1. INTRODUCTION

1.1. Overview

The expression of heterologous genes in bacteria is by far the simplest and most inexpensive means available for obtaining large amounts of desired proteins for research and industrial applications. However, this technique has a major limitation due to deposition of a fraction or all (for example, BMP-2) of the recombinant protein in insoluble form. These misfolded proteins aggregate into inclusion bodies (IBs). IBs are refractile, intracellular protein aggregates with porous structure (Marston, 1986; Mitraki and King, 1989). They are known to be in non-native form and hence biologically inactive (Goloubinoff et al., 1999; Carrió and Villaverde, 2001; Hoffmann and Rinas, 2004b).

Considerable research has been dedicated to the goal of achieving a fundamental understanding of protein aggregation, and much progress has been achieved in several systems *in vitro* (London et al., 1974; Zettlmeissl et al., 1979; Kiefhaber et al., 1991). On the contrary, relatively little is known about how this process occurs *in vivo*. Strategies adopted in reducing aggregation and to optimise the *in vivo* production of soluble, active form of proteins include the use of lower cultivation temperatures (Schein, 1989; Valax and Georgiou, 1993), protein engineering by site-directed mutations (Fane and King, 1991; Wetzel et al., 1991; Rinas et al., 1992), low level of protein synthesis by limited induction (Kopetzki et al., 1989; Bowden and Georgiou, 1990), gene fusion technology (LaVallie et al., 1993; Sachdev and Chirgwin, 1998; Choi et al., 2000) and coexpression of chaperones and foldases (Goloubinoff et al., 1989; Blum et al., 1992; Amrein et al., 1995; Winter et al., 2000). Alternatively, IBs formed can be purified and refolded to obtain active protein. However, this *in vitro* recovery of functional protein from IBs involves trial-and-error and it can be expensive and yields low productivity.

Molecular chaperones are ubiquitous and highly conserved proteins that shepherd other polypeptides to fold properly and are not themselves components of the final functional structures (Ellis, 1987; Baneyx and Palumbo, 2003). There are ~ 20 families of this class of proteins which have different molecular weights, structures, cellular locations and functions (Radford, 2000). They were originally identified by their increased abundance as a result of heat shock (Bukau and Horwich, 1998) and hence named heat shock proteins (HSPs). The major classes of these highly conserved proteins are Hsp100 (Clp), Hsp90 (HtpG), Hsp70

(DnaK), Hsp60 (GroEL), Hsp40 (DnaJ) and sHsps (IbpA/IbpB) (Hoffmann and Rinas, 2004b).

Although co-overexpression of molecular chaperones in heterologous production of proteins in *Escherichia coli* may not always lead to increased solubility (Nishihara et al., 1998; Wickner et al., 1999; Schlieker et al., 2002), enormous evidences exist in the literature which show that chaperones are helpful in proper folding of some proteins (see Table 1). In addition, chaperones for disaggregation form an interlinked network, which include the DnaK-chaperone system, the AAA+ protein ClpB and the sHsps – IbpA/IbpB (Mogk et al., 2003b). DnaK can solubilize small aggregates *in vitro*, but cooperation with ClpB is necessary for larger aggregates (Goloubinoff et al., 1999; Zolkiewski, 1999; Diamant et al., 2000).

Some examples of economic important proteins that are being produced as recombinant products are: hormones – growth factors, insulin, human factor VIII; enzyme – cellulases, streptokinase, urokinase; physical active substances – interferons, interleukins. The growth factors market as of 2003 was twenty billion dollars and this has been predicted to exceed thirty billion dollars by 2007 (http://www.the-infoshop.com/press/kl15845_en.shtml). Most proteins with therapeutic potential are found in very low amounts in their natural sources. Genetic engineering permits the introduction of the encoding genes into easily cultivable recipient cells. *E. coli* was the first host used to produce a recombinant DNA (rDNA) pharmaceutical, enabling the approval of Eli Lilly's rDNA human insulin in 1982 (Swartz, 2001).

