimize protein production while keeping in mind the need to link this up with a simplified purification strategy which would help in the comprehensive design of a production process.

1.5 AIMS & OBJECTIVES

The aim of this work was to design a strategy for optimal expression of recombinant asparaginase in its biologically active form and further studies to optimize its production in a lab scale bioreactor.

Two alternative expression strategies were envisaged for the production of recombinant asparaginase in *E. coli*.

1. Cytoplasmic expression of asparaginase in the form of inclusion bodies and subsequent refolding to get the soluble and active protein from purified inclusion bodies.

2. Extracellular/periplasmic expression of asparaginase to get soluble expression *in vivo* and its subsequent purification.
The work was further sub-divided into the following components:

- PCR amplification and cloning of the asparaginase gene (i.e. the region coding for the mature asparaginase protein without the signal sequence) from *E. coli* K-12 strain (JM109).

- Development of various host-vector combinations and expression studies at the shake-flask level.

- Study of various culture conditions to optimize expression (media composition, temperature, induction strategy, antibiotic concentration, time of induction, carbon source, and additives) at the shake-flask level.

- Optimization studies for high-level production of recombinant asparaginase using high cell density cultivation (HCDC) techniques in a lab-scale bioreactor.

- Design of a purification strategy followed by the characterization of the purified recombinant asparaginase.