CHAPTER I

INTRODUCTION

1.1 Introduction

Advances in genetic engineering have made possible the production of therapeutics and vaccines for human and animals in the form of recombinant proteins (Ghosh, 1998; Padh, 1999). These biotechnology derived recombinant proteins form a new class of drugs for many ailments like genetic disorders, cancer, hypertension and AIDS for which we have no better treatment or cure. Unlike chemical drugs, biologicals are our own molecules and hence more compatible with biological systems. At present there are more than 120 biotechnology derived therapeutics and vaccines approved by US FDA (Food and Drug authority) for medical use and over 800 additional drugs and vaccines are in various phases of clinical trials. In addition, use of DNA, proteins and enzymes in diagnostics is increasing exponentially. Industrial uses of enzymes in food, textile, leather, detergent, medicinal chemistry sectors is also increasing rapidly. The growing need of therapeutic and other applications of enzymes and proteins could only be met by heterologous synthesis of recombinant proteins (Ghosh, 1998; Padh, 1999). **Table I** indicates key therapeutic recombinant biotech products with their recent market sales and names of key manufacturers. **Table II** shows how rapidly many new therapeutic products have recently entered the market and the packed pipeline of product development and clinical trials. Many of these products will be approved in the near future adding to the growing biotech product line. Indian share in consumption of biotech products is very insignificant totaling only about 1.2 billion USD (Ghosh, 1998), most of which is by trading of imported goods **(Table III)**. Recently new companies like **Shanta Biotech** and **Bharat Biotech** have come up in Hyderabad producing hepatitis B vaccine with few other products in the pipeline. Similarly few pharma companies like Cadila Pharma, Zydus Cadila, Dr.Reddy's Lab, INTAS, etc. have also initiated work in production of recombinant therapeutic proteins.

1.2 Heterologous production of proteins

Protein over expression refers to the directed synthesis of large amounts of desired proteins. The heterologous production of proteins and enzymes involves two major steps:

Product	**1998 Sales ($)**	**Makers**
H-Insulin	3.0 Billion	Eli Lily, Novo Nordisk Hoechst, Yamanouchi
G-Factors	2.2 Billion	Eli Lily, Novo NordiskGenentech,Pharmcia
Blood Factors	5.5 Billion	Amgen, J & J, Sankyo Chugai,Sandoz
mABs	0.5 Billion	J & J, Cytogen
Interferons	3.8 Billion	Schering-Plough, Roche Daiichi, Wellcome.

1.1 TABLE I - KEY THERAPEUTIC PRODUCTS

Year	Approved	Pipeline
1994	17	-
1995	33	450
1996	79	700
1998	>100	1200

1.1 TABLE II - THERAPEUTIC PIPELINE

	Biotech sales (Rs. in crores)	
Category	**1995**	**2000**
Healthcare	1959	3532
Agriculture	154	385
Industrial Products	570	1500
Others	30	130
TOTAL	**2793**	**5547**
USD in million	**635**	**1260**

(Ghosh, 1998)

1.1 TABLE III - BIOTECHNOLOGY SALES IN INDIA

1. Introduction of foreign DNA into the host cell. This step has three major considerations.
    a. Identification and isolation of the DNA to be introduced
    b. Identification of the vector and construction of recombinant vector
    c. Identification of the suitable expression system to receive rDNA.
2. Factors affecting the expression of foreign DNA for protein synthesis in the chosen expression system.

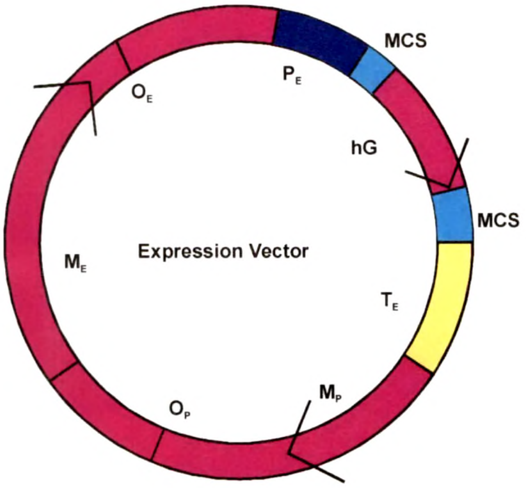
Briefly, at present a variety of vectors are available to ferry DNA in and out of cells: plasmids, lambda phage, cosmids, phagmids, artificial chromosomes from bacteria, yeast or human origin [BAC, YAC and HAC (Ikeno, 1998) respectively]. The vectors could either be integrating (becomes part of host's chromosomes) or extrachromosomal. They could be in copies varying from one to several hundred.
In general, expression vectors have the following attributes **[Fig. I(a)]:**

- **Ori:** sequence that allow their autonomous replication within the cell.
- **Promoter:** a tightly regulated promoter, that is one which can be switched on and off easily, is desirable.
- **Selection Marker(s):** sequences encoding a selectable marker that assures maintenance of the vector in the host.
- **Terminator:** a strong transcriptional terminator should be used with a strong promoter to ensure that the RNA polymerase disengages and does not continue to transcribe downstream genes.
- **Polylinker:** to simplify the insertion of the heterologous gene in the correct orientation within the vector.

In addition, artificial chromosomes have centromeric and telomeric sequences which are host specific (Ikeno, 1998) **[Fig. I(b)]**. These attributes permit artificial chromosome vectors to be faithfully replicated and distributed to daughter cells during cell division. Use of artificial chromosomes in commercial production of heterologous proteins remains unexplored.