Among the many systems available for heterologous protein production, the Gram-negative bacterium *Escherichia coli* remains one of the most attractive hosts because of its ability to grow rapidly and at high density on inexpensive substrates, its well-characterized genetics and the availability of an increasingly large number of cloning vectors and mutants host strains (Baneyx, 1999). Although *E. coli* does not appear to be good candidate as an industrial organism when the secretion of foreign proteins is desired, application of fusion partners (LaVallie et al., 1993; Sachdev and Chirgwin, 1998; Choi et al., 2000) can be used to circumvent this problem.

Table 1: Enhancing soluble production by coexpression of molecular chaperones

Chaperone	Protein product	Results	References
GroEL/GroES	Rubisco	Increased production of assembled and active Rubisco proteins from various species is observed.	Goloubinoff et al., 1989
	Protein-tyrosine kinase P^{50csk}	>50% of P^{50csk} is soluble following GroELS overexpression.	Amrein et al., 1995
Trigger factor (TF)	Endostatin	>80% of Endostatin is soluble following TF overexpression.	Nishihara et al., 2000
DnaK	Human growth factor	Coexpression of DnaK inhibits human growth factor IB formation and increases the amount of soluble product from 5% to >85%.	Blum et al., 1992
GroEL/GroES and TF	ORP150	86% of ORP150 is soluble following GroELS/TF overexpression.	Nishihara et al., 2000
GroEL/GroES and DnaK	Cryj2	Coexpression of GroELS/DnaK resulted in marked stabilization and accumulation of Cryj2 without extensive aggregation	Nishihara et al., 1998

The human basic fibroblast growth factor (hFGF-2) is a small protein with 155 amino acids and a molecular weight of 18 kDa. Because of its ability in modulating cell proliferation and differentiation (Gospodarowicz et al., 1985; Burguess and Maciag, 1989), it is a potential therapeutic agent of great industrial importance (Robinson, 1992). While Rinas et al. (1992) replaced the four cysteines with serines by site-directed mutation on the hFGF-2, which altered the partitioning of the protein into soluble and insoluble cell fractions, Seeger et al. (1995) produced the hFGF-2 in *E. coli* using two different induction systems and showed that formation of IBs and soluble hFGF-2 occurs simultaneously and starts immediately after temperature upshift from 30 to 42^{o}C. Both reported 30% of the recombinant hFGF-2 was deposited as insoluble protein when non-mutated form of the hFGF-2 was studied. In addition, studies were carried out on this protein during high-cell density cultivations in *E. coli* to determine the metabolic burden generated during this process (Hoffmann and Rinas, 2001b). *In vitro* studies have also been carried out to determine the refolding conditions after solubilization of IBs formed by hFGF-2 (Estapé et al., 1998), however, this process is slow.

Glucosyltransferase-S (GTF-S) is a large protein with 1430 amino acids and a molecular weight of 155 kDa (Hanada and Kuramitsu, 1989; Honda et al., 1990) produced by *Streptococcus mutans*. This *Streptococcus* has been implicated as a primary etiological agent of dental caries in animals and humans, and it is also involved in plaque formation (Loesche, 1986). The enzymes forming polymers (glucans) from dietary sucrose, particularly the GTFs making water-soluble and water-insoluble glucans, have attracted considerable interest as potential targets for inhibition of the processes that can lead to caries and as a means for directing therapeutic agents to plaque (Shah and Russell, 2002). The GTFs are of commercial interest because when suitable acceptors like maltose are added to the reaction medium, low molecular weight oligosaccharides are synthesized (Koepsell et al., 1953; Demuth et al., 2002). For example, the GTF from *Leuconostoc mesenteroides* synthesizes oligosaccharides (containing ζ-1,6 linked glucosyl residues and a terminal ζ-1,2 linked residues), known as glucooligosaccharides (GOS), through an acceptor reaction with maltose and sucrose (Remaud-Simeon et al., 2000). These are industrial products used in cosmetic and nutritional applications (Remaud-Simeon et al., 2000). The GTF-S synthesizes water-soluble ζ-1,6-linked glucans (Honda et al., 1990). The *gtfD* gene with its native signal sequence has been cloned and expressed in *E. coli*, however, it was not secreted as extracellular protein (Hanada and Kuramitsu, 1989).

