Development of a lipase gene based expression and secretion system for the protein over-production in *Bacillus licheniformis*

by

Faranani Ramagoma

Submitted in fulfilment of the requirements for the degree of

Magister scientiae

in the

Department of Microbial, Biochemical and Food Biotechnology

Faculty of Natural and Agricultural Sciences

University of the Free State

Bloemfontein

Republic of South Africa

January 2006

Supervisor: Dr. Mulalo. B. Nthangeni

Co-Supervisor: Dr. Evodia. M. Setati

Contents

Acknowledgements

My profound and belated gratitude goes to the almighty GOD, who gave me strength, courage and the energy to carry out this research. I also wish to express my sincere gratitude and thanks to my supervisor, Dr Nthangeni MB for believing in me. The thesis is dedicated to my family, for their unprecedented support and eternal love. Last but not least, my appreciation goes to my colleagues, for their contribution to the success of this study. The National Research Foundation (NRF) of South Africa provided financial support for this research, and their support is gratefully acknowledged.

CHAPTER 1

Literature review

1.1 General introduction

The genus *Bacillus* constitutes a diverse group of rod-shaped, Gram-positive aerobic or facultative bacteria that are characterized by their ability to produce robust endospores in response to adverse environmental conditions (Slepecky, 1992). These bacteria are ubiquitous in nature, and are relatively easy to isolate from a wide variety of sources including soil and water. Discovered by Cohn in 1872 (Cohn, 1872), the genus has undergone considerable taxonomic changes. Initiated by two prominent and truly endospore-forming species, *Bacillus anthracis* and *Bacillus subtilis* (until the early 190Os some taxonomists did not restrict the genus to endospore-forming bacteria), the number of species allocated to this genus increased to 146 in the 5th edition of Bergey's Manual of Determinative Bacteriology (Bcrgey, 1939).

There is great diversity in physiology among members of the genus, whose collective features include degradation of many substrates derived from plant and animal sources, including cellulose, starch, pectin, proteins, agar, hydrocarbons; antibiotic production; nitrification; denitrification; nitrogen fixation; facultative lithotrophy; autotrophy; acidophily; alkaliphily; psychrophily; thermophily; and parasitism (Senesi *et al*., 2001). Spore formation, universally found in the genus, is believed to be a strategy for survival in the soil environment, wherein the bacteria predominate (Smith *et al*., 1952). Certain species from the genus are of significant medical importance, *Bacillus anthracis,* the

causative agent of anthrax, and *Bacillus cereus*, which causes food poisoning. However, some *Bacillus* species have remarkable industrial applications (Harwood, 1992) e.g. *Bacillus subtilis*, *Bacillus licheniformis* and *Bacillus amyloliquefaciens* have a long history of safe commercial application in the food, detergents and pharmaceuticals industries (Harwood, 1992; Priest, 1993; Bron *et al*., 1998).

Bacillus species of industrial importance are vastly applied in the production of several biological products (Schmidt, 2004). These species are important organisms for both fundamental research and industrial applications. Bacilli currently account for 60% of the commercially available proteins synthesized on an economical scale (Bron *et al*., 2004). Majority of these proteins are homologous proteins that are naturally secreted into the growth medium, such as alkaline proteases and amylases (Schweder, 2001; Quax, 2003). Certain species from the genus are applied in the development of expression systems for recombinant protein production (Schallmey *et al*., 2004). The demand for expression systems capable of overexpresing both homologous and heterologous proteins is rapidly increasing (Hazawa-Cho, 1999).

Production of heterologous proteins using bacterial expressions systems is commonly achieved by *Escherichia coli* as a host (Wong and Wu, 1999). However, *Escherichia coli* expressions systems are being overlooked as tools to produce recombinant products due to certain disadvantages. Drawbacks that accompany *Escherichia coli* include its possession of lipopolysaccharides often referred to as endotoxins, which are pyrogenic to humans and animals (Schallmey *et al*., 2004). In commercial applications, *Escherichia*

coli is not ideal due to its inability to secrete proteins into the surrounding medium since they localize proteins in the periplasm space resulting in the formation of inclusion bodies (Chen, 1989). The formation of protein aggregates within the cell imposes metabolic burdens on the host and also complicates downstream processing which result in poor yield of proteins (Li, 2004). Alternative organisms as expression hosts for recombinant products include fungal species *Aspergillus* and *Pichia pastoris* and yeast species such as *Yarrowia lipolytica* and *Saccharomyces cerevisiae* as well as *Bacillus* species. In contrast to fungal expression systems, *Bacillus* offer more rapid fermentation and in general have a much greater secretion capacity for large proteins (i.e. those in excess of 20 KDa) than yeasts (Lam, 1998).

Bacilli are well-known high-level producers of a variety of extracellular proteins (Aiba *et al*., 1983; Vasantha *et al*., 1984; Fahnestock and Fisher, 1987; Gartner *et al*., 1988; Yasuhiko *et al*., 2000). Their capacity to produce and secrete large quantities (20-25 g/l) of extracellular proteins has placed them amongst the most important industrial protein producers (Chang *et al*., 1982; Schallmey *et al*., 2004). These organisms continue to be dominant bacterial workhorses in microbial fermentations and have been extensively applied in the production of useful biochemicals, antibiotics, insecticides and industrial enzymes. Certain species from this genus have been engineered and developed to be commercial producers of nucleotides, riboflavin, ribose and poly-γ-glutamic acid (Wong and Wu, 1999).

Several advantages accompany the use of *Bacillus* species in the production of proteins and other biological products on an industrial scale (Behnke, 1992; Harwood, 1992, Nargarajan, 1993; Sarvas, 1995; Wong, 1995, Bron *et al*., 1998). First, these bacteria have a huge ability to secrete enormous quantities of proteins directly into the growth medium, a factor that greatly facilitates downstream processing. Extensive application of Bacilli in the production of industrial proteins has resulted in vast knowledge accumulation in terms of fermentation technologies (Palva, 1982). The genetics of these bacteria is gradually advancing, stemming from the completion of the genome sequencing projects of a number of *Bacillus* species, *Bacillus subtilis* (Kunst *et al*., 1997); *Bacillus licheniformis* (Rey *et al*., 2004); and *Bacillus cereus* (Ivanova *et al*., 2003).

These factors have made *Bacillus* species amenable hosts from which a variety of genetic tools have been developed (Wong, 1995). Other features include the non-pathogenicity and absence of endotoxins in some industrially important species such as *Bacillus subtilis* and *Bacillus licheniformis* (Wong and Wu, 1999); absence of significant codon bias (Brown, 1990) and extensively studied genetics as well as comprehensively documented properties essential to gene expression; and facility of large scale genetic manipulations using standard protocols (Simonen and Palva, 1993). Some industrially important organisms from this genus, including *Bacillus subtilis* and *Bacillus licheniformis*, bear the GRAS (generally regarded as safe) status (Lam, 1998).

Bacillus species also present certain bottlenecks in the production of commercial proteins. Bacilli are prolific producers of proteases, which usually render heterologous proteins vulnerable and therefore minimize their yield (Lim, 2003). Moreover, the ability of *Bacillus* species to form spores when environmental conditions become unfavorable present challenges with respect to initiating spore mediated diseases (Wong, 1995). Contrary to their *Escherichia coli* counterpart, whose gene expression signals have been extensively explored and characterized, *Bacillus* gene expression signals still need extensive research (Lam *et al*., 1998).

1.2 Basic features of a *Bacillus* expression-secretion system.

A basic expression-secretion unit to express proteins is usually composed of the following components: promoter, ribosome binding site (RBS) within which the Shine-Dalgarno sequence is embedded, a signal peptide (SP), a multiple cloning site, a transcription terminator and selectable marker (Figure 1.1). In cases where a protein tagging sequence is to be introduced, it is inserted in frame either at the N-or C-terminus of the heterologous or homologous gene-coding region (Doi, 1984; Wong and Chang, 1986).

Figure 1.1. A schematic representation of the arrangement of the different components of an expression-secretion system. SP is the signal peptide; RBS, the ribosome binding site; MCS, the multiple cloning site; TT, the transcription terminator and the selection marker.

1.3. DNA elements involved in transcription

The available evidence indicates that the transcription machinery of *Bacillus* species is capable of utilizing most regulatory regions originating from any of the species from this genus and other closely related Gram-positive bacteria (Patek *et al*., 2003), but only a small minority of those from Gram-negative bacteria (Lovett and Schoner, 1983). Transcription plays a major role in the expression of all genes and can be divided into three stages: initiation, elongation and termination. There are generally two different DNA sequences and a multicomponent enzyme that are involved in transcription: the promoter, the transcription terminator and RNA polymerase (Schumann *et al*., 2004). In most cases, gene expression is regulated or controlled at the step of initiation, since the synthesis of RNA requires much energy (ATP equivalents) and it is therefore imperative for the cell to regulate gene expression (Schallmey *et al*., 2004).

1.3.1. RNA polymerase

The presence of a family of RNA polymerase holoenzymes in *Bacillus* species has been well documented (Losick and Pero, 1981; Doi and Wang, 1986; Stragier and Losick, 1990). Different forms of RNA polymerase holoenzyme exist depending on the growth situation; and this makes it necessary to understand and control the growth condition so that proper transcription signals are present to turn on genes under specific conditions. The elaboration of a complex transcription machinery has also imparted a high degree of specificity to *Bacillus* gene expression systems (Wang and Doi, 1992). The core enzyme RNA polymerase is composed of four subunits and is capable of basic polymerization in vivo. The sigma factor is essential in recognizing and binding to the promoter region of

the gene to initiate transcription. After transcription initiation, the sigma factor dissociates from the transcription complex (Uptain *et al*., 1997; Mooney *et al*., 1998). The major difference in RNA polymerases between *Bacillus* species and *Escherichia coli* lies in their sigma factors. *Escherichia coli* contains four major sigma factors (σ^{70} σ^{54} σ^{32} and σ^{38}), while *Bacillus* species have 17 different sigma factors employed at different stages of growth (Mooney *et al*., 1998). Growing cells can make use of six different sigma factors: σ^{A} (housekeeping factor), σ^{B} (general stress response), σ^{C} (unknown postexponential gene expression), σ^{D} (Chemotaxis/autolysin/flagellar genes) and σ^{L} (degradative enzyme gene expression) (Haldenwang, 1995). RNA polymerase can be used by multiple sigma factors. Different classes of sigma factors recognize different promoter sequences, and this regulates gene expression by altering the pattern of RNA polymerase (Huang and Helman, 1998).

1.3.2. Promoters

A promoter is a specific region just upstream from a gene that acts as a binding site for transcription factors and RNA polymerase during transcription initiation. They are characterized by sequences at positions -10 and -35 base pairs upstream from the transcription start site in prokaryotic organisms. The consensus sequence 5′-TTGACA-17nt-TATAAT-3′ that is recognized by σ^{A} in *Bacillus* species is identical to the consensus recognized by σ^{70} in *Escherichia coli* (Kazuo and Ogasawara, 2002). There are at least three intrinsic parameters that affect promoter engagement during transcription: the hexamer centered around -10 box, the hexamer around the -35 box, and the region of DNA between the two boxes (spacer region). The strength of the promoter

refers to the frequency of the promoter to initiate transcription. A strong promoter is therefore one that has sequences conforming to the consensus sequences, and ensures frequent transcription initiation, whereas weak promoters have substitutions resulting in infrequent initiations. Strong promoters are usually applied to improve the level of gene expression (Wang and Doi, 1984).

The suitability of promoters for high-level gene expression is controlled by several criteria. First, the promoter needs to be strong, capable of protein production in excess of 10–30% of the total cellular proteins. Second, the promoter should exhibit a minimal level of basal transcription; a highly repressible promoter is particularly important for cases in which the protein is toxic or detrimental to the growth and development of the host cell. Third, an ideal promoter should also be capable of induction in a simple and cost–effective manner (Chang *et al*., 1982).

One of the main problems associated with the delay in the development of stable, efficient *Bacillus* expression systems was the lack of well-characterized, strong controllable promoters. An ideal promoter should satisfy at least two major points, which may act in a contradictory fashion and hence, have to be determined experimentally. The chosen promoter should be actively expressed under desired conditions, allowing high-level expression of the target gene products. At the same time, the promoter should also be compatible with the target gene so that a high level of expression can be attained (Provvedi *et al*., 2005).

With the exception of promoters derived from *Bacillus* species and other Gram- positive bacteria, majority of foreign promoters are not well utilized in *Bacillus* expression systems (Patek *et al*., 2003). It is therefore imperative that well-characterized promoters compatible with *Bacillus* expression systems are employed to drive the expression of both heterologous and endogenous genes (Wang, 1995). There are two types of promoters that are commonly used in the development of *Bacillus* expression systems, the first involving promoters that naturally show growth phase dependent expression (Sloma *et al*., 1992; Doi *et al*., 1996; Furutani *et al*., 1997; Wilson *et al*., 1999), such as those used in the expression of genes that encode extracellular enzymes.

The other group involves promoters that are inducible. For expression of structural genes that are under control of such promoters certain molecules, e.g. xylose for the xylose operon (Bhavsar *et al*., 2001) needs to be present to induce the expression of such genes. Generally, several classes of *Bacillus* species native promoters together with engineered promoters have been employed for the expression of genes with *Bacillus* species as host cells. These include constitutive promoters, temporally regulated promoters, sporulation genes promoters and genetically engineered promoters (Schallmey *et al*., 2004).

1.3.2.1 Constitutive promoters

Constitutive promoters express genes at all times. Most of these genes encode proteins that are essentially required by the organism for growth and development, generally referred to as housekeeping genes. Expression vectors that are based on constitutive

promoters have been developed by different groups for expression of homologous and heterologous genes (Harry *et al*., 1994). Expression of genes during the vegetative phase has the major advantage of less proteolytic degradation since most *Bacillus* proteases are produced during the stationary phase (Priest, 1977, 1989).

However, there are also shortcomings associated with the employment of strong constitutive promoters. High-level expression of certain genes may inevitably exert metabolic pressure on the host cells, which may decrease the overall accumulation of biomass. Strong transcription read-through is one of the major causes of the instability of some recombinant clones (Ehrlich *et al*., 1986; Gentz *et al*., 1981).

1.3.2.2 Temporally regulated promoters

Promoters for most extracellular proteins are temporally regulated in *Bacillus* species. The best-characterized promoters are those for the α-amylase (Furutani *et al*., 1997) and alkaline protease genes (*aprE*), (Wilson *et al*., 1999). Although the exact mechanism of temporal regulation is virtually not yet well understood, in the case of the alkaline protease gene, it has been demonstrated that the promoter is recognized by σ^A RNA polymerase both in vivo and in vitro (Park *et al*., 1989). Temporally regulated promoters are predominantly switched on at the onset of the post-exponential stage of growth. This type of promoter is very attractive for expression of secreted gene products since all the required signals, including the promoter, ribosome binding site, and signal peptide can be derived from a single DNA segment. This does not only simplify expression systems

construction, but also facilitate easy genetic manipulations of these elements, which have been optimized through evolution (Doi *et al*., 1996)

However, one major drawback that accompanies these promoters is that they are relatively weak. Prolonged stationary phase expression is required to achieve high-level accumulation of the desired products. Most of the heterologous gene products are very sensitive to proteolytic degradation during this stage (Doi *et al*., 1986). Several different approaches, including development of protease deficient strains, have been followed to optimize the production of heterologous gene products during the stationary phase when using these promoters, since they are more prone to proteolytic degradation than homologous proteins (Wong and Wu, 1999).

1.3.2.3. Inducible promoters

Inducible promoters' activity requires an inducing agent for transcription to commence. The expression of genes under control of inducible promoters is usually repressed by a repressor molecule that binds to the operator thereby inhibiting the binding of RNA polymerase. Genes whose expression is driven by inducible promoters can only be expressed in the presence of an inducer, which bind to the repressor molecule and make it dissociate from the operator such that transcription can be initiated. These promoters have gained widespread usage in the development of many *Bacillus* expression systems (Crutz *et al*., 1990; Bhavsar *et al*., 2001).

The use of inducible promoters present at least two advantages. Under repressed conditions these promoters permit stable maintenance of the recombinant clones throughout the growth phase. Induced synthesis of the target gene product over a limited period of time minimizes its exposure to proteases. This also cut the fermentation time and the costs in industrial biotechnology. Several inducible *Bacillus* promoters have been characterized, which, after proper adaptation, proved useful for the expression of both homologous and heterologous genes (Sa-Nogueira *et al*., 1988). These include promoters of genes involved in tryptophan and arginine synthesis (Shimotsu *et al*., 1986; Smith *et al*., 1986); sucrose, xylose, arabinose, and gluconate utilization (Crutz *et al*., 1990; Gartner *et al*., 1988; Sa-Nogueira *et al*., 1988; Sa-Nogueira and de Lencastre, 1989; Fujita and Fujita, 1986; Fujita and Fujita, 1987); and lysogeny in phage θ105 (Van Kaer *et al*., 1987).

Controlled inducible expression in the *Bacillus* genus is well illustrated by the xylose inducible expression system (Rygus and Hille, 1991; Bhavsar *et al*., 2001). Xylose utilization in Bacilli requires the production of xylose isomerase (XylA) and xylulose kinase (XylB) and is tightly regulated at the level of transcription by a xylose responsive repressor protein encoded by xylR. The xylR and xylAB genes are divergently transcribed from a common intergenic region containing xyl operator sequences which are bound by xylR in the absence of an inducer, xylose. In the presence of the inducer, it reacts with xylR and makes it dissociate from the transcription complex facilitating synthesis of mRNA (Rygus and Hillen, 1991).

1.3.2.4 Sporulation promoters

Certain promoters are expressed very weakly or not at all during growth, but are strongly expressed during early and late stages of sporulation (Blaskovic and Barak, 2002). A number of these early stationary phase promoters are transcribed by σ^H, a form of holoenzyme that is present at low levels during growth, but increases dramatically between early and late stationary phase. These σ^H promoters include PEP4 genes promoters of the σ^A operon. Other sporulation related promoters include those recognized by σ^E, σ^G and σ^K (Bin Zhan *et al*., 1997). Examples of sporulation promoters include the sigG promoter (Evans *et al*., 2004) and SpoOA gene promoter (Stragier and Losick, 1996). Some of these promoters have also been applied in the development of an expression system (Zhu *et al*., 2003)

1.3.2.5 Genetically engineered promoters

Hybrid promoters have been constructed that can be controlled and expressed (or derepressed) under certain physiological conditions. These include the *Pspac* promoter, which consists of a phage SPO1 promoter containing the *Eshecrichia coli lac* operator sequence adjacent to its 3' end (Bhavsar *et al*., 2001). If the system contains the *lac* repressor gene (*lacI*) and lac repressor is synthesized, then the *spacI* promoter controlled gene is repressed until an inducer such as isopropyl-thiogalactoside (IPTG) is added. A similar promoter construct, *pac-I*, contains the *Bacillus licheniformis* penicillinase gene promoter with a 3′-adjacent *lac* operator sequence that is controlled by the *lac* repressor and induced by IPTG. In both these cases if the *lacI* gene product is absent, constitutive expression of the promoter occurs (Kobayashi *et al*., 1987; Zhu *et al*., 1990).

1.3.3. Transcription terminators

Much less attention has been given to transcription termination than to transcription initiation in heterologous and homologous protein expression. This does not imply the insignificance of terminators for maximal gene expression. There exists well-documented evidence indicating that transcription read-through is probably one of the major causes of plasmid instability (Ehrlich *et al*., 1986; Mountain *et al*., 1984; Mountain, 1989; Gentz *et al*., 1981). Depending on chromosomal location and orientation of integration, transcription read-through may also play a role in determining the stability and expression level of recombinant genes integrated in the host chromosome. Incorporation of an efficient transcription terminator is, therefore, highly recommended in the construction of expression vectors. This does not only increase plasmid stability, but also decreases metabolic load by reducing the transcription (and translation) of other plasmid-encoded genes whose high level expression is not essential for the maintenance and expression of the gene of interest (Mountain, 1989).

It has also been reported that certain terminator sequences can increase the level of expression by increasing the mRNA stability (Wong and Chang, 1986), probably due to the formation of a stable stem-loop structures that protects the mRNA from exonuclease attack. A number of *Escherichia coli p*-independent terminators were shown to be able to function in *Bacillus* species (Peschke *et al*., 1985). Doi and Wang (1992) utilized *rrnB* (an operon for constitutively expressed rRNA genes) terminators in the construction of the expression vector pWE1, which was used for the expression and secretion of human atrial natriuretic α-factor (Wang *et al*., 1988). Sets of probe plasmids, pWT18 and

pWT19, were constructed so that the terminators of *Bacillus subtilis* could be isolated and characterized by measuring and comparing their termination efficiencies *in vivo*. Different generated DNA sequences had a variety of effects on the level of expression (Wang and Doi, 1987).

1.4. DNA elements involved in translation

Another barrier for efficient expression of foreign or native genes in any organism resides at the level of mRNA translation. Additionally, translation machineries of *Bacillus* species are quite specific and require homologous ribosome binding sites (RBS). The RBS usually contains a sequence GGAGG and has an average free energy of about -17 kcal/mol for binding between the 3′ end of the 16S ribosomal RNA and the RBS region of the mRNA (Band and Henner, 1984). Supplying an efficient *Bacillus* RBS sequence to the gene of interest is one solution, but it does not always work because the secondary structure around the translation initiation site also plays a pivotal role in determining translation efficiency (Ganoza *et al*., 1987).

1.4.1. Ribosome binding site

It has been well established that *Bacillus* species require a "stringent" RBS for efficient translation initiation (McLaughlin *et al*., 1981; Band and Henner, 1984). For efficient expression of cloned genes in *Bacillus* species, it is indispensable to use promoters and RBS optimal for the host species (Ohashi *et al*., 2003). Since most heterologous and homologous genes are expressed in *Bacillus* species by direct fusion of the gene of interest to the transcription and translational signals of a native *Bacillus* species gene, not

much effort has been invested in optimizing the RBS conditions. Flock *et al* (1984) reported the expression of the human urogastrone gene in *Bacillus subtilis* using two different synthetic RBS sequences.

There was no significant difference observed in the level of expression when these two RBS sequences were compared, whereas the same set of sequences showed an eightfold difference when expressed in *Escherichia coli*. Nevertheless, there appear to be a certain general criteria necessary for a good *Bacillus* RBS (Doi, 1984; Band and Henner, 1984): the Shine-Dalgarno sequence should have extensive complementarity to the 3′ end of the 16S rRNA with a free energy of interaction around -17 to -18 kcal/mol; the sequence GGAGG in the RBS is usually highly conserved; and the spacer region between the GGAGG sequence and initiation codon is approximately eight bases long and rich in A and U nucleotides (Duvall *et al*., 1983; Bechhofer and Dubnau, 1987).

1.4.2 Start codons

Although GUG and UUG have been reported to function as initiation codons in *Bacillus* (Wong *et al*., 1984; Wang and Doi, 1986; Smith *et al*., 1986; Shields and Sharp, 1987; Mountain, 1989), the fact that highly expressed *Bacillus subtilis* genes (such as *rpoA*, *rpoD*, *rpmA*, etc.) generally use AUG as the initiation codon may indicate the existence of subtle differences in initiation efficiencies for these codons. Based on the comparison of the relative efficiencies of AUG with GUG and UUG, it is advisable to use AUG as an initiation codon for heterologous gene expression in *Bacillus* species. The employment

of AUG as the initiation codon simplifies hybrid gene construction since most prokaryotic and eukaryotic genes initiate with the AUG codon (Ganoza *et al*., 1987).

1.5. Translocation

This refers to the movement of extracellular proteins across the cytoplasmic membrane and it is the most critical step during secretion. In contrast to Gram-negative bacteria, proteins that are secreted to the extracellular environment by Gram-positive bacteria only need to travel through a single membrane to enter the surrounding medium (van Wely *et al*., 2001). While targeting a protein for export appears to be straightforward through the use of signal peptides, subtle factors need to be understood to achieve maximum efficient translocation of a particular protein (Simonen and Palva, 1993). To differentiate cytosolic proteins from extracellular ones, proteins destined for secretion are synthesized as precursors with a cleavable amino terminal signal peptide (Bron *et al*., 1998). These signal peptides ensure proper targeting of the polypeptide to the translocation machinery at the cytosolic membrane. The general secretory pathway mediates the translocation of proteins in an unfolded conformation (Bron *et al*., 2004).

Translocation occurs through a confined aqueous channel composed of a set of integral membrane proteins (Manting and Driessen, 2000; Manting *et al*., 2000). The general principle of this channel is highly conserved among bacteria, archae and eukaryotes (Pohlschroder *et al*., 1997). In addition, the channel polypeptides provide a binding site for cytoplasmic components that drive targeting and translocation. Cytosolic chaperones are essential in that they prevent tight folding or aggregation of the precursor protein

(Fekkes and Driessen, 1999). There are generally two translocation mechanisms; co-translational translocation and post-translational translocation. In co-translational translocation the site of synthesis for the precursor protein is brought into contact with the translocation channel and couples translocation with translation of the secretory protein at the ribosome. The other possibility, post-translational translocation is to complete synthesis of the secretory protein and then direct it to the translocation channel. The protein is then transported across the membranes and maturation to its final folded conformation is ensured (Figure 1.2). Folding of translocated proteins usually compete with proteolytic degradation by extracellular proteases (Fink, 1999).

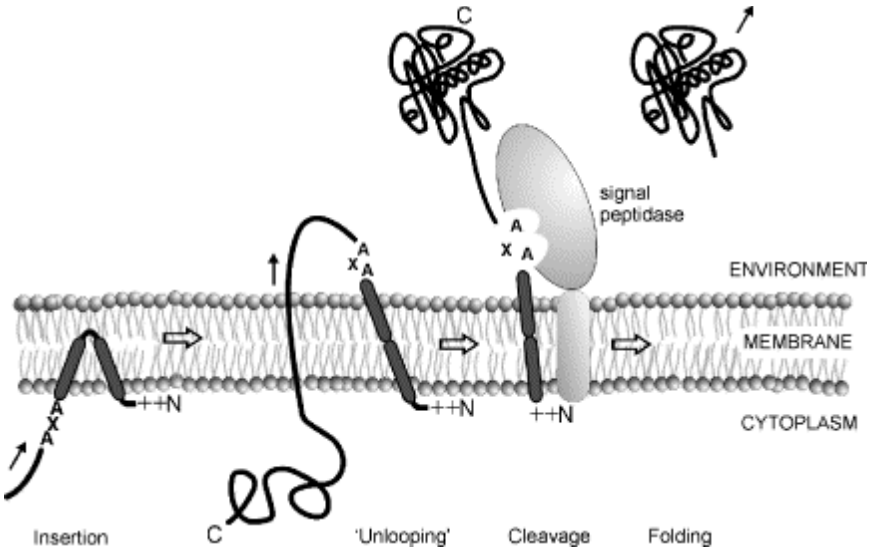

Figure 1.2. Model for signal peptide insertion into the cytoplasmic membrane during translocation and cleavage by type I SPase. First, the positively charged N-domain of the signal peptide interacts with negatively charged phospholipids in the membrane, after which the H-domain inserts loopwise into the membrane. Next, the H-domain unloops, whereby the first part of the mature protein is pulled through the membrane. During or shortly after translocation by a protein transport machinery, the signal peptide is cleaved by type I SPase and, thereby, the mature protein is released from the membrane. After its translocation across the membrane, the mature protein folds into its native conformation (Taken from Bron *et al*., 2004)

1.5.1. Signal peptides and signal peptidases

1.5.1.1 Signal peptides

The signal peptide serves at least three functions in the export of extracellular proteins. First, it is recognized by receptors of the secretion machinery and transferred to the translocation machinery that catalyzes the transportation from across the membranes (Hartl *et al*., 1990; Dalbey *et al*., 2000). The signal peptide also serves as a topological determinant for the preprotein in the membrane. These polypeptides initiate translocation of the C-terminal hydrophilic regions of precursor proteins whilst the N-terminal remains associated with the *cis* side of the membrane (Andersson *et al*., 1992; Dalbey *et al*., 1990). The folding of nascent chains is also ensured by signal peptide, thereby ascertaining avoidance in the activation of potentially harmful secretory enzymes inside the cells and concurrently retains translocation competence (von Heijne and Abrahmsen, 1989; Nielsen *et al*., 1997; Edman *et al*., 1999).

In general, proteins destined for export are synthesized as precursors equipped with a signal peptide that is proteolytically removed by signal peptidases during or shortly after translocation (Paetzael, 2000). A signal peptide is usually 14-25 amino acids long and consists of three identifiable domains, the amino (N-), hydrophobic (H-), and carboxy-terminal (C-) regions (Fig 2). The N-region is rich in positively charged amino acids, and is followed by a hydrophobic region that tends to organize into an α-helical conformation when brought into contact with the membrane lipid phase. The C-terminal is hydrophilic in nature and contains the signal peptide cleavage site that is recognized by signal peptidases. This site conforms to the -3, -1 rule (for signal peptidase cleavage) (von

Heijne and Abrahmsen, 1989), and in many cases corresponds to an Ala-X-Ala cleavage site.

In the general secretion pathway, two major categories of signal peptides can be identified; the general signal peptides (type I) and the lipoprotein signal peptides (type II). Other classes of signal peptides are categorized based on the pathway they mediate with respect to protein translocation. These classes include the twin-arginine and type IV pilin export signal peptides (Bron *et al*., 2000). Von Heijne and Abrahmsen (1989) reported that on average, type I signal peptides of *Bacillus* species are five to seven amino acids longer than those from *Escherichia coli*. Extension was observed to occur in all three regions (N-, H- and C-regions). Additionally, the N-region usually contains a higher number of positively charged lysine and arginine residues. In Gram-positive bacteria, cleavage by signal peptidase prefentially occurs seven to nine amino acid residues from the C-terminal end of the H-region, whereas in *Escherichia coli*, processing takes place three to seven residues from the same position (Dalbey *et al*., 1997).

Type II signal peptides largely resemble type I signal peptides, but differ only because their C-region contain a so called lipoprotein box with a Leu-Ala-Gly-Cys consensus sequence at position -4. In the cytosolic membrane, the cysteine in this sequence is covalently linked to a lipid (Tokunaga *et al*., 1982; Hayashi *et al*., 1986). Just after lipomodification, the type II signal peptide is recognized by signal peptidases type II and cleaved. A higher positive charge (average of three versus two in Gram-negatives) is found within the N-region of Gram-positive type II signal peptides (Bron *et al*., 1999).

Several *Bacillus* signal peptides that carry only two basic residues at the N-terminus have been reported to mediate efficient secretion (e.g. *Bacillus amyloliquefaciens* neutral protease, Vasantha *et al*., 1984; *Bacillus subtilis* β-glucanase, Murphy *et al*., 1984; Robson and Chambliss, 1987; *Bacillus brevis* middle and outer wall protein, Tsuboi *et al*., 1988). Furthermore, mutational alteration of the net charge within the N-region of the levansucrase signal peptide to +1 or even zero still allowed secretion to proceed as long as there was both a positively and a negatively charged residue together at the N-terminus. Absence of any charge, nevertheless, prevented secretion of levansucrase (Borchet and Vasantha, 1991). This is in contrast to *Escherichia coli*, where signal peptides have been reported to function even in the absence of N-terminal charged amino acids (Oliver, 1985).

Another interesting difference from *Escherichia coli* is the frequent use (almost 30%) of TTG or GTG start codons in the coding regions for *Bacillus* signal peptides. It is tempting to speculate that this reflects a special adaptation of translational speed to the requirements of the secretion process. Within the hydrophobic core of *Bacillus* signal peptides, leucine, isoleucine, valine and phenylalanine prevail with alanine and, in particular, glycine, tryptophan and proline less frequent (Table 1.1A and B). This is consistent with the hypothesis that the hydrophobic core mainly assumes a α-helical conformation (Emr and Silhavy, 1983; Briggs and Gierasch, 1986). In general, the hydrophobic cores of *Bacillus* signal peptides appear to be longer than those of exported proteins from Gram–negative bacteria (Tjalsma *et al*., 1997; Bron, 1998).