1.3 Available expression systems

With the ability to clone and express a foreign gene in the heterologous host, came



1.2 Fig. I(a) - Design of typical shuttle vector for heterologous gene expression

MCS - multiple cloning sequence

O_p - bacterial origin of replication

M_p - marker gene for selection in bacteria

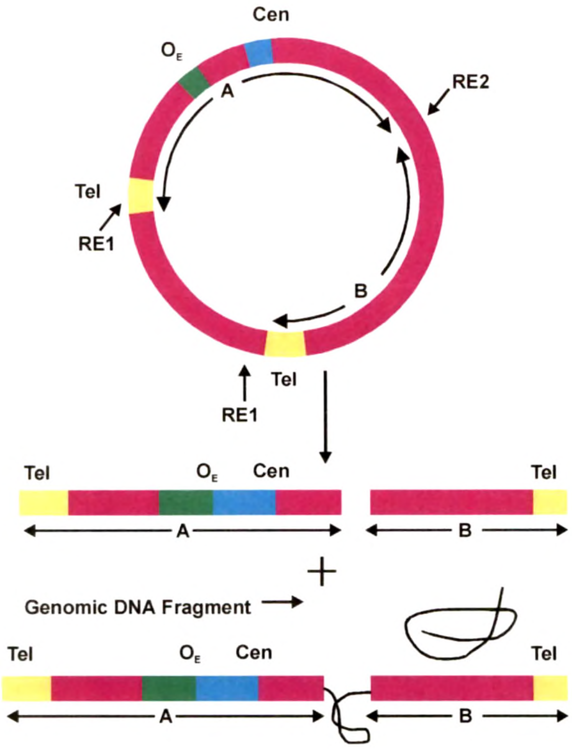
M_E - marker gene for selection in eukaryotes

O_E - eukaryotic origin of replication

P_E - eukaryotic promoter sequence

T_E - eukaryotic terminator sequence

hG - heterologous gene for expression



1.2 Fig. I(b) - Design of an Artificial Chromosome

RE1 and RE2 - restriction enzymes cleavage sequences

O$_E$ - eukaryotic origin of replication

CEN - respective centromeric sequence

TEL - respective telomeric sequence

A and B - fragments generated by cleavage of the vector by RE1 and RE2.

Cleavage of the vector by a mixture of RE1 and RE2 generates fragments A and B and another fragment which is discarded. A and B fragments have TEL sequences at one end. Larger fragments of targeted genomic DNA to be cloned can be bracketed by ligation with fragments A and B as shown. Since the construct has centromeric and telomeric sequences it can replicate as a chromosome in appropriate eukaryote.

a remarkable capability to make almost any protein in abundant quantity to be used as therapeutic or diagnostic agents. Prokaryotic and eukaryotic systems are the two general categories of expression systems.

1.3.1 Bacterial system

E. coli is by far the most widely employed host, provided post translational modifications of the product are not essential. Its popularity is due to the vast body of knowledge about its genetics, physiology and complete genomic sequence which greatly facilitates gene cloning and cultivation (Shatzman and Rosenberg, 1987; Casadaban *et al.,* 1983). High growth rates combined with the ability to express high levels of heterologous proteins i.e. strains producing up to 30% of their total protein as the expressed gene product result in high volumetric productivity. Furthermore, *E. coli* can grow rapidly to high densities in simple and inexpensive media. Strains used for recombinant production have been genetically manipulated so that they are generally regarded as safe for large scale fermentation. Purification have been greatly simplified by producing recombinant fusion proteins which can be affinity purified e.g. glutathione-S- transferase and maltose-binding fusion proteins (Maina *et al.,* 1988). However expression in bacteria does have some serious disadvantages. It poses significant problems in post translational modifications of proteins. Common bacterial expression systems such as *E. coli* have no capacity to glycosylate proteins in either *N* or O-linked conformation. Although other bacterial strains such as *Neisseria meningirulls* have recently been shown to O-glycosylate some of their endogenous proteins, the trisaccharide added is different from O-linked sugars found in eukaryotes. Protein expressed in large amounts often precipitates into insoluble aggregates called inclusion bodies, from which they can only be recovered in an active form by solubilization in denaturing agents followed by careful renaturation (Marston and Harttley, 1990). Lysis to recover the cytoplasmic proteins often results in the release of endotoxins, which must be removed from the final product. Currently strategies to secrete the target proteins by translocation into the periplasmic space or to release the target proteins by linking to existing excretory systems are being developed (Robinson *et al.,* 1984). Additionally, the efficiency of expression will also depend on differences of codon utilization by bacteria (Schein *et al.,* 1989). At times the original sequence of the heterologous gene has to be modified to reflect the codon usage by the chosen expression system. *E. coli* has toxic cell wall pyrogens and hence products need to be tested more extensively before use.

1.3.2 Yeast

Yeast is the favoured alternative host for expression of foreign proteins for research, industrial or medical use (Hitzeman *et al.,* 1981). As a food organism, it is highly acceptable for the production of pharmaceutical proteins. In contrast, *E. coli* has toxic cell wall pyrogens and mammalian cells may contain oncogenic or viral DNA. Compared to mammalian cells, yeast can be grown relatively rapidly (doubling time about 90 minutes) on simple media and to high cell density, and its genetics is more advanced than any other eukaryote, so that it can be manipulated as readily. Added advantages are the availability of complete genomic sequence, the nuclear stable high copy plasmids, and ability to secrete the target protein (Hitzeman *et al.,* 1990). *Saccharomyces* strains have high copy stably inherited plasmid of 6.3 kb known as 2 micron plasmid which codes for 4 genes FLP, REP1, REP2, and D. It also contains an ORF, STB locus (required in *cis* for stabilization) and two 599 bp inverted repeat sequences. FLP encodes a site-specific recombinase which promotes flipping about the FLP recombination targets (FRT) within the inverted repeats, so that cells contain two forms of 2 micron plasmids, A and B. The simple 2 micron shuttle vectors contain the 2 micron ORI-STB, a yeast selectable marker and bacterial plasmid sequence (*ori* and selection markers) and are used in host strains which supplies REP1 and REP2 proteins.