1.2. Objectives

In this work, two proteins, i.e. a small (hFGF-2) and a large (GTF-S) protein, both of therapeutic and commercial interest, will be employed in carrying out the following studies:

Heterologous production of hFGF-2 with co-overexpression of different combinations of molecular chaperones will be studied to determine the effect of chaperone coexpression on inclusion body formation and dissolution. In addition, mutant strains deficient in chaperones will be used for the heterologous production of hFGF-2 to ascertain their effect.

By using an expression system carrying pASK12-gtfD with the *gtfD* gene (without its native signal sequence) and the signal sequence of *E. coli* OmpA, secretion of GTF-S into the periplasmic space of *E. coli* cells will be investigated. Also, the influence of co-overexpression of the major molecular chaperones (e.g. DnaK and GroEL systems) on secretion of GTF-S and on prevention/dissolution of IBs of GTF-S will be studied.

Optimization studies will be carried out on the heterologous production of GTF-S using three different host strains ΨTG1, BL21(DE3), MC4100β and four different culture media at shake-flask level. In addition, in order to understand how the overproduction of GTF-S influences the synthesis of critical cellular proteins, proteomic analysis will be performed using both control strain MC4100:pASK-IBA12 (i.e. without *gtfD* gene) and the producing strain MC4100:pASK12-gtfD after inducing them under the same conditions. Identification of protein spots from 2D-PAGE will be carried out by MALDI-TOF-MS technique.

Finally, based on the results obtained above, large-scale heterologous production of GTF-S will be carried out in a bioreactor under various process conditions.

2. LITERATURE REVIEW

2.1. The human Basic Fibroblast Growth Factor (hFGF-2)

The basic fibroblast growth factor (FGF-2) is a member of the growth factor family, which comprised of the Vascular Endothelial Growth Factor (VEGF), Transforming Growth Factor ζ (TGF-ζ), Transforming Growth Factor η (TGF-η), Platelet-Derived Growth Factor (PDGF), Neurotropic Factor (NF), Keratinocyte Growth Factor (KGF), Insulin-like Growth Factor (IGF), Fibroblast Growth Factor (FGF) and Epidermal Growth Factor (EGF) (Walsh, 2002).

Up to date, twenty distinct fibroblast growth factors have been discovered, they are numerically assigned 1 to 20 (Powers et al., 2000). The basic fibroblast growth factor (bFGF; FGF-2) is a member of this class of growth factors whose discovery was based on the presence of a substance in pituitary extracts capable of stimulating the proliferation of fibroblasts (Holley and Kiernan, 1968). This was subsequently isolated from cow brain and bovine pituitary as a non-glycosylated protein (Gospodarowicz, 1975). It has since been isolated from a wide range of other tissues/organs and various tumours, namely, bovine kidney, adrenal glands, corpus luteum, thymus, hypothalamus, liver; human brain and placenta; human hepatoma cell line (Gospodarowicz, 1975; Esch et al., 1985; Abraham et al, 1986b). The FGF-2 isolated from human organs is also known as the human basic fibroblast growth factor (hFGF-2).