TABLE 1.1A Type I Signal peptides of *Bacillus* species

Protein	Species of origin	Signal peptide	Reference
a-Acetolactate decarboxylase	*B. brevis*	MKKNIITSITSLAIVAGLSLTAFA4AITT;A*TV	Diderichsen *et al*., 1990
Alkaline cellulase	*Bacillus sp.*	MLRKKTKQLISSILILVLLLSLFPTALAA*EG	Fukumori *et al*., 1986
Alkaline phosphatase	*B. subtilis*	MFAKRFKTSLLPLFAGFLLLFYLVLAGPAAASA*ET	Bookstein *et al*., 1990
a-Amylase	*B. amyloliquefaciens*	LKKFPKKLLPIAVLSSIAFSSLASGSVPEASAI*QE	Takase *et al*., 1988
a-Amylase	*B. licheniformis*	MIQKRKRTVSFRLVLMCTLLFVSLPITKTSAI*VN	Takidnen *et al*., 1983
α-Amylase	*B. licheniformis*	MKQHKRLYARLLPLLFALIFLLPHSAAAAI*AN	Sibakov , 1984
a-Amylase	*B. stearothermophilus*	MKQQKRLYARLLTLLFALIFLLPHSAAAAN*AN	Zagorec *et al*., 1989
α-Amylase	*B. stearothermophilus*	MLTFHRIIRKGWMFLLAFLLTALLFCPTGQPAKA*AA	Nakajima *et al*., 1985
a-Amylase	*B. stearothermophilus*	MLTFHRIIRKWVFLLAFWLTASLFCPTGQPAKA*AA	Gray *et al*., 1986
a-Amylase	*B. polymyxa*	MKKKTLSLFVGLMLLIGLLFSGSLPYNPNAAEAI*SS	Diderichsen, 1988
P-Amylase	*B. megaterium*	MTLYRSLWKKGCMLLLSLVLSLTAFIGSPSNTASA*AV	Kawazu *et al*., 1987
Amylase	*Bacillus species*	MKGKKWTALALTLPLAASLSTGVDAETI*VH	Metz *et al*., 1988
Amylase	*B. subtilis*	MKMRTGKXGFLSILLAFLLVITSIPFTLVDVEA*HH	Tsukamoto *et al*., 1988
Bacillopeptidase F	*B. circulans*	MRKKTKNRLISSVLSTVVISSLLFPGAAGA*SS	Sloma *et al*., 1990
Chitinase Al	*Bacillus sp.*	MINLNKHTAFKKTAKFFLGLSLLLSVIVPSFAPLQPATAEA*AD	Watanabe *et al*., 1990
Cyclodextrin	*B. licheniformis*	MKRFMKLTAVWTLWLSLTLGLLSPVHA*AP	Kimura *et al*., 1987
Glucanotransferase	*B. macerans*	MFQMAKRVLLSTTLTFSLLAGSALPFLPASA*IY*	Hill *et al*., 1990
Cyclodextrin	*B. subtilis*	MKSRYKRLTSLALSLSMALGISLPAWP*SP	Takano *et al*., 1986
Glucosyltransferase	*B. subtilis*	MKNMSCKLVVSVTLFFSFLTIGPLAHA*QN*	Sloma *et al*., 1988
Cyclodextrin	*B. subtilis*	MKRSISIFITCLLITLLTMGGMIASPASA*AG	Mackay *et al*., 1986
Glucanotransferase	*B. subtilis*	MPYLKRVLLLLVTGLFMSLFAVTATASAI*QT	Murphy *et al*., 1984
Extracellular protease	*B. polymyxa*	MPYLKRVLLLLVTGLFMSLFAVTSTASP*QT	Tezuka *et al*., 1988
1-Glucanase	*B. lautus*	NKKKGLKKTFFVIASLVMGFTLYGYTPVSADA*AS*	Baird *et al*., 1990
,-Glucanase	*B. cereus*	MKKRRSSKVILSLAIVVALLAAVEPNAALA*AP*PP*	Joergensen, 1990
P-Glucanase	*B. cereus*	MKNKRMLKIGICVGILGLSITSLEAI*FT	Mezes *et al*., 1985
P-Glucanase	*B. cereus*	MKNKKMLKIGMCVGILGLSITSLVTI*FT	Wang *et al*., 1985
1-Glucanase	*B. cereus*	MKNTLLKLGVCVSLLGITPFVSTISSVQAI*ER	Lim *et al*., 1988
,-Lactamase	*B. subtilis*	MKKNTLLKVGLCVGLLGTIQFVSTISSVQAI*SQ	Hussain *et al*., 1985
P-Lactamase	*B. subtilis*	MKKRLIQVMIMFTLLLTMAFSADA*AD	Schorgendorfer *et al*., 1987
j-Lactamase	*Bacillus sp.*	MNIKKFAKQATVLTFTTALLAGGATQAFA*KE	Steinmetz *et al*., 1985
P-Lactamase	*B. subtilis*	MKVYKVAFVMAFIMFFSVLPTISMS*SE	Akino *et al*.,1989
Levanase	*B. subtilis*	MKLVPRFRKQWFAYLTVLCLALAAAVSFGVPAKAI*AE	Sloma *et al*., 1990
Levansucrase	*B. subtilis*	MKVYKVAFVMAFIMFFSVLPTISMS*SE	Tsuboi *et al*., 1983-
Mannanase	*Bacillus sp.*	MKNTLLKLGVCVSLLGITPFVSTISSVQAI*ER	Vasantha, 1986
Metalloprotease	*B. subtilis*	MNIKKFAKQATVLTFTTALLAGGATQAFA*KE	Shimada *et al*., 1985
Middle wall protein	*B. brevis*	MKKVVNSVLASALALTVAPMAFA*AE	Yang *et al*., 1984
Neutral protease	*B. amyloliquefaciens*	MGLGKKLSVAVAASFMSLTISLPGVQA*AQ	Takagi *et al*., 1985
Neutral protease	*B. amyloliquefaciens*	MGLGKKLSSAVAASFMSLTISLPGVQA*AE	Nishiya et al.,1990
Neutral protease	*B. subtilis*	MGLGKKLSVRVAASFMSLSISLPGVQA*AE	Tsuboi *et al*., 1986
Neutral protease	*B. stearothermophilus*	MNKRAMLGAIGLAFGLLAAPIGASA*KG	Paddon *et al*., 1989

TABLE 1.1A continued.

Protein	Species of origin	Signal peptide	Reference
Neutral protease	*B. stearothermophilus*	MKRKMKMKLVRFGLAAGVAAQVFFLPYNALAISTI*EH	Yamada *et al*., 1988
Outer wall protein	*B. brevis*	MNKXVVLSVLSTTLVASVAASAFA*AP	Yong and Doi, 1986
RNase	*B. amyloliquefaciens*	MKKRLSWISVKLLVLVSAAGMLFSTA*AR	Vasantha and Thompson,1986
Sphingomyelinase	*B. cereus*	MKGKLLKGVLSLGVGLGALYSGTSAQAP*EA	Jacobs *et al*., 1985
Subtilisin E	*B. subtilis*	MRSKKLWISLLFALTLIFTMAFSNMSAQA*AG	Fukusaki, 1984
Subtilisin	*B. amyloliquefaciens*	MRGKKVWISLLFALALIFTMAFGSTSSAQA*AG	Hamamoto *et al*., 1987
Subtilisin Carlsberg	*B. licheniformis*	MMRKKSFWLGMLTAFMLVFTMAFSDSASAI*AQ	Fukusaki, 1984
Xylanase	*B. pumilus*	MNLRKLRLLFVMCIGLTLILTAVPAIAL*RT	Hamamoto *et al*., 1987
Xylanase	*Bacillus sp.*	MITLFRKPFVAGLAISLLVGGGIGNVA*AQ	Fukusaki, 1984

*, signal peptidase cleavage site (Putative signal peptidase cleavage site).

TABLE 1.1B. Comparison of the signal peptide of Braun's lipoprotein with those of *Bacillus* species

Protein	Source	Signal peptide	Reference(s)
Braun's lipoprotein	*E. coli*	MKATKLVLGAVILGSTL**LAG*C**S	Nakamura *et al*., 1980
P-Lactamase	*B. lichenifonnis*	MKLWFSTLKLKKAAAVLLFSCVA**LAG*C**A	Nielsen and Lampen, 1982
1-Lactamase	*B. cereus*	MFVLNKFFTNSHYKKIVPVVLLSCAT**LIG*C**S	Hussain *et al*., 1987
P-Lactamase	*Bacillus sp.* (alkalophilic)	MIVPKKFFHISHYKKMLPVVLLSCVT**LIG*C**S	Kato *et al*., 1989
PrsA	*B. subtilis*	MKKIAIAAITATSILA**LSA*C**S	Hemina *et al*., 1991
PAL-related protein[b]	*B. subtilis*	MRYBAVFPMLIIVFA**LSG*C**T	Kato *et al*., 1989

*, cleavage site for signal peptidase II (the consensus cleavage sequence has been written in bold); +, positively charged residues.

[b] The lipoprotein nature of the protein is deduced from the sequence only. Gram-positive bacteria have lipoprotein 1-lactamases on the outer surface of the cytoplasmic membrane.

1.5.2.2 Signal peptidases

Removal of signal peptides during or following translocation is accomplished by specific processing enzymes called signal peptidases. *Bacillus* possesses two types of signal peptidases, type I and type II signal peptidases. Type I signal peptidases are responsible for the cleavage of signal peptides in general whilst type II signal peptidases are aimed at lipoprotein signal peptides (Tjalsma *et al*., 1997). The latter, which are also denoted SPase II or prolipoprotein signal peptidases (LSp), cleave off signal peptides from precursors of diacylglycerol-modified prolipoproteins (Pragai *et al*., 1997).

Type I signal peptidases remove typical signal peptides from the majority of exported extracellular proteins (Dalbey *et al*., 1997). On the contrary, *Bacillus* species contain multiple type I signal peptidases (van Dijl *et al*., 1992; Bolhuis *et al*., 1996; Tjalsma *et al*., 1997). There are at least five genome-encoded type I signal peptidases in *Bacillus subtilis*. These enzymes, denoted, SipS, SipT, SipU, SipV and Sip W are made up of about 168 to 193 amino acids. Additionally two sip genes were found on certain *Bacillus subtilis* strains (Meijer *et al*., 1995). SipW is different from other members of the enzyme family and strikingly related to type I signal peptidases of eukaryotic organisms. The other SPases are highly related with an amino acid identity between 40-70% and are also functionally active (Tjalsma *et al*., 1997).

All these enzymes possess the characteristic five domains of type I signal peptidases (van Dijl *et al*., 1992) and the serine and lysine residues, which are essential for the formation of the catalytic dyad (van Dijl *et al*., 1995). All type I signal peptidases are important for efficient processing of precursor protein molecules prior release into the surrounding medium and to ascertain cell viability. This is an indication that these enzymes probably complement each other. Mutations inactivating some of the sip genes have confirmed the essentiality of certain signal peptidases in terms of cell viability. Most strains containing mutations inactivating more than one sip gene appeared to be still viable. However, no strain lacking either *SipT* or *SipS* was viable indicating that at least one member of these two genes should be present in a functional form for continuous growth of the cell (Tjalsma *et al*., 1997).

A Gram-positive type II signal peptidase gene was first cloned from *Staphylococcus aureus* by complementation of a conditionally lethal *Escherichia coli* mutant (Zhao and Wu, 1992). The molecular weight of its membrane protein is 18.3 kDa. On the contrary to the vast number of type I signal peptidases, the *Bacillus subtilis* chromosome was observed to contain only a single signal peptidase type II gene, *lsp* (Pragai *et al*., 1997). This gene accumulates lipomodified proteins with molecular masses corresponding to precursor but also to mature forms of *PrsA*, a lipoprotein involved in maturation of some extracellular proteins (Tjalsma *et al*., 1999). Disruption of the gene that encodes prelipoprotein diacylglycerol transferase, *lgt*, results in secreted *PrsA* and β-lactamase proteins that are not lipomodified. These proteins were, however, normally processed and released into the external medium (Leskela *et al*., 1999).

1.6. Secretion

Extracellular proteins produced by Gram-positive bacteria are efficiently secreted as they traverse through a single membrane due to lack of an additional outer layer. It is advantageous to produce proteins of interest in bacteria in secreted form since exported proteins usually maintain their native conformation, in contrast to intracellular production, which, in many cases results in aggregation of the proteins produced. Another advantage is that secretion facilitates simplified downstream processing (Archibald *et al*., 1993, Bron *et al*., 1998). *Bacilli* appear to be endowed with efficient secretory apparatus that is amenable for the secretory overexpression of commercially valuable homologous and heterelogous proteins (Behnke, 1992).

The are basically two types of exported *Bacillus* proteins, the first being the 'true' soluble exoproteins, mainly exoenzymes, which are often secreted in large quantities (several grams per liter) directly into the growth medium with very little, if any, remaining cell associated. Typical examples include extracellular degradative enzymes such as amylases and proteases. The second category comprises a heterogeneous group of proteins which to some extent remain associated with the cell wall and/or cytoplasmic membrane, including peptidoglycan-associated proteins such as staphylococcal protein A, and lipoproteins such as *Bacillus licheniformis* β-lactamase (Saier *et al*., 1989).

The majority of extracellular proteins secreted by *Bacillus* species appear to be exported through the cytoplasmic membrane via the Sec pathway (Figure 1.3), but several other alternative export pathways exist (Jiang *et al*., 2000). The first pathway which has been identified and denoted the twin-arginine translocation (Tat) is mediated by signal peptides with the RR-motif and is comprised of conserved components (Figure 1.3) (Bron *et al*., 2000). The assembly of extracellular prepilin-like structures depends on components that are most likely not involved in Sec-dependent protein secretion. Certain small prepeptides contain signal peptides without the hydrophobic domain. These peptides are transported across the membrane by ABC transporters (Bron *et al*., 2004).

1.6.1 The Sec pathway

In the major secretion pathway, integral membrane proteins secY, secE, secG and secDF are the main components of the secretion machinery (Fig 1.3) (Bolhuis *et al*., 1998; Markus *et al*., 1999). The energy required by this translocator for preprotein translocation is supplied by SecA. SecA is a peripheral membrane-associated ATPase, with high affinity for both the precursor/chaperone complex and the translocase. Repeated cycles of binding of SecA to the precursor, followed by its release from the translocase complex is a process that requires ATP binding and hydrolysis, during the initial stages of secretion (Meyr *et al*., 1999). Majority of the proteins that are channeled through the Sec pathway are translocated via the membrane in a more or less unfolded conformation to ensure easy passage through the cytoplasmic membrane (Bron *et al*., 2000).

1.6.2 The Tat pathway

It is presumed that this pathway specifically evolved for the export of folded preproteins since it transports unfolded proteins (Dalbey and Robinson, 1999; Berks *et al*., 2000). The pathway was first discovered in chloroplasts, where it is involved in pH-dependent protein import into the thylakoid lumen (Robinson *et al*., 1994; Settless *et al*., 1997). For the chloroplast system, it was observed that on the contrary to Sec-dependent translocation, proteins can be translocated in a folded conformation with this pathway (Hynds *et al*., 1998). Additionally, it was demonstrated that two adjacent arginines combined with a hydrophobic determinant (preferably leucine) at position +2 or +3, relative to the twin arginines, are essential for the N-domain of signal peptides to ensure direction of precursors towards this pathway (Brink *et al*., 1997; Cristobal *et al*., 1999).

Although the exact mechanism of protein transport via the Tat pathway is yet to be unraveled, five components of the Tat pathway have been described in *E. coli*. These include TatA (a putative membrane-bound receptor, homologous to the maize Hcf106 protein) (Settles *et al*., 1997), TatB (a TatA paralogue) (Sargent *et al*., 1998), TatC (the putative translocase), TatD (a predicted soluble protein) and TatE (a TatA paralogue) (Bron *et al*., 2004). Interestingly, *Bacillus* species contain three TatA/B/E homologues (encoded by the ydiI, yczB and ynzA genes), two homologues of TatC (encoded by the ydiJ and ycbT genes) and another TatD homologue, which is encoded by the yabD gene (Kunst *et al*., 1997; Bron *et al*., 2004). Of all the three proteins, TatA, TatB and TatE are structurally related. TatA and TatE are functionally redundant, and this means the presence of one of these components is not essential for the translocation of twin-arginine

signal peptides. On the other hand, TatB and TatC are indispensable for translocation activity (Robinson and Bolhuis, 2001; van Dijl *et al*., 2002). It was recently discovered that TatB and TatC are involved in twin-arginine signal peptide reception. Additionally, TatB and TatC, after forming a complex with TatA, formed a protein-conducting channel (Alami *et al*., 2003). It is yet to be confirmed if the TatA protein of *Bacillus* species is functionally equivalent to TatA and TatB of *E. coli* (Bron *et al*., 2004).

1.6.3 Type IV pilin export

Another class of proteins that are secreted via a Sec-independent fashion consists of type IV pilin like proteins that are encoded by the *comGC*, *comGD*, *comGE* and the *comGG* genes. Their corresponding products are involved in the development of genetic competence. These proteins resemble type IV pilins of various Gram-negative bacteria that are formed as precursors carrying cleavable signal peptides. Although prepilin signal peptides demonstrate certain similarities to signal peptides of secretory proteins and lipoproteins, the prelipin proteins are believed to bypass the Sec and Tat pathways, as their translocation is dependent on a cleavage event at the cytoplasmic side of the membrane (Chung *et al*., 1998; Lory *et al*., 1994; Nunn *et al*., 1991). *ConC*, the signal peptidase that cleaves products of comG, ComGF and ComGA is believed to be involved in the assembly of pilin-like *ComG* proteins (Chung *et al*., 1998). Processing of *ComG* products is required for the assembly and anchoring of the pilin-like structures to the membrane, which in turn is needed for DNA binding during transformation (Dubnau, 1997).

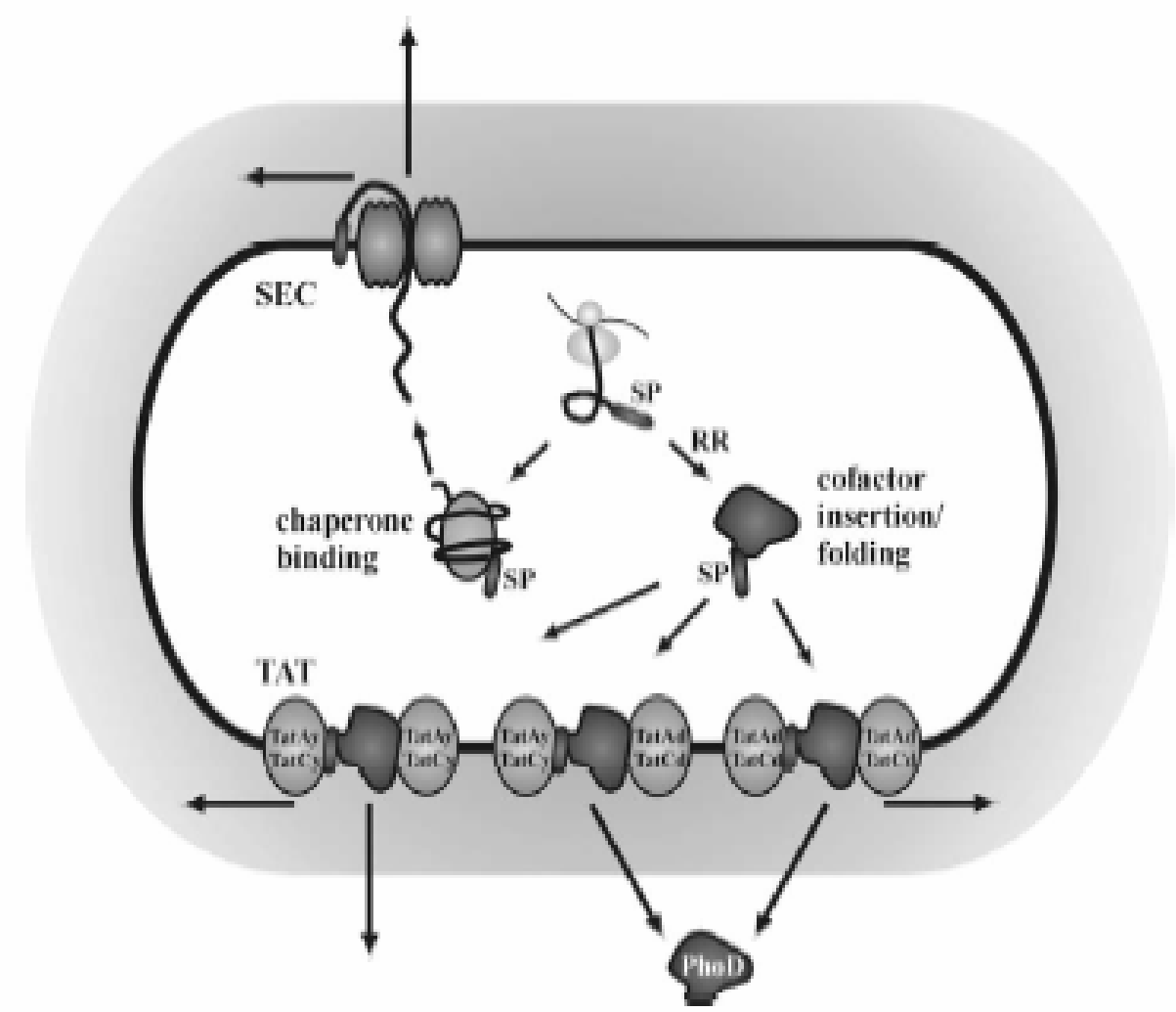

Figure 1.3. Model for protein secretion via the Sec and Tat pathways of *Bacillus*. The Sec pathway is made up of SecAYEG and it secretes unfolded precursors with the use of energy from ATP. Precursors with a twin-arginine (RR-) signal peptide have the potential to fold in the cytoplasm before their translocation by the Tat machinery in the membrane. Upon translocation, processing by signal peptidase, and cell wall passage, the folded mature proteins are secreted into the growth medium. SP, signal peptide; SEC, Sec pathway; TAT, Tat pathway (taken from van dijl *et al*., 2002).

1.6.4 Protein folding

The folding of precursor proteins in their journey to the external environment is mediated by chaperones (Tjalsma *et al*., 2004). *Bacillus* species have two classes of chaperones, intracellular and extracytoplasmic molecular chaperones. GroES, GroEL, DnaK, DnaJ and GrpE make up the intracellular chaperones. Indigenous proteins are efficiently prevented from folding in their passage through the cell membrane in *Bacillus* species (Bron *et al*., 1999). However, heterologous proteins may result in the formation of insoluble aggregates in the cytoplasm due to limited activity of intracellular molecular chaperones. Amongst the functions performed by chaperones include their ability to minimize aggregation of preproteins and ensure their translocation-competent conformation (Yuan *et al*., 1995).

PrsA is the only known extracytoplasmic folding factor in *Bacillus* species. Leskela *et al* (1999) illustrated that PrsA, a lipoprotein composed of a 33-KDa lysine-rich protein part and the N-terminal cysteine with a thiol-linked diacylglycerol anchoring the protein, is essential for efficient protein secretion. The folding of the mature protein molecule into a stable and active conformation following its secretion from the secretion machinery usually requires PrsA. Mutants lacking the PrsA gene confirmed the indispensability of PrsA in the secretion of extracellular proteins, whilst increased expression of the PrsA gene enhanced the secretion of exoproteins (Kontinen and Sarvas, 1993; Kontinen *et al*., 1993).

For the production of extracellular proteins, a higher cellular level of endogenous molecular chaperones and PrsA may result in an increased production. Wu *et al* (1998) proved that chaperones are essential in the improvement of the secretory production of an antidigoxin single chain antibody (SCA) fragment from *Bacillus subtilis*. Some of the constructed strains reduced the formation of insoluble SCA by 45% and increased the secretory production yield by 60%. In general, increased production levels meant increased reduction of the formation of inclusion bodies and enhanced the secretion of secretory proteins (Wu *et al*., 1993).

1.7. Proteolysis

Proteases secreted by *Bacillus* species severely affect the production and secretion of both homologous and foreign proteins by these bacteria. Some exoenzymes from certain *Bacillus* species appear to be insensitive to their secreted proteases, but most heterologous proteins are degraded. It has been necessary, therefore, to construct strains which have low protease activities but which nevertheless grow to a high biomass. Proteolytic degradation of heterologous proteins is one of the major obstacles in developing *Bacillus* into efficient secretory expression systems suitable for biotechnological applications. The two major enzymes are neutral proteases (or metaloproteases) that require divalent cations for activity and are therefore sensitive to EDTA; and alkaline proteases (or serine proteases) that are inhibited by phenomethylsulfonyl fluoride (PMSF) (Himanen *et al*., 1990). In addition, several so-called minority proteases are also exported by *Bacillus* species (Simonen and Palva, 1993).

Since the production of secreted proteases occurs primarily in stationary phase, it may be possible to use controllable promoters to induce a burst of synthesis and secretion during the transition from growth to stationary phase in order to minimize exposure to proteases. A more radical alternative may be possible for proteins whose degradation cannot be avoided using a combination of protease-defective *Bacillus* strains, appropriate media and timing of expression. This would involve developing expression-secretion systems for alternative *Bacillus* species that produce much lower protease levels. A particular strain of *B. brevis*, for example, has been reported to produce no detectable intracellular or extracellular protease activity in some media (Mezes and Lampen *et al*., 1985, Takao *et al*., 1989). In order to overcome the problem of proteolytic degradation of heterologous proteins by *Bacillus* species hosts, the cloned protease genes have been used to genetically engineer *Bacillus* strains to generate host strains stably deficient in exoprotease formation. The most appropriate way of inactivating protease genes is the deletion of upstream expression signals with part or the entire prepro–sequences. In addition to inactivating the protease gene, this approach also avoids the expression of truncated proteases that, although inactive, could cause interference in the secretion pathway (Stephenson and Harwood, 1998).

1.8. Production of heterologous proteins

Bacillus species are ideal candidates as hosts for the manufacture of heterologous proteins on a commercial scale due to their ability to efficiently secrete large quantities of proteins directly into the surrounding environment. On the contrary, Gram-negative bacteria results in the accumulation of the desired protein product intracellularly in the

protoplasm or the periplasmic space, and such accumulation may be toxic or lethal to the host cell. Intracellular accumulations often result in problems related to the formation of insoluble inclusion bodies, incorrect protein folding, and inefficient disulfide bond formation. The most exploited Gram-negative bacterial host, *E. coli*, usually produce endotoxins, which are toxic to human and animals, and can thus generate problems if they are used in the production of recombinant proteins such as biopharmaceuticals (Schallmey *et al*., 2004).

Exoprotein genes of Gram-positive bacteria are usually expressed in *Bacillus* species with their own promoters. The same proteins can be secreted to the medium by the aid of their own secretion signals. Proteins of Gram-negative bacteria origin, on the contrary, are generally secreted by the aid of secretion vectors based on promoters and ribosome binding sites originating from Gram-positive bacteria, since most of these elements are nonfunctional in *Bacillus*. The joint between the vector and the foreign gene is usually made at or near the signal peptide cleavage site, since a joint close to the promoter or the ribosome binding site may interfere with their functions or cause unfavorable modifications in the 5' end of the mRNA (Wu and Wong, 1999).

A great number of extracellular proteins from different organisms have been cloned and expressed in *Bacillus* (Table 1.7). In the examples listed in the table, the entire signal peptide or a substantial part of it is derived from a *Bacillus* extracellular protein. However, there is no evidence that the signal peptides of gram-negative bacteria would

be nonfunctional in *Bacillus* species, although there are indications that they would not be optimal for protein export in *Bacillus* (Bron *et al*., 2004).

Many periplasmic and extracellular proteins of Gram-positive and Gram-negative bacteria are efficiently secreted by *Bacillus* species (Table 1.7). The yield of the secreted protein depends mainly on the expression system applied and on the efficiency of the mechanism used to protect foreign protein against protease attack of the host (Lam *et al*., 1998). Several bottlenecks have been encountered when attempts were made to cause secretion of eukaryotic proteins in bacilli. Many of these proteins are poorly exported despite being secretory proteins by nature, and some of them appear to be toxic to the producer cell. The toxic effect may also be due to the inefficient export of the heterologous protein. Production of eukaryotic proteins is further hampered by proteolytic degradation, and so far, only few of them have been secreted in *Bacillus* species with reasonable yields (Tjalsma *et al*., 1997).

Table 1. Some foreign proteins expressed in *Bacillus* species.

Protein	Origin	Host	P/SS source	Secretion	Yield/Comment	References
α-amylase	Human	*B. pumilus*	mwp (*B.bre)*	+	40 mg/l	Udaka *et al*., 1989
Prochymosin	Calf	*B. lich*	amy (*B. lich*)	-	Efficient expression	Vos *et al*., 1988
Proinsulin	Rat	*B. subtilis*	penP (*B. lich*)	+	7-10μg/l	Vosbach *et al*., 1983
EGF	Human	*B. brevis*	mwp (*B. bre*)	+	240 mg/ml	Yamagata *et al*., 1989
Staphylokinase	*S. aureus*	*B. subtilis*	Intact	+	50 mg/ml	Gerlach *et al*., 1988
α-amylase	*B. stearo*	*B. brevis*	Intact	+	3 g/l	Takagi *et al*., 1985
β-lactamase	*B. cereus*	*B. subtilis*	Intact	+	40 mg/ml	Wang *et al*., 1985
Serum albumin	Human	*B. subtilis*	amy (*B. amy*)	-	Not secreted	Schatz, 1991
Pepsinogen	Swine	*B. brevis*	mwp (*B. bre*)	+	11 mg/ml	Takao *et al*., 1989
Growth hormone	Human	*B. subtilis*	nuc (*S. aures*)	-	Not expressed	Miller *et al*., 1985
Insulin	Human	*B. subtilis*			1 g/l	Olmos-soto and Contreras-Flores, 2003
IL-3	Human	*B. lich*	amy (*B. lich*)	+	100 mg/ml	Van leen *et al*., 1991
Toxin	*C. dipht*	*B. subtilis*	amy (*B.amy*)	±	4 mg/ml	Hemila *et al*., 1989
Staphylokinase	*S. aureus*	*B. subtilis*			340 mg/ml	Ye *et al*., 1999
Endoglucanase A	*C. thermo*	*B. stearo*	celA(*C. thermo*)	+	108 mg/ml	Joliff *et al*., 1989
Taka-amylase A	*A. oryzae*	*B. brevis*	HWP (*B. bre*)	+	22 mg/ml	Ebisu *et al*., 1993
Pectinase	*E. carato*	*B. subtilis*	amy (*B. amy*)	+	1000mg/ml	Saris *et al*., 1990
Protein G	*S. strain* G148	*B. subtilis*	apr (*B. lich*)	±	50 mg/l	Egnell *et al*., 1989
PHA depolymerase A	*P. lemoi*	*B. subtilis*	P43/*sacB*	+	1.9 mg/l	Braaz *et al*., 2002
Penicillin G acyclase	*B. megaterium*	*B. subtilis*	P43/*sacB*	+	6 U/ml	Yang *et al*., 1999
Interferon-α_2	Human	*B. subtilis*	amy (B. amy)	+	1.2 mg/l	Schein *et al*., 1986

Note: *P. Lemoi*, *Paucimonas lemoignei; C. dipht, Corynebacterium diphtheriae; C. thermo, Clostridium thermocellum; B. stearo, Bacillus stearothermophilus; E. carato, Erwinia caratovora; S. strain, Streptococcus strain; B. lich, Bacillus licheniformis; B. bre, Bacillus brevis; B. amy, Bacillus amyloliquefaciencs; S. aureus, Staphylococcus aureus; A. oryzae, Aspergillus oryzae* P, Promoter; SS, Signal sequence amy, amylase; mwp, molecular weight protein; penP, β-lactamase; nuc, nuclease; apr, alkaline protease; celA, endoglucanase A; EGF, epidermal growth factor; IL-3, Interleukin-3;HWP, cell wall protein; Intact, the gene is under control of its own native promoter with its own signal sequence.

1.9. Vectors that are commonly used in the development of *Bacillus* expression-secretion systems.

Bacillus expression systems are developed using both integrative and autonomously replicating vectors. For convenience purposes, majority of the vectors are shuttle vectors that can replicate in two dissimilar organisms. Such vectors are usually furnished with a selection marker for selective pressure; the expression cassette and the elements for maintenance in the host cell (Wu and Wong, 1999). In most cases, the vectors are introduced into *Bacillus* species by transformation, though some strains are difficult to transform. Conventionally, this is usually circumvented using protoplast transformation, which is more efficient and independent of the natural transformation pathways and allow easy introduction and maintenance of foreign genes in the host (Wang and Doi, 1992). Other techniques that ensure stable maintenance of recombinant clones are through electroporation transformation of the hard-to-transform *Bacillus* strains (Brigidi *et al*., 1990).

A number of endogenous plasmids have been found in many *Bacillus* species. These plasmids, however, do not permit direct selection of transformants carrying the plasmids of some *Bacillus* species and control no known host functions such that they are cryptic plasmids. Since it has been difficult to obtain useful plasmid vectors in *Bacillus*, autonomously replicating vectors have been secured from other bacteria that carry selectable markers, replicate and are stably maintained in *Bacillus* species. Staphylococcal plasmids (pUB110 and pE194) can replicate autonomously and are stably

maintained in *Bacillus* species, and have formed the basis for the development of various cloning and expression vectors (Wu and Wong, 1999). Cloning studies in *Bacillus* have been performed utilizing pUB110, pC194 or pE194 or shuttle vectors achieved by joining these with *Escherichia coli* replicating plasmids.

1.9.1. Autonomously replicating vectors

1.9.1.1. pUB110

This plasmid was initially identified in *S. aureus* as a small, high copy extrachromosomal element that specifies resistance to kanamycin and neomycin (Gryczan, 1982). Both resistances are due to a single gene found in this plasmid that encodes the enzyme kanamycin (neomycin) nucleotidytransferase. The molecular mass of pUB110 is 4548 base pairs as determined by nucleotide sequencing of the whole plasmid. pUB110 is stably maintained in *Bacillus* for at least 80 generations in the absence of selective pressure at a copy number estimated at about 50 chromosomes per equivalent.