These lower eukaryotic systems are able to glycosylate the target proteins, but it has been shown that both *N*- and *0*-linked oligosaccharide structures are however significantly different from their mammalian counterparts (Kukuruzinska *et al.,* 1987; Kornfeld and Kornfeld, 1985). Hypermannosylation (addition of a large number of mannose residues to the core oligosaccaride), is a common feature in yeast hindering proper folding and therefore the activity of protein. At the moment yeast provides a good compromise between bacteria on one side and mammalian cell-lines on other.

1.3.3 Insect

The baculoviruses have emerged as a popular system for overproducing recombinant proteins in eukaryotic cells (Kitts and Possee, 1993; O' Reilly *et al.,* 1992). Several factors have contributed to its popularity. Being a eukaryote it uses many of the protein modifications, processing, and transport systems present in higher eukaryotic cells (Matsuura et. al., 1987). It uses a helper-independent virus that can be propagated to high titers in insect cells adapted for growth in suspension cultures, making it possible to obtain large amounts of recombinant proteins with relative ease. Expressed proteins are usually expressed in the proper cellular compartment i.e. membrane proteins are usually localized to the membrane, nuclear proteins to nucleus, and secreted proteins secreted into the medium. Majority of the overproduced proteins remain soluble in insect cells. Viral genome is large (130 kb) and thus can accommodate large fragments of foreign DNA. Baculoviruses are non-infectious to vertebrates, and their promoters have been shown to be inactive in mammalian cells which gives them a possible advantage over other systems when expressing oncogenes or potentially toxic proteins. Also the process development time is short. Expression using baculoviral vectors also have some limitations. Since baculoviruses infect invertebrates, it is possible that the processing of proteins produced by vertebrates is different and this seems to be the case for some post translational modifications e.g. internal proteolytic cleavages at arginine- or lysine-rich sequences are highly inefficient. The glycosylation capability is generally limited to producing only high mannose type and not processed to complex type oligosaccharide containing fucose, galactose and sialic acid.

1.3.4 Mammalian cells

(Kaufman, 1990a; Kaufman, 1990b)

Ideally, proteins requiring mammalian post translational modifications should be expressed in mammalian cells. If product authenticity is absolutely essential for clinical efficacy, then despite the many short comings, a mammalian host is the only choice, as it offers the greatest degree of product fidelity. It should however be noted that oligosaccaride processing is species and cell type dependent among mammalian cells. Differences in glycosylation pattern are reported in rodent cell lines and human tissues. Even the use of human cell line is not perfect, since the transformation event required in most cases to produce a stable cell line may itself result in altered glycosylation profiles. Also mammalian expression techniques are time consuming and much more difficult to perform on large scale. Complex nutrient requirement and low product concentration have meant that the end product must be highly value added for this approach to be commercially viable.

1.4 Factors affecting intracellular expression

Having the target DNA in an appropriate vector in the expression system of choice is the first step in optimising production of the heterologous proteins. Within a given system the transcription and translation processes leading to the heterologous protein production are complex set of reactions Each process is carried out and controlled by several enzymes/factors. In recent years we have learned that the following few key steps or reactions are critical in determining ultimate outcome.

1.4.1 Initiation of transcription

Gene expression is most frequently regulated at the level of transcription, and it is generally assumed that the steady-state mRNA level is a primary determinant of the final yield of a foreign protein. The mRNA level is determined both by the rate of initiation and the rate of turnover. In most cases the yield of a foreign protein expressed using a host promoter has been much lower than the yield of the homologous protein using the same promoter (~50%) (Mellor *et al.,* 1985; Chen et.al. 1984). Many factors could account for these differences e.g. downstream activating sequences (DAS) and upstream activating sequences (UAS) (Purvis.*et al.,* 1987). If DASs are characterized, it may be possible to incorporate them into upstream promoter fragments in order to create more efficient expression vectors.

Alternatively if DAS proves to be strongly position dependent, they could be placed within an intron which could be excised prior to translation. If neither of these options work, then maximal transcription will only be possible using fusion proteins.

1.4.2 RNA Elongation

The elongation of transcripts is not thought normally to affect the overall rate of transcription, but the yield of full length transcripts could be affected by fortuitous sequences in foreign genes which cause pausing or termination. These could either act in the same way as natural host terminator or else by a different mechanism e.g. in yeast, though not widely recognized, this problem could be a very common reason for low yields or complete failure of expression of foreign genes. At present the only solution is to increase the AT/GC of offending section of genes by chemical synthesis.

1.4.3 RNA stability

(Humphrey et. al., 1991; Romanos et. al., 1991)

There is evidence that subtle changes in mRNA sequence affect the stability of mRNA and low mRNA stability being a primary factor in poor yields of foreign proteins. Where mRNA instability is diagnosed as a problem, overall yield might be improved by (i) using a more powerful promoter, (ii) using a promoter with more rapid induction kinetics, or (iii) chemically synthesizing the gene with altered codons or deleting the 3' untranslated region in the hope that instability determinant will be removed. Degradation of mRNA is also more pronounced under adverse growth conditions.

1.4.4 Gene dosage

Since the target gene is often incorporated into a plasmid vector system, gene dosage is dependent on plasmid copy number. As can be expected, an increase in copy number results in concomitantly higher recombinant protein productivity, but not indefinitely. Plasmid copy number is affected by plasmid and host genetics and also by cultivation conditions such as growth rates, media and temperature (Hughes and Welker, 1989).

1.4.5 Initiation of translation

(Kozak, 1989)

Translational efficiency is a function of either translational initiation or elongation rates. Translational efficiency is controlled primarily by the rate of initiation. Initiation in eukaryotes is thought to follow a scanning mechanism whereby the 40S ribosomal subunit plus co-factors bind the 5' cap of the mRNA and then migrate down the untranslated leader scanning for the first AUG codon. Any part of this process, which is affected by the structure of the leader and the AUG content, could limit the initiation rate. AUG is recognised efficiently as initiation codon only when it is in right context and optimal content is found to be GCC(A or G)CC**AUGG**. The purines (A or G) three bases before AUG and G immediately following it are found to be the most important, influencing translation to the tune of 10-fold. The following factors have also been found to be important for prokaryotes (i) the ribosome binding nucleotide sequence or Shine-Dalgarno (S-D) sequence; (ii) the distance between the initiation codon and S-D & (iii) the secondary structure of mRNA.