The pituitary and brain FGF-2 is a basic protein with an isoelectric point (pI) of 9.6 (Gospodarowicz, 1987). The amino acid sequence of hFGF-2 is 99% homologous to that of bovine FGF-2 (Abraham et al., 1986a) and has 155 amino acids with a molecular weight of 18 kDa. The FGF-2 was first identified as a 146-amino acid protein isolated from pituitary with a molecular weight of 16 kDa (Böhlen et al., 1984). It has been demonstrated that the FGF-2 exists in multiple forms (Prats et al., 1989). The human hepatoma cell line SK-HEP-1 coexpresses four molecular forms (17.8, 22.5, 23.1 and 24.2 kDa) of the hFGF-2 (Florkiewicz and Sommer, 1989). The 18 kDa form is a result of translational initiation at the 5' AUG start codon, while the others are a result of translation beginning at upstream, in-frame, CUG codons (Prats et al., 1989), thus the larger forms are colinear amino-terminal extensions of the 18 kDa form.

The AUG-initiated form (18 kDa) is localized primarily in the cytoplasm, while the forms initiated by the CUG codons are predominantly localized in the nucleus (Renko et al., 1991). Like FGF-1, hFGF-2 does not contain a signal sequence for secretion. The unique feature of the high molecular weight FGF-2 forms, which distinguishes them from 18 kDa FGF-2, is the amino-terminal extension (Bikfalvi et al., 1997).

The three-dimensional structure of the 146 residues form of hFGF-2 was determined via X-ray crystallography. The hFGF-2 contains entirely of η-sheet structure (12 antiparallel η-sheets), organized as a trigonal pyramid with topology identical to that of interleukin-1η (Eriksson et al., 1991). The hFGF-2 contains four cysteine residues at amino acids 26, 70, 88 and 93, however, there are no intramolecular disulfide bonds (Thompson, 1992). While the cysteine residues at 26 and 93 are conserved, those at 70 and 88 are absent or located elsewhere in other FGFs (Arakawa et al., 1989). The *fgf-2* gene has been expressed in *Escherichia coli* (Fox et al., 1988) and the polypeptide formed though active, displayed a high degree of instability towards oxidative processes (Caccia et al., 1992). In addition, the recombinant production of FGF-2 in *E. coli* using a temperature-sensitive expression system leads to formation of inclusion bodies, IBs (Rinas et al., 1992; Seeger et al., 1995). Hence, various combinations of site-directed mutagenesis of the four cysteine residues were carried out to study their effects on stability as well as on inclusion body formation (Fox et al., 1988; Caccia et al., 1992; Rinas et al., 1992). The mutation of all four cysteine residues to serines results in an active protein with the same secondary structure. The FGF-2 is the primary inducer of mesodern formation in embryogenesis, modulating both cell proliferation and differentiation *in vitro* and *in vivo* (Burguess and Maciag, 1989). It is both mitogenic and chemotactic for capillary endothelial cells (Gospodarowicz et al., 1985) and has been shown to induce angiogenesis in the cluck chorioallantoic membrane and rabbit corneal assays (Esch et al., 1985). Because of its wound healing activity, FGF-2 is a potential therapeutic agent of great industrial importance (Robinson, 1992). It is undergoing clinical trials for treating chronic soft tissue ulcers and pressure sores (Meyer-Ingold, 1993; Walsh, 2002).

The biological effects of FGF-2 depend on its binding to specific receptor complexes at the surface of the target cell (Lee et al., 1989). Receptor activation is followed by a series of tightly regulated molecular interactions, which result in signal being transduced to the interior of the cell (Meyer-Ingold, 1993). Four major receptor families have been identified; FGFR1 (flg), FGFR2 (bek), FGFR3, FGFR4 (Baird et al., 1988; Jaye et al., 1992). These receptors

share common features including a cytoplasmic conserved tyrosine kinase domain, a transmembrane domain, and an extracellular ligand-binding domain (Bikfalvi et al., 1997). Phosphorylation of the receptor itself is the first step in the intracellular transduction pathway (Yarden and Ullrich, 1988).

Fig. 1. Structure of the human basic fibroblast growth factor.
This is the 154-amino-acid form of hFGF-2 after Kastrup et al. (1997), drawn with the aid of software package from Protein Explorer (Martz, 2002). The orange marks correspond to the cysteine positions.