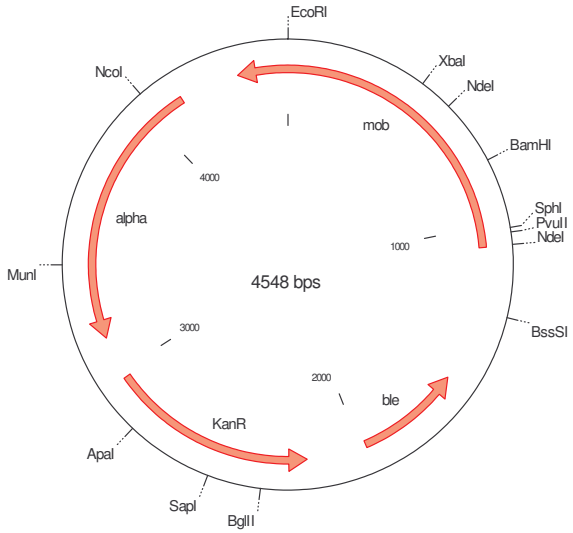

Figure 1.6. A schematic diagram of pUB110, a *Bacillus* replicating plasmid from *Staphylococcus aureus*.

1.9.1.2. pE194

This plasmid was originally detected in *S. aureus* and transformed into *B. subtilis* (Gryczan, 1982). It has been extensively studied and its primary structure has been determined. It contains extremely active recognition sequences for transcription and translation systems, which can be used to synthesize methylase, which gives resistance to erythromycin. pE194 is 3728 (Figure 1.7) base pairs in size and is maintained at a copy number of about 10 in *Bacillus* at 32ºC. This plasmid also specifies resistance to macrolide, lincosamide and streptogramin (MLS) antibiotics. All these resistances are due to a single gene, *erm*, whose expression is erythromycin–inducible and specifies an enzyme involved in methylating adenine residues in 23S rRNA. This methylation reduces affinity of the ribosome for MLS antibiotics. pE194 exhibit two properties not shown by pUB110 or pC194. First, pE194 is naturally temperature–sensitive for replication. The plasmid replicates at 32ºC, but at temperatures above 45ºC (up to 50ºC), replication ceases and the plasmid is lost from *Bacillus*. The second unusual feature of pE194 was detected when *Bacillus subtilis* was grown on erythromycin-containing media at temperatures, which were nonpermissive for pE194. The result was random integration of copies of pE194 into the host chromosome at various sites (Kwak and Weisblum, 1994).

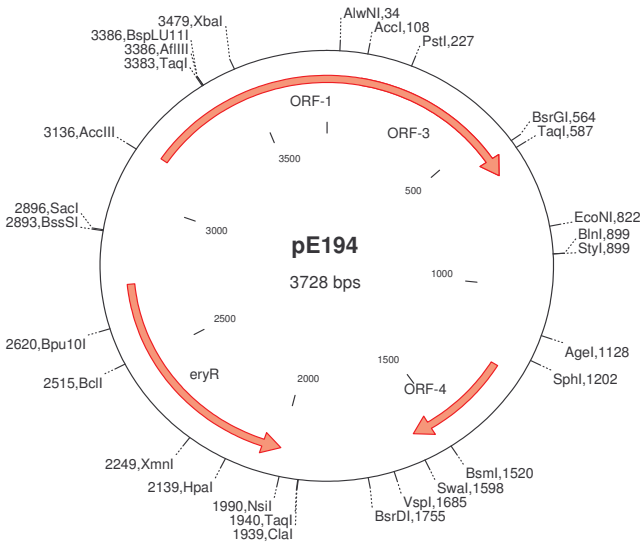

Figure 1.7.A schematic representation of pE194., a *Bacillus* replicating thermosensitive plasmid.

1.9.1.3. pC194

The third commonly used plasmid in genetic and recombinant DNA experiments in *Bacillus* species is pC194. The complete nucleotide sequence of pC194 reveals that the plasmid consists of 2910 base pairs. pC194 is stably maintained in *B. subtilis* at a copy number of about 15 per cell, and this plasmid specifies resistance to the antibiotic chloramphenicol due to the presence of a gene for chloramphenicol acetyltransferase. The expression of this *cat* gene is inducible by chloramphenicol, as are all *cat* genes of Gram–positive origin. pC194 originated in *Staphylococcus aureus*; however, it has been observed that it has the ability to replicate not only in Gram-positive bacteria but also in *Saccharomyces* and *Escherichia coli* (Byeon and Weisblum, 1984, Goursort *et al*., 1982).

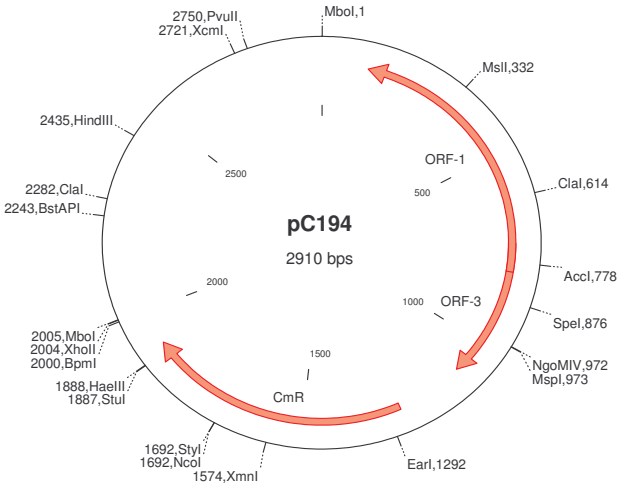

Figure 1.8. A diagram of pC194, a chloramphenicol resistant vector which replicates in *Bacillus.*

1.9.3. *Bacillus* integrative vectors

Several integrative vectors have also been applied in the development of *Bacillus* expression systems (Moqk *et al*., 1996, Kim *et al*., 1997). These vectors are exclusively integrated into the genome of host cells either randomly or at single specific loci through homologous recombination. Several bacilli integrative vectors have been extensively applied in the development of expression and secretion systems for both heterologous and homologous protein expression with efficient results. Majority of these vectors are stably maintained in the host cell. (Schumann *et al*., 1996).

These vectors allow the integration of any expression cassette together with the gene of interest, coding for its authentic protein, at the locus relative to the sequence homologous to the sequence of the vector. The controllable expression cassette is usually composed of the encoding gene, the promoter and ribosome binding site region, sandwiched between the 5'- and 3'-ends the sequence homologous to the locus required for integration for double crossover of the cassette. This thereby allows insertion of in vitro constructed transcriptional fusions at that particular specific locus of the *Bacillus* chromosome (Moqk *et al*., 1996; Kim *et al*., 1997). Examples of integrative vectors that have been used to date include Φ105 (Ho and Kim, 2003); vectors based on the amyE locus of the host cell (Moqk *et al*., 1997) and others based on the protease locus (Kim *et al*., 1996).

1.9.2 Autonomously replicating versus integrative vectors

Majority of the *Bacillus* species based expressions-secretion systems developed to date were constructed using autonomously replicating sequences (ARS). Autonomous replication of the vector is advantageous since it presents greater flexibility and compatibility, and if a vector is a high copy number plasmid, the dosage of the gene of interest is automatically increased (Wong and Wu, 1999). However the only major shortcoming associated with ARS-based expression systems is the requirement for maintenance of a selective pressure that in certain cases is not always compatible with efficient management (Bron *et al*., 1999).

Integrative vectors are exclusively introduced into *Bacillus* species by homologous recombination which can either be a single or a double crossover depending on the integration cassette (Kim *et al*., 1996). Integrated vectors exhibit a very high stability and can be stably maintained in the host for a number of generations. In addition to the advantages they offer over replicative vectors, integrative *Bacillus* vectors present the possibility of multiple integrations and of a correlated increase in gene expression (Lim and Ho, 2003).

1.10. Conclusions

The opportunities for genetic engineering in several species of *Bacillus*, other than *B. subtilis* are now becoming a reality. A number of species from the *Bacillus* genus presents advantages in the development of expression systems. While a major commercial application of *Bacillus* is in industrial enzymes, significant contributions have also been made in other areas. The powerful techniques of genetic and protein engineering have dramatically increased our basic understanding of *Bacillus* genetics and biochemistry, and have opened up new possibilities for commercial exploitation of *Bacillus* species. Advances in gene expression, gene regulation, and vector development have also offered new approaches to strain improvement.

It is now possible to increase production levels in many *Bacillus* strains that are already high producers of amylases, proteases, penicillinases, and other useful pharmaceutical and food products, by the introduction of such genes on high copy number plasmids. Possible problems in gene regulation, expression, limits on protein production and

secretion are encountered, but recent reports on comparisons of expression of bacterial and eukaryotic genes in *Bacillus* species suggest that *Bacillus* species prove to be better expression hosts. Large varieties of genetic tools have also been developed to facilitate proper adaptations for *Bacillus* species improvement in terms of expression systems construction (Harwood and Cutting, 1990).

A number of *Bacillus* species have been genetically engineered for heterologous gene expression. The majority of these species gives high extracellular expression levels with minimal loss of product due to proteolytic cleavage subsequent to secretion. However, notwithstanding their inherent advantages, the secretion of various heterologous proteins by bacilli, in particular proteins of eukaryotic origin, is frequently inefficient, which limits the application potential of these organisms

Vectors for *Bacillus* include both integrating as well as autonomously replicating forms. The essential elements of these vectors are the strong promoter, efficient terminator, suitable signal sequence, and a selection marker. Extensive studies on plasmid stability have shown the vectors to be recombinationally stable during fermentation in the absence of selective pressure (Schallmey *et al*., 2004).

Evidently, *Bacillus* species have several attractive properties with respect to the development of expression systems. *Bacillus licheniformis* have also been used for several decades for the industrial production of alpha-amylase and various proteases. With the classification of GRAS (Generally Recognized As Safe) and a huge capacity for protein secretion, this organism is perfectly suited as an amenable host for heterologous protein expression. Some strains of this *Bacillus* species have a transparent genome,

because the complete sequence is known (Rey *et al*., 2004), and this factor makes *Bacillus licheniforms* accessible in terms of genetic manipulation.

With this background, the objective of the study was to develop *a Bacillus licheniformis* lipase-gene based expression-secretion systems that will facilitate heterologous expression. Desirable properties would be a strong controllable promoter, a signal peptide for secretion, a transcription terminator for stability of the expression system, a selection marker to select plasmid-containing strains.

In the first chapter of the study, the cloning of the complete extracellular lipase gene is described, as well as the procedure applied to clone the gene. The complete extracellular gene encoding the lipolytic activity of *Bacillus licheniformis* was cloned and compared with gene sequences from other related *Bacillus* species. Chapter 3 describes the development of a lipase-gene based expression-secretion system and its application in the expression of Taq polymerase from *Thermus aquaticus* and carboxylesterase from *Bacillus pumilus*, which were used as models.

Aiba S., Kitai K., Imanaka T. (1983). Cloning and expression of thermostable alpha-amylase gene from *Bacillus stearothermophilus* in *Bacillus subtilis*. *Appl. Environ. Microbiol.* **46**: 1059-1065.

Akino T, Kato C, Horikoshi K. (1989). Two *Bacillus* beta-mannanases having different COOH termini are produced in *Escherichia coli* carrying pMAH5. *Appl. Environ. Microbiol.* **55**:3178-3183.

Andersson, H., E. Bakker and G. von Heijne, (1992). Different positively charged amino acids have similar effects on the topology of a polytopic transmembrane protein in *Escherichia coli*. *J. Biol. Chem.* **267**:1491-1495.

Archibald, A. R., I. C. Hancock, and C. R. Harwood. (1993). Cell wall structure, synthesis, and turnover, p. 381-410. *In* A. L. Sonenshein, J. A. Hoch, and R. Losick (ed.), *Bacillus subtilis* and other gram-positive bacteria. American Society for Microbiology, Washington, D.C.

Baird S.D., Johnson D.A., Seligy V.L. (1990). Molecular cloning, expression, and characterization of endo-beta-1,4-glucanase genes from *Bacillus polymyxa* and *Bacillus circulans*. *J. Bacteriol.* **172**:1576-1586.

Band L., Henner D.J. (1984) *Bacillus subtilis* requires a "stringent" Shine-Dalgarno region for gene expression. *DNA* **3**: 17-21.

Band L., Henner D.J., Ruppen M. (1987). Construction and properties of an intracellular serine protease mutant of *Bacillus subtilis*. *J. Bacteriol.* **169**: 444-446.

Band L., Yansura D.G., Henner D.J. (1983). Construction of a vector for cloning promoters in *Bacillus subtilis*. *Gene* **26**: 313-315.

Bcrgey, D. H., Breed, R. S., Murray, E. G. D., and Ilitchens, A. P. (eds.) 1939. Bergey's Manual of Determinative Bacteriology. 5th ed. Williams and Wilkins, Baltimore.

Bechhofer D.H., Dubnau D. (1987). Induced mRNA stability in *Bacillus subtilis*. *Proc Natl. Acad. Sci USA* **84**: 498-502.

Behnke, D., (1992). Protein export and the development of secretion vectors. In: Doi, R.H., McGloughlin, M. (Eds.), Biology of Bacilli: Applications to industry. MA: Butterworth-Heinemann, pp. 143-188.

Bellini A.V., Gali A., Fascietti G., Branduzzi P., Luchasse G., Grandi G. (1991). Production process of recombinant 1L – 1 from *Bacillus subtilis*: comparison between intracellular and extracellular expression. *J. Biotechnol.* **18**: 177-192.

Benson SA, Hull MN, Silhavy TJ (1985). Genetic analysis of protein export in *Escherichia coli* K12. *Annu. Rev. Biochem.* **54**:101-34.

Bergquist P.L., Love D.R., Croft J.E., Streiff M.B., Daniel R.M., Morgan W.H., (1987). Genetics and potential biotechnological applications of thermophilic and extremely thermophilic micro-organisms. *Biotechnol. Genet. Eng. Rev.* **5**:199-244.

Berks, B.C. (1996). A common export pathway for proteins binding complex redox cofactors? *Mol. Microbiol* .**22**: 393-404.

Bolhuis, A., Broekhuizen, C.P., Sorokin, A., van Roosmalen, M.L., Venema, G., Bron, S., Quax, W.J. and Van Dijl, J.M., (1998). SecDF of *Bacillus subtilis*, a molecular Siamese twin required for the efficient secretion of proteins. *J. Biol. Chem.* **273**: 21217-21224.

Bolhuis, A., Mathers, J.E., Thomas, J.D., Barrett, C.M. and Robinson, C., (2001). TatB and TatC form a functional and structural unit of the twin-arginine translocase from *Escherichia coli*. *J. Biol. Chem.* **276**: 20213-20219.

Bolhuis, A., Tjalsma, H., Smith, H.E., de Jong, A., Meima, R., Venema, G., Bron, S. and van Dijl, J.M., (1999). Evaluation of bottlenecks in the late stages of protein secretion in *Bacillus subtilis*. *Appl. Environ. Microbiol.* **65**: 2934-2941.

Bookstein C, Edwards C.W, Kapp N.V, Hulett F.M. (1990). The *Bacillus subtilis* 168 alkaline phosphatase III gene: impact of a phoAIII mutation on total alkaline phosphatase synthesis. *J Bacteriol.* **172**:3730-3737.

Bookstein, C., C. W. Edwards, N. V. Kapp, and M. Hulett. (1990). The *Bacillus subtilis* 168 alkaline phosphatase III gene:impact of a phoAIII mutation on total alkaline phosphatasesynthesis. *J. Bacteriol.* **172**: 3730-3737.

Borchert T.V. and V. Nagarajan, (1991). Effect of signal sequence alterations on export of levansucrase in *Bacillus subtilis*. *J. Bacteriol.* **173**: 276-282.

Braaz, R., S.L. Wong and D. Jendrossek, (2002). Production of PHA depolymerase A (PhaZ5) from *Paucimonas lemoignei* in *Bacillus subtilis*. *FEMS Microbiol. Lett.* **209**: 237-241.

Brammar WJ, Muir S, McMorris A (1980) Molecular cloning of the gene for the beta – lactamase of *Bacillus licheniformis* and its expression in *Escherichia coli*. *Mol. Gen. Gen.* **178**: 217-224.

Braun, P., Gerritse, G., van Dijl, J.M. and Quax, W.J., (1999). Improving protein secretion by engineering components of the bacterial translocation machinery. *Curr. Opin. Biotech.* **10**: 376-381.

Briggs, M.S., D.G. Cornell, R.A. Dluhy and L.M. Gierasch, (1986). Conformations of signal peptides induced by lipids suggest initial steps in protein export. *Science* **233**: 206-208.

Brink, S., Bogsch, E.G., Edwards, W.R., Hynds, P.J. and Robinson, C., (1998). Targeting of thylakoid proteins by the delta pH-driven twin-arginine translocation pathway requires a specific signal in the hydrophobic domain in conjunction with the twin-arginine motif. *FEBS Lett.* **434**: 425-430.

Bron, S., A. Bolhuis, H. Tjalsma, S. Holsappel, G. Venema and J.M. van Dijl, (1998). Protein secretion and possible roles for multiple signal peptidases for precursor processing in bacilli. *J. Biotech.* **64**: 3-13.

Bron, S., Meima, R., van Dijl, J.M., Wipat, A. and Harwood, C.R., (1999). Molecular biology and genetics of *Bacillus* species. In: Demain, A.L. and Davies, E., Editors, 1999. *Manual of Industrial Microbiology and Biotechnology* (second ed.), ASM press, Washington, DC, pp. 392-416.

Brown, T.A., (1990). Genomes and genes. In: Brown, T.A. (Ed.), Molecular Biology Labfax. BIOS Scientific Publisher, Oxford, pp. 235-254.

Bruckner R, Doi R.H. (1987). Characterization of a gene for a minor extracellular protease from *Bacillus subtilis* In: 4th International Conference on Genetics and Biotechnology of *Bacilli*, San Diego, July 1987, Abstract.

Byeon W.H., Weisblum B. (1984). Posttrasnscriptional regulation of chloramphenicol acetyltransferase. *J. Bacteriol.* **158**: 543-550.

Chamberlin M.J., (1974). The selectivity of transcription. *Annu. Rev. Biochem.* **43**:721-75.

Chen, M. and Nagarajan, V., (1994). Effect of alteration of charged residues at the N termini of signal peptides on protein export in *Bacillus subtilis*. *J. Bacteriol.* **176**: 5796-5801.

Chung, Y. J., M. T. Steen, and J. N. Hansen., (1992). The subtilin gene of *Bacillus subtilis* ATCC 6633 is encoded in an operon that contains a homolog of the hemolysin B transport protein. *J. Bacteriol.* **174**:1417-1422.

Chung, Y.S., F. Breidt and D. Dubnau, (1998). Cell surface localization and processing of the ComG proteins, required for DNA binding during transformation of *Bacillus subtilis*. *Mol. Microbiol.* **29**: 905-913.

Conn, F. 1872. Untersuchungen über Bakterien in: Beiträge zur Biologie der Pflanzen 1, 1875 (Heft 1), J. U. Kern's Verlag (Max Müller), Breslau. Pages:127-224

Connors M..J, Mason M.J., Setlow P., (1986) Cloning and sequencing of genes for three small and soluble proteins of *Bacillus subtilis*. *J. Bacteriol.* **166**: 417-425.

Cornelles P, Dignette C, Willemot K.m (1982) Cloning and expression of *a Bacillus coagulans* amylase gene in *Escherichia coli*. *Mol. Gen Genetic.* **186**: 507-511.

Cristóbal, S., de Gier, J.W., Nielsen, H. and von Heijne, G., (1999). Competition between Sec- and Tat-dependent protein translocation in *Escherichia coli*. *EMBO Journal* **18**: 2982-2990.

Crutz A, Steinmetz M, Aymerich S, Richter R, Le coq D (1990). Induction of levansucrase in *Bacillus subtilis*: an antitermination mechanism negatively controlled by the phosphotransferase system. *J. Bacteriol.* **172**:1043-1050.

Dalbey, R.E. (1990). Positively charged residues are important determinants of membrane protein topology. *Trends Biochemical Science* **15**: 253-257.

Dalbey, R.E. and Robinson, C., (1999). Protein translocation into and across the bacterial plasma membrane and the plant thylaloid membrane. *Trends Biochem. Sci.* **24**: 17-22.

Dalbey, R.E., Lively, M.O., Bron, S. and van Dijl, J.M., (1997). The chemistry and enzymology of the type I signal peptidases. *Prot. Science* **6**: 1129-1138.

den Blaauwen T., J.P.W. van der Wolk, C. van der Does, K.H.M. van Wely and A.J.M. Driessen, (1999). Thermodynamics of nucleotide binding to NBS-I of the *Bacillus subtilis* preprotein translocase subunit SecA. *FEBS Lett.* **458**: 145-150.

den Blaauwen, T., P. Fekkes, J.G. de Wit, W. Kuiper and A.J.M. Driessen, (1996). Domain interactions of the peripheral preprotein translocase subunit SecA. *Biochem.* **35**: 11994-12004.

Diderichsen B, Wedsted U, Hedegaard L, Jensen BR, Sjøholm C. (1990). Cloning of aldB, which encodes alpha-acetolactate decarboxylase, an exoenzyme from *Bacillus brevis*. *J Bacteriol.* **172**:4315-4321.

Diderichsen, B., and L. Christiansen. (1988). Cloning of a maltogenic alpha-amylase from *Bacillus stearothermophilus*. *FEMS Microbiol. Lett.* **56**:53-60.

Doi RH (1984) Genetic engineering in *Bacillus subtilis* In: Russel EG (editor) Biotechnology and Genetic engineering reviewed, volume 2, Intercept, Newcastle Upon Tyne, UK, pp 121-153.

Doi RH, Wang LF (1986) Multiple prokaryotic ribonucleic acid polymerase sigma factors. *Microbial. Rev.* **50**: 227-243.

Doi RH, Wong S, Kawamura F (1986) Potential use of *Bacillus subtilis* for secretion and production of foreign proteins, *Tends Biotech.***4**: 232-235.

Doi, R. H. (1991). Proteolytic activities in *Bacillus*. *Curr. Biotech.* **2**: 682-684.

Doi R.H., Wang L-F, Brucker R (1996). Extracellular serine protease and a *Bacillus subtilis* alkaline neutral serine protease mutant strain. United States Patent No. 5 585 0253.

Driessen AJ, van Wely KH, Swaving J, Freudl R, (2001). Translocation of proteins across the cell envelope of Gram-positive bacteria. *FEMS Microbiol. Rev.* **4**: 437-454.

Driessen, A.J., Manting, E.H. and van der Does, C., (2001). The structural basis of protein targeting and translocation in bacteria. *Nat. Struct. Biol.* **8**: 492-498.

Duvall E.J., Williams, D.M., Mongkolsuk, S. and Lovett, P.S., (1984). Regulatory regions that control expression of two chloramphenicol-inducible cat genes cloned in *Bacillus subtilis*. *J. Bacteriol.* **158**: 784-790.

Ebisu S, Mori M, Takagi H, Kadowaki K, Yamagata H, Tsukagoshi N, Udaka S. (1993). Production of a fungal protein, Taka-amylase A, by protein-producing *Bacillus brevis* HPD31. *J Ind Microbiol.* **11**:83-88.

Ehrlich SD, Noirot P, Petit MA *et al* (1986) in Genetic engineering volume 8 (Setlow JK and Hollaender S, editors) Plenum Press, New York, pp 71-83.

Emr SD, Silhavy TJ (1983). Importance of secondary structure in the signal sequence for protein secretion. *Proc Natl Acad Sci U S A.* **80**:4599-4603.

Evans L, A. Feucht and J. Errington (2004). Genetic analysis of the *Bacillus subtilis sigG* promoter, which controls the sporulation-specific transcription factor σ^G. *Microbiol.* **150**: 2277-2287.

Fahnestock RS, Fisher KE (1987) Protease–deficient *Bacillus subtilis* host strains for production of staphylococcal protein A. *Appl. Environ. Microbiol.* **52**: 379-384.

Fahnestock SR, Fisher KE (1986) Expression of the Staphylococcal protein A gene in *Bacillus subtilis* by gene fusions utilizing the promoter from a *Bacillus amyloliquefaciens* alpha – amylase gene. *J. Bacteriol.* **165**: 796-804.

Fekkes P. and A.J.M. Driessen, (1999). Protein targeting to the bacterial cytoplasmic membrane. *Microbiol. Mol Biol. Rev.* **63**: 161-173.

Ferrari E, Henner D, Perego M, Hoch J (1987) Transcription of *Bacillus subtilis* subtilisin and expression of subtilisin in sporulating mutants. *J. Bacteriol.* **170**: 162 -167.

Flock J., Fotheringham I., Light J., Bell L., Derbyshire R., (1984). Expression in *Bacillus subtilis* of the gene for human urogastrone using synthetic ribosome binding sites. *Mol. Gen. Genet.* **195**:246-251.

Franke, C. M., J. Tiemersma, G. Venema, and J. Kok. (1999). Membrane topology of the lactococcal bacteriocin ATP-binding cassette transporter protein LcnC. Involvement of LcnC in lactococcin a maturation. *J. Biol. chem.* **274**: 8484-8490.

Fritze, D. (2002). identification-traditional approachesin: Applications and Systematics of *Bacillus* and Relatives. Blackwell Publishing, Oxford. . Pg 8-22.

Fujii M., Takagi M., Imanaka T., Aiba S., (1983) Molecular cloning of a thermostable neutral protease gene from *Bacillus stearothermophilus* in a vector plasmid and its expression in *Bacillus stearothermophilus* and *Bacillus subtilis*. *J. Bacteriol.* **154**: 831-837.

Fujita Y., Fujiita T., Miwa Y., Nihashi J.I., Aratani Y., (1986). Organization and transcription of the gluconate operon, *gnt*, of *Bacillus subtilis*. *J. Biol. chem.* **261**: 13744-13753.

Fujita Y., Fujita T., (1986). Identification and nucleotide sequence of the promoter region of the *Bacillus subtilis* gluconate operon. *Nucleic Acids Res.* **14**:1237-52.

Fujita Y., Fujita T., (1987). The gluconate operon gnt of *Bacillus subtilis* encodes its own transcriptional negative regulator. *Proc Natl Acad Sci U S A*. **84**:4524-4528.

Fukumori, F., T. Kudo, Y. Narahashi, and K. Horikoshi. (1986). Molecular cloning and nucleotide sequence of the alkaline cellulase gene from the alkalophilic *Bacillus sp.* strain 1139. *J. Gen. Microbiol.* **132**:2329-2335.

Fukusaki, E., W. Panbangred, A. Shinmyo, and H. Okada. (1984). The complete nucleotide sequence of the xylanase gene(xynA) of *Bacillus pumilus*. *FEBS Lett.* **171**:197-201.

Ganoza MC, Kofoid EC, Marlere P, Louis B (1987). Potential secondary structure at translation-initiation sites. *Nucleic Acids Res.* **15**:345-60.

Gartner D, Geissendofer M, Hiller W (1988). Expression of the *Bacillus subtilis* xylose operon is expressed at the level of transcription and is induced by xylose. *J. Bacteriol.* **170**: 1302-1309.

Geiduschek EP, Ito J (1982) Regulatory mechanisms in the development of lytic phages in *Bacillus subtilis* In: Dubnau D (editor) The molecular biology of the *Bacilli*, Academic press, New York, pp 203-245.

Gentz R, Langner A, Chang ACY Chang ACY, Cohen SN, Bujard H (1981) Cloning and analysis of a strong promoters is made possible by the downstream placement of an RNA termination signal. *Proc. Natl. Acad. Sci. USA* **78**: 4936-4940.

Gerlach D, Kraft R, Behnke D. (1988). Purification and characterization of the bacterial plasminogen activator staphylokinase secreted by a recombinant *Bacillus subtilis*. *Zentralbl Bakteriol Mikrobiol Hyg [A].* **269**:314-322.

Gerlach D, Kraft R, Behnke D. (1988). Purification and characterization of the bacterial plasminogen activator staphylokinase secreted by a recombinant *Bacillus subtilis*. *Zentralbl Bakteriol Mikrobiol Hyg* **269**:314-322.

Gierasch LM (1989). Signal sequences. *Biochem.* **28**:923-930.

Gonzy-Treboul G, Zagorec M, Rain-Guion MC, Steinmetz M. (1989). Phosphoenolpyruvate:sugar phosphotransferase system of *Bacillus subtilis*: nucleotide sequence of ptsX, ptsH and the 5'-end of ptsI and evidence for a ptsHI operon. *Mol Microbiol.* **3**:103-112.

Goursot R, Goze A, Niaudet B, Ehrlich SD (1982) Plasmids for Staphylococcus aureus replicate inn yeast *Saccharomyces cerevisiae*. *Nature* **298**: 488-490.

Grand G, Del Blue, Palla M, Mele A, Colleti E, Tome S (1985) New plasmid expression vectors for *Bacillus subtilis*. *Plasmid* **16**: 1-4.

Gray G.L., Mainzer S.E., Rey M.W., Lamsa M.H., Kindle K.L., Carmona C, Requadt C. (1986). Structural genes encoding the thermophilic alpha-amylases of *Bacillus stearothermophilus* and *Bacillus licheniformis*. *J. Bacteriol.* **166**:635-643.

Grossman M.J., Lampen J.O. (1987) Purification and DNA binding properties of the *bla* 1 gene product, repressor for the beta-lactamase gene, *bla*P of *Bacillus licheniformis*. *Nucleic Acids Res.* **15**: 6049-6062.

Gryczan T.J., (1982) Molecular cloning in *Bacillus subtilis* In: The Molecular Biology of the Bacilli, volume 1 (Dubnau D, editor), Academic Press, New York, pp 307-329.

Gryczan T.J., Contente S., Dubnau D., (1978) Charachterrizarion of *Staphylococcus aureus* plasmids introduced by transformation into *Bacillus subtilis*. *J. Bacteriol.* **134**: 318-329.

Guan T., Ghosh A., Ghosh B.K., (1985). Immunoelectron microscopic double labelling of alkaline phosphatase and penicillinase with colloidal gold in frozen thin sections of *Bacillus licheniformis* 749/C. *J. Bacteriol.* **164**:107-113.

Guzman P., Westphelling J., Youngman P. (1988) Characterization of the promoter region of the *Bacillus subtilis* spo11E operon. *J. Bacteriol.* **170**: 1598-1609.

Hahn J, Dubnau D (1985) Analysis of plasmid deletional instability in *Bacillus subtilis*. *J. Bacteriol.* **1162**: 1014-1023.

Haldenwang W.G., (1995). The sigma factors of *Bacillus subtilis*. *Microbiol. Rev.* **59**: 1-30.

Hamamoto, T., and K. Horikoshi. (1987). Alkalophilic *Bacillus* xylanase A, a secretable protein through outer membrane of *Escherichia coli*. *Agric.Biol. Chem.* **51**:3133-3135.

Harold, T., A. Bolhuis, J.D.H. Jongbloed, S. Bron and J.M. van Dijl, (2000). Signal peptide-dependent protein transport in *Bacillus subtilis*: A genome-based survey of the secretome. *Microbiol. Mol. Biol. Rev.* **64**: 515-547.

Hartl, F.U., S. Lecker, E. Schiebel, J.P. Hendrick and W. Wickner, (1990). The binding cascade of SecB to SecA to SecY/E mediates preprotein targeting to the *Escherichia coli* plasma membrane. *Cell* **63**: 269-279.

Harwood CR, Williams DM, Lovett PS (1983) Nucleotide sequence of a *Bacillus pumilus* gene specifying chloramphenicol aceyltransferase. *Gene* **24**: 163-169.

Harwood, C. R., and S. M. Cutting. (1990). Molecular biological methods for *Bacillus*. John Wiley and Sons, Chichester, United Kingdom.

Harwood, C.R., (1992). *Bacillus subtilis* and its relatives: molecular biological and industrial workhorses. *Trends Biotech.* **10**: 247-256.

Harwood, C.R., M. Sarvas, S. Bron and J. M. van Dijl (2004). Post-translocational folding of secretory proteins in Gram-positive bacteria *Biochim. Biophysic. Acta* **1694**: 311-327.

Havarstein, L. S., D. B. Diep, and I. F. Nes. (1995). A family of bacteriocin ABC transporters carry out proteolytic processing of their substrates concomitant with export. *Mol. Microbiol.***16**:229-240.

Hazawa-Cho, Kanagawa-ku, Yokohama-shi and Kanagawa, (1999). Expression system for foreign genes using the fission yeast *Schizosaccharomyces pombe*. *Biotech. Appl. Biochem.* **30**: 235-244.

He, X.-S., Shyu, Y.-T., Nathoo, S., Wong, S.-L. and Doi, R.H., 1991. Construction and use of a *Bacillus subtilis* mutant deficient in multiple protease genes for the expression of eucaryotic genes. *Ann. NY Acad. Sci.* **646**: 69-77.