1.4.6 Translational elongation

Translational elongation does not effect the yield or quality of polypeptide normally, but it can become limiting with very high mRNA levels. Codon usage is considered as a potential factor affecting the product yield. Despite the degeneracy of the genetic code, a non-random codon usage is found in most organisms. The codon usage of most genes reflects the nucleotide composition of the genome, highly expressed genes shows a strong bias towards a subset of codons (Bennetzen and Hall, 1982). This major codon bias, which can vary greatly between organisms, is thought to be a growth optimization strategy such that only a subset of tRNAs and aminoacyl-tRNA synthetases are needed at high concentration for efficient translation of highly expressed genes at fast growth rates (Kurland, 1987). Rare codons, for which the cognate tRNA is less abundant, are translated at a slower rate but this will not normally affect the level of product from an mRNA since initiation is usually rate limiting (Pedersen, 1984). A ribosome finishing translation of one mRNA molecule is most likely to initiate translation of a different mRNA species, unless the original species comprises a large proportion of the total mRNA. Thus, the overall rate of translation of an rnRNA is not usually affected by a slower elongation rate unless ribosome's become limiting, which would affect all transcripts in the cell. In contrast to the normal situation, there is evidence that codon usage may affect both the yield and quality of a protein when a gene is transcribed to very high levels. With very high levels of mRNA containing rare codons, aminoacyl-tRNAs may become limiting, increasing the probability of mistranslation (Scorer *et al.,* 1991), which is the incorporation of an amino acid which does not correspond to the codon being translated, and possibly causing ribosomes to drop off. Thus the codon content of a foreign gene may influence the yield of protein where the mRNA is produced at very high levels. This may be more likely to occur on growth in minimal medium, when the cell produces a wide variety of biosynthetic enzymes, encoded by genes containing rare codons. The effect on product quality has been difficult to measure but requires further attention since it has further implications for therapeutic proteins. Proteins containing amino acid mis-incorporation are difficult to separate and may affect the activity and antigenicity of the product. Since small genes are now frequently synthesized chemically, they may be easily and perhaps profitably engineered to contain optimal codons for high level expression. mRNA secondary structure, in addition to codon usage may affect translational elongation (Baim *et al.,* 1985).

1.4.7 Polypeptide folding

During or following translation, polypeptide must fold so as to adopt their functionally active conformation. Since many denatured proteins can be refolded *in vitro*, it appears that the information for correct folding is contained in the primary polypeptide structure (Gething and Sambrook, 1992). However, folding comprises rate-limiting steps during which some molecules may aggregate, particularly at high rates of synthesis and at higher temperature. There is evidence that certain heat shock proteins act as molecular chaperones in preventing the formation and accumulation of unfolded aggregates, while accelerating the folding reactions. Due to the intrinsic nature of polypeptide folding and low specificity of chaperons, it is very unlikely that foreign cytosolic proteins will accumulate in non-native conformations, but when fragments of proteins or fusion proteins are expressed, however, normal folding domains may be perturbed resulting in an insoluble product. Insoluble proteins can often be renatured *in vitro* though the techniques for this can be complex and unpredictable (Marston, 1986). In contrast to intracellular proteins, naturally secreted proteins encounter an abnormal environment in the cytoplasm; disulfide bond formation is not favoured and glycosylation cannot occur. In *E. coli,* foreign proteins are frequently insoluble but low temperature has been found to increase solubility in some cases (Schein and Noteborn, 1988). This may be due to a decreased translation rate or to the fact that hydrophobic interactions, such as occurring in aggregates, become less favourable. A dramatic increase in the yield of active, soluble protein is observed on reducing the rate of induction (Kopetzki *et al.,* 1989).

1.4.8 Post-translational processing

Prokaryotic expression systems are generally useful for producing heterologous proteins from cloned eukaryotic cDNA. In some cases, however eukaryotic proteins that have been synthesized in bacteria are either unstable or lack biological activity. The inability of prokaryotic organisms to produce authentic versions of proteins is for the most part, due to the absence of appropriate mechanisms for generating certain post-translational modifications. In eukaryotes there are number of modifications that may occur at the post translational stage, after protein synthesis is complete.

- **Amino-terminal modifications** of polypeptides are the commonest processing events and occur on most cytosolic proteins (Kendall *et al.,* 1990). Two types of events normally occur: removal of the N-terminal Met residue, catalyzed by Met aminopeptidase (MAP), and acetylation of the N-terminal residue, catalyzed by N-acetyltransferase (NAT). Both enzymes are associated with ribosomes and act on nascent polypeptide. In most cases the structure of N-terminus should not affect the biological activity of a protein, but there may be exceptions, e.g. the response of hemoglobin to physiological modifiers involves the N-terminus, and correct folding of alpha and beta globins is therefore advantageous. Similarly N-acetylation of melanocyte-stimulating hormone is required for full biological activity.

- **Disulfide bond formation:** In eukaryotes formation of disulfide bond (cys-s-s-cys) occurs in the lumen of RER and is mediated by an enzyme called disulfide isomerase (Freedman, 1989). Disulfide bond is confined to secretory proteins and exoplasmic membrane proteins. This is important in stabilisation of tertiary structure. An improperly folded protein is unstable and lacks activity.

- **Proteolytic cleavage** of a precursor form is required in some cases. Selected segments of amino acid sequences are removed to yield a functional protein (Thim *et al.,* 1986).