2.2. The Glucosyltransferase-S (GTF-S)

Glucosyltransferases (GTFs; EC 2.4.1.5) belong to the family of glucansucrases. These are extracellular enzymes produced by *Streptococcus* species from the oral flora and the soil bacterium *Leuconostoc mesenteroides* (dextransucrases, DSRs). They catalyse the synthesis of high molecular weight polysaccharides (known as glucans, D-glucose polymers) by using sucrose as substrate and, when suitable acceptors like maltose, isomaltose, catechol and methyl ζ-D-glucoside are added to the reaction medium, low molecular weight oligosaccharides are synthesized (Koepsell et al., 1953; Robyt and Martin, 1983; Demuth et al., 2002). About 16 different GTF enzymes have been identified in oral *Streptococci* (Table 1). *Streptococcus mutans* has been implicated as the principal etiological agent in the development of human dental caries (Loesche 1986), because they have the ability to tenaciously colonize the surface of the teeth in the presence of dietary sucrose. This process

depends on the synthesis of water-insoluble glucans (GTF-I) by the GTFs (Shimamura et al., 1994). Various strains of *Streptococci* synthesize more than one GTF (Ciardi et al., 1977).

Table 1: Properties of glucansucrases from *Streptococci* sp. and *Lactococci* sp.

Strain	Gene	Glucan	Size (aa)	M (10^3)	Primer
S. mutans GS-5	*gtf-B*	87% ζ(1-3) 13% ζ(1-6)	1475	150	-
	gtf-C	85% ζ(1-3) 15% ζ(1-6)	1375	140	-
	gtf-D	30% ζ(1-3) 70% ζ(1-6)	1430	155	+
S. mutans LM7	*gtf-C*	nd	1375	150	+
S. downei Mfe28	*gtf-I*	88% ζ(1-3) 12% ζ(1-6)	1556	160	+
	gtf-S	10% ζ(1-3) 90% ζ(1-6)	1328	147	-
S. sobrinus 6715 (serotype *g*)	*gtf-Ia*	nd	1592	160	+
S. sobrinus OMZ176 (serotype *d*)	*gtf-T*	27% ζ(1-3) 73% ζ(1-6)	1556	163	-
	gtfIs	nd	1590	175	+
S. salivarius ATTC25975	*gtf-J*	90% ζ(1-3) 10% ζ(1-6)	1522	168	+
	gtf-K	100% ζ(1-3)	1599	176	+/-
	gtf-L	50% ζ(1-3) 50% ζ(1-6)	1490	157	-
	gtf-M	5% ζ(1-3) 95% ζ(1-6)	1576	171	-
S. gordonii (S. sanguis)	*gtf-G*	40% ζ(1-3) 60% ζ(1-6)	1578	170	-
S. oralis	*gtf-R*	nd	1575	173	-
L. reuteri 121	*gtf-A*	?% ζ(1-4) ?% ζ(1-6)	1781	198	nd
L. mesenteroides NRRL B-1299	*drs-A*	15% ζ(1-3) 85% ζ(1-6)	1290	146	-
	drs-B	5% ζ(1-3) 95% ζ(1-6)	1508	167	-
L. mesenteroides NRRL B-512F	*drs-S*	5% ζ(1-3) 95% ζ(1-6)	1527	170	-

nd – not determined, M – molecular weight. (Monchois et al., 1999a).

Guggenheim and Newbrun (1969) were the first to show that a single strain of *S. mutans* could produce a number of electrophoretically distinct GTFs. Both biochemical and genetic analysis showed that *S. mutans* strains carry three distinct *gtf* genes on the chromosome (Aoki et al., 1986; Hanada and Kuramitsu, 1988; 1989). These genes are labelled *gtfB*, *gtfC*, and *gftD* and they encode the GTFs that synthesize water-insoluble, water-soluble/insoluble and