Hemilä H, Glode LM, Palva I. (1989). Production of diphtheria toxin CRM228 in *B. subtilis*. *FEMS Microbiol. Lett.* **53**:193-198.

Hemilä H, Sibakov M. (1991). Production of heterologous proteins in *Bacillus subtilis*: the effect of the joint between signal sequence and mature protein on yield. *Appl Microbiol. Biotechnol.* **36**:61-64.

Hemila, H. (1991). Sequence of a PAL-related lipoprotein from *Bacillus subtilis*. *FEMS Microbiol. Lett.* **82**: 37-42.

Henner DJ, Yang M, Ferrari E (1988). Localization of *Bacillus subtilis* sacU(Hy) mutations to two linked genes with similarities to the conserved procaryotic family of two-component signalling systems. *J Bacteriol.* **170**:5102-5109.

Hill DE, Aldape R, Rozzell JD. (1990). Nucleotide sequence of a cyclodextrin glucosyltransferase gene, cgtA, from *Bacillus licheniformis*. *Nucleic Acids Res.* **18**:199.

Himeno T, Imanaka T, Aiba S (1986) Protein secretion in *Bacillus subtilis* as influenced by the combination of signal sequence and the following mature portion. *FEMS Microbiol. Lett.* **35**: 17-21.

Hoch JA (Editors) Molecular cloning and gene regulation in *Bacilli*. Academic press, New York, NY, pp 159-169.

Holland IB, Mackman N, Nicaud JM (1986) Secretion of proteins from bacteria. *Biol/Technol* **4**: 427-431.

Horinouchi S, Weisblum B (1982) Nucleotide sequence and functional map of pE194, a plasmid that specifies inducible resistance to macrolide, lincosamide and streptogramin type B antibiotics. *J. Bacteriol.* **150**: 804-814.

Hussain M, Carlino A, Madonna MJ, Lampen JO. (1985). Cloning and sequencing of the metallothioprotein beta-lactamase II gene of *Bacillus cereus* 569/H in *Escherichia coli. J Bacteriol.* **164**:223-229.

Hussain, M., F. I. J. Pastor, and J. 0. Lampen. (1987). Cloning and sequencing of the blaZ gene encoding P-lactamase III, a lipoprotein of *Bacillus cereus* 569/H. *J. Bacteriol.* **169**:579-586.

Hynds, P.J., Robinson, D. and Robinson, C., (1998). The sec-independent twin-arginine translocation system can transport both tightly folded and malfolded proteins across the thylakoid membrane. *J. Biol. chem.* **273**: 34868-34874.

Imanaka T, Himeno T, Aiba S (1987). Cloning and nucleotide sequence of the penicillinase antirepressor gene *penJ* of *Bacillus licheninformis*. *J. Bacteriol.* **169**: 3867-3872.

Imanaka T, Tanaka T, Tsunekawa H, Aiba S (1981) Cloning of the genes for penicillinase, *penP* and *penI*, of *Bacillus licheniformis* in some vector plasmids and their expression in *Escherichia coli*, *Bacillus subtilis* and *Bacillus licheniformis*. *J. Bacteriol.* **147**: 776-786.

Ivanova N, Sorokin A, Anderson I, Galleron N, Candelon B, Kapatral V, Bhattacharyya A, Reznik G, Mikhailova N, Lapidus A, Chu L, Mazur M, Goltsman E, Larsen N, D'Souza M, Walunas T, Grechkin Y, Pusch G, Haselkorn R, Fonstein M, Ehrlich SD, Overbeek R, Kyrpides N. (2003). Genome sequence of *Bacillus cereus* and comparative analysis with *Bacillus anthracis*. *Nature* **6935**:87-91.

Jackson, P. J., Hill, K. K., Laker, M. T., Ticknor, L. O., and Keim, P. (1999). Genetic comparison of *B. anthracis* and its close relatives using AFLP and PCR analysis. *J. Appl. Microbiol.* **87**:263-269.

Jacobs, M., M. Eliasson, M. Uhlen, and J.-I. FlocL (1985). Cloning, sequencing and expression of subtilisin Carlsberg from *Bacillus licheniformis*. *Nucleic Acids Res.* **13**:8913-8926.

Jeong, S.M., H. Yoshikawa and H. Takahashi, (1993). Isolation and characterization of the *secE* homologue gene of *Bacillus subtilis*. *Mol. Microbiol.***10**: 133-142.

Joergensen, P. L., and C. K. Hansen. 1990. Multiple endo-,-1,4-glucanase-encoding genes from *Bacillus lautus* PL236 and characterization of the celB gene. *Gene* **93**:55-60.

Joliff, Gwennael.; Edelman, Alex.; Klier, Andre.; Rapoport, Georges. (1989). Inducible Secretion of a Cellulase from *Clostridium thermocellum* in *Bacillus subtilis*. *Appl Environ Microbiol.* **55**:2739-2744.

Jongbloed, J.D.H., Martin, U., Antelmann, H., Hecker, M., Tjalsma, H., Venema, G., Bron, S., van Dijl, J.M. and Müller, J., (2000). TatC is a specificity determinant for protein secretion via the twin-arginine translocation pathway. *J. Biol. chem.* **275**: 41350-41357.

Kaneko R, Koyama N, Tsai YC *et al* (1989). Molecular cloning of the structural gene for alkaline elastase YaB, a new subtilisin produced by an alkalophilic *Bacillus* strain. *J Bacteriol.* **171**:5232–5236.

Kato, C., Y. Nakano, and K. Horikoshi. (1989). The nucleotide sequence of the lipo-penicillinase gene of alkalophilic *Bacillus sp.* strain 170. *Arch. Microbiol.* **151**:91-94.

Kawazu T, Nakanishi Y, Uozumi N, Sasaki T, Yamagata H, Tsukagoshi N, Udaka S. (1987). Cloning and nucleotide sequence of the gene coding for enzymatically active fragments of the *Bacillus polymyxa* beta-amylase. *J Bacteriol.* **169**:1564-1570.

Kazuo K. and N. Ogasawara, (2002). Genome biology of a model bacterium, *Bacillus subtilis*. *International Congress Series* **1246**: 15-25.

Khan SA, Novick RP (1983) Complete nucleotide sequence of pT181, a tetracycline resistance plasmid from *Staphylococcus aureus*. *Plasmid* **10**: 25 -259.

Kimura K, Kataoka S, Ishii Y, Takano T, Yamane K. (1987). Nucleotide sequence of the beta-cyclodextrin glucanotransferase gene of alkalophilic *Bacillus sp.* strain 1011 and similarity of its amino acid sequence to those of alpha-amylases. *J Bacteriol.* **169**:4399-4402.

Kobayashi T, Zhu YF, Nicholas NJ, Lampen JO (1987). A second regulatory gene, *blar1*, encoding a potential penicillin – binding protein required for induction of beta – lactamase in *Bacillus licheniformis*. *J. Bacteriol.* **169**: 3873-3878.

Kobayashi, H., Y. Ohashi, H. Nanamiya, K. Asai and F. Kawamura, (2000). Genetic analysis of SecA-SecY interaction required for spore development in *Bacillus subtilis*. *FEMS Microbiol. lett.* **184**: 285-289.

Kontinen, V. P., and M. Sarvas. (1993). The PrsA lipoprotein is essential for protein secretion in *Bacillus subtilis* and sets a limit for high-level secretion. *Mol. Microbiol.* **8**:727-737.

Kunst, F., N. Ogasawara, I. Moszer, A.M. Albertini, G. Alloni, V. Azevedo, M.G. Bertero, P. Bessieres, A. Bolotin, S. Borchert, R. Borriss, L. Boursier, A. Brans, M. Braun, S.C. Brignell, S. Bron, S. Brouillet, C.V. Bruschi, B. Caldwell, V. Capuano, N.M. Carter, S.K. Choi, J.J. Codani, I.F. Connerton, A. Danchin *et al.*, (1997). The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* **390**: 249-256.

Kuroda M, Henner D, Yanofsky C (1988) *Cis* – acting sites in the transcript of the *Bacillus subtilis trp* operon regulates expression of the operon. *J. Bacteriol.* **170**: 3080-3088.

Lam, K.H., K.C. Chow and W.K. Wong, (1998). Construction of an efficient *Bacillus subtilis* system for extracellular production of heterologous protein. *J. Biotech.* **63**:167-177.

Lampen JO, Pastor FI, Hussain M (1986) in Microbiology (Scholesinger, editor), American Society for Microbiology, Washington DC, pp 279-282.

Laplace F, Egerer R, Gumpert J, Kraft R, Kostka S, Malke H (1989). Heterologous signal peptide processing in fusion interferon synthesis by engineered L-forms of *Proteus mirabilis*. *FEMS Microbiol Lett.* **50**:59-63.

Le Grice SFG, Beuch V, Mous J (1987) Expression of biologically active human T-cell lymphotrophic virus type III reverse transcriptase in *Bacillus subtilis*. *Gene* **55**: 95-103.

Lee G, Talkington C, Pero J (1980) Nucleotide sequence of a promoter recognized by *Bacillus subtilis* RNA polmerase. *Mol. Gen. Gene.* **180**: 57-65.

Lehtovaara P, Ulmanen I, Palva I (1984) *In vitro* transcription initiation and termination sites of an alpha – amylase gene from *Bacillus amyloliquefaciens* cloned in *Bacillus subtilis*. *Gene* **30**: 11-16.

Leskela, S., E. Wahlstrom, V.P. Kontinen and M. Sarvas, (1999). Lipid modification of prelipoproteins is dispensable for growth but essential for efficient protein secretion in *Bacillus subtilis*: characterization of the *lgt* gene. *Mol. Microbiol.* **31**: 1075-1085.

Li W., Zhou X., Lu P. (2004). Bottlenecks in the expression and secretion of heterologous proteins in *Bacillus subtilis*. *Res. Microbiol.* **8**:605-610.

Lim HM, Pène JJ, Shaw RW. (1988). Cloning, nucleotide sequence, and expression of the *Bacillus cereus* 5/B/6 beta-lactamase II structural gene. *J Bacteriol.* **170**:2873-2878.

Lory, S. (1994). Leader peptidase of type IV prepilins and related proteins, p. 17-29. *In* G. von Heijne (ed.), Signal peptidases. R. G. Landes Company, Austin, Tex.

Losick R, Pero J (1981). Cascades of Sigma factors. *Cell* **25**: 582-584.

MacKay RM, Lo A, Willick G, Zuker M, Baird S, Dove M, Moranelli F, Seligy V. (1986). Structure of a *Bacillus subtilis* endo-beta-1,4-glucanase gene. *Nucleic Acids Res.* **14**:9159-9170.

Manting E.H. and A.J.M. Driessen, (2000). *Escherichia coli* translocase: the unravelling of a molecular machine. *Mol. Microbiol.* **37**: 226-238.

Markus, H., M. Klein, E.H. Manting, A.J.M. Driessen and R. Freudl, (1999). Temporal expression of the *Bacillus subtilis secA* gene, encoding a central component of the pre-protein translocase. *J. Bacteriol.* **181**: 493-500.

McGlaughlin JR, Chang SY, Chang S (1982). Transcriptional analyses of the *Bacillus licheniformis penP* gene. *Nucleic Acids Res.* **10**:3905-3919.

Meijer, W.J.J., A. de Jong, G.B.A. Wisman, H. Tjalsma, G. Venema, S. Bron and J.M. van Dijl, (1995). The endogenous *Bacillus subtilis* (natto) plasmids pTA1015 and pTA1040 contain signal peptidase-encoding genes: identification of a new structural module on cryptic plasmids. *Mol. Microbiol.* **17**: 621-631.

Metz, R. J., L. N. Allen, T. M. Cao, and N. W. Zeman. (1988). Nucleotide sequence of an amylase gene from *Bacillus megaterium*. *Nucleic Acids Res.* 16:5203.

Meyer, T.H., J.F. Menétrét, R. Breitling, K.R. Miller, C.W. Akey and T.A. Rapoport, (1999). The bacterial SecY/E translocation complex forms channel-like structures similar to those of the eukaryotic Sec61p complex. *J. Mol. Biol.* **258**: 1789-1800.

Mezes P, Lampen O (1985) Secretion of proteins by *Bacilli* In: Dubnau D (editor) The molecular Biology of the Bacilli, Academic Press, New York, pp 151-183.

Mezes, P. S. F., R. W. Blacher, and J. 0. Lampen. (1985). Processing of *Bacillus cereus* 569/H (3-lactamase I in *Escherichia coli* and *Bacillus subtilis*. *J. Biol. Chem.* **260**:1218-1223.

Miller JR, Kovacevic S, Veal LE. (1987). Secretion and processing of staphylococcal nuclease by *Bacillus subtilis. J Bacteriol.* **169**:3508-3514.

Mogk, A., A. Völker, S. Engelmann, M. Hecker, W. Schumann and U. Völker, (1998). Nonnative proteins induce expression of the *Bacillus subtilis* CIRCE regulon. *J. Bacteriol.* **180**: 2895-2900.

Mongklsuk S, Duval EJ, and Lovett PS (1985) Transcription termination signal for the *cat – 86* indicator gene in a *Bacillus subtilis* promoter-cloning plasmid. *Gene* **37**: 83-90.

Mooney, R., I. Artsimovitch and R. Landick, (1998). Information processing by RNA polymerase: Recognition of regulatory signals during RNA chain elongation. *J. Bacteriol.* **180**: 3265-3275.

Moran CP Jr (1989). In Regulation of prokaryotic development (Smith I, Slepecky R, and Setlow P, editors). American Society for Microbiology, Washington DC, pp 167-184.

Moran CP Jr (1992). Transcriptional regulation of gene expression In: Biology of Bacilli: Applications to industry (Doi RH and McGoughlin M, editors) Butterworth-Heirneman, USA, pp 39-62.

Moran CP, Lang N, Le Grice SFJ, Lee G, Stephens M, Soneshein A, Pero J, Losick R (1982). Nucleotide sequences that signal the initiation of transcription and translation in *Bacillus subtillis*. *Mol. Gene Genet.* **186**: 339-346.

Mountain A, Mann NH, Munton RV, Baumberg S (1984) Cloning of a *Bacillus subtilis* restriction fragment complementing auxotrophic mutants of eight *Escherichia coli* genes of Arginine biosynthesis. *Mol. Gen. Genet.* **197**: 82-89.

Mountain, A., (1989). Gene expression system for *Bacillus subtilis*. In: Harwood, C.R. (Ed.), Biotechnology Handbooks, vol. 2: *Bacillus*. Plenum Press, New York, pp. 73-114.

Murphy N, McConnell DJ, Cantwell BA. (1984). The DNA sequence of the gene and genetic control sites for the excreted *B. subtilis* enzyme beta-glucanase. *Nucleic Acids Res.* **12**:5355-5367.

Nagarajan, V., (1993). Protein secretion. In: Sonenshein, A.L., Hoch, J.A. and Losick, R., Editors, 1993. *Bacillus subtilis* and other Gram-positive Bacteria, American Society for Microbiology, Washington, DC, pp. 713-726.

Nakajima R, Imanaka T, Aiba S. (1985). Nucleotide sequence of the *Bacillus stearothermophilus* alpha-amylase gene. *J Bacteriol.* **163**:401-406.

Nakamura K, Nakamura A, Takamatsu H, Yoshikawa H, Yamane K (1990). Cloning and characterization of a *Bacillus subtilis* gene homologous to *E. coli* secY. *J Biochem* **107**:603-607.

Nakamura K, Pirtle RM, Pirtle IL, Takeishi K, Inouye M. (1980). Messenger ribonucleic acid of the lipoprotein of the *Escherichia coli* outer membrane. II. The complete nucleotide sequence. *J Biol Chem.* **255**:210-216.

Neugbauer K, Sprengel R, Schaeller H (1981). Penicillinase from *Bacillus licheniformis*: nucleotide sequence of the gene and implications for the biosynthesis of a secretory protein in Gram-positive bacteria. *Nucleic Acids Res.* **9**: 2577-2588.

Nicaud JM, Markman M, Holland IB (1986). Current status of secretion of foreign proteins products by microorganisms. *J. Biotech.* **3**: 255-270.

Nicholson W, Park Y, Chambliss GH (1987) Catabolite repression-resistant mutants of the *Bacillus subtilis* alpha – amylase promoter affect transcription levels and are in an operator like sequence. *Journal of Molecular Biology* **198**: 609-618.

Nicholson WN, Chambliss GH (1985) Isolation and characterization of a *cis*-acting mutation conferring catabolite repression resistance to alpha – amylase synthesis in *Bacillus subtilis* 168. *J. Bacteriol.* **170**: 875-881.

Nielsen JB, Lampen JO. (1982). Glyceride-cysteine lipoproteins and secretion by Gram-positive bacteria. *J Bacteriol.* **152**:315-322.

Nielsen, H., J. Engelbrecht, S. Brunak and G. von Heijne, (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering* **10**:1-6.

Nielsen, J. B. K., and J. 0. Lampen. (1982). Membrane-bound penicillinases in gram-positive bacteria. *J. Biol. Chem.* **257**:4490-4495.

Nishiya Y, Imanaka T. (1990). Cloning and nucleotide sequences of the *Bacillus stearothermophilus* neutral protease gene and its transcriptional activator gene. *J Bacteriol.* **172**:4861-4869.

Nishiya, Y., and T. Imanaka. (1990). Cloning and nucleotide sequences of the *Bacillus stearothermophilus* neutral protease gene and its transcriptional activator gene. *J. Bacteriol.* **172**:4861-4869.

Novick RP, Adler GK, Majumder S, Khan SA, Carleton S, Rosenblum WD, Iordanescu S (1982) Coding sequence for the pT181 *repC* product: A plasmid coded protein uniquely required for replication. *Procl. Natl. Acad. Sci. USA* **79**: 4108-4112.

Nyberg, P.M.A. (1984). Production of α-amylase by *Bacillus subtilis* containing a cloned α-amylase gene- experiences in the laboratory and semi industrial scale. *European Congreg. Biotechnol.* **3**: 409-413.

Oliver D.B., J. Beckwith, (1982). Identification of a new gene (*secA*) and gene product involved in the secretion of envelope proteins in *Escherichia coli*. *J. Bacteriol.* **150**: 686–691.

Olivier D (1985). Protein secretion in *Escherichia coli*. *Ann. Rev. Microbiol.* **39**: 615-648.

Olmos-Soto J., R. Contreras-Flores, (2003). Genetic system constructed to overproduce and secrete proinsulin in *Bacillus subtilis*. *Appl. Microbiol. Biotechnol.* **62**: 369-373.

Ortlepp SA, Ollington JF, McConell DJ (1983) Molecular cloning in *Bacillus subtilis* of a *Bacillus licheniformis* gene encoding a thermostable alpha amylase. *Gene* **23**: 267–276.

Paddon CJ, Vasantha N, Hartley RW. (1989). Translation and processing of *Bacillus amyloliquefaciens* extracellular RNase. *J Bacteriol.* **171**:1185-1187.

Paetzel, M., A. Karla, N.C. Strynadka and R.E. Dalbey, (2002). Crystal structure of a bacterial signal peptidase apoenzyme: implications for signal peptide binding and the Ser-Lys dyad mechanism. *J. Biol. chem.* **277**: 9512-9519.

Palva I (1989) in Society for General Microbiology Symposium 44 (Bamberg S, Hunter I, Rhodes M, editors), Cambridge University Press, Cambridge, pp 255-269.

Palva I, Petterson RF, Kalkinen N, Lehtovaara P, Sarvas M (1981). Nucleotide sequence of the promoter and NH2-terminal signal peptide region of the alpha-amylase gene from *Bacillus amyloliquefaciens. Gene* **15**:43-51.

Palva, I. (1982). Molecular cloning of α-amylase gene from *Bacillus amyloliquefaciens* and its expression in *Bacillus subtilis*. *Gene* **19**: 81–87.

Park SS, Wong SL, Wang LF, Doi RH (1989). *Bacillus subtilis* subtilisin gene (aprE) is expressed from a sigma A (sigma 43) promoter in vitro and in vivo. *J Bacteriol.* **171**:2657-2665.

Perego M, Hoch JA (1988). Sequence analysis and regulation of the hpr locus, a regulatory gene for protease production and sporulation in *Bacillus subtilis*. *J Bacteriol.* **170**: 2560-2567.

Perkins JB, Youngman P (1983) *Streptococcus* plasmid pAM∝ is a composite of two separable replicons, one of which is closely related to *Bacillus* plasmid pBC16. *J. Bacteriol.* **155**: 607-615.

Perlman D, Halvorson HO (1983). A putative signal peptidase recognition site and sequence in eukaryotic and prokaryotic signal peptides. *J Mol Biol.* **167**: 391-409.

Peschke U, Beuch V, Biyard H, Gentz R, Le Grice S (1985) Efficient utilization of *Escherichia coli* transcriptional signals in *Bacillus subtilis*. *Journal of Molecular Biology* **186**: 547-555.

Pohlschröder, M., W.A. Prinz, E. Hartmann and J. Beckwith, (1997). Protein translocation in the three domains of life: variation on a theme. *Cell* **91**: 563-566.

Prágai, Z., H. Tjalsma, A. Bolhuis, J.M. van Dijl, G. Venema and S. Bron, (1997). The signal peptidase II (*lsp*) gene of *Bacillus subtilis*. *Microbiol.* **143**:1327-1333.

Priest F.G. (1989). Products from Bacilli. In: C.F. Harwood, Editor, *Handbooks of Biotechnology*, Plenum Press, New York pp. 293-315.

Priest FG (1977) Extracellular enzyme synthesis in the genus *Bacillus*. *Bacteriology Reviewed* **41**: 711-753.

Priest FG (1989) in *Bacillus* (Harwood CR, editor) Plenum Press, New York, pp 293–320.
Pulido D, Jimenez A, Salas M, Mellado PR (1987) A *Bacillus subtilis* phage theta29 transcription terminator is efficiently recognized in *Streptomyces lividans*. *Genes* **56**: 221-227.

Quax W.J., Lidia W., H. Westers (2004). *Bacillus subtilis* as cell factory for pharmaceutical proteins: a biotechnological approach to optimize the host organism. *Biochim Biophys Acta* **1694**:299–310.

Reid S.J., Sugrue, J.A., and Thompson J.A. (1986). Industrial applications of a cloned neutral protease gene in *Bacillus subtilis*. *Appl. Microbiol. Biotechnol.* **24**: 311-318.

Rey MW, Ramaiya P, Nelson BA, Brody-Karpin SD, Zaretsky EJ, Tang M, Lopez de Leon A, Xiang H, Gusti V, Clausen IG, Olsen PB, Rasmussen MD, Andersen JT, Jorgensen PL, Larsen TS, Sorokin A, Bolotin A, Lapidus A, Galleron N, Ehrlich SD, Berka RM (2004). Complete genome sequence of the industrial bacterium *Bacillus licheniformis* and comparisons with closely related *Bacillus* species. *Genome Biol.* **10**:R77.

Robinson, C., Bolhuis, A., (2001). Protein targeting by the twin-arginine translocation pathway**.**. *Mol. Cell Biol.* **2**: 350–356.

Robson LM, Chambliss GH (1987). Endo-beta-1,4-glucanase gene of *Bacillus subtilis* DLG. *J Bacteriol.* **169**: 2017-2025.

Saier MH, Werner PK, Muller M (1989). Insertion of proteins into bacterial membranes: mechanism, characteristics, and comparisons with the eucaryotic process. *Microbiol Rev.* **53**: 333-366.

Sambrook, J., Fritsch, E.F., Maniatus, T., (1989). Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sa-Nogueira I, de Lancastre H (1989). Cloning and characterization of araA, araB, and araD, the structural genes for L-arabinose utilization in *Bacillus subtilis*. *J Bacteriol.* **171**:4088–4091.

Sa-Nogueira I, Paveia H, de Lancastre H (1988). Isolation of constitutive mutants for L-arabinose utilization in *Bacillus subtilis*. *J Bacteriol.* **170**: 2855–2857.

Saris P, Taira S, Airaksinen U, Palva A, Sarvas M, Palva I, Runeberg-Nyman K. (1990). Production and secretion of pertussis toxin subunits in *Bacillus subtilis*. *FEMS Microbiol Lett.* **56**:143-148.

Sarvas M (1986) Protein secretion in *Bacilli*. *Curr. Topics Microbiol. Immunol*. **125**: 103-125.

Sarvas, M., (1995). Gene expression in recombinant *Bacillus*. *Bioprocess Technology* **22**: 53-120.

Sasamoto H, Nakazawa K, Tsutsumi K, Takese K, Yamane K (1989). Signal peptide of *Bacillus subtilis* alpha-amylase. *J. Biochem.* **106**: 376 - 382.

Saunders CW, Schmidt BJ, Mirot MS, Thompson LD, Guyer MS (1984) Use of chromosomal integration in the establishment and expression of *blaZ*, a *Staphylococcus aureus* beta-lactamase gene from *Bacillus subtilis*. *J. Bacteriol.* **157**: 718-726.

Saunders CW, Schmidt RJ, Mallonee RL, Guyer MS (1987) Secretion of human serum albumin from *Bacillus subtilis*. *J. Bacteriol.* **169**: 2917 – 2925.

Schallmey M., Singh A., Ward O.P. (2004). Developments in the use of *Bacillus* species for industrial production. *Canadian J. Microbiol.* **1**:1-17.

Schatz G. (1991). A protein translocation machine in yeast. *Curr Biol.* **1**:43–44.

Scheer-Abramowitz J, Gryczan TJ, Dubnau D (1981) Origin and mode of replication of plasmids pE194 and pUB110. *Plasmids* **6**: 67-77.

Schmidt, F.R. (2004) Recombinant expression systems in the pharmaceutical industry. *Appl. Microbiol. Biotechnol.* **4**:363-372.

Schörgendorfer K, Schwab H, Lafferty RM. (1987). Nucleotide sequence of a cloned 2.5 kb *Pst*I--*Eco*RI *Bacillus subtilis* DNA fragment coding for levanase. *Nucleic Acids Res.* **15**:9606.

Schumann W., Ferreira L.C.S. (2004). Production of recombinant protein in *Escherichia coli*. *Genetics Mol. Biol.* **3**:442-453.

Senesi, S., Celandroni, F., Tavanti, A., and Ghelardi, E. (2001). Molecular characterization and identification of *Bacillus clausii* strains marketed for use in oral bacteriotherapy. *Appl. Environ. Microbiol.* **67**:834-839.

Setlow P (1973). Inability of detect cyclic AMP in vegetative or sporulating cells or dormant spores of *Bacillus megaterium*. *Biochem Biophys Res Commun.* **52**:365–372.

Sharp, P.M. (1994). Identification of genes encoding ribosomal protein L33 from *Bacillus licheniformis*, *Thermus thermophilus* and *Thermotoga maritima*. *Gene* **139**: 135–136.

Shields DS, Sharp PM (1987) Synonymous codon usage in *Bacillus subtilis* reflects both translational selection and mutational biases. *Nucleic Acids Res.* **15**: 8023 – 8040.

Shimada, H., M. HoDio, I. Mita, A. Nakayama, A. Akaoka, K. Manabe, and Y. Furutani. (1985). The nucleotide sequence and some properties of the neutral protease gene of *Bacillus amyloliquefaciens*. *J. Biotechnol.* **2**:75-85.

Shimotsu H, Henner DJ (1984). Characterization of the *Bacillus subtilis* tryptophan promoter region. *Proc Natl Acad Sci U S A.* 1984 **81**:6315-6319.

Shimotsu H, Kuroda MI, Yanofsky C, Henner DH (1986) Novel form of transcription attenuation regulates expression of the *Bacillus subtilis* tryptophan operon. *J. Bacteriol.* **166**: 461-471.

Shiroza T, Nakazawa K, Tashiro M et al (1985). Synthesis and secretion of biologically active mouse interferon-B using a *Bacillus subtilis* alpha–amylase secretion vector. *Gene* **34**: 1-8.

Sibakov M (1986) High expression of *Bacillus licheniformis* alpha–amylase with a *Bacillus* secretion vector. *J. Biochem.* **155**: 577 – 581.

Sibakov M, Palva I. (1984). Isolation and the 5'-end nucleotide sequence of *Bacillus licheniformis* alpha-amylase gene. *Eur J. Biochem.* **145**:567-572.

Simonen, M. and Palva, I. (1993). Protein secretion in *Bacillus* species. *Microbiol. Rev.* **57**:109–137.

Slepecky RA (1992). What is *Bacillus*? In: The biology of bacilli: Application to industry (Doi RH and MacGloughlin M, editors) Butterworth-Heineman, USA, pp 1–21.

Sloma A, Ally A, Ally D, Pero J. (1988). Gene encoding a minor extracellular protease in *Bacillus subtilis*. *J Bacteriol.* **170**:5557-5563.

Sloma A, Gross M (1983) Molecular cloning and nucleotide sequence of the type I beta – lactamase gene from *Bacillus cereus*. *Nucleic Acids Res.* **11**: 4997 – 5004.

Sloma A, Rudolph CF, Rufo GA Jr, Sullivan BJ, Theriault KA, Ally D, Pero J. (1990). Gene encoding a novel extracellular metalloprotease in Bacillus subtilis. *J Bacteriol.* **172**:1024-1029.

Sloma, A., G. A. Rufo, Jr., C. F. Rudolph, B. J. Sullivan, K. A. Theriault, and J. Pero. (1990). Bacillopeptidase F of *Bacillus subtilis*: purification of the protein and cloning of the gene. *J. Bacteriol.* **172**:1470-1477.

Sloma, A., Rufo, G.A. Jr., Rudolph, C.F., Sullivan, B.J., Theriault, K.A., Pero, J., (1990). Cloning and deletion of the genes for three minor extracellular proteases of *Bacillus subtilis*. In: Zukowski, M.M., Ganesan, A.T., Hoch, J.A. (Eds.), Genetics and Biotechnology of *Bacilli*, vol. 3. Academic Press, New York, pp. 295–302.

Sloma A., Hannett; N. M., Stephens; M. A., Rudolph C. F., Rufo, Jr., Gerald A., Pero; J. (1992). Expression of heterologous DNA using the *Bacillus coagulans* amylase gene. United States Patent No. 5,171,673

Smith MCM, Mountain A, Baumberg S (1986) Sequence analysis of the *Bacillus subtilis argC* promoter region. *Gene* **49**: 53-60.

Smith, N. R., Gordon, R. E., and Clark, F. E. (1952). Aerobic Spore Forming Bacteria. U.S. Dep. Agric. Monogr. 16. Washington, DC.

Spizizen J., Karunairatnam M.C., Gest H (1958). Preparation and properties of protoplasts of *Rhodospirillum rubrum. Biochim. Biophys. Acta.* **29**:649-50.

Stahl ML, Ferrari E (1984) Replacement of *Bacillus subtilis* subtilisin in structural gene with an in vitro derived deletion mutation. *J. Bacteriol.* **158**: 411-418.

Steinmetz M, Le Coq D, Aymerich S, Gonzy-Tréboul G, Gay P. (1985). The DNA sequence of the gene for the secreted *Bacillus subtilis* enzyme levansucrase and its genetic control sites. *Mol Gen Genet* **200**:220-228.

Stephens M.A., Ortlepp S.A., Ollington J.F., McConnell D.J. (1984). Nucleotide sequence of the 5' region of the *Bacillus licheniformis* alpha-amylase gene: comparison with the *B. amyloliquefaciens* gene. *J Bacteriol.* **158**:369-372.

Stephenson K. and C.R. Harwood, (1998). Influence of a cell-wall-associated protease on production of α-amylase by *Bacillus subtilis*. *Applied and Environmental Microbiology* **64**: 2875-2881.

Stragier P. and Losick R (1996). Molecular genetics of sporulation in *Bacillus subtilis. Annu. Rev. Genet.* **30**: 297-341.

Stragier P., Losick P. (1990). Cascades of sigma factors revisited. *Mol Microbiol.* **4**:1801-1806.

Sugimoto K, Sugisaki H, Okamoto T, Takanami M (1977). Studies on bacteriophage fd DNA. IV. The sequence of messenger RNA for the major coat protein gene. *J. Mol. Biol.* **111**:487-507.

Takagi M, Imanaka T, Aiba S (1985). Nucleotide sequence and promoter region for the neutral protease gene from *Bacillus stearothermophilus*. *J Bacteriol.* **163**:824-831.

Takahara, M., H. Sagai, S. Inouye, and M. Inouye. (1988). Secretion of human superoxide dismutase in *Escherichia coli*. *Bio/Technology* **6**:195-198.

Takano T, Fukuda M, Monma M, Kobayashi S, Kainuma K, Yamane K. (1986). Molecular cloning, DNA nucleotide sequencing, and expression in *Bacillus subtilis* cells of the *Bacillus macerans* cyclodextrin glucanotransferase gene. *J Bacteriol.* **166**:1118-1122.

Takase, K., H. Mizuno, and K. Yamane. (1988). NH2-terminal processing of *Bacillus subtilis* a-amylase. *J. Biol. Chem.* **263**:11548-11553.