- **Glycosylation:** Glycosylation is the most extensive of all the post-translational modification and has important function in secretion, antigenicity and clearance of glycoproteins (Rademacher *et al.,* 1988). Oligosaccarides can attach to proteins in three ways (i) via an *N*-glycosidic bond to the R-group of an *Asn* residue within the consensus sequence *Asn-X-Ser/Thr* (N -glycosylation) (Kornfeld and Kornfeld, 1985). All mature *N*- linked glycan structures have a common core of *Man.GlcNAc,* which can form part of simple oligomannose structures or be extensively modified by other residues such as fucose, galactose and sialic acid. Hybrid structures also exist where one or more arms of the glycan are modified and the remaining arms contain only mannose, (ii) via an *O*-glycosidic bond to the R group of the *Ser* or *Thr* (*O*-glycosylation). *O*-linked glycosylation is extensive in structural proteins such as proteoglycans. Small glycan structures can also be *O*-linked to the side chain of hydroxylysine or hydroxyproline (iii) carbohydrates are also components of the glycophosphotidylinositol anchor used to secure some proteins to cell membrane. The presence of these consensus sequences by no means guarantees their glycosylation. They show varying degrees of occupancy with oligosaccarides (macroheterogeneity) depending on their position within the protein and its conformation, the host cell type used for expression, and its physiological status. These three factors also determine the extent of variation in the type of sugar residues found within each oligosaccaride (microheterogeneity). Glycosylation is both organism and cell type specific and therefore expression of a protein in a heterologous system will almost certainly result in a product with different modification from the native protein. This may affect the function and immunogenicity of the protein (Parekh *et al.,* 1989; Kukuruzinska *et al.*, 1987)

- **Modification of amino acid** within proteins: Modifications of this type include phosphorylation, acetylation, sulfation, acylation, (carboxylation, myristylation, and palmitylation) (James and Olson, 1990).

Of these modifications, prokaryotic host cells are least likely to carry out either proper glycosylation or additions to specific amino acid within the heterologous protein.

1.4.9 Stability of intracellular proteins

So far processes affecting the rate of synthesis of proteins have been considered, but the ultimate yield is equally affected by the rate of degradation (Dice, 1987). Yields might logically be improved by the following measures: (i) fusion to a stable protein (Lees *et al.,* 1984; Cousens et. al., 1987) (ii) secretion to segregate the product from intracellular proteases (Itoh *et al.,* 1986) (iii) using a more rapidly induced promoter (iv) using additional protease inhibitors to minimise degradation during extraction (v) inducing at lower temperature (vi) harvesting cells in the exponential growth phase.

1.4.10 Stability of plasmid

Plasmid instability is a major problem in continuous and large scale fermentation, since these cultures go through many generations. The resulting effects are lower productivity and increased production cost because of the build-up of non-productive plasmid free cells. Plasmid instability are categorized as segregational instability (Cashmore *et al.,* 1986; Scott, 1984) and structural instability (Novick, 1987; Ahem *et al.,* 1988). Segregational instability is the loss of plasmid from one of the daughter cells during division because of defective partitioning. Structural instability is attributed to deletion, insertion and rearrangement in plasmid structure resulting in a loss of desired gene function. Plasmid stability is influenced by the plasmid vector and host genotypes; the same plasmid in different hosts exhibit different degree of stability and vice versa. The origin and size of foreign DNA has been observed to affect the plasmid stability. Plasmid stability is also a function of physiological parameters that affect the growth rate of the host cell, which includes pH, temperature, aeration rate, medium components and heterologous protein accumulation. Mathematically structured and unstructured kinetic models of plasmid stability have been developed which are ultimately useful for the design of recombinant processes.

1.5 *Dictyostelium discoideum* as an expression system

Prokaryotic systems are generally easier to handle and are satisfactory for most purposes. However, there are serious limitations in using prokaryotic cells for the production of eukaryotic proteins. For example, many of the eukaryotic proteins undergo a variety of post-translational modifications like proper folding, glycosylation, phosphorylation, formation of disulphide bridges etc. There is no universal expression system for heterologous protein production. All expression systems have some advantages as well as some disadvantages that should be considered in selecting which one to use. Choosing the best one requires evaluating the options- from yield to glycosylation, to proper folding, to economics of scale up. Moreover no single eukaryotic host cell system is capable of performing all the possible post-translational modifications for every potential heterologous protein. Therefore, if a particular protein requires a specific set of modifications, then it may be necessary to examine different eukaryotic expression systems to find the one that can produce a biologically authentic product. The soil amoeba *D. discoideum* is an organism that provides an attractive alternative for heterologous expression of recombinant proteins. Though it can be grown and transformed with the same ease as the yeast *saccharomyces*, it has some complex features that resemble mammalian cells, such as glycosylation and chemotaxis. It is capable of expressing functional heterologous proteins, which are glycosylated, secreted or inserted into the membrane.

1.5.1 Classification

(Kessin, 2001)

Dictyostelium, often referred to as "slime mold" or "social amoeba", was first described by the mycobiologist Oskar Brefeld in 1867 as *D. mucoroides* and initially classified as a fungus. He named the species as *Dictyostelium* (Dicty means net like and stelium means tower) because the aggregation territories he observed looked like nets and the fruiting bodies like towers. He added the qualifier *mucoroides* because the new organism resembled the fungus *Mucor.* It soon became clear that *D. discoideum* is not a true fungus since the germination of spores led not to hyphae and a mycelium but to distinctly amoeboid cells. Also it did not have the same sporangial walls as the fungus, but instead the spores were suspended in a drop of liquid. Phylogenetic analysis of the sequence of its proteins led to the classification together with *Physarum* and *Planoprotostelium*, as a monophyletic group which is different from animals, fungi, plants and protists but more closely related to animals and fungi.

1.5.2 Life cycle

Dictyostelium is a multicellular organism that feeds on bacteria (Depraitere and Darmon, 1978; Raper, 1937) and divides by simple binary fission every 8-10 hours. Under starvation, cells become chemotactically sensitive to gradients of cAMP, which is released from aggregation centres in pulses. Cells eventually converge into aggregates of ~10^5 cells. A complex extracellular matrix of protein, cellulose and polysaccharides surround the cells to form a sheath that isolates the developing structure (Hohl and Jehli, 1973). In laboratory, *D. discoideum* cells can be grown in axenic media that allows the recovery of large quantities of cells. Development can be induced synchronously by washing the cells free of nutrients, and laying them on agar plates or nitrocellulose filters supported by buffered pads.

After aggregation, several cell types arise that will organize themselves spatially in the developing structure. A tip is formed in the aggregate that elongates to give rise to a finger shaped structure. This finger may initiate a transient period of migration named slug stage. The anterior 20% region of the finger (and also the slug) is composed of prestalk cells that will eventually form the stalk at culmination. Prestalk cells are not a homogenous population (Early *et al.,* 1993; Gomer *et al.,* 1986; Jermyn *et al.,* 1989; McRobbie *et al.,* 1888). The gene encoding the extracellular matrix protein EcmA defines two subpopulations of prestalk (pst) cells: pst A cells, located at the most anterior part, and pst O cells, that lie at the posterior part of the stalk region. A central core of cells in the prestalk region expresses another extracellular matrix protein, EcmB. The expression of this protein, together with EcmA defines the pstAB. Another type of cell, the ant-like cell (ACL) is dispersed through out the structure (Devine and Loomis, 1985; Sternfeld and David, 1982). ACLs are very heterogeneous in their pattern of gene expression, and subpopulations of ACLs express either *EcmA* or *EcmB*, one, two or none of the two markers A or B (Gaskell *et al.,* 1992). During terminal differentiation, some anterior-like cells move upwards to form the upper cup whereas other ALCs migrate downwards to form the lower cup that surrounds the sorus. This cell type also contributes to the formation of the basal disc that attaches the stalk to the substratum (Jermyn et. al., 1996). Cells fated to become spores (prespore cells) are located in the posterior region of the finger. Prespore cells express specific gene products such as the spore coat proteins encoded by the *cotA, cotB* and *cotC* genes.

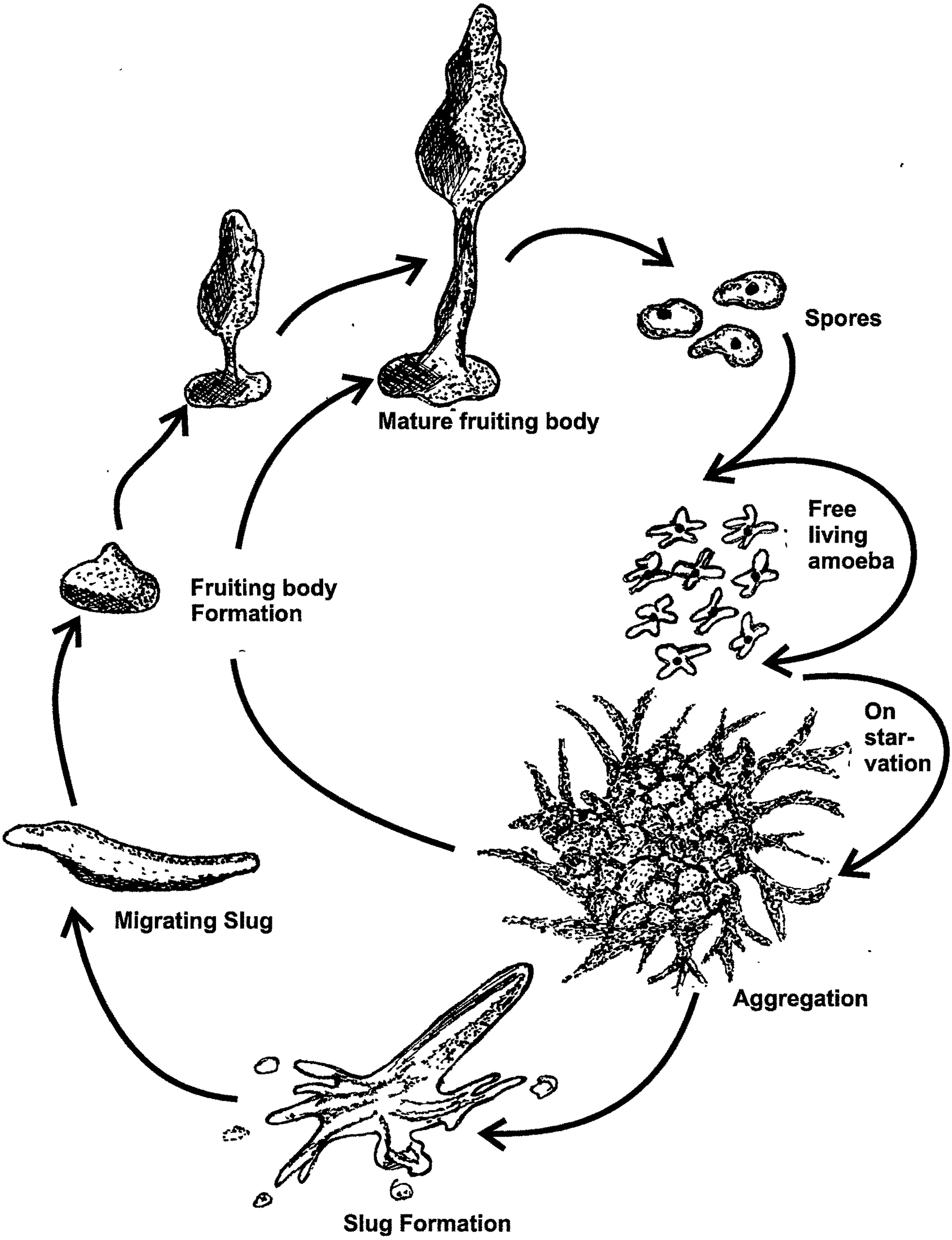
Depending on environmental conditions at the finger stage *D. discoideum* structures must choose between two developmental pathways: either direct culmination at the site of aggregation or transient formation of migrating slugs, the latter allowing culmination in a different spot. Slugs are phototactic and thermotactic (Bonner *et al.,* 1950; Bonner, 1998) and their migration is achieved by the coordinated movement of individual cells that makes traction on the surface sheath, which is continuously produced and left behind as the slug moves. Cells in the posterior region have linear trajectories, but the prestalk cells in the anterior zone rotate around the long axis. Orchestrating culmination requires coordination of cell movement and terminal differentiation of each cell type. The prestalk cells from the tip build the stalk tube, which grows from the top downwards to lift the spore head over the substratum, thus favouring the dispersal of the spores (Chen *et al.,* 1998; Sternfeld, 1998) **[Fig. I(C)]**.