Takidnen, K., R. F. Pettersson, N. Kalkkinen, I. Palva, H. Soderlund, and L. Kilariainen. (1983). Amino acid sequence of a-amylase from *Bacillus amyloliquefaciens* deduced from the nucleotide sequence of the cloned gene. *J. Biol. Chem.* **258**:1007-1013.

Tanaka T, Kawata M, Nagami Y, Uchiyama H (1987). *prtR* enhances of the mRNA level of the *Bacillus subtilis* extracellular proteases. *J. Bacteriol.* **169**: 3044 – 3050.

Tezuka, H., T. Yuuki, and S. Yabuuchi. (1989). Construction of a ,3-glucanase hyperproducing *Bacillus subtilis* using the cloned ,B-glucanase gene and a multi-copy plasmid. *Agric. Biol.Chem.* **53**:2335-2339.

Tjalsma, H., Bolhuis, A., Jongbloed, J.D.H., Bron, S. and van Dijl, J.M., (2000). Signal peptide-dependent protein transport in *Bacillus subtilis*: a Genome-based survey of the secretome. *Microbiol.Mol. Biol. Rev.* **64**: 515–547.

Tjalsma, H., Bolhuis, A., van Roosmalen, M.L., Wiegert, T., Schumann, W., Broekhuizen, C.P., Quax, W.J., Venema, G., Bron, S. and van Dijl, J.M., (1998). Functional analysis of the secretory precursor processing machinery of *Bacillus subtilis*: identification of a eubacterial homolog of archaeal and eukaryotic signal peptidases. *Genes Develop.* **12**: 2318–2331.

Tjalsma, H., J. van den Dolder, W.J.J. Meijer, G. Venema, S. Bron and J.M. van Dijl, (1999). The plasmid-encoded signal peptidase SipP can functionally replace the major signal peptidases SipS and SipT of *Bacillus subtilis*. *J. Bacteriol.* **181**: 2448–2454.

Tjalsma, H., M.A. Noback, S. Bron, G. Venema, K. Yamane and J.M. van Dijl, (1997). *Bacillus subtilis* contains four closely related type I signal peptidases with overlapping substrate specificities. Constitutive and temporally controlled expression of different *sip* genes. *J. Biol. chem.* **272**: 25983–25992.

Tokunaga, M., J.M. Loranger, P.B. Wolfe and H.C. Wu, (1982). Prolipoprotein signal peptidase in *Escherichia coli* is distinct from the M13 procoat protein signal peptidase. *J. Biol. chem.* **257**: 9922–9925.

Tsuboi, A., R. Uchihi, T. Adachi, T. Sasaki, S. Hayakawa, H. Yamagata, N. Tsukagoshi, and S. Udaka. (1988). Characterization of the genes for the hexagonally arranged surface layer proteins in protein-producing *Bacillus brevis* 47: completenucleotide sequence of the middle wall protein gene. *J. Bacteriol.* **170**:935-945.

Tsukamoto, A., K. Kimura, Y. Ishii, T. Takano, and K. Yamane. (1988). Nucleotide sequence of the maltohexaoseproducing amylase gene from an alkalophilic *Bacillus sp.* #707 and structural similarity to liquefying type a-amylases. *Biochem. Biophys. Res. Commun.* **151**:25-31.

Udaka S (1970). Isolation of the arginine repressor in *Escherichia coli*. *Nature.* **228**:336-338.

Udaka S, Tsukagoshi N, Yamagata H (1989). *Bacillus brevis*, a host bacterium for efficient extracellular production of useful proteins. *Biotechnol. Genet. Eng. Rev*. **7**:113-46.

Udaka S. and H. Yamagata, (1993). Protein secretion in *Bacillus brevis*. *Antonie Van Leeuwenhoek* **64**:137–143.

Uozumi N, Sakurai K, Sasaki T, Takekawa S, Yamagata H, Tsukagoshi N, Udaka S. (1989). A single gene directs synthesis of a precursor protein with beta- and alpha-amylase activities in *Bacillus polymyxa*. *J Bacteriol.* **171**:375-82.

Uptain, S., C. Kane and M. Chamberlin, (1997). Basic mechanisms of transcript elongation and its regulation. *Ann. Rev. Biochem.* **66**: 117–172.

Van Dijl JM (1990). Ph. D. Thesis, Rijksuniversiteit Groningren.

van Dijl, J.M., A. de Jong, G. Venema and S. Bron, (1995). Identification of the potential active site of the signal peptidase SipS of *Bacillus subtilis*: Structural and functional similarities with LexA-like proteases. *J. Biol. chem.* **270***:* 3611–3618.

van Dijl, J.M., A. de Jong, H. Smith, S. Bron and G. Venema, (1991). Signal peptidase I overproduction results in increased efficiencies of export and maturation of hybrid secretory proteins in *E. coli*. *Mol. Gen. Genet.* **227**: 40–48.

van Dijl, J.M., A. de Jong, J. Vehmaanperä, G. Venema and S. Bron, (1992). Signal peptidase I of *Bacillus subtilis*: patterns of conserved amino acids in prokaryotic and eukaryotic type I signal peptidases. *EMBO Journal* **11**: 2819–2828.

van Dijl, J.M., Bolhuis, A., Tjalsma, H., Jongbloed, J.D.H., de Jong, A. and Bron, S., (2001). Protein transport pathways in *Bacillus subtilis*: a genome-based road map. In: Sonenshein, A.L., Hoch, J.A. and Losick, R., Editors, 2001. *Bacillus subtilis* and its Closest Relatives: from Genes to Cells, American Society for Microbiology, Washington, DC, pp. 337–355.

van Dijl, J.M., P.G. Braun, C. Robinson, W. J. Quax, H. Antelmann, M. Hecker, J.Müller, H. Tjalsma, S. Bron and J.D.H. Jongbloed (2002). Functional genomic analysis of the *Bacillus subtilis* Tat pathway for protein secretion. *J. Biotechnol.* **98**: 243-254.

van Kaer L, Van Montagu M, Dhaese P (1987). Transcriptional control in the *Eco*r1–F immunity region of the Bacillus subtilis phage theta 105. *J. Mol. Biol.* 197: 55 – 67.

van Leen RW, Bakhuis JG, van Beckhoven RF, Burger H, Dorssers LC, Hommes RW, Lemson PJ, Noordam B, Persoon NL, Wagemaker G. (1991). Production of human interleukin-3 using industrial microorganisms. *Biotechnol. (N Y).* **9**:47-52.

van Roosmalen, M.L., J.D. Jongbloed, J.Y. Dubois, G. Venema, S. Bron and J.M. van Dijl, (2001). Distinction between major and minor *Bacillus* signal peptidases based on phylogenetic and structural criteria. *J. Biol. chem.* **276**: 25230–25235.

van Wely K. H. M., J. Swaving, R. Freudl and A.J. M. Driessen (2001). Translocation of proteins across the cell envelope of Gram-positive bacteria. *FEMS Microbiology Reviews* **25**: 437-454.

van Wely, K.H.M., J. Swaving, C.P. Broekhuizen, M. Rose, W.J. Quax and A.J.M. Driessen, (1999). Functional identification of the product of the *Bacillus subtilis yvaL* gene as a SecG homologue. *J. Bacteriol.* **181**:1786–1792.

Vasantha N, Thompson LD, Rhodes C, Banner C, Nagle J, Filipula D (1984) Genes for alkaline protease and neutral protease from *Bacillus amyloliquefaciens* contain a large open reading frame (orf) between the regions coding for signal sequence and mature protein. *J. Bacteriol.* **159**: 811 – 819.

Vasantha N, Thompson LD. (1986). Secretion of a heterologous protein from *Bacillus subtilis* with the aid of protease signal sequences. *J Bacteriol.* **165**:837-842.

Von Heijne G (1983). Patterns of amino acids near signal-sequence cleavage sites. *Eur. J. Biochem.* **133**:17-21.

Von Heijne G (1985). Signal sequences. The limits of variation. *J. Mol. Biol.* **184**:99-105.

Von Heijne G (1986). A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res.* **14**:4683-4690.

von Heijne G. and L. Abrahmsen, (1989). Species-specific variation in signal peptide design. Implications for protein secretion in foreign hosts. *FEBS Lett.* **244**: 439–446.

Wang L.F, Wong S.L, Lee S.G, Kalyan N.K, Hung P.P, Hilliker S, Doi R.H. (1988). Expression and secretion of human atrial natriuretic alpha-factor in *Bacillus subtilis* using the subtilisin signal peptide. *Gene.* **69**:39-47.

Wang L.F., Doi R.H. (1987). Developmental expression of three proteins from the first gene of the RNA polymerase sigma 43 operon of *Bacillus subtilis*. *J Bacteriol.* **169**:4190-4195.

Wang L.F., Doi R.H. (1990). Complex character of senS, a novel gene regulating expression of extracellular-protein genes of *Bacillus subtilis*. *J Bacteriol.* **172**:1939-1947.

Wang P.Z. and R.H. Doi, (1984). Overlapping promoters transcribed by *Bacillus subtilis* σ^{55} and σ^{37} RNA polymerase holoenzymes during growth and stationary phases. *J. Biol. chem.* **259**: 8619–8625.

Wang W, Mézes PS, Yang YQ, Blacher RW, Lampen JO. (1985). Cloning and sequencing of the beta-lactamase I gene of *Bacillus cereus* 5/B and its expression in *Bacillus subtilis. J Bacteriol.* **163**:487-492.

Wang, L.-F, Doi, R.H., (1992). Heterologous gene expression in *Bacillus subtilis*. In: Doi, R.H., McGloughlin, M. (Eds.), Biology of Bacilli: Applications to Industry. Stoneham, MA: Butterworth–Heinemann, pp. 63-104.

Wang, L.-F. and Doi, R.H., (1987). Promoter switching during development and the termination site of the sigma 43 operon of *Bacillus subtilis*. *Mol. Gen. Genet.* **207**: 114–119.

Wang, L.F., R. Bruckner and R.H. Doi, (1989). Construction of a *Bacillus subtilis* mutant deficient in three extracellular proteases. *J. Gen.l Appl. Microbiol.* **35**: 487–492.

Wargner B, Wargner M, Wollweber L, Behnke D (1989). Immunoelectron microscopy of *Bacillus subtilis* cells secreting human interferon alpha 1 or staphylokinase. *FEMS Microbiol. Lett.* **53**:327-332.

Watanabe T, Suzuki K, Oyanagi W, Ohnishi K, Tanaka H. (1990). Gene cloning of chitinase A1 from *Bacillus circulans* WL-12 revealed its evolutionary relationship to *Serratia chitinase* and to the type III homology units of fibronectin. *J. Biol. Chem.* **265**:15659-15665.

Weisblum, B., Graham, M.Y., Gryczan, T. and Dubnau, D., (1979). Plasmid copy number control: isolation and characterization of high-copy-number mutants of plasmid pE194. *J. Bacteriol.* **137**: 635-643.

Wells JA, Ferrari E, Henner DJ, Estell DA, Chen EY (1983) Cloning, sequencing and secretion of *Bacillus amyloliquefaciens* subtilisin in *Bacillus subtilis*. *Nucleic Acids Research* **11**: 7911-7925.

Wiggs JL, Bush JW, Chamberlin MJ (1979) Utilization of promoter and termination sites on bacteriophage and DNA by RNA polymerase from a variety of bacterial orders. *Cell* **16**: 97-109.

Williams D, Schoner. RG, Duvall EJ, Preis LH, Lovett PL (1981) Expression of *Escherichia coli trp* genes and the mouse dihydrofolate reductase gene cloned in *Bacillus subtilis*. *Gene* **16**: 199 -206.

Wong HC, Chang S (1986). Identification of a positive retroregulator that stabilizes mRNA's in bacteria. *Proc. Natrl. Acad. Sci. U.S.A*. **83**: 3233-3237.

Wong S (1989). Development of an inducible and enhancible expression and secretion system in *Bacillus subtilis*. *Gene* **83**: 215-223.

Wong S.L. and R.H. Doi, (1986). Determination of the signal peptidase cleavage site in the preprosubtilisin of *Bacillus subtilis*. *J. Biol. chem.* **26**:176-181.

Wong S.L., R. Ye and S. Nathoo, (1994). Engineering and production of streptokinase in a *Bacillus subtilis* expression-secretion system. *Appl. Environ. Microbiol.* **60**: 517–523.

Wong SL (1995) Advances in the use of *Bacillus subtilis* for the expression and secretion of heterologous proteins. *Curr. Opin. Biotech.* **6**: 517 – 522.

Wong SL, Doi RH (1986). Determination of the signal peptidase cleavage site in the preprosubtilisin of *Bacillus subtilis*. *J. Biol. Chem.* **261**:10176-10181.

Wong SL, Kawamura F, Doi RH (1986) Use of the *Bacillus subtilis* subtilisin signal peptide for efficient secretion of TEM beta–lactamase during growth. *J. Bacteriol.* **168**: 1005-1009.

Wong, S.-L., and R. H. Doi. (1986). Determination of the signal peptidase cleavage site in the preprosubtilisin of *Bacillus subtilis*. *J. Biol. Chem.* **261**:10176-10181.

Wong, S.-L., Wang, L.-F. and Doi, R.H., (1988). Cloning and nucleotide sequence of *senN*, a novel *Bacillus natto* gene that regulates expression of extracellular protein genes. *J. Gen. Microbiol.* **134**: 3269–3276.

Wong, S.-L., Wu, X.-C., Yang, L.-P., Ng, S.-C., Hudson, N., (1995). Production and purification of antibody using *Bacillus subtilis* as an expression host. In: Wang, H.Y., Imanaka, T. (Eds.), Antibody Expression and Engineering. Washington, DC: American Chemical Society, pp. 100–107.

Wong S.L. and Wu W.C., (1999). Development of improved pUB110-based vectors for expression and secretion studies in *Bacillus subtilis*. *J. Biotech* **72**:185-95.

Wu X.C., Ng S.C., Near R.I., Wong S.L. (1993). Efficient production of a functional single – chain antidigoxin antibody fragment via an engineered *Bacillus subtilis* expression – secretion system. *Bio/Technol.* **11**: 71 – 76.

Wu, J. J., R. Schuch, and P. J. Piggot. (1992). Characterization of a *Bacillus subtilis* sporulation operon that includes genes for an RNA polymerase sigma factor and for a putative DD-carboxypeptidase. *J. Bacteriol.* **174**:4885-4892.

Wu, L.-F., Ize, B., Chanal, A., Quentin, Y. and Fichant, G., (2000). Bacterial twin-arginine signal peptide-dependent protein translocation pathway: evolution and mechanism. *Mol. Microbiol. Biotechnol.* **2**: 170–189.

Wu, S.C., J.C. Yeung, Y. Duan, R. Ye, S.J. Szarka, H.R. Habibi and S.L. Wong, (2002). Functional production and characterization of a fibrin-specific single- chain antibody fragment from *Bacillus subtilis*: effects of molecular chaperones and a wall-bound protease on antibody fragment production. *Appl. Environ. Microbiol.* **68**: 3261–3269.

Wu, S.C., R. Ye, X.C. Wu, S.C. Ng and S.L. Wong, (1998). Enhanced secretory production of a single-chain antibody fragment from *Bacillus subtilis* by coproduction of molecular chaperones. *J. Bacteriol.* **180**: 2830–2835.

Wu, S.C., W. Lee, L. Tran and S.L. Wong, (1991). Engineering a *Bacillus subtilis* expression-secretion system with a strain deficient in six extracellular proteases. *J. Bacteriol.* **173**:4952–4958.

Yagi Y, McLellan TS, Frez WA, Clewell DB (1978) Characterization of a small plasmid determining resistance to erythromycin, ljncomycin and vernamycin B in a strain of streptococcus sanguis isolated form dental plaque. *Antimicrobial agents Chemotherapy* **13**: 884 – 890.

Yamada A, Tsukagoshi N, Udaka S, Sasaki T, Makino S, Nakamura S, Little C, Tomita M, Ikezawa H. (1988). Nucleotide sequence and expression in *Escherichia coli* of the gene coding for sphingomyelinase of *Bacillus cereus*. *Eur. J. Biochem.* **175**:213-220.

Yamagata H, Nakahama K, Suzuki Y, Kakinuma A, Tsukagoshi N, Udaka S. (1989). Use of *Bacillus brevis* for efficient synthesis and secretion of human epidermal growth factor. *Proc Natl Acad Sci U S A*. **86**:3589–3593.

Yamasaki H, Ohmura K, Nakayama A, Takichi Y, Otokai K, Yamasaki M *et al* (1983) alpha –amylase genes (amy R2 and amy E+) from alpha – amylase – hyperproducing *Bacillus subtilis* strains: Molecular cloning and nucleotide sequences. *J. Bacteriol.* **126**: 327 – 337.

Yang M, Galezzi A, Henner D (1983). Nucleotide sequences of the amylase gene from *Bacillus subtilis. Nucleic Acids Research* **11**: 237 – 249.

Yang M.Y, Shimotsu H, Ferrari E, Henner D.J (1987). Characterization and mapping of the *Bacillus subtilis* prtR gene. *J Bacteriol.* **169**:434-7.

Yang MY, Ferrari E, Chen E, Henner DJ (1986) Identification of the pleiotropic *sacQ* gene of a *Bacillus subtilis*. *J. Bacteriol.* **166**: 113-119.

Yang MY, Ferrari E, Henner DJ. (1984). Cloning of the neutral protease gene of *Bacillus subtilis* and the use of the cloned gene to create an in vitro-derived deletion mutation. *J Bacteriol.* **160**:15-21.

Yang S, Huang Y.H, Huang X.D, Li S.Y, Yuan Z.Y.High Expression of Penicillin G Acylase Gene from *Bacillus megaterium* in *Bacillus subtilis*. Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao 31:601-603.

Yansura DG, Henner DJ (1984). Use of the *Escherichia coli lacZ* operator and repressor to control gene expression in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA* **81**: 439-443.

Yasuhiko M, Atsuo N, Atsushi O, Mami Y, Yasutaro F (2000) Evaluation and characterization of catabolite – responsive elements (*cre*) of *Bacillus subtilis*. *Nucleic Acids Res.* **28**: 1206 -1210.

Ye, R., J.H. Kim, B.G. Kim, S. Szarka, E. Sihota and S.L. Wong, (1999). High-level secretory production of intact, biologically active staphylokinase from *Bacillus subtilis*. *Biotechnol. Bioengin.* **62**: 87–96.

Yoshimura K, Myazaki T, Nakahama K, Kikuchi M (1986) *Bacillus subtilis* secretes a foreign protein by the signal sequence of *Bacillus amyloliquefaciens* neutral protease. *Appl. Microbiol. Biotechnol.* **23**: 250 – 256.

Young RCA, Mackenzie CR, Bilon D, Narang SA (1989). *Appl. Microbiol. Environ. Biotechnol.* **55**: 568 – 572.

Zagorec, M., M. Steinmetz, and H. Heslot. (1989). Secretion in *Bacillus subtilis*: interaction of levansucrase::P-galactosidase hybrids with secretory apparatus, p. 519-527. In L. 0. Butler, C. Harwood, and B. E. B. Moseley (ed.), Genetic transformation and expression. Intercept, Andover, England.

Zhao X.J. and H.C. Wu, (1992). Nucleotide sequence of the *Staphylococcus aureus* signal peptidase II (*lsp*) gene. *FEBS Letters* **299**: 80–84.

Zhu C.G, Sun M, Yu Z.N. (2003). Expression of gene aiiA carrying the promoter of gene cry3Aa in *Bacillus thuringiensis. Sheng Wu Gong Cheng Xue Bao.***4**:397-401.

Zhu YF, Curran IHA, Joris B, Ghuysen JM, Lampsen JO (1990). Identification of BlaR, the signal transducer for beta-lactamase production in *Bacillus licheniformis*, as a penicillin-binding protein with strong homology to the OXA-2 beta-lactamase (class D) of *Salmonella typhimurium*. *J Bacteriol.* 1**72**:1137-1141.

Zukowski M (1992). Production of commercially valuable products In: Biology of Bacilli: application to industry, (Doi RH and McGloughlin M, editors) Butterworth–Heirneman, USA, pp 311-331.

Zukowski M.M, Miller L, Cosgwell P, Chen K, Aymerich S, Steinmetz M. (1990). Nucleotide sequence of the sacS locus of *Bacillus subtilis* reveals the presence of two regulatory genes. *Gene.* **90**:153-155.

Zukowski MW, Gaffney FD, Speck D, Kauffman M, Findell A, Wiseman A, Lecocq JP (1983). Chromogenic identification of genetic regulatory signals in *Bacillus subtilis* based on expression of a cloned *Pseudomonas* gene. *Proc. Natl. Acad. Sci. USA* **80**: 1101-1105.

CHAPTER 2

Cloning of the complete lipase gene from *Bacillus licheniformis* by an improved cassette ligation-mediated PCR

2.1. Introduction

There are various approaches that are available to isolate unknown DNA fragments flanking a known locus and these procedures are commonly referred to as "genome walking". The traditional approach consists of the construction and screening of genomic DNA libraries using the known DNA as the probe, followed by the isolation, subcloning, and sequencing of putative positive clones. The PCR-based genome walking techniques are popular since the methods are relatively fast and not labour intensive (Harrison *et al*., 1997).

Inverse PCR (Figure 2.1), described by Ochman *et al* in 1988, is one method for amplification of DNA sequences that flank a region of known sequence. The target genomic DNA is digested and ligated in diluted DNA concentrations to enhance self-ligation to make circular DNA molecules. The circularized DNA molecules are used as templates in the PCR reaction using primers designed based on the known sequence. The primers are designed to direct divergent amplification of the template DNA strands. A double stranded PCR product containing the targeted unknown region amplicon is eventually obtained given the fact that the template strand is circular (Yoshitomo-Nakagawa *et al*., 1997).

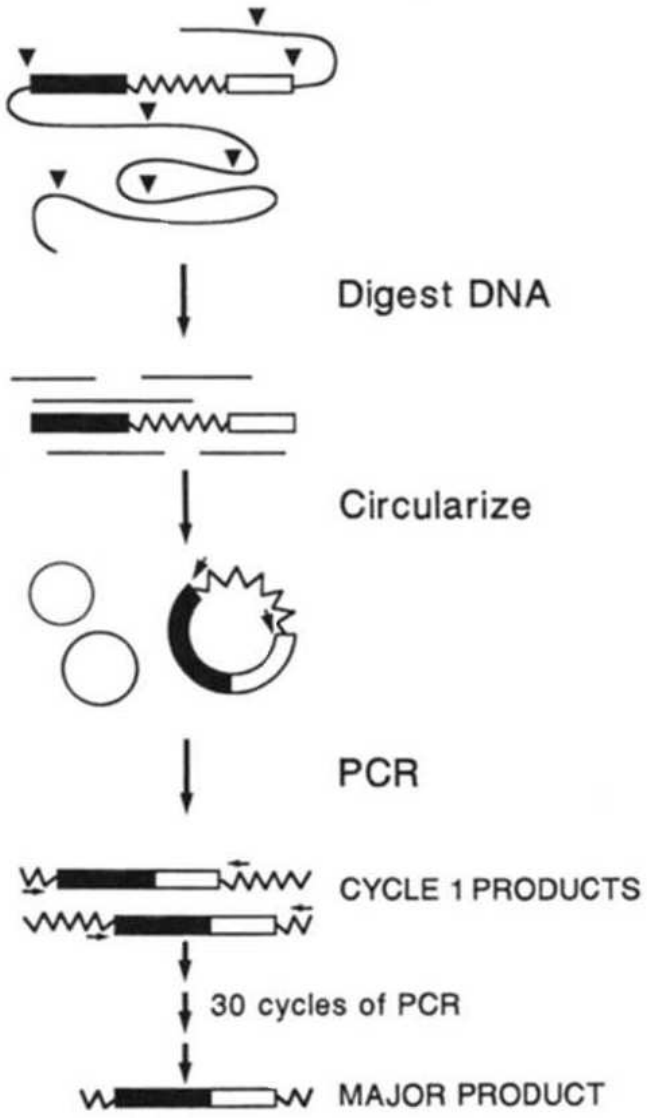

Figure 2.1. Schematic diagram of the inverse PCR procedure. The known region is depicted as a jagged line. Filled and open boxes represent upstream and downstream flanking regions, respectively. DNA is digested with a restriction enzyme (restriction sites denoted by triangles), circularized under conditions that favor the formation of monomeric circles, and amplified using PCR. Oligonucleotide primers (constructed to anneal to the known region) and the direction of DNA synthesis are shown by arrows (Taken from Ochman *et al*., 1988).

Vectorette PCR (Figure 2.2) is a method that enables the amplification of specific DNA fragments in situations where the sequence of only one primer is known. It extends the application of PCR to stretches of DNA where the sequence information is only available at one end (Arnold and Hodgson, 1991). A vectorette is a partially double-stranded DNA cassette that is phosphorylated in its 5′-ends, and after ligation of the vectorette unit to a mixture of genomic DNA fragments, each strand has a vectorette unit attached to both ends (Allen *et al*., 1994; Hengen, 1995). This library is used as a template for amplification with a genome-specific primer together with a vectorette-specific primer, which is identical but not complementary to a central mismatched region in the vectorette cassette. The central mismatched region is included to avoid first-strand synthesis by the vectorette primer. In the second cycle, the vectorette primer can bind to the complementary region present on the product of the first synthesis cycle (Kilstrup and Kristiansen, 2000).

The other genome walking technique is the cassette/adapter/linker-mediated PCR method (Siebert *et al*., 1995), which involves the digestion of genomic DNA with a restriction enzyme and followed by respective ligation to cassette, linker, or to an adaptor. This results in the unknown locus being flanked by known DNA sequences enabling the design of convergent primers that use the ligated product as the template for conventional PCR amplification (Isegawa *et al*., 1992, Siebert *et al*., 1995, Padegimas and Reichert, 1998, Kilstrup and Kristiansen, 2000 and Rishi *et al*., 2004). With this genome walking approach, DNA- and cassette/linker/adaptor-specific nested primers are used for amplifying flanking regions. The technique is rendered more specific by the use of two-round PCR, where nested locus-specific primers are used in the second PCR, thereby increasing the selectivity of amplification of the desired flanking DNA sequence (Rishi *et al*., 2004).

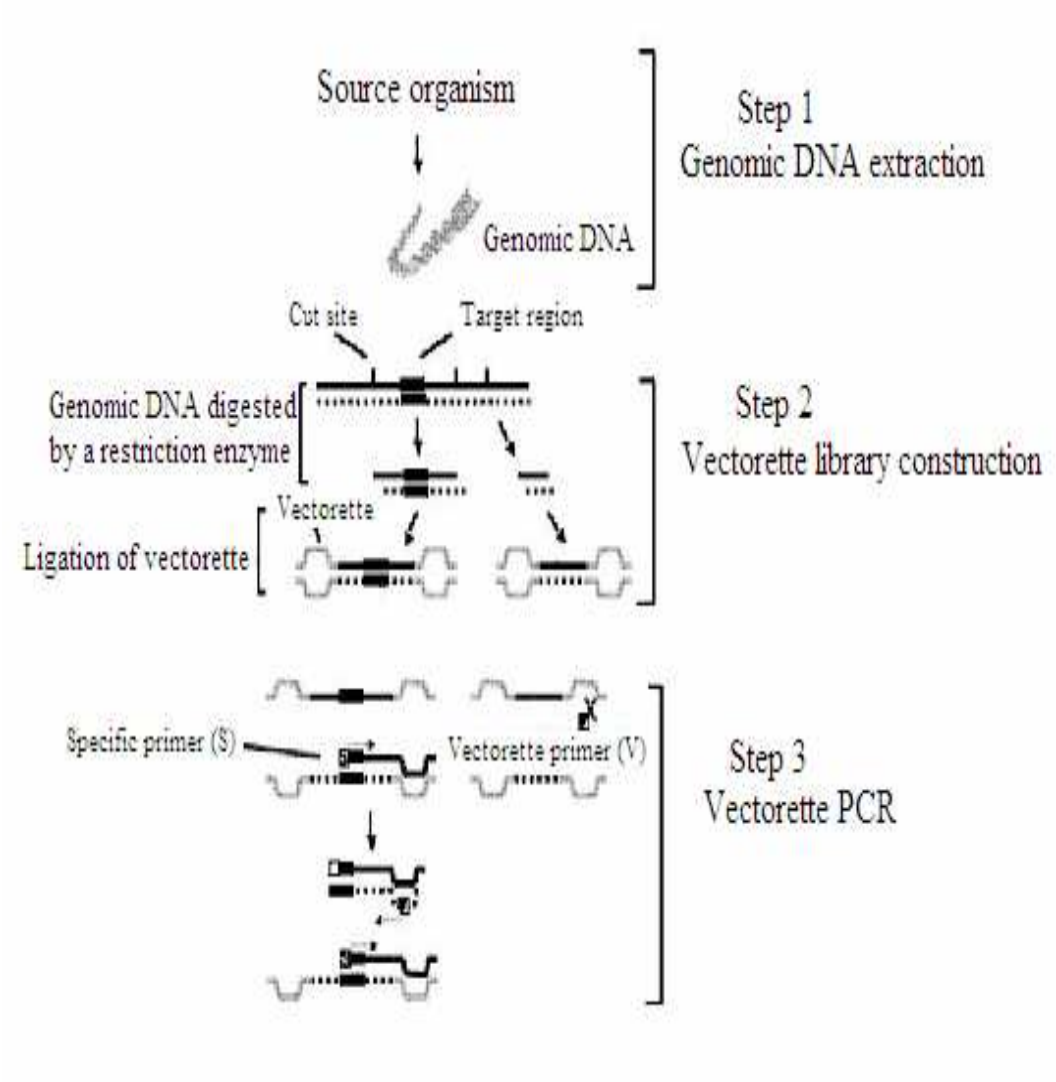

Figure 2. 2. Vectorette PCR. Genomic DNA is extracted (step 1) and digested with a restriction enzyme followed by the ligation of annealed synthetic oligonucleotides (vectorettes)(step 2). PCR amplification from vectorette libraries employs a specific primer that recognizes the region of interest and one primer that anneals to the vectorette. In the first cycle of amplification, primer extension only proceeds from the specific primer (S). Amplification from vectorettes (in gray) does not occur because the vectorette primer (V) only anneals to the complement of the bottom strand of the vectorette. In the second and subsequent cycle of PCR, priming occurs from both the specific primer and the vectorette primer (Taken from Arnold and Hodgson, 1991).

The other genome walking technique is the cassette/adapter/linker-mediated PCR method (Siebert *et al*., 1995), which involves the digestion of genomic DNA with a restriction enzyme followed by respective ligation to cassette, linker, or to an adaptor. This results in the unknown locus being flanked by known DNA sequences enabling the design of convergent primers that use the ligated product as the template for conventional PCR amplification (Isegawa *et al*., 1992, Siebert *et al*., 1995, Padegimas and Reichert, 1998, Kilstrup and Kristiansen, 2000 and Rishi *et al*., 2004). With this genome walking approach, known locus- and cassette/linker/adaptor-specific primers are used for amplifying the flanking regions. The technique is rendered more specific by the use of two-round PCR, where a nested locus-specific primer is used in the second PCR to increase the selectivity of the desired flanking DNA sequence (Rishi *et al*., 2004).

The strategies described above have particular limitation in amplifying the desired size of flanking DNA when the restriction site is too close or far away from the targeted region. In certain cases, it may be possible to overcome this problem by using different restriction enzymes, so that there is a greater chance of obtaining templates where the location of the restriction site is within the amplifiable distance by common commercial thermostable DNA polymerases (Padegimas and Reichert, 1998). Since the genome sequence information is not available for the flanking region, it is difficult to predict the right choice of restriction enzymes to be used and this necessitated prior experiments requiring restriction of the genomic DNA with different restriction enzymes and deducing the appropriate restriction fragments by Southern blotting (Rishi *et al*., 2004).

Genome walking experiments using cassette/adapter-based techniques also suffer from unspecific PCR products being obtained. It was observed that the specificity of adaptor

ligation-based walking techniques could be significantly enhanced by using specifically blocked adaptors along with removal of unligated genomic DNA by exonuclease III digestion (Padegimas and Reichert, 1998). In the other method, locus specific primers are synthesized with the biotin tag. Conventional PCR is done using cassette specific primers and the biotin tagged locus-specific primer. After the first round of amplification, the PCR product is purified using streptavidin affinity chromatography column, which selectively binds biotin tagged PCR products. The biotin tagged PCR products are eluted from the column and used as template using nested locus- and cassette specific primers (Rosenthal and Jones, 1990, Rosenthal, 1992).

The main contributing factors that are responsible for the high failure rates of ligation-mediated PCR genome walking techniques include dependence on the availability of favourable restriction map around the known locus, reduced annealing efficiency and recovery of the annealed cassette, reduced ligation efficiency between the cassette and the restricted target DNA fragments, mispriming, costs, and technical difficulties involved. We hypothesized that a perfectly annealed dephosphorylated double-stranded DNA oligonucleotide with various sticky end restriction enzyme recognition sites would serve as an efficiently ligatable cassette. Linear ligation reaction products would predominantly occur between the compatibly restricted cassette and target DNA molecules when the ligation reaction is done in dilute target DNA concentration in the presence of excess cassette DNA molecules, resulting in suitable PCR templates. The PCR product specificity would be greatly enhanced by firstly performing the first round PCR with a single primer that extends and accumulates the template strand as specified by the locus-specific primer. The single-strand PCR product would then serve as the template for the second round PCR with nested locus-specific primer and the anchoring cassette primer.