1.5.3 Genome

The *D. discoideum* genome is haploid, and a size of 34 Mb (Cox *et al.,* 1990; Kuspa and Loomis, 1996) organized in six chromosomes. Copies of extra chromosomal DNA, which comprises nearly 18% of the nuclear DNA mass, encode the ribosomal RNAs. The mitochondrial genome also comprises a third of the total cell DNA and has already been sequenced. The A+T content of the DNA in *D. discoideum* is very high. In protein coding regions the base composition is skewed to an average of 60-70% (A+T) but the regions that do not code for proteins contain a much higher A+T content (80-95%) (Firtel and Bonner, 1972; Sussman and Rayner, 1971). The regions of high A+T content include the 5' and 3'- untranslated sequences of mRNA and introns. Introns are usually short (100-200 bp) and the average number of introns per gene is about 1 to 2 (Wu and Franke, 1990).

1.5.4 Plasmids

Circular plasmids are common in prokaryotes, but only a few eukaryotes have been identified and studied for having circular nuclear plasmids and *D. discoideum* is one of them. *D. discoideum* species have a variety of extrachromosomal plasmids, some of which have been exploited as transformation vectors (Firtel *et al.*, 1985; Hughes *et al.,* 1988; Leiting and Noegel, 1988). About 20% of wild isolates harbour nuclear plasmids, which can be divided into four families, based on sequence and structural similarities. Plasmids in the Ddp1 and Ddp2 families are the best studied (Gonzales *et al.*, 1999). Plasmid Ddpp1 and Ddpp3 from *D. purpureum* define the



1.5.2 Fig. I(c) - Life cycle of *Dictyostelium discoideum* other two families, based on sequence and structural homology (Kiyosawa *et al.*, 1993). Families are defined by sequence homology and do no connote incompatibility groups.

Ddp1

Ddp1 plasmids are found in the wild type isolates NC4 and V12. Ddp1 is a 13.7 Kb plasmid, which is present at an estimated copy number of 50-100 per cell and encodes at least five growth specific (G1-G5) and five development specific (D1-D5) transcripts, in addition to an origin of replication. None of the known transcripts of Ddp1 are essential for its replication. All the plasmid carried genes expressed during growth are essential for long term maintenance, while deletion of the Ddp1 genes expressed during development had no detectable effect on long-term maintenance (Hughes *et al.,* 1994). The origin of replication of Ddp1 has been localized to a 543 bp region (Kiyosawa *et al.,* 1995). All essential replication factors are encoded on the host cell chromosomes. The Ddp1 elements necessary for extra-chromosomal replication have been utilized to drive expression of heterologous genes.

Ddp2

Ddp2like plasmids are the best-characterized family. These plasmids are small (4.4-5.8 kb) and present at a copy number of 300 per cell. Ddp2 family of plasmids contains a conserved family of *rep* genes that code for a protein that is required for plasmid replication (Slade *et al.,* 1990). The *rep* gene and a characterized inverted repeat are necessary for plasmid maintenance (Rieben *et al.*, 1998).

There are sequence relationship among these families, but there is no relationship to other nuclear plasmids as budding yeast 2 micron circle. Different plasmids can coexist in the same nucleus, indicating that there are different replication systems preventing incompatibility (Kiyosawa *et al.,* 1994). The *Dictyostelium* plasmids show a clear species specificity. The plasmids share the AT- richness of their host and are packaged in nucleosomes. The complete plasmids are stable over time in the absence of any selection (Hughes and Welker, 1989)

1.5.5 Model organism

Multicellular morphogenesis of this simple eukaryote provides excellent opportunities to investigate several basic principles of developmental biology. Cell movement, cell type determination, spatial patterning, co-ordination between mophogenesis and gene expression are among the central aspects that are shared by any development process. The availability of an extensive repertoire of well-developed biochemical and molecular genetic techniques has permitted an elucidation of the molecular mechanisms that co-ordinate these aspects of multicellular development. Chromosome 2, the largest among six, has been sequenced and complete genome sequence is nearing completion. The genome sequence will also provide opportunity to study function of genes that are shared with more complex organisms. *D. discoideum* has also been extensively used as a model organism to study the various other aspects of cellular and molecular biology, some of which are described below.

1.5.5.1. Cytokinesis

(Wolf *et al.,* 1999; Chen *et al.,* 1994; Chen *et al.,* 1995)

The cellular slime mold *D. discoideum* is amenable to biochemical, cell biological, and molecular genetic analyses, and offers a unique opportunity for multifaceted approaches to dissect the mechanism of cytokinesis. Amoeba cells of the *D. discoideum* typically divide by a mechanism referred to as cytokinesis. Cytokinesis of *D. discoideum* involves constriction of a contractile ring or a cleavage furrow, in a manner similar to higher animal cells. Contractile rings of both types of cells contain parallel filaments of actin and myosin II, and it is believed that active sliding between the two filament systems drives the constriction. They also share many of the intracellular signalling components to regulate the cytoskeletal system. *D. discoideum* being haploid, allows phenotypic manifestation of recessive mutations. Targeted disruption of genes and reintroduction and expression of mutated genes in the absence of the wild type copy is straightforward. It is, therefore, relatively easy in *D. discoideum* to isolate novel genes involved in cytogenesis and to characterise their functions.

1.5.5.2. Differentiation

(Bonner and Cox, 1995; Inouye, 1992; Kay, 1992)

Fundamental problem of embryology is to describe how embryonic fields are established and subdivided into different tissues. Several formal models that describe this process have been presented but the chemical basis of spatial differentiation is unknown. The development of *D. discoideum* displays many of the features of embryonic development. They go through both individual and social phases in their life cycle. This makes it possible to study single cell properties of importance for multicellular differentiation. The evolutionary ancestors of the Dictyostelids were very likely free-living soil amoeba and even in the case of advanced members of the group, under certain circumstances single amoeba can undergo terminal differentiation. Differentiation and pattern formation in *D. discoideum* involves a large number of similar events.

1.5.5.3. Chemotaxis and signal transduction

Chemotaxis is a process by which cells display directional movement towards the source of diffusible chemicals and plays important roles in a wide variety of phenomenon, including the migration of mammalian phagocytic cells such as neutrophils and migration of macrophages, axonal targeting and morphogenesis. The molecular basis of how eukaryotic cells sense and interpret a chemical gradient is poorly understood. Free-living *D. discoideum* amoebae must be able to find their way toward pray or, in face of starvation, towards each other. *D. discoideum* shares a chemotactic capacity with leukocytes and many other motile cells, and employ many of the same mechanisms during the detection of the chemotactic molecule, the activation of signal transduction pathways, and the mobilisation of the cytoskeleton (Devreotes and Zigmond, 1988; Parent *et al.*, 1998). Chemotactic molecules bind to cell surface receptors and stimulate G protein mediated signal transduction pathways in amoebae and mammalian cells. Agonists are degraded to steepen the gradients and to overcome the effects of adaptation. Despite evolutionary distance, the cytoskelotons of leukocytes and *D. discoideum* employ similar cytoskeletal rearrangements to move in the right direction. The advantage of *D. discoideum* in the study of chemotaxis, mobility and aggregation is that the gene products involved in each event can be eliminated by mutation, and the contribution of each element can be studied. To study the second messenger responses to cAMP, this experimental system takes advantage of the fact that starving cells, once they have elaborated all of the molecules that are important to chemotaxis, secrete and respond to cAMP in suspension (Gerisch and Hess, 1974). The light scattering properties of the cells change as they detect a pulse of cAMP and mobilise their cytoskeleton elements. The change in shape occurs whether the cells are on a solid substratum or in suspension. In cell suspension, the contractile properties of the cells can be followed as a periodic change in light scattering. The phase of oscillation of cells in suspension can be shifted with exogenous cAMP. The advantage of suspended cells signalling in synchrony is that large number of amoebae can be rapidly sampled for biochemical estimations (Gerisch *et al.,* 1975).

1.5.5.4. Cell motility

D. discoideum is one of the few organisms with impressive mobility (10-15 µm/min.) (Varnum and Soll, 1984) and tractable genetics. The cells move toward folate during growth and to cAMP during development. All the cytoskeletal elements of *D. discoideum* are similar to those found in mammalian cells. The amoebae are useful for optical observation, so that with a few tricks, the movements of macromolecules within the cells can be observed by a variety of microscopic methods. Above all, the cells can be manipulated genetically and the mutational analysis of components of the actin and myosin cytoskeleton has produced several important observations. The cytoskeleton forms from many elements and these are dynamic, constantly being redeployed in the formation of macropinosomes, phagosomes, filopodia, pseudopodia, ruffles, substrate contacts, the mitotic spindle, or the contractile ring of cytokinesis. Many proteins are reconfigured to do different tasks, and this presents a problem for genetic analysis because the effects of a mutation may be pleiotropic- affecting several processes. Despite partial redundancy and pleiotropy, the *D. discoideum* cytoskeleton presents opportunities that have been exploited by many laboratory so that the organism has become one of the most used in studies of motility (Zigmond *et al.,* 1997; Podolski, *et al.,* 1990; Mckeown *et al.,* 1978; Romans et. al., 1985).

1.5.5.5. Expression system

Recently, the cellular slime mold *D. discoideum* has been developed as an alternative eukaryotic system for expressing recombinant proteins (Dittrich *et al.,* 1994; Tiltscher *et al.,* 1993). The development of reliable transformation systems for *D. discoideum* has provided the possibility of expressing heterologous genes in this microbe (Firtel *et al.,* 1985; Hughes *et al.*, 1992). They grow to a high cell density without the serum factors or special aeration needed by animal cell cultures. Lack of cell wall provides convinent downstream processing. High copy number plasmid vectors allow the expression of protein in cell-associated, membrane attached, or secreted form under the control of regulatable promoters (Manstein *et al.,* 1995). The cells of *Dictyostelium* can do both *0-* as well as *N-*glycosylation (Jung *et al.,* 1995; Jung *et al.,* 1997. The major advantages of this system includes a very simple and inexpensive growth medium and the potential for large scale production of proteins.Several mammalian glycoproteins have been expressed in *D. discoideum*, including rotavirus VP7, human muscarinic receptors m2 and m3, antithrombin III, a soluble form of mast cell IgE receptor, hCG, hFSH, insulin like growth factors etc., it still needs to be optimized to maximize the production of heterologous proteins for commercialization.

1.6 OBJECTIVES

The overall objective of the present study is to develop *D. discoideum* as an heterologous expression system for production of recombinant proteins. More specifically it involves :

1. Designing of an ideal secretory expression vector for *D. discoideum*
2. Construction of the secretory expression vector.
3. Expression of reporter gene: Green fluorescent protein (GFP).
4. Expression of test genes: human DNaseI and Proteinase K.