To test this hypothesis, the partially known mature lipase gene fragment from *Bacillus licheniformis* was used as the known locus from which upstream and downstream genome walking experiments were initiated. The interest in this region was based on reports which indicated that the lipase gene is expressed and secreted at relatively high levels by *Bacillus licheniformis* during the exponential phase of growth (Nthangeni *et al*, 2001). This suggested that the lipase gene from *Bacillus licheniformis* is preceded by a secretion signal which directs the translocation of the lipase to the extracellular medium. The lipase gene from *Bacillus licheniformis* is transcribed from a promoter whose activity is detectable mainly during the late phase of growth with maximal expression being achieved during the late exponential phase of growth as judged by growth kinetics studies monitoring lipase production versus cell growth (Nthangeni *et al*, 2001). The promoter and the secretion signal regions of the lipase gene from *Bacillus licheniformis* could therefore be useful tools as expression/secretion signals for heterologous production of proteins in *Bacillus licheniformis*.

2.2. Materials and methods

2.2.1. Bacterial strains, DNA manipulations, and culture conditions

Bacillus licheniformis MBB01 (obtainable from the University of the Free State Microbial Cultures Collection) was used. *Escherichia coli* JM109 host cells and the pGemT-Easy vector for PCR fragment cloning were purchased from Promega (Madison, USA). The plasmid pMBE-082 is a derivative of pUC18 and contains additional restriction enzyme recognition sequences within the multiple cloning site. The bacterial strains were grown in Luria–Bertani (LB) media (Sambrook *et al*., 1989) at 37 °C with shaking in 50- to 250-ml shake flasks. $CaCl_2$ competent *E. coli* cells were prepared and transformed with DNA as described by Sambrook *et al*. (1989). TLB media contained sonicated emulsion of 10 ml tributyrin and 10

g gum arabic in 1 L of LB medium. When solid growth media was required, agar (15 g/l) was added and the incubation was done in a 37 °C incubator. The media was supplemented with ampicillin (100 µg/ml) for the growth of plasmid-carrying strains. Genomic DNA isolation was done according to the method described by Shyamala and Ames (1993). Standard microbial and recombinant techniques using commercially available molecular biology grade enzymes and reagents were followed as described by Sambrook *et al*. (1989). Oligonucleotide primers (Table 1 and Figure 2.3) were purchased from Integrated DNA Technologies (USA). The DNA fragments for subcloning were recovered and purified from agarose gel using the GFX PCR and Gel band purification kit (Amersham, UK).

Table 1. The list of primers used in the PCR amplification of *Bacillus licheniformis* extracellular lipase gene. F and the R refer to the forward and reverse primers, respectively. [b] The corresponding position of the underlined bold nucleotide of the primer within the gene sequence as submitted to the nucleotide sequence databases. The underlined sequences are recognition sequences for *Sfi*I or *Hin*dIII used for subcloning.

Primer	Primer sequence (5′– 3′)	Position[b]
BUmlip1F	**G**CTGAGCATAATCCGGTCGTCATG	683
BUmlip220F	**G**TGGCGCATAGTATGGGCGGAGC	902
BUmlip543R	**T**TAATTCGTATTCTGTCCTCCTCC	1228
Blm82R	**C**AACCAAATAACTTTTAATCGAAGC	764
Bl-lip0F	GTATGGCCAAGTCGGCC**A**AGATTTACAGAGCGGT	1 (*Sfi*I)
Blmlip717R	CCTTTAAGCTT**G**TTTGCACAAACGCTCACC	1588 (*Hin*dIII)

2.2.2 Construction of a ligation cassette

The ligation cassette was constructed by PCR with lambda DNA as the template and LAMD11203F: 5′-GGC GAATTC GAGCTC GGTACC CGG GGATCC TGACGAGGTCGCTGAAGCC-3′ (the underlined sequences indicate recognition sequences for *Eco*R1, *Sac*I, *Kpn*I, or *Bam*HI) and LAMB11402R: 5′-CCC AAGCTT CTCGAG CTGCAG GACGAT AT CTCTAG AGGAGGCGGCGGGTC-3′ (the underlined sequences indicate recognition sequences for *Hind*III, *Xho*I, *Pst*I, *Eco*RV, and *Xba*I) as forward and reverse primers, respectively. PCR amplification was done according to standard conditions with general purpose *Taq* DNA polymerase (Roche, Germany; Promega, USA) under the following conditions: 1 denaturation cycle (94 °C, 2 min) and 25 cycles of amplification (94 °C, 30 sec; 55 °C, 30 sec; 72 °C, 1 min). The PCR product generated (Figure 2.4) was digested with *Hind*III and *Eco*R1 and ligated into similarly restricted pUC18 plasmid giving rise to a plasmid denoted pLigCas. The ligation mixture was used to transform the *Escherichia coli* host and the transformed cells were selected on LB ampicillin agar medium. The appropriate recombinant plasmids were identified by the isolation of the plasmid from the host followed by restriction enzyme analysis.

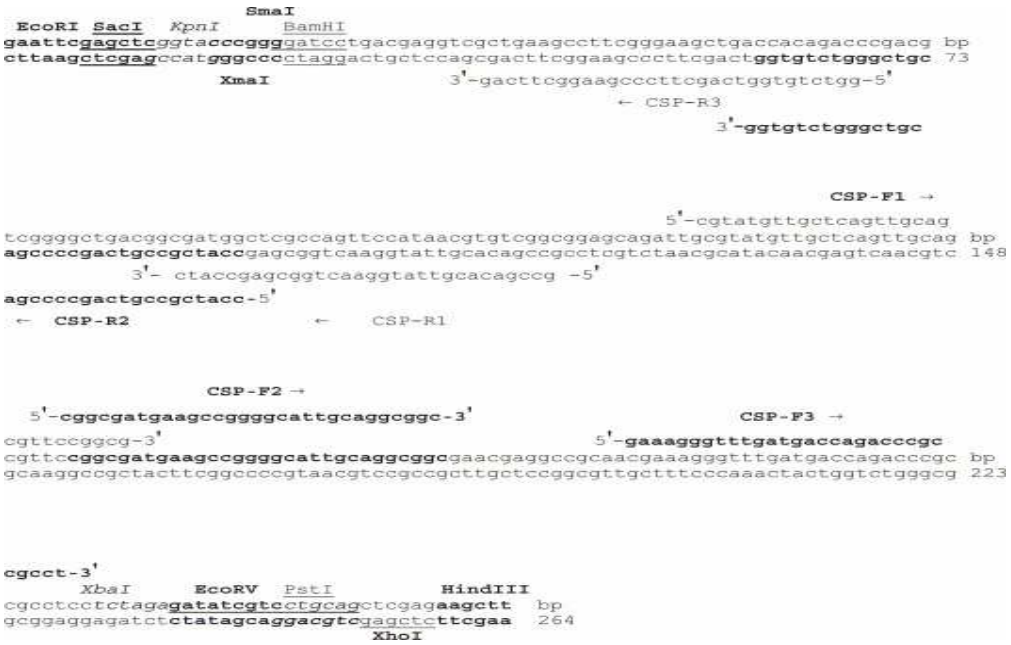

Figure 2.3. The nucleotide sequence of the cassette for genome walking PCR as contained within the pLigCas plasmid. The cassette represents lambda DNA nucleotide sequences 11203–11402 flanked by the indicated recognition sites for restriction enzymes. The cassette-specific primer sequences used in this study have been shown with arrows indicating the 5′–3′ direction of the primer.

2.2.3. Cassette-ligation mediated PCR principle

The ligation cassette is released from pLigCas by digestion with restriction enzymes located on opposite ends of the cassette. The restricted DNA cassette is treated with calf intestinal alkaline phosphatase and run on agarose gel electrophoresis followed by excision and extraction from agarose gel. The genomic DNA is overdigested with the selected restriction enzyme and purified with a DNA cleaning kit. The purified is ligated with the compatibly restricted genomic DNA at a mass ratio of 4:1 in at least 30 µl reaction volume. An aliquot of the ligation mixture is used as the template for the Single-Strand Amplification PCR (SSA-PCR) with the known locus-specific primer-1 (LSP-1) (Fig 2.4) as the lone primer at

annealing temperatures that were 5 °C below the calculated T_m of the LSP-1 primer. The obtained PCR product is used as the template in the second round PCR which is done using nested locus specific primer LSP2 (Figure 2.3) and one of the cassette specific primers shown in Figure 2.3. When *EcoR*1, *Sac*I, *Kpn*I, *Xma*I or *Bam*HI sticky end of the cassette is used, CSP-R1, CSP-R2 or CSP-R3 primer is selected to pair locus-specific primer (LSP2). When *Xba*I, *Pst*I, *Xho*I or *Hind*III sticky end of the cassette is used, CSP-F1, CSP-F2 or CSP-F3 primer (Figure 2.3) is used to pair LSP-2. The second round PCR is done according to the conventional PCR conditions.

2.2.4. Cloning of the regions bordering the mature lipase gene of *Bacillus licheniformis*

The PCR primers for genome walking were constructed based on the 546 bp DNA fragment encoding the mature lipase from *Bacillus licheniformis* (Nthangeni *et al*., 2001). For the amplification of the downstream flanking region, the pLigCas plasmid was digested with *Eco*R1 and *Hind*III restriction enzymes. The *Hind*III sticky end-containing cassette was ligated with *Hind*III-digested genomic DNA from *Bacillus licheniformis* MBB01. BUmlip1F was used as LSP1-A primer in SSA-PCR while BUmlip220F (LSP2-A) and CSP-F3 were used in the second round PCR. The upstream border was amplified with the *Hind*III ligation mixture of cassette: genomic DNA as the template using BUmlip543R as LSP1-B in SSA-PCR, and the second round PCR was performed with BlmLip82R and CSP-F3 as LSP2-B and cassette-specific primers, respectively (Figure 2.4). The PCR was done in the Eppendorf Master gradient thermocycler under the following conditions: one initial denaturation step (94 °C, 2 min) and 30 amplification cycles (94 °C, 30 sec; 55 °C, 30 sec; 72 °C, 4 min).

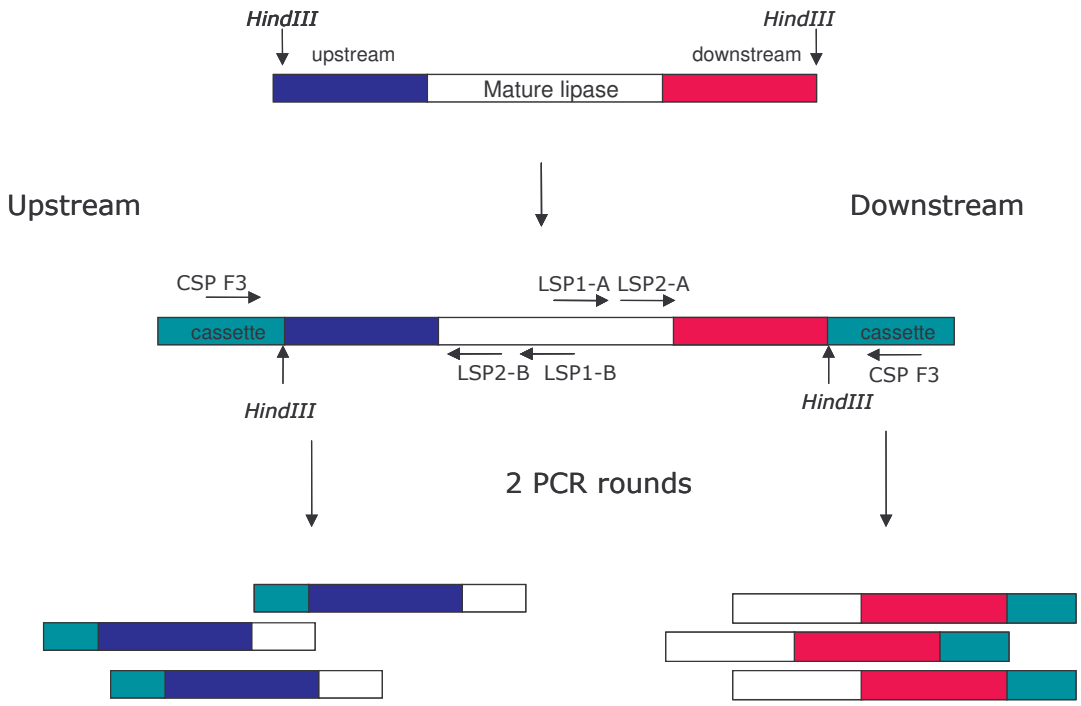

Figure 2.4. A schematic representation of ligation-cassette mediated PCR used to amplify the regions bordering the mature lipase gene from *Bacillus licheniformis*. The mature lipase gene was used as the known locus to design specific primers, which were paired with the cassette-specific primer to amplify both the upstream and downstream regions, flanking the lipase gene.

2.2.5. DNA sequence determination

The DNA fragments obtained by PCR amplifications were ligated into pGemT-Easy and sequenced using T7 and Sp6 universal promoter primers and custom-made primers that became designable after initial sequencing with universal promoter primers. Sequencing was performed with the BigDye Terminator Cycle sequencing kit (Perkin Elmer), in a 377 Perkin Elmer DNA sequencer. The complete nucleotide sequence was deposited in the nucleotide databases as AJ297356 (for the 3040 bp DNA fragment that inclusively code for the extracellular lipolytic activity from *Bacillus licheniformis* MBB01). The different open

reading frames (ORFs) found within the sequence after analyses were aligned with other open reading frames using CLUSTALW tools (Oliver *et al*., 2005), and homology was ascertained through Basic Local Alignment Search Tools (BLAST) (Altschul *et al*, 2005). Signal peptide identification was performed using the SignalP V2.0 software (Dyrlov Bendtsen *et al*., 2004).

2.2.6. Functional expression of the lipolytic gene

The primers Bl-lip0F and Blmlip717R were used to amplify the 1517 bp fragment within the genomic DNA of *Bacillus licheniformis* MBB01. The PCR product was digested with *Sfi*I and *Hind*III and ligated into pMBE-082 cut with similar enzymes. The ligation mixture was transformed into *Escherichia coli* host cells and selected on LB ampicillin agar plates. *Escherichia coli* cells containing appropriate recombinant plasmids were patched on TLB agar plates and incubated at 37 °C for 2 to 3 days.

2.3. Results

2.3.1. Genome walking PCR

A 266 bp lambda DNA fragment was amplified by PCR to serve as an adapter for cassette ligation-mediated PCR (Fig 2.5). A number of commonly used restriction enzyme sites were incorporated to flank the ligation-cassette and to facilitate cloning into pUC18 during the construction of the pLigCas plasmid. Digestion of the pLigCas plasmid with restriction enzymes positioned on opposite sides of the ligation-cassette yielded efficiently annealed cassette (Fig 2.3).

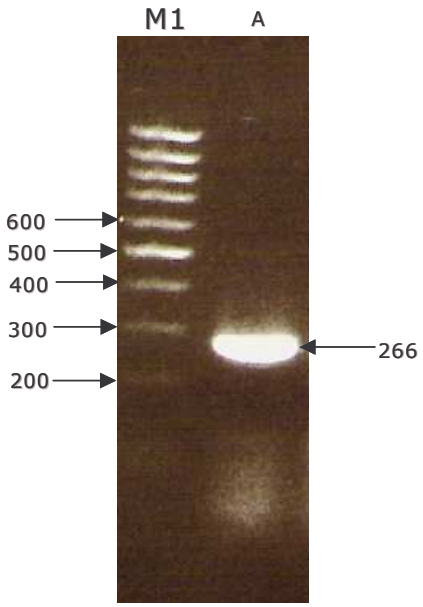

Figure 2.5. A 1% agarose gel showing the cassette DNA fragment after PCR amplification. Lane A is the ~ 266 bp long fragment and M1 is the 100 bp DNA molecular weight ladder.

2.3.2. Cloning of the complete extracellular *Bacillus licheniformis* MBB01 lipase gene

DNA fragments of 800 and 2200 bp respectively corresponding to the upstream and downstream regions flanking the mature extracellular lipase from *Bacillus licheniformis* were amplified using the SSA-PCR-based genome walking procedure (Fig 2.6). The distance between the LSP-2 and the position of the restriction enzyme used to digest the target DNA molecule defined the size of the PCR fragment. The 3040 bp DNA fragment could be amplified from *Bacillus licheniformis* MBB01 genomic DNA by conventional PCR with primers designed based on the bordering *Hind*III sites.

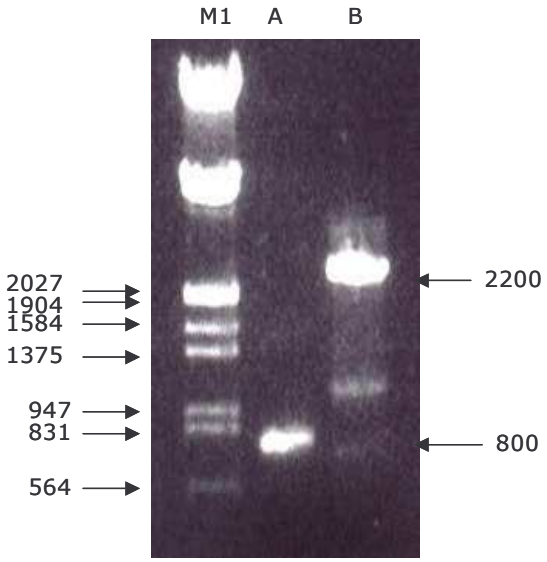

Figure 2.6. SSA-PCR-based genome walking DNA fragments run on 1% agarose gel. (A) Upstream region and (B) downstream border flanking *Bacillus licheniformis* mature lipase gene.. M1 is the lambda DNA (digested with *Hind*III and *Eco*RI) molecular weight marker.

Nucleotide sequence analysis revealed the presence of three open reading frames (ORFs) (Figure 2.7). The extracellular lipase ORF was located between nucleotides 186 and 826 and encoded a lipolytic protein consisting of 213 amino acids, which included the region encoding the mature lipase from *Bacillus licheniformis* (Nthangeni *et al*., 2001). The nucleotide sequence corresponding to a typical signal peptide was present at the 5′-end of the ORF [nucleotides 587–682] (Figure 2.8). Downstream to the lipase gene, there was an ORF (position 1581–2144) encoding a protein of 187 amino acids, which was revealed by similarity searches to belong to the isochorismatase protein family. There was also an additional ORF located upstream to the lipase gene on the complementary strand. The ORF

encoded a small protein of 118 amino acids, which aligned with hypothetical conserved proteins of unknown functions.

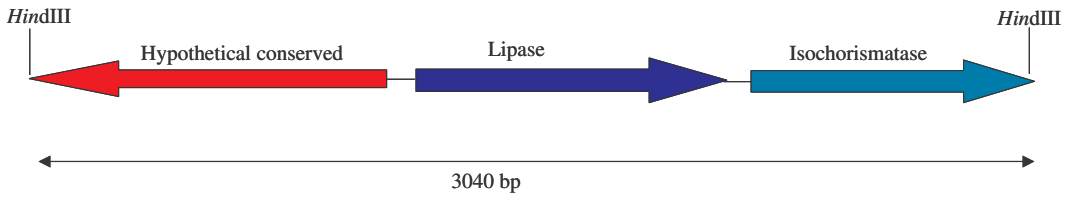

Figure 2.7. A schematic diagram representing the arrangement and orientation (as indicated by arrows) of the ORFs within the 3040 bp fragment amplified from the genomic DNA of *Bacillus licheniformis*.

2.3.4. Nucleotide sequences analysis

Two ORFs within the amplified 3040 bp region are transcribed in the same direction and the other one was located on the complementary strand. Potential promoter sequences were observed at nucleotide sequences immediately upstream to putative translation start codons of all the reported ORFs. The start codon is ATG for the hypothetical conserved protein, and GTG for isochorismatase ORF. For the extracellular lipase ORF, although there is an ATG codon immediately downstream of the RBS (Figure 2.8), translation initiation is likely to begin at the GTG codon located about 6 bp downstream of the RBS due to the requirement for a 6-9 nucleotide sequence between the RBS and the initiation codon for efficient translation (Mountain, 1989). The promoter region of *Bacillus licheniformis* lipase is 96% identical to that of *Bacillus pumilus* (Accession Number A34992), suggesting a common source during evolution and similar mechanisms of regulation. The start codons were separated from the putative −10 regions (TATAAT) by AG-rich regions, presumably

representing the RBS (Figures 2.8, 2.9 and 2.10). Inverted repeats could also be located at the 3′ ends of the lipase (834-861) and isochorimatase (position 698-733) genes (Figure 2.10). These nucleotides are probably involved in the formation of stem–loop structures that function as transcription termination signals. The characteristic sequence encoding a peptide that conforms to *Bacillus* signal peptides could be identified within the deduced lipase protein sequence and contained the characteristic Ala-X-Ala signal peptidase cleavage site (Figure 2.9).

2.3.5 Similarity searches of ORFs.

The amino acid sequences of the encoded proteins were aligned with similar proteins from other bacterial species. *Bacillus licheniformis* lipase displayed 96% amino acid sequence identity with lipase from *Bacillus pumilus* (Figure 2.11A). There was also a good alignment with the amino acid sequences of lipases from the other species, particularly the conserved regions of the gene. The first 32 amino acids occupying the N-terminal regions of the lipases do not display any significant homology. These are the signal peptide regions with the cleavage site located between the two consecutive Ala residues. The mature portions of the lipase exhibit a sequence homology of more than 80%. The isochorismatase amino acid sequence of *Bacillus licheniformis* exhibited 55% sequence similarity with isochorismatases from other *Bacillus* species (Figure 2.13B). The hypothetical conserved protein from *Bacillus licheniformis* also aligned relatively well with hypothetical conserved proteins from other *Bacillus* species (Figure 2.13C).

A. Lipase encoding nucleotide sequence

```
TTTCATATTATATATGAAAGAAAGAAAGTCTCGTAGTCTTTTTGTTTTAATTTAAAAAAG          461
TCCAAGGTGCTTTATGATTATTTATATTTCTGTAAAATCATCTCATAAACATTACCTTGT          521
TCACTTTTCTGACATATTTTTCTTGTATAAAATAGAGTCGAATAAGATGAATAAGGGGGA          581
-35                     -10                          rbs
ATGAAAGTGTCGTTATTGATGAAAAGGAGTTTGCAGATTCTCGTTGCATTTGCATTGGTG          641
M  K  V  S  L  L  M  K  R  S  L  Q  I  L  V  A  F  A  L  V
ATTGGTTCAATGGCTTTTATCCAGCCTAAAGAGGTGAAGGCGGCTGAGCACAATCCGGTC          701
I  G  S  M  A  F  I  Q  P  K  E  V  K  A  A  E  H  N  P  V
GTCATGGTACATGGTATTGGAGGAGCGTCTTATAACTTTGCTTCGATTAAAAGTTATTTG          761
V  M  V  H  G  I  G  G  A  S  Y  N  F  A  S  I  K  S  Y  L
GTTGGACAAGGCTGGGATCGAAACCAATTATTTGCTATCGATTTCATAGACAAAACAGGG          821
V  G  Q  G  W  D  R  N  Q  L  F  A  I  D  F  I  D  K  T  G
AATAACCGCAACAATGGTCCTCGTTTATCTAGATTCGTCAAAGATGTGCTAGACAAAACG          881
N  N  R  N  N  G  P  R  L  S  R  F  V  K  D  V  L  D  K  T
GGTGCCAAAAAAGTAGATATTGTGGCGCATAGTATGGGCGGGGCGAACACGCTATACTAT          941
G  A  K  K  V  D  I  V  A  H  S  M  G  G  A  N  T  L  Y  Y
ATTAAGAATCTAGATGGCGGCGATAAAATTGAAAACGTCATCCCCATTGGTGGAGCAAAC         1001
I  K  N  L  D  G  G  D  K  I  E  N  V  I  P  I  G  G  A  N
GGACTCGTTTCAAGCAGAGCATTACCAGGAACAGATCCAAATCAAAAAATTCTTTACACA         1061
G  L  V  S  S  R  A  L  P  G  T  D  P  N  Q  K  I  L  Y  T
TCTGTTTATAGCTCGGCAGATCTCATCGTCGTCAACAGCCTTTCTCGGTTAATTGGTGCA         1121
S  V  Y  S  S  A  D  L  I  V  V  N  S  L  S  R  L  I  G  A
AGAAACGTCCTGATCCATGGCGTTGGCCATATCGGTCTATTAACCTCAAGCCAAGTGAAA         1181
R  N  V  L  I  H  G  V  G  H  I  G  L  L  T  S  S  Q  V  K
GGGTATATTAAAGAAGGACTAAACGGTGGAGGACAGAATACGAATTAAAAATGAAAAAGA         1241
G  Y  I  K  E  G  L  N  G  G  G  Q  N  T  N  *
GCGACTTGTGCTGTCTTTTTTGATGTCCTCGCTATATGGCAGCTACTAATTGGTTAAGGA         1301
GGCAACACGTT
```

Figure 2.8. The nucleotide sequence of the complete extracellular *Bacillus licheniformis* lipase gene and the deduced amino acid sequence. The putative promoter region (−35 and −10 regions) and rbs are underlined. The arrowhead is representing the signal peptide-processing site. The palindromic inverted repeat sequence is shown by horizontal arrows under the sequence. The stop codon is indicated by an asterisk.

B. Nucleotide sequence coding for isochorismatase

```
ATCTATTACA TCATCGCTAG TGTGTGGTTC GTCTTTTTGA TGAAAAAGCA GTAAAAGCGT    1525
GTACAAAAAA ATCACCATTC CTTACACTAC ATGTAGTGT  ATAAGAATCT TCGGAGGTGA    1585
           -35                          -10             rbs
      M Q   T  S  Q   K  A  L   I  I  V  D   V  Q  K   A  F  E
GCGTTTGTGC AAACGAGTCA AAAGGCTTTG ATCATTGTTG ATGTACAGAA GGCGTTCGAG    1645
H  E  K  W   G  E  R   N  N  L   E  A  E  N   N  I  S   Q  L  L
CATGAAAAGT GGGGCGAACG AAATAATTTA GAGGCAGAGA ATAATATCAG CCAGTTATTG    1725
R  F  W  R   E  K  G   W  K  V   I  H  I  Q   H  T  A   D  N  P
AGGTTTTGGA GAGAGAAAGG CTGGAAAGTC ATTCATATTC AGCATACAGC CGACAATCCA    1785
D  S  L  F   H  P  A   N  E  G   H  L  F  K   D  I  A   R  P  L
GACTCTTTGT TTCACCCGGC AAATGAGGAC ATTTGTTCAA AGATATTGCC AGACCGTTAG    1845
E  E  E  T   V  I  Q   K  K  V   N  S  S  F   I  G  T   N  L  E
AGGAAGAAAC GGTGATCCAA AAGAAAGTGA ACAGCAGCTT CATTGGTACG AACTTGGAAG    1925
E  Q  L  R   S  N  H   M  D  T   V  V  I  A   G  L  T   T  P  H
AACAACTAAG ATCAAACCAT ATGGACACGG TCGTGATCGC AGGGCTCACA ACACCGCATT    1985
C  V  S  T   T  T  R   M  S  G   N  L  G  F   D  T  Y   L  I  T
GTGTATCCAC AACCACAAGA ATGAGCGGAA ATCTTGGCTT TGATACGTAT CTGATTACGG    2045
D  A  T  A   A  F  G   L  T  D   Q  T  G  T   Y  F  D   P  E  T
ACGCAACAGC TGCCTTTGGG CTAACAGATC AAACCGGTAC ATATTTTGAC CCAGAAACAA    2125
I  H  R  L   S  L  A   T  L  H   D  E  F  A   T  V  L   T  T  E
TCCATCGGCT CTCGCTTGCC ACGCTCCATG ATGAATTTGC GACTGTGCTG ACAACAGAGC    2185
Q  L  L  T   N  W  A   K  S  *
AACTCCTCAC AAATTGGGCA AAATCATAAT TGAACGTTAA AAAGCTCCTA GAAAAATATC    2245
TAGGAGCTTT TTTATGCCAA ATCAGATGTG TCCAGTGATT GTTCTCGTTT TTTTCTTGAG    2325
```

Figure 2.9. The DNA sequence encoding isochorismatase and the deduced amino acid sequence. The –10 and –35 sequences and rbs are underlined. The stop codon is represented by * and the inverted repeat sequence is depicted by arrows.

C. Nucleotide sequence encoding hypothetical conserved protein

```
ATCATAAAGCACCTTGGACTTTTTTAAATTAAAACAAAAAGACTACGAGACTTTCTTTCT          60

TTCATATATAATATGAAAGACAAAAGGAGGTTCTACTAG**ATG**TTATTTACTCAAACACCT       120

       -10             rbs            M  L  F  T  Q  T  P

GCAACGCCTTATTATGCCGTGATTTTCACATCAAAACGAACAAATGTCGAGCAGAAGGAA          180

 A  T  P  Y  Y  A  V  I  F  T  S  K  R  T  N  V  E  Q  K  E

TACGCTGCGGCTGCTCAAAAAATGGAAGAATTAGCGAAAGAACTAGACGGCTTTCTTGGA         240

 Y  A  A  A  A  Q  K  M  E  E  L  A  K  E  L  D  G  F  L  G

ATAGAAAGTGTAAGAGATGAACATGGCGTTGGAATCACTGTTTCTTACTGGAATTCATTA          300

 I  E  S  V  R  D  E  H  G  V  G  I  T  V  S  Y  W  N  S  L

GCAGCCATCAAAGAATGGAAAGCACATACAGCACATCAGCATGTTCAAAAAAAAGGGAAA         360

A  A  I  K  E  W  K  A  H  T  A  H  Q  H  V  Q  K  K  G  K

AAGGATTGGTATGAAGCCTATACAACCCGAATTTGTAAGGTAGAACATGCTTATACAAAC         420

K  D  W  Y  E  A  Y  T  T  R  I  C  K  V  E  H  A  Y  T  N

GAAGATTCACTTCTTGAACAAGAAGCAGACAAATGAGAGAAACCGCTCTGTAAAGCTT           480

E  D  S  L  L  E  Q  E  A  D  K  *
```

Figure 2.10. The nucleotide sequence encoding the hypothetical conserved protein and the deduced amino acid sequence. The promoter sequences are underlined and the terminator is represented by arrows. The translation start codon is in bold and the stop codon is represented by *..

A.

```
B.pumilus           MKVILFKKRSLQILVALALVIGSMAFIQPKEVKA|AEHNPVVMVHGIGGASYNFASIKSYL 60
B.licheniformis     --VSLLMKRSLQILVAFALVIGSMAFIQPKEVKA|AEHNPVVMVHGIGGASYNFASIKSYL 58
Bacillus            MKVIRFKKKSLQILVALALVLGSIAFIQPKEAKA AEHNPVVMVHGMGGASYNFASIKRYL 60
B.subtilis          --MKFVKRRIIALVTILMLSVTSLFALQP-SAKA AEHNPVVMVHGIGGASFNFAGIKSYL 57
                      :  . :: : ::. : * : *:  :** ..** ***********:****:***.** **

B.pumilus           VGQGWDRNQLYAIDFIDKTGNNRNNGPRLSRFVKDVLDKTGAKKVDIVAHSMGGANTLYY 120
B.licheniformis     VGQGWDRNQLFAIDFIDKTGNNRNNGPRLSRFVKDVLDKTGAKKVDIVAHSMGGANTLYY 118
Bacillus            VSQGWDQNQLFAIDFIDKTGNNLNNGPRLSRFVKDVLAKTGAKKVDIVAHSMGGANTLYY 120
B.subtilis          VSQGWSRDKLYAVDFWDKTGTNYNNGPVLSRFVQKVLDETGAKKVDIVAHSMGGANTLYY 117
                    *.***.:::*:*:** ****.* **** *****:.** :*********************

B.pumilus           IKNLDGGDKIENVVTIGGANGLVSSRALPGTDPNQKILYTSVYSSADLIVVNSLSRLIGA 180
B.licheniformis     IKNLDGGDKIENVIPIGGANGLVSSRALPGTDPNQKILYTSVYSSADLIVVNSLSRLIGA 178
Bacillus            IKNLDGGDKIENVVTLGGANGLVSLRALPGTDPNQKILYTSVYSSADLIVVNSLSRLIGA 180
B.subtilis          IKNLDGGNKVANVVTLGGANRLTTGKALPGTDPNQKILYTSIYSSADMIVMNYLSRLDGA 177
                    *******:*: **:.:**** *.: :***************:*****:**:* **** **

B.pumilus           KNVLIHGVGHIGLLTSSQVKGYIKEGLNGGGQNTN 215
B.licheniformis     RNVLIHGVGHIGLLTSSQVKGYIKEGLNGGGQNTN 213
Bacillus            RNVLIHGVGHIGLLTSSQVKGYVKEGLNGGGQNTN 215
B.subtilis          RNVQIHGVGHIGLLYSSQVNSLIKEGLNGGGQNTN 212
                    :** ********** ****:. :************
```

B.

```
B.thuriengensis     MMNNQIALIIIDVQKAFQLPYWGERNNLFAEENMRILLAEWRKRKRPVFHIQHVNKENDQ 60
B.anthracis         -MNNQIALILIDVQKAFQLPYWGERNNLFAEENMRILLAEWRKRKRPVFHIQHVNKENDQ 59
B.cereus            MMTNQIALIIIDVQKAFQLPDWGERNNIFAEENMRILLEEWRKRERPVFHIQHVNKENAQ 60
Bacillus sp         VQTSQKALIIVDVQKAFEHEKWGERNNLEAENNISQLLRFWREKGWKVIHIQHTA-DNPD 59
B.licheniformis     -MEEKKALIIVDVQKAFDDKKWGERNNVKAEENISKILELWREKGWTVIYIQHTS-DKPH 58
                       .: ***::******:   ******: **:*:  :*  **::   *::***.  :: .

B.thuriengensis     SMFYEGAETVNFKEEVKPLSCEVVIQKTVNSAFIGTNLEEQLRENKCNAVVIVGLTTNHC 120
B.anthracis         SMFYEGAETVNFKEEVKPLSGEVVIQKTVNSAFIGTNLEEQLRENKCNAVVIVGLTTNHC 119
B.cereus            SMFYERAETVNFKEEVQPLLGEVVIQKFVNSAFIGANLEEQLRKNKCNAVVVVGLTTNHC 120
Bacillus sp         SLFHPANEGHLFKDIARPLEEETVIQKKVNSSFIGTNLEEQLRSNHMDTVVIAGLTTPHC 119
B.licheniformis     SLFHPKNEGFAIKEIVKPMDEEVIITKTVNSSFIGTNLEEFLKLNEITTVVITGLTTPHC 118
                    *:*:   *   :*: .:*:  *.:* * ***:***:**** *: *.  :**:.**** **

B.thuriengensis     VETTTRMAGNLGFTTYLVRDATATFNRKGLDGKEYSAEDIHNMTLVNLHEEFATIVTTKE 180
B.anthracis         VETTTRMAGNLGFTTYLVRDATATFNRKGLDGKEYSAEDIHNMTLVNLHEEFATIVTTKE 179
B.cereus            VETTTRMAGNLGFTTYLVSDATATFNRKGLDGKEYSAEDIHNMTLVNLHEEFATIVTTKE 180
Bacillus sp         VSTTTRMSGNLGFDTYLITDATAAFGLTDQTGTYFDPETIHRLSLATLHDEFATVLTTEQ 179
B.licheniformis     VSTTTRMSGNLGFDTYLISDATAAFGMRDQNDTYYDAATIHNISLATLHDEFATILTTDQ 178
                    *.*****:***** ***: ****:*.  .  .. :..  **.::*..**:****::**.:

B.thuriengensis     ALKLF---- 185
B.anthracis         ALKLF---- 184
B.cereus            VLKLF---- 185
Bacillus sp         LLTNWAKS- 187
B.licheniformis     LINDFIKTH 187
```

C.

```
B.cereus            MKNNLQKTSHYYAVIFTSNLSN-DTTDYNAASDKMEELAKQQPGFLGVESARDHSGLGIT 59
B.anthracis         MASNLEKNNSYYAVIFTSNLSN-DTADYSTVADKMEELAKQQPGFLDVESVRDHSGLGIT 59
B.licheniformis     MLFTQTPATPYYAVIFTSKRTNVEQKEYAAAAQKMEELAKELDGFLGIESVRDEHGVGIT 60
                    *  .    . ********: :* :  :* :.::********:  ***.:**.**. *:***

B.cereus            ISYWESLEAIENWKQNALHKEAKKRGREQWYENFHLRICLVEKEFKFHRGTL------ 111
B.anthracis         ISYWESLEAIENWKQNVLHKEAKKRGREQWYENFHLRICLVEKEFKFHRNTL------ 111
B.licheniformis     VSYWNSLAAIKEWKAHTAHQHVQKKGKKDWYEAYTTRICKVEHAYTNEDSLLEQEADK 118
                    :***:** **::** :. *:..:*:*:::*** :  *** **: :. . . *
```

Figure 2.11. Alignment of amino acid sequences of (A) lipases, (B) isochorismatases, and (C) hypothetical conserved proteins with proteins from various *Bacillus* species. The asterisks (*) represent identical amino acid residues in all sequences in the alignment. The vertical arrow indicates the predicted cleavage site of the lipase signal peptide.

2.3.6. Functional expression of the lipolytic gene

The subcloning of the DNA fragment by conventional PCR to amplify the lipase gene with its regulatory regions including the promoter, signal peptide and the terminator generated *Escherichia coli* host cells containing recombinant plasmid becoming surrounded by zones of clearance when grown on TLB plates (Fig 2.12). This indicated that lipolytic gene, together with its promoter region, was indeed cloned.

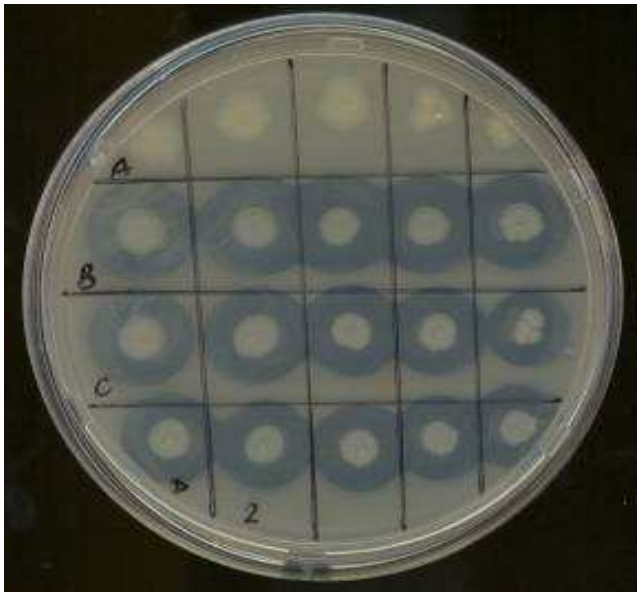

Figure 2.12. A Photograph showing lipase production by *Escherichia coli* transformed with plasmid containing the complete lipase gene. A is the negative control. The experiment was done using three independent transformants (B, C and D) and the picture was taken after 48 hours of incubation at 37°C.

2.4. Discussion

The study described the construction and application of a DNA cassette that allows genome walking from a known locus into its flanking regions. The study used only one cassette-specific primer, but many other primers with different sequences and T_m are possible to accommodate PCR conditions that require different parameters (Nthangeni *et al*., 2005). Various restriction enzyme sites that generate sticky ends to improve the efficiency of ligation with compatibly restricted target DNA fragments flanked the cassette. The availability of various recognition restriction enzyme sites reduced the dependence of the technique on the availability of a favourable restriction map around the known locus, which was found to be a major limitation factor in genome walking experiments. It was indeed, possible to walk upstream and downstream of the known locus sequences without any prior knowledge of the restriction map by using the various possible compatible restriction enzymes.

When cassettes with different sticky ends are required, it is a common practice to design and anneal separately synthetic pairs of complementary oligonucleotides (Zhang and Gurr, 2000 and Reddy *et al*., 2002). The annealing process of synthetic oligonucleotides does not result in perfectly annealed cassettes. The single-stranded DNA molecules that remain during the annealing process reduce the ligation efficiency of the cassette to the target DNA; otherwise, they are removed by alcohol precipitation, which has the disadvantage of compromising the recovery of the cassette. The subcloning of the cassette to create the pLigCas plasmid ensured optimal restriction digestion and availability of the cassette in almost unlimited quantities. The size of the cassette is small enough (approximately 260 bp) to be sufficiently recovered from agarose gel by standard commercial kits after restriction digestion and allowed efficient ligation on both ends of the restricted target DNA molecule to form inverted repeats.

Depending on the choice of restriction enzyme, sticky ends or blunt ends are generated on one or both ends of the cassette. The cassette was treated with alkaline phosphatase to remove the phosphate group from the 5′-ends of the cassette to prevent self-ligation. Digestion of the cassette with restriction enzymes that generate sticky ends of more than one base is preferred to ensure efficient ligation to a compatibly restricted target genomic DNA molecule. The ligation reaction, however, results in a nick between the 5′-end of the cassette and the 3′-end of the target DNA since the 5′-end of the cassette is not phosphorylated. The cassette/genome DNA ligation reaction is theoretically optimal in a relatively large volume that ensures dilute concentrations of genomic DNA fragments while maintaining higher concentration of the cassette molecules.

Under these conditions, self-ligation between genomic DNA molecules is minimal favouring linear ligation reaction products with the cassette. SSA-PCR is done with a lone primer designed based on the known DNA sequence. The PCR selectively amplifies the template strand through the site of ligation into the cassette DNA sequence. The non-template strand, which contains the nick, is not amplified, firstly due to the absence of a cassette-specific primer and secondly due to the presence of the nick. The SSA-PCR results in the amplification and accumulation of the targeted DNA strand, some of which are tagged with a DNA sequence of one strand of the ligated cassette.

The SSA-PCR involved an enrichment step that results in the selective extension of the template DNA strand of the known locus-specific primer, through the ligation site into the cassette DNA sequence. This results in the synthesis of the complementary strand of the cassette strand that becomes the template of the cassette-specific primer during the second round PCR. It is only after the locus-specific primer has been extended in the first round PCR

that cassette-specific primer can find suitable template to bind and be successfully extended in the amplification cycles of the PCR.

A novel method for targeted DNA strand enrichment involved the use of biotin capture PCR where the first round biotinylated PCR product is purified by the biotin/streptavidin affinity chromatography for subsequent use as template with nested primers in the second round PCR (Mishra *et al*., 2002). While the SSA-PCR step achieved the same objective as the targeted strand enrichment step using biotinylated primers, it was cost-effective, rapid, and simple to perform. We have done PCR amplifications with and without the initial SSA-PCR. We observed that while the SSA-PCR product readily resulted in distinct bands in the second round of conventional PCR with a nested known locus-specific primer, the PCR without the SSA-PCR required extensive optimisation in the second round and, in some instances, required the use of the second nested primer to obtain specific products.

The potential of this method for walking into the regulatory regions of partially known genes has been demonstrated. We successfully walked into upstream and downstream border regions of the extracellular *Bacillus licheniformis* lipase gene and in the process cloned two new genes (isochorismatase and hypothetical conserved genes).

The isochorismatase gene, which was previously unknown, was walked into by SSA-PCR. Isochorismatase from *Bacillus licheniformis* MBB01 displayed 55% sequence identity with isochorismatases from other *Bacillus* species. In *Escherichia coli*, this enzyme has been implicated in the conversion of chorismate to 2,3 dihydroxybenzoate, a precursor in the formation of enterobactins. Although putative isochorimatase genes have been identified in bacterial genomes, it is only recent that they are starting to attract some interest owing to their

ability to stereoselectively convert chorismate to 2,3 dihydroxybenzoate which has been shown to be of importance as a chiral starting material for the synthesis of bioactive substances, especially for the synthesis of carbasugars (Franke *et al*, 2003). Isochorismatase is also one of the enzymes that take part in the shikimate biosynthetic pathway, which is found in plants, bacteria, fungi and some parasites. The final product of the shikimate pathway, chorismate, is the branchpoint for the production of folate, aromatic acids, vitamins K and E, coenzyme Q, enterobactin, chloramphenicol, plastoquinones, phenoxazinones and other metabolite (He *et al*., 2004). The central importance of the chorismate-dependent pathways and their absence in mammals make them very attractive targets for the development of antimicrobials and herbicides.

The other ORF encoded a hypothetical conserved protein from *Bacillus* with unknown functions. The nucleotide sequence encoded a 119 amino acid residues long polypeptide. BLAST searches grouped the encoded protein with hypothetical proteins from *Bacillus cereus* and *Bacillus anthracis* whose functions are not known. Alignment of the hypothetical-conserved protein from *Bacillus licheniformis* with hypothetical conserved proteins from other *Bacillus* species exhibited a 55 % sequence similarity.

The third lipase encoding ORF is 641 bp long and encodes a protein that is 213 amino acid residues long. The putative cleavage site of the signal peptide was located between Ala-2 and Ala-3 residues in consensus with signal peptides from other *Bacillus* extracellular proteins. The mature lipase gene-encoding region could also be located within this ORF. The characteristic signal peptide region of extracellular proteins from *Bacillus* was also observed preceding the region encoding the mature lipase. The deduced signal peptide of lipase from

Bacillus licheniformis MBB01 was 22 amino acid residues long with the cleavage site Ala-Ala that is common amongst *Bacillus* exoproteins (Simonen and Palva, 2003).

From the partial gene sequence, we successfully used the cassette ligation-mediated PCR to amplify the complete extracellular lipase gene (together with its promoter and terminator regions) encoding for lipolytic activity. *Escherichia coli* cells harboring the complete lipolytic gene subcloned into pMBE-082 became surrounded by zones of clearance when plated on TLB agar plates, an indication that the complete functional gene was indeed amplified. The availability of the promoter region upstream of the lipase gene reported here could facilitate studies aimed towards understanding their transcriptional regulation and function. The putative TATAAT region could be identified within the promoter region of the lipolytic gene but the −35 region could not be predicted with ease. There was no major nucleotide sequence variation within the promoter region of *Bacillus licheniformis* and *Bacillus pumilus* lipase genes, suggesting similar mechanisms of gene transcription regulation.

The promoter regions of extracellular proteins produced by some *Bacillus* species have been used as expression and secretion signals for heterologous protein production. The cloned upstream fragment of *Bacillus licheniformis* mature lipase gene possesses attractive features that warrant further investigation as tools for heterologous protein expression. First, the promoter results in relatively high-level transcription of the lipase gene during late exponential to stationary phases of growth of the organism. The region contains the signal peptide which could efficiently translocate heterologous genes to the extracellular medium. It is therefore presumable that the lipase gene promoter and the signal peptide when contained in a multiple copy plasmid could result in enhanced overproduction of endogenous and heterologous proteins.

2.5. References

Allen M.J., A. Collick and A.J. Jeffreys, (1994). Use of vectorette and subvectorette PCR to isolate transgene flanking DNA. *PCR Methods. Appl.* **4**:71–75.

Altschul SF, Wootton JC, Gertz EM, Agarwala R, Morgulis A, Schaffer AA, Yu YK. (2005). Protein database searches using compositionally adjusted substitution matrices. *FEBS J.* **272**:5101-5109.

Arpigny J.L. and K.E. Jaeger, (1999). Bacterial lipolytic enzymes: classification and properties. *J. Biochem.* **343**: 177-183.

Bron, S., Meima, R., van Dijl, J.M., Wipat, A. and Harwood, C.R., (1999). Molecular biology and genetics of *Bacillus* species. In: Demain, A.L. and Davies, E., Editors, 1999. Manual of Industrial Microbiology and Biotechnology (second ed.), ASM press, Washington, DC, pp. 392–416.

Dyrlov Bendtsen J., H. Nielsen, G. Von Heijne and S. Brunak, (2004). Improved prediction of signal peptides: signalP 3.0, *J. Mol. Biol.* **340**:783–795.

Franke D, Lorbach V, Esser S, Dose C, Sprenger GA, Halfar M, Thommes J, Muller R, Takors R, Muller M. (2003). (S,S)-2,3-Dihydroxy-2,3-dihydrobenzoic acid: microbial access with engineered cells of *Escherichia coli* and application as starting material in natural-product synthesis. *Chem.* **9**: 4188-4196.

Furutani; Y., Shimada H., Mita I,; Manabe K., Honjo M., Tomioka N., (1997). DNA base sequence containing regions involved in the production and secretion of a protein, recombinant DNA including the whole or a part of the DNA base sequence, and method of producing proteins by use of the recombinant DNA. United States Patent No. 5 612 192.

He Z, Stigers Lavoie KD, Bartlett PA, Toney MD. (2004). Conservation of mechanism in three chorismate-utilizing enzymes. *J. American Chem. Soc.* **126**: 2378-2385.

Hengen P.N., (1995). Vectorette, splinkerette and boomerang DNA amplification, *Trends Biochem. Sci.* **20**: 372–373.

Isegawa Y., J. Sheng, Y. Sokawa, K. Yamanishi, O. Nakagomi and S. Ueda, (1992). Selective amplification of cDNA sequence from total RNA by cassette-ligation mediated polymerase chain reaction (PCR): application to sequencing 6.5 kb genome segment of hantavirus strain B-1. *Mol. Cell. Probes* **6**: 467–475.

Kilstrup M. and K.N. Kristiansen, (2000). Rapid genome walking: a simplified oligo-cassette mediated polymerase chain reaction using a single genome-specific primer. *Nucleic Acids Res.* **28**: E55.

Kobayashi G, Fujii K, Serizawa M, Yamamoto H, Sekiguchi J. (2002). Simultaneous display of bacterial and fungal lipases on the cell surface of *Bacillus subtilis*. *J. Biosci. Bioeng.* **93**:15-19.

Mishra R.N., S.L. Singla-Pareek, S. Nair, S.K. Sopory and M.K. Reddy, (2002). Directional genome walking using PCR. *Biotechniq.* **33**: 830–834.

Mountain, A., (1989). Gene expression system for *Bacillus subtilis*. In: Harwood, C.R. (Ed.), Biotechnology Handbooks, vol. 2: *Bacillus*. Plenum Press, New York, pp. 73-114.

Moore J.C. and F.H. Arnold, (1996). Directed evolution of a *para*-nitrobenzyl esterase for aqueous-organic solvents. *Nat. Biotech.* **14**: 458–467.

Nthangeni M.B., H. Patterton, A. van Tonder, W.P. Vergeer and D. Litthauer, (2001). Over-expression and properties of a purified recombinant *Bacillus licheniformis* lipase: a comparative report on *Bacillus* lipases. *Enzyme Microbiol. Technol.* **28**: 705–712.

Nthangeni M.B., Ramagoma F, Tlou M.G, Litthauer D. (2005). Development of a versatile cassette for directional genome walking using cassette ligation-mediated PCR and its application in the cloning of complete lipolytic genes from *Bacillus* species. *J. Microbiol. Methods.* **61**:225-234.

Oliver T, Schmidt B, Nathan D, Clemens R, Maskell D. (2005). Using reconfigurable hardware to accelerate multiple sequence alignment with ClustalW. *Bioinformatics.* **21**:3431-3432.

Padegimas L.S. and N.A. Reichert, (1998). Adaptor ligation-based polymerase chain reaction-mediated walking. *Analytic. Biochem.* **260**: 149–153.

Reddy M.K., S. Nair and S.K. Sopory, (2002). Global amplification of cDNA from limiting amounts of tissue. An improved method for gene cloning and analysis. *Mol. Biotech.* **22**: 223–230.

Rishi A.S., N.D. Nelson and A. Goyal, (2004). Genome walking of large fragments: an improved method. *J. Biotechnol.* **111**: 9–15.

Rosenthal A. and D.S. Jones, (1990). Genomic walking and sequencing by oligo-cassette mediated polymerase chain reaction. *Nucleic Acids Res.* **18**: 3095–3096.

Rosenthal A., (1992) PCR amplification techniques for chromosome walking. *Trends Biotech.l* **10**: 44–48.

Sambrook J., E.F. Fritsch and T. Maniatis, (1989). Molecular cloning: a laboratory manual, Cold Spring Harbor Laboratory Press, New York.

Shyamala V. and G.F. Ames, (1993). Genome walking by single specific primer–polymerase chain reaction. *Methods Enzymol.* **217**: 436-446.

Siebert P.D., A. Chenchik, D.E. Kellogg, K.A. Lukyanov and S.A. Lukyanov, (1995). An improved PCR method for walking in uncloned genomic DNA. *Nucleic Acids Res.* **23**: 1087-1088.

Simonen, M. and Palva, I., (1993). Protein secretion in *Bacillus* species. *Microbiol. Rev.* **57**:109-137.

Thompson J.D., T.J. Gibson, F. Plewniak, F. Jeanmougin and D.G. Higgins, (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**: 4876-4882.

Tjalsma H., M.A. Noback, S. Bron, G. Venema, K. Yamane and J.M. van Dijl (1997). *Bacillus subtilis* contains four closely related type I signal peptidases with overlapping substrate specificities. Constitutive and temporally controlled expression of different *sip* genes. *J. Biol. Chem.* **272**: 25983-25992.

Wilson C., Ron T.M.R., Christianson T., Maurer K-H, Weiss A. (1996). Expression systems for commercial production of cellulase and xylanase in *Bacillus subtilis* and *Bacillus licheniformis*. United States Patent No. 5 888 800.

Yoshitomo-Nakagawa K., M. Muramatsu and S. Sugano, (1997). Cloning of the promoter regions of mouse TGF-beta receptor genes by inverse PCR with highly overlapped primers, *DNA Research* **4**:73-75.

Zhang Z. and S.J. Gurr, (2000). Walking into the unknown: a 'step down' PCR-based technique leading to the direct sequence analysis of flanking genomic DNA. *Gene* **253**:145-150.

Arnold C, Hodgson IJ. (1991). Vectorette PCR: a novel approach to genomic walking. *PCR Methods Appl.* **1**:39-42.

Kilstrup M and Kristiansen K. N. (2000). Rapid genome walking: a simplified oligo-cassette mediated polymerase chain reaction using a single genome-specific primer. *Nucleic Acids Res.* **11**: e55.

Harrison R.W., Miller, J.C., D'Souza, M.J. and Kampo, G. (1997). Easy gene walking. *Biotechniques*. **22**:650-653.

Ochman H, Gerber AS, Hartl DL. (1988). Genetic applications of an inverse polymerase chain reaction. *Gen* **3**:621-623.

CHAPTER 3

The development of a *Bacillus licheniformis* lipase-gene based expression-secretion system.

3.1. Introduction

Bacillus species are notable producers and secretors of extracellular proteins. Due to the GRAS status of some species, some members of this genus (e.g. *Bacillus subtilis*, *Bacillus licheniformis*, *Bacillus amyloliquifaciens*) have been used for large-scale production of heterologous proteins using the regulatory elements driving the extracellular protein production. Examples of extracellular proteins whose regulatory elements have been used for heterologous production of proteins include the alpha-amylase (Furutani *et al*., 1977), serine protease (Sloma *et al*., 1992), levansucrase (Wilson *et al*., 1996) from *Bacillus* species. The selection of these transcription signals was based on the level of expression of the respective endogenous genes within the host. The extracellular proteins are synthesized in *Bacillus* species as precursor proteins with signal peptides that translocate proteins to the extracellular medium (Tjalsma *et al*., 1997). The signal peptides are cleaved by membrane bound signal peptidases and the mature proteins are released to the extracellular. The signal peptide regions of extracellular proteins are attractive tools for the development of protein secretion systems were the heterologous protein is cloned just downstream of the signal peptide with the hope that the cloned gene would be expressed and efficiently translocated to the extracelluar medium to simplify the downstream processing of the expressed protein.

A fragment encoding an extracellular mature lipase gene from *Bacillus licheniformis* has been cloned previously (Nthangeni *et al*., 2001). Physiological studies revealed that the gene is controlled by a growth-dependent promoter, which has pronounced transcriptional activity as cells enter the stationary phase of growth. This phenomenon has also been observed with many other extracellular proteins from *Bacillus* species, and it has been suggested that increased expression during the stationary phase of growth is a result of preferred nutrient exhaustion in the media and the cells respond by releasing extracellular proteins such as amylases, proteases and lipases "to search" for alternative energy sources. The extracellular lipase from *Bacillus licheniformis* is produced to high levels as judged by lipase activity assays during stationary phase growth on nutrient broth supplemented with Tween 80.

The objective of this study was to evaluate the extracellular lipase promoter and signal peptide regions of *Bacillus licheniformis* as tools for production of homologous and heterologous proteins. To achieve this, an *Escherichia coli*/*Bacillus* shuttle vector comprising of replicative elements from pUB110 and pUC18 plasmids was constructed. The mature lipase from *Bacillus lichenifomis*, the carboxylesterase from *Bacillus pumilus* and DNA polymerase from *Thermus aquaticus* were used as model enzymes in the evaluation of the efficiency of expression and secretion of endogenous and heterologous proteins.

3.2. Materials and Methods

3.2.1 Bacterial strains, culture conditions and DNA sources and isolations

Escherichia coli JM109 (Promega, Madison, USA), *Bacillus licheniformis* MBB01 and *Bacillus pumilus* MBB02 (the MBB strains are obtainable from the University of the Free State Microbial Cultures Collection) were used. The pUC18 and pUB110 plasmids were purchased from commercial sources while pMBE-082 is a derivative of pUC18 and contain

additional restriction enzyme recognition sequences within the multiple cloning site (Figure 3.1)**.** The bacterial strains were grown in Luria–Bertani (LB) [10g tryptone, 5g yeast extract, 10g NaCl, per liter] media at 37°C with shaking in 5-, 50- or 250-ml shake flasks. TLB media contained sonicated emulsion of 10 ml tributyrin and 10 g gum arabic in 1L of LB medium. When solid growth media was required, agar (15 g/l) was added and the incubation was done in a 37°C incubator. The media was supplemented with ampicillin (100 µg/ml) or kanamycin (10 µg/ml) for the selection of plasmid-carrying strains in *Escherichia coli* and *Bacillus licheniformis*, respectively.

Genomic DNAs from *Escherichia coli* and *Bacillus species* were isolated according to the method described by Shyamala and Ames (1993), which comprises of suspending of pelleted cells in buffer containing lysozyme and proteolytic enzymes followed by several extractions with phenol/chloroform/isoamyl organic solvents and precipitation with ethanol. The purified DNA was re-suspended in sterile distilled water. Isolation of plasmids from *Bacillus licheniformis* was done using the GFX Plasmid purification kit after initial treatment of the cells with lysozyme. Plasmids in *Escherichia coli* were isolated using the lysozyme/boiling procedure, which briefly involved inoculation of a single colony in 5 ml LB media containing ampicillin. The culture was grown at 37°C overnight. Cells were harvested by centrifugation at maximum speed (14000 rpms) followed by re-suspension in STET (8% sucrose w/v; 0.1% triton X-100; 50 mM Tris-HCl, pH 8.0; 50 mM EDTA) buffer. Lysozyme was added to a final concentration of 10 mg/ml, incubated at 37°C for approximately 20 minutes and placed in a 100^{o}C heating block for 1 minute and quickly immersed on ice for 5 minutes. The supernatant was separated from cell debris by centrifugation at maximum speed for 10 minutes. The supernatant was transferred into a new tube and plasmid DNA was precipitated

using either isopropanol or ethanol. The precipitated DNA was re-suspended in 5 mM Tris-HCl buffer, pH 8.5.

3.2.2 Transformation

Competent cells of *Escherichia coli* were prepared by the Rubidium chloride / Calcium chloride method. Briefly, the method consisted of the following steps: a single colony was used to inoculate 5 ml LB to prepare a preinoculum. The culture was incubated at 37˚C with shaking overnight. One ml of the overnight culture was inoculated into 100 ml Psi broth (5 g yeast extract, 20 g tryptone and 5 g magnesium sulphate per litre, pH 7.6) and the culture was grown to an OD_{550} of 0.48. The culture was put on ice for 15 minutes and cells harvested by centrifugation at 5000 rpm for 5 minutes. The cell pellet was re-suspended in TFBI (0.58 g Potassium acetate; 2.42 g rubidium chloride; 0.294 g calcium chloride; 2.0 g manganese chloride; 30 ml glycerol per 200 ml; pH 5.8) buffer and put on ice for 15 minutes. The suspension was again centrifuged and the cell pellet was resuspended in TFBII (0.21 g MOPS; 1.1 g calcium chloride; 0.121 g rubidium chloride; 15 ml glycerol per 100 ml; pH 6.5) buffer. Aliquots of the competent cells were prepared and stored at -70°C.

Bacillus licheniformis transformation was performed according to the method of Xue *et al*., 1999. An overnight culture (1 ml) of *Bacillus licheniformis* was diluted 20-fold in growth media (LB containing 0.5 M sorbitol). The culture was grown at 37 ºC until OD_{600} of 0.85-0.95. The cells were cooled on ice for 10 minutes, harvested by centrifugation at 5000rpm, for 5 minutes at 4ºC and washed four times with ice-cold EP buffer (0.5 M sorbitol, 0.5 mannitol and 10 % glycerol). The cell pellet was suspended in 1/40 of the culture volume of the EP buffer. The cells were either used immediately or stored at -80ºC. Transformation was done by an electroporation technique using a Gene pulser (Bio rad), where 60 µl of the

competent cells was mixed with plasmid DNA and then transferred to an electroporation cuvette (1mm gap), under the following settings; 2.0 kV voltage, 25μF capacitance, 200 ohms resistance, 21 kV/cm field strength and 4.5-5.0 of pulse length. The transformed cells were plated on LB plates supplemented with kanamycin and incubated at 37°C overnight.

3.2.3 Recombinant DNA techniques

Recombinant DNA techniques using commercially available molecular biology grade enzymes and chemical reagents were essentially followed as described by Sambrook *et al*. (1989) or manufacturers of the restriction and DNA modifying enzymes. Oligonucleotide primers were purchased from Integrated DNA Technologies (USA) and are written in the .5'-3' orientation, unless stated otherwise. Oligonucleotide cassettes were prepared by annealing the complementary oligonucleotides at equimolar concentrations in a 200 ml boiling water, and the temperature of the water was allowed to drop gradually to room temperature followed by brief centrifugation of the mixture and incubation on ice for 15 minutes. The PCR techniques were done using Expand DNA polymerase, (Roche, Germany) under the following general conditions. 94 ^{o}C for 2 min, 30 cycles (94 ^{o}C for 30 sec, 55 ^{o}C for 30 sec and 72 ^{o}C for 2-4 min). Agarose gel electrophoresis was done using 1% agarose and the DNA fragments were purified using the GFX PCR and Gel band purification kit (Amersham, UK). DNA sequencing was done at Inqaba Biotec, Pretoria, South Africa).

3.2.4 Construction of an *Escherichia coli*/*Bacillus* shuttle vector

The pUB110 and pUC18 plasmids were digested with *Eco*RI and *Nde*I. The digested mixtures were run on 1.5% agarose gel, and their respective large plasmid backbones were excised from agarose gel, cleaned, ligated together and transformed into *Escherichia coli* JM109 cells.

Transformants were selected on LB agar plates containing ampicillin, followed by plasmid isolation and restriction analysis. The correct plasmid was denoted pSV1 (Figure 3.1).

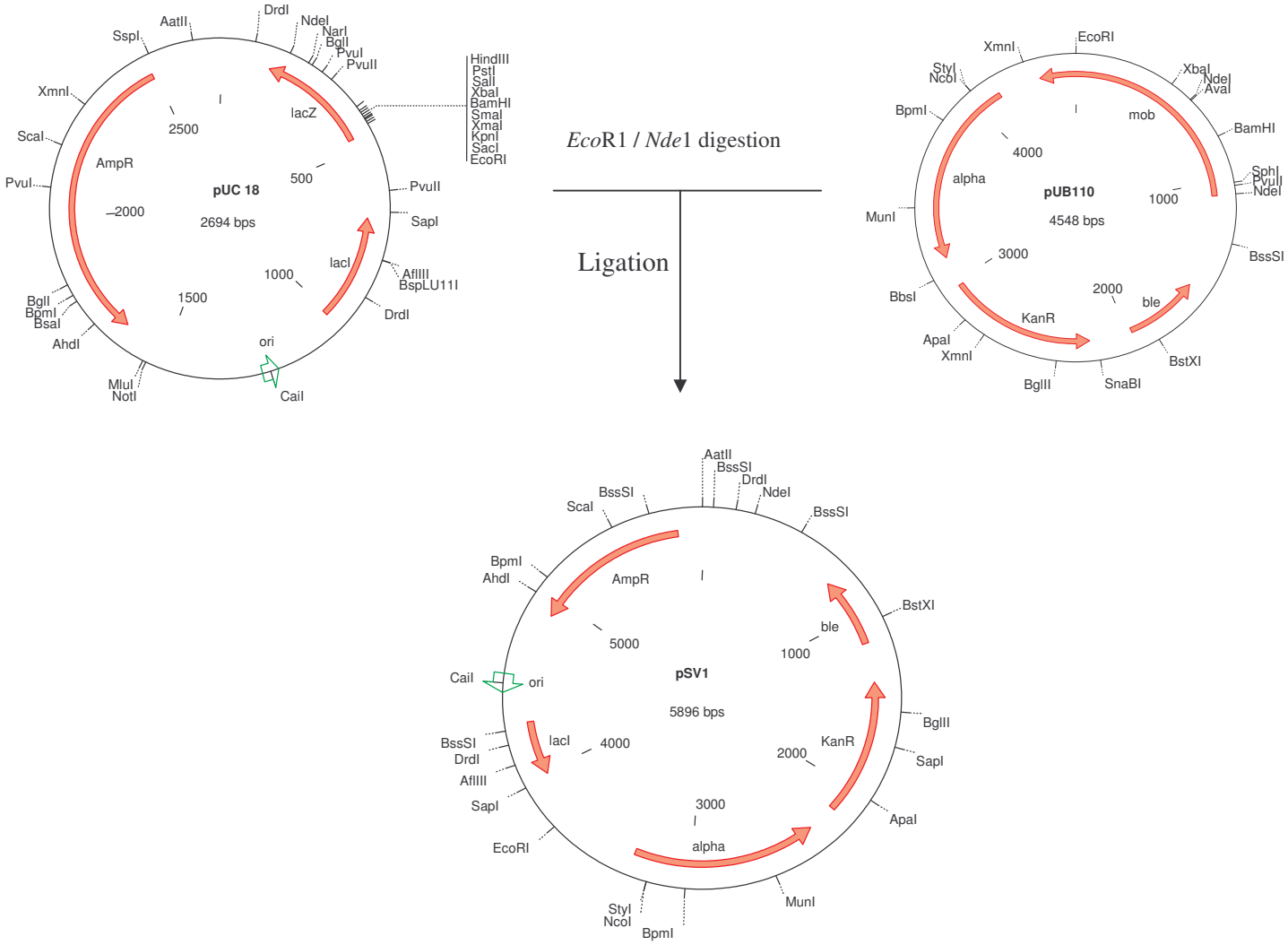

Figure 3.1. A schematic representation illustrating the mechanism used to construct pSV1. pUC18 and pUB110 were digested with *Eco*RI and *Nde*I and their respective large backbone fragments were ligated.

3.2.5 Introduction of the multiple cloning site, the 6XHis tag and the transcription terminator

Primers pSVE (GATTACACTAGTCTTAAGGAACGTACAGACGGC) and pSVF (GTATTGGGCTCTAGACCGCTTCCTCGCTCACTG) respectively corresponding to A and B on Figure 3.3 were used to amplify the pSV1 plasmid. The PCR product was digested with *Spe*1 and *Xba*1 contained within the respective primers (underlined sequences) and self-ligated to create a plasmid denoted pSV2 (i) that lacked the lacI and *Eco*R1 site of the pSV2 plasmid. Another set of primers designated pSVA (GGAGGGTTTTGCTAGCGTGCTCGT and pSVB (GATGCCTCACATTTGTGCCACCTA) respectively corresponding to D and C on Figure 3.3 were used to amplify pSV2 (i) plasmid to yield a PCR product which was digested with *Nhe*1 and *Nde*1 restriction enzymes. The *Nhe*1 enzyme site was introduced as part of the pSVA primer (underlined sequence) while the *Nde*1 restriction site was amplified as part of the template plasmid. The digested DNA fragment was ligated to a linker-cassette constructed by annealing two complementary oligonucleotides (Figure 3.3). The annealed cassette has sticky ends complementary to the *Nde*1 and Nhe1 restricted DNA, although upon ligation, the *Nde*1 site is not regenerated. The cassette contained an internal *Sfi*1 site which became introduced to the pSV2 (iii).

The multiple cloning site was introduced into the shuttle vector firstly by subcloning into the pMBE082 multiple cloning site (Figure 3.4 (A)] an annealed cassette (Figure 3.4B) which contained a sequence encoding 6X His tag. The pMBE082 plasmid was digested with *Xho*1 and *Nhe*1 followed by ligation with the cassette to create pMBE082(ii) plasmid. Primer pair RrnBT1T2F, and RrnBT1T2R specifying the amplification of the rrnB terminator from the wild type *Escherichia coli* strain was used to amplify the T1T2 terminator region followed by

restriction digestion with *Mlu*1 and *Nhe*1. The digested PCR fragment was ligated into pMBE082(ii) digested with similar enzymes. The ligation mixture was transformed into *Escherichia coli* JM 109 cells followed by selection on LB ampicillin agar plates, plasmid isolation and restriction analysis. The correct plasmid was denoted pMBE082 (iii). The pMBE082 (iii) plasmid was digested with *Sfi*1 and *Nhe*1 to release the DNA fragment containing the multiple cloning site, the 6XHis-encoding fragment and the T1T2 terminator region. This fragment was ligated into pSV2 (iii) to create a shuttle vector denoted pSV4 (Figure 3.5A).

3.2.6 Introduction of the promoter and signal peptide sequences into the shuttle vector

Primer pair Bllip183F and BLlip689R2 (Figure 3.2) were used to amplify the promoter and signal peptide encoding sequence from *Bacillus licheniformis* genomic DNA. The PCR product was digested with *Sfi*1 and *Eco*R1 followed by ligation into pSV4 digested with similar restriction enzymes to create pSV6 expression plasmid (Figure 3.5B)

Figure 3.2: Nucleotide sequence of the promoter and signal peptide encoding region of *Bacillus licheniformis* lipase gene that was subcloned into the pSV4 plasmid. The primers used to amplify the fragment (Bllip183F and (BLlip689R2, respectively corresponding to the Forward and reverse primers) have been aligned with their template. The restriction enzymes that have been introduced within the primers are depicted close to their corresponding sequences. The signal sequence-encoding region has been translated into amino acid sequence with the GTG encoding the putative initial Met underlined and bolded. An arrow indicates the predicted cleavage site of the signal peptide by the signal peptidase after translocation. Transcriptional signal sequences are in bold and underlined and indicated as –35 or –10.

(i)

*Sfi*I

5′ -TATTTCCGGCAACCACGGCCAAGTCGGCCATAACACG -3′

3′-AAAGGCCGTTGGTGCCGGTTCAGCCGGTATTGTGCGATC -5′

(ii)

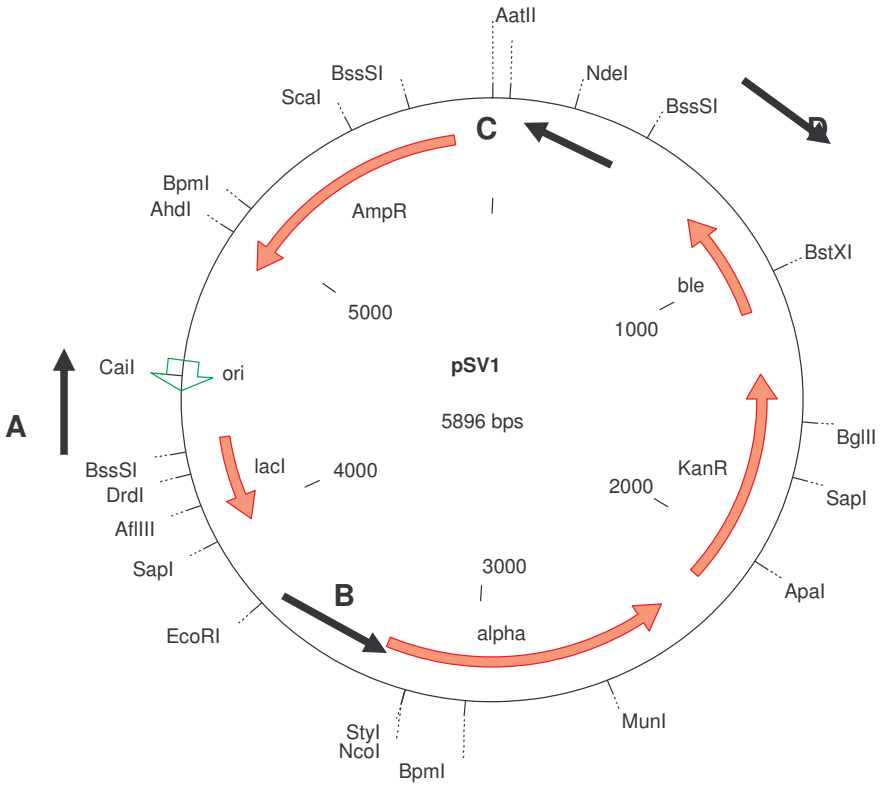

Figure 3.3. Construction of pSV2 was done by ligating a cassette containing a *Sfi*I restriction site and sticky ends for *Nde*I and *Nhe*I at the 5′ and 3′ ends respectively. The *Nde*I site is not regenerated during ligation. (A, B) and (C, D) are representing the oligonucleotides used to eliminate the region inclusively containing the *Eco*RI site and the lac I region on pSVI and the primers used in the construction of pSV2 by insertion of the cassette respectively. *Nde*I and *Nhe*I were used to cut pSVI and ligated with pSVI to introduce the *Sfi*I site thereby generating pSV2.

A

B

C

Figure 3.4. (A) The multiple cloning site as contained within the pMBE082 plasmid, a derivative of pUC18. (B) The oligocassette containing the sequence encoding 6X His, the stop codon indicated by the ***, and the sticky ends compatible with *Xho*1 and *Nhe*1 digested DNA molecules. (C) The terminator rrnB transcription terminator region (T1T2) sequence of *Escherichia coli* together with the primers (RrnBT1T2F, and RrnBT1T2R2) used to amplify the PCR product shown with the introduced restriction enzymes sites.

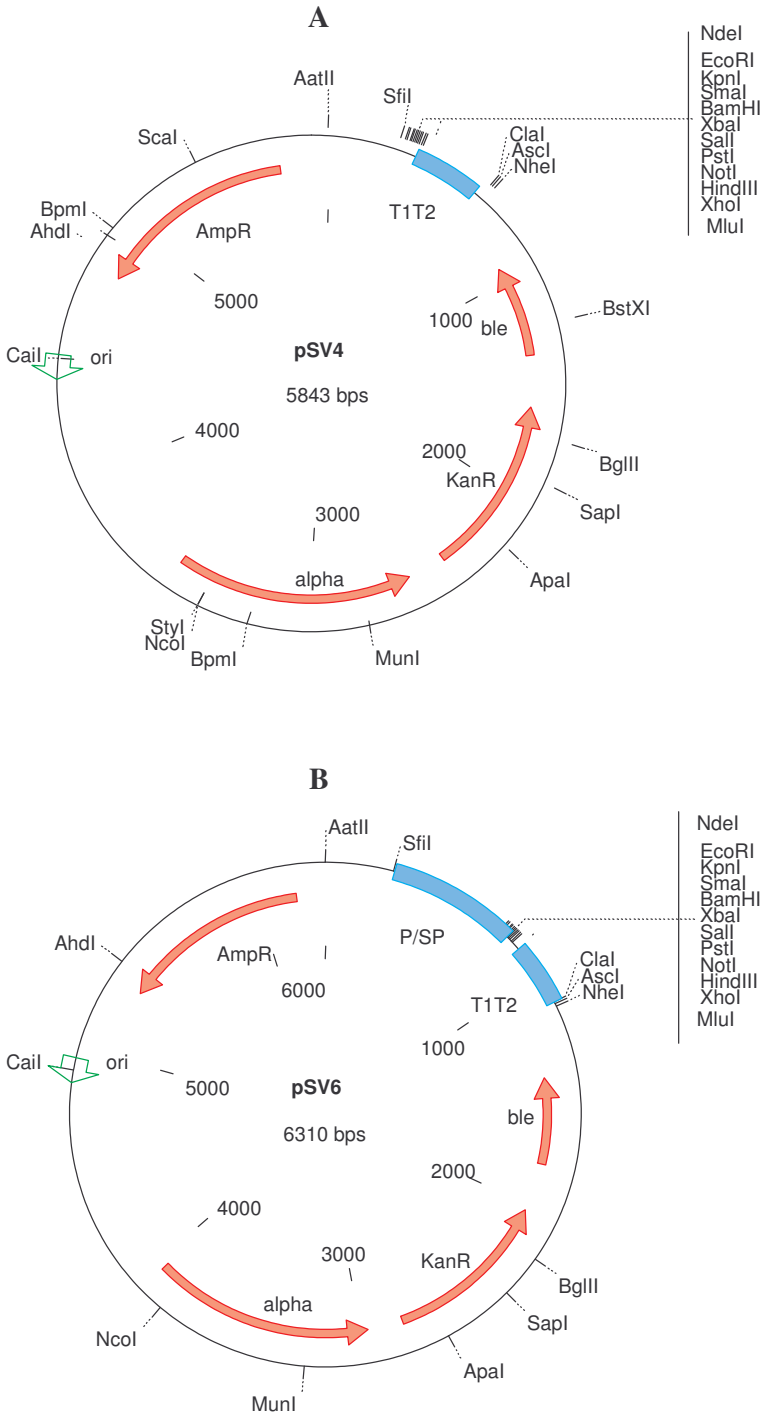

Figure 3.5. Schematic diagrams illustrating pSV4 and pSV6 plasmids. (A) The pSV4 plasmid, which was constructed by subcloning the 6Xhis fragment and the rrnBT1T2 terminator into pSV2 as described in section 3.2.5. (B) The pSV6 shuttle vector was constructed by subcloning the promoter and signal sequence into pSV4 by digestion with *Sfi*I and *Eco*RI.

3.2.7. Sub-cloning of the mature lipase gene into pSV6

The mature lipase gene from *Bacillus licheniformis* was rescued from pET-Blich-Lip (Nthangeni *et al*., 2001) by digestion with *NdeI* and *XhoI*. The 560 bp released fragment was excised from the gel, purified and ligated into pSV6 digested with similar enzymes. The correct construct was confirmed by transformation of *Escherichia coli* followed by plasmid isolation and restriction analysis, and also by patching transformed independent clones of *Escherichia coli* clones on TLB agar plates and incubation at 37 ^{o}C to observe zones of clearance around the growing colonies. The correct plasmid was designated pSV6-lip (Figure 3.6).

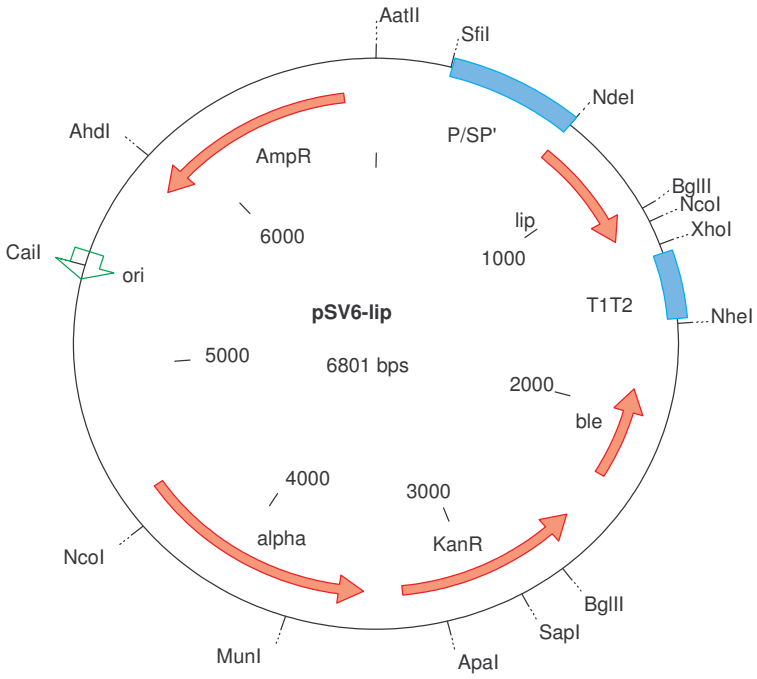

Figure 3.6. A schematic representation of pSV6-lip, which was constructed by subloning the mature lipase gene from *Bacillus licheniformis* into pSV6 downstream of the promoter and signal peptide following digestion with *Nde*I and *Xho*I.

3.2.8 Sub-cloning of the *Bacillus pumilus* carboxylesterase gene.

A set of two primers, Cest1F (GTGAGAGAACCATATGACTCATCAAATAGTAACG ACTC) and Cest1R (CATATGTCGACTTCTCCTTTTGAAGGGAATAGCTTCTG) respectively corresponding to the forward and reverse primers, were designed based on the sequence of the carboxylesterase gene from *Bacillus pumilus* MBB02 (accession number AY692083). The underlined sequences within the primers represent the *Nde*1 (forward primer) and the *Sal*1 (reverse primer). The PCR product was subjected to complete digestion with *Sal*1 and partially digested with *Nde*1. The fragment of about 1500 bp was excised from agarose gel and subcloned into pSV6 digested with *Nde*1 and *Xho*1 to create pSV-6-Cest expression plasmid (Figure 3.7). The correct plasmid was confirmed by restriction analysis and plating of the transformed *Escherichia coli* clone on TLB agar plate for observation of zones of clearance.

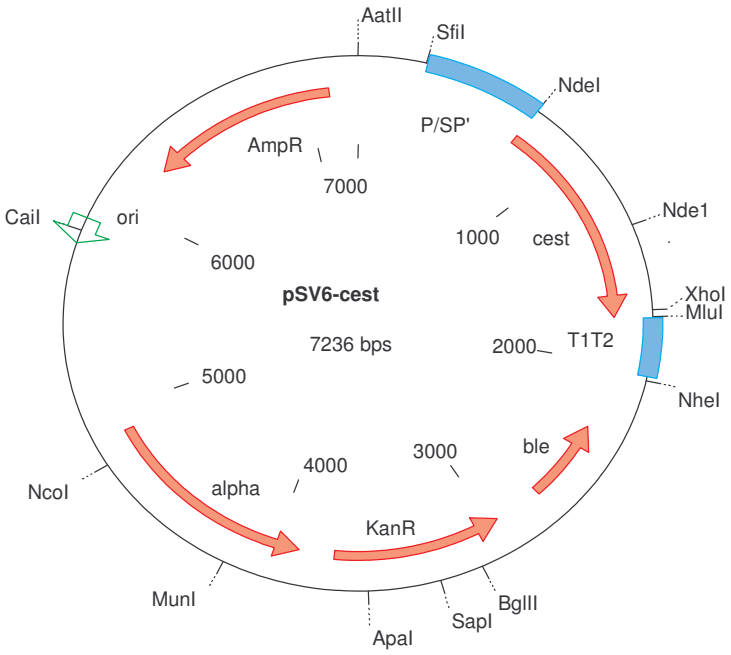

Figure 3.7. A diagram illustrating the construction of the pSV6-carboxyesterase recombinant which was done by subcloning the carboxyesterase gene from *Bacillus pumilus* into pSV6 with *Nde*I and *SalI*.

3.2.9 Sub-cloning of *Thermus aquaticus* (Taq) DNA polymerase gene

The forward Taq1F (GATATACATATGAGGGGCATGCTGCCCCTCTTTGAGC) and reverse Taq1R (GTTTGTCGACCTCCTTGGCGGAGAGCCAGTCCTCGCC) primers were constructed based on the sequence of the *Taq* DNA polymerase gene as submitted in nucleotide databases (accession number DD170163). The open reading frame encoding this gene was contained in the pGEM-Taq Pol plasmid donated by Michel Labuschagne (CSIR, Modderfontein). The plasmid was used as the template in the PCR amplification. The PCR product was digested with *Nde*1 and *Sal*1 and ligated into the pSV6 expression plasmid digested with *Nde*1 and *Xho*1 restriction enzymes. Positive constructs were confirmed after selection of transformed *Escherichia coli*, plasmid isolation and restriction analysis. The correct clone was designated pSV6-taq (Figure 3.8).

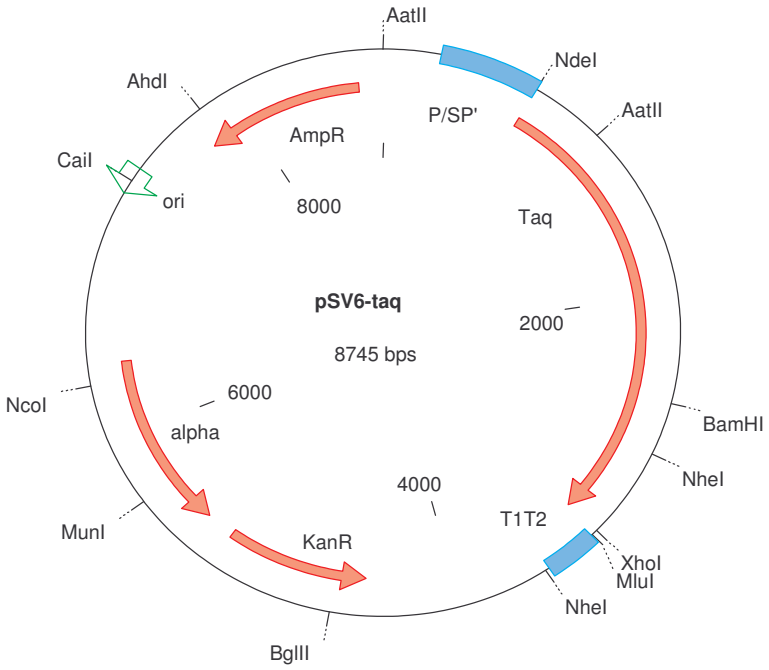

Figure 3.8. A schematic representation of the Taq DNA polymerase gene construct. pSV6 contains the 2500 bp Taq polymerase open reading frame, in frame with the lipase promoter.

3.2.10. Overproduction of lipase, carboxylesterase and *Taq* DNA polymerase by *Bacillus licheniformis*.

The plasmids pSV6, pSV6-lip, pSV6-cest and pSV6-taq were transformed into competent *Bacillus licheniformis* cells as described in Section 3.2.2. Single colonies were used to prepare a 5 ml LB/kanamycin pre-inoculum. The pre-inoculum was harvested by centrifugation as the cells entered the exponential phase of growth (OD_{600} of ~0.4) and used to inoculate 50 ml nutrient broth (2 g/l) containing Tween 80 (0.1 g/l). The cultures were grown on 500 ml conical flasks on shaker incubators (200 rpm) at 30 °C for 36 hours. The cultures were harvested by centrifugation and 50 µl of the supernatant was placed in a speed vacuum dryer to dry and the dried samples were re-suspended in 15 µl loading dye for the SDS-PAGE gel. SDS-PAGE gels were stained with Coomasie-brilliant blue stain to detect the protein bands.

3.2.11. Protein purification

The pH values of the supernatants obtained in Section 3.2.10 were adjusted to 7.5 and imidazole was added to a final concentration of 10 mM followed by mixing with 5 ml of the 50% Ni^{2+}-chelated nitroacetic acid (Ni-NTA) resin slurry (Qiagen, CA USA). The crude protein extract-NTA mixture was loaded into a chromatographic column and washed with 50 ml of 50 mM sodium phosphate buffer, pH 7.5. Elution of the bound proteins was effected with sodium phosphate buffer containing 250 ml imidazole. The resultant solutions were loaded on SDS-PAGE.

3.2.12. Assays for Taq DNA polymerase

Assays for Taq DNA polymerase activity were performed using the supernatant fractions of the cultures or the partially purified protein sample. Aliquots were used to perform PCR by initial dilution of the solutions according to the method described by Desai and Pfaffle, 1995. The oligonucleotides used to amplify the region inclusively containing the promoter and signal peptide regions (section 3.2.6) within the genomic DNA from *Bacillus licheniformis* were used to assay the activity of Taq DNA polymerase. The PCR was performed to amplify a 557 bp fragment from the genomic DNA of *Bacillus licheniformis* under the following parameters: 1 initial denaturation at 94°C step for 2 minutes; 30 cycles at 94°C for 30 seconds, annealing at 58°C for 30 seconds, extension at 72°C for 1 minute, except for the final cycle where extension proceeded for 10 minutes.

3.2.13. Assay for lipolytic activity

Lipase activity was measured spectrophotometrically using supernatant fractions (Section.3.2.10) with *p*-nitrophenyl palmitate (pNPP) as the substrate according to the method of Winkler and Stuckman (1979). Proteins samples (50 µL) were added to the substrate mixture containing phosphate buffer (50 mm, pH 8.5), 0.2% (w/v) sodium deoxycholate and 0.1% (w/v) gum arabic *p*NPP (0.30 mM final concentration). Lipase activity was determined by the rate of *p*-nitrophenol production (*p*NP), measured at 410 nm in a model DU600 spectrophotometer (Beckman Coulter, Fullerton, CA) at 37^{o}C. Lipolytic activity was determined, using enzyme free substrate blanks as control. One unit of enzyme activity was defined as the amount of enzyme forming 1 µmol of *p*NP min^{-1} under the mentioned conditions.

3.2.14. Assay for carboxylesterase activity

Esterase activity was determined by measuring the amount of *p*-nitrophenol released by esterase-catalyzed hydrolysis as described (Rhee *et al.*, 2005). The production of *p*-nitrophenol was continuously monitored at 405 nm by use of a DU-600 spectrophotometer with a thermal controller (Beckman) at 37°C**,** with 1 mM *p*-nitrophenyl caproate as a substrate in 50 mM morpholineethanesulfonic acid (MES; pH 6.0) containing 1% acetonitrile. The reaction was started by the addition of the enzyme solution (50 µl). Blank reactions were performed with every measurement using the substrate mixture without the enzyme solution. The activity was determined by measuring the rate of hydrolysis of *p*-nitrophenyl caproate**.** One unit of enzyme activity was defined as the amount of activity required to release 1 µmol of *p*-nitrophenol/min from the *p*-nitrophenyl ester.

3.2.15 Electrophoresis

SDS-PAGE was performed on 12.5% running gels as described by Laemli (1970) and resolved proteins visualized by Coomassie staining following standard procedures. A broad range protein standard (Fermentas) was used as molecular mass markers.

3.3 Results

3.3.1 Construction of an expression-secretion vector

A shuttle vector capable of replicating and propagating in both *Escherichia coli* and *Bacillus licheniformis* was constructed based on pUC18 and pUB110 replicative and selective elements. A multiple cloning site derived from pMBE082 was introduced with the sequence encoding 6X His tag incorporated downstream to the multiple cloning site. A 270 bp fragment (Figure 3.9A) corresponding to the terminator region of the rrnBT1T2 terminator was amplified from the genomic DNA of *Escherichia coli*. A DNA fragment of 556 bp (Figure 3.9B) was amplified from the genomic DNA of *Bacillus licheniformis* MBB01 and consisted of the promoter and a sequence encoding the signal peptide of the extracellular lipase. The DNA fragment was subcloned to construct an expression vector denoted pSV6 (Figure 3.5), with the multiple cloning site region and sequence as shown on Figure 3.10.

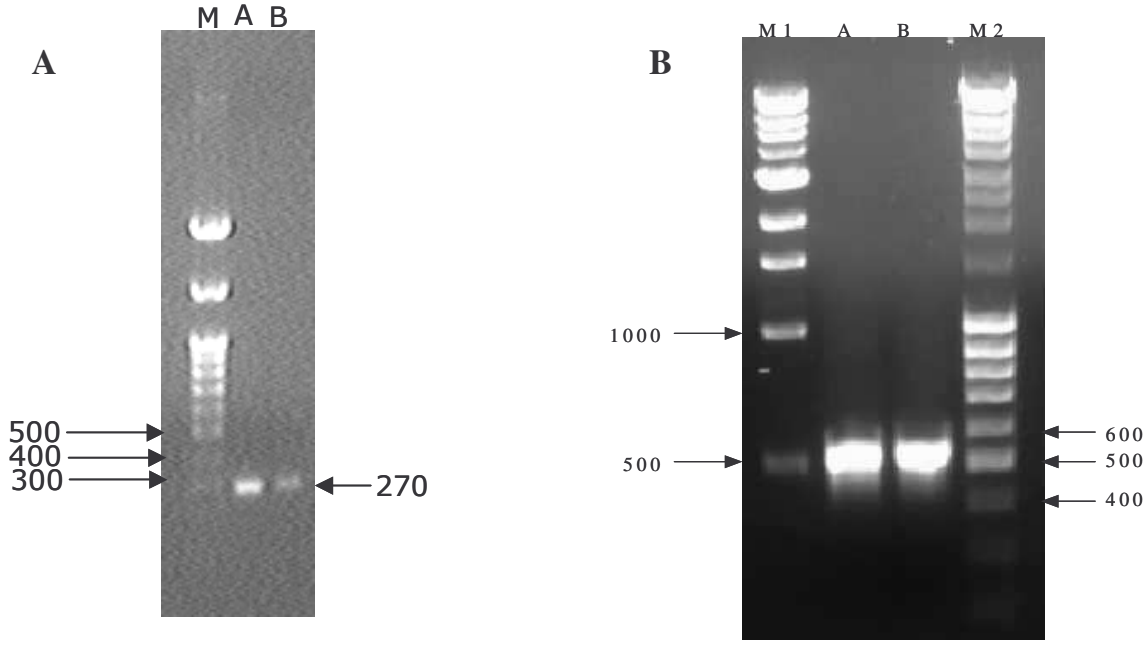

Figure 3.9. (A): Agarose gel electrophoresis showing DNA fragments generated by PCR to amplify (A) the *Escherichia coli* rrnBT1T2 terminator and (B) the promoter and signal sequence from *Bacillus licheniformis.* A and B are the obtained PCR products (in duplicate) while M1 and M2 are standard DNA molecular weight markers in bp.

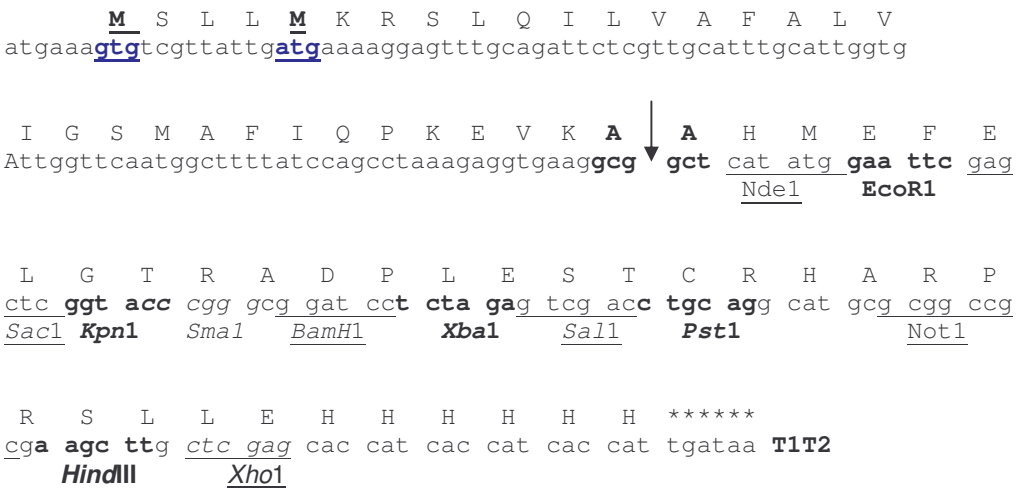

Figure 3.10. The multiple cloning site (MCS) region of the pSV6 *Bacillus* expression plasmid showing unique restriction sites for cloning and the deduced amino acid sequences for in-frame subcloning of genes. The MCS region is flanked by sequences encoding the signal peptide and the 6X His tag. The stop codons are represented by ***. T1T2 represents the position of rrnBT1T2 terminator sequence. The putative initiation Met residues are highlighted within the signal peptide in underlined bolded letter M.

3.3.2 Sub-cloning and functional expression of model genes in *Escherichia coli*

The open reading frames of the carboxylesterase gene from *Bacillus pumilus* MBB02, and DNA polymerase from *Thermus aquaticus* were amplified by PCR from their respective templates to generate products of 1500 and 2500 bp, respectively (Fig 3.11). The *Bacillus pumilus* carboxylesterase and the Taq DNA polymerase PCR products, together with the DNA fragment encoding the mature lipase from *Bacillus licheniformis* MBB01 were subcloned into the pSV6 expression plasmid to construct expression vectors denoted pSV6-cest, (Fig 3.7), pSV6-taq (Fig.3.8) and pSV6-lip, (Fig3.6), respectively. The construction of the correct expression vectors was confirmed by restriction analysis and expected banding patterns [5631 (corresponding to the shuttle vector backbone fragment), 800 bp (promoter/signal peptide region) and the 270 bp fragment (terminator region); Additional unique bands of 980 and 400 bp for the pSV6 cest, 1600 and 460 bp for the pSV6-Taq and 600 and 80 bp for the pSV-lip plasmid] were observed for the plasmids (Figure 3.12). The genes were subcloned in manner that after expression, the corresponding proteins would contain a His tag (6X His) at the C-termini.

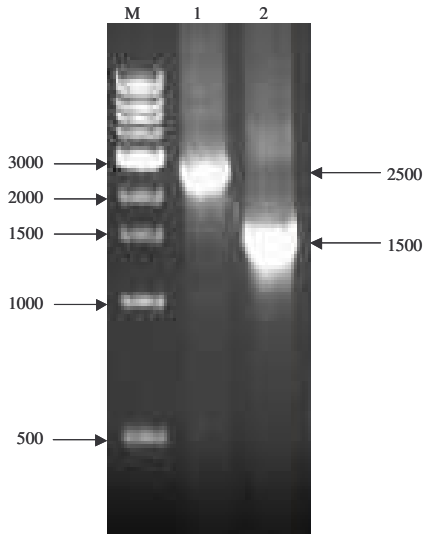

Figure 3.11. Agarose gel electrophoresis of PCR products consisting of *Thermus aquatics* DNA polymerase (lane 2) and *Bacillus pumilus* carboxylesterase genes (lane 2). M is the 1000 bp DNA molecular weight marker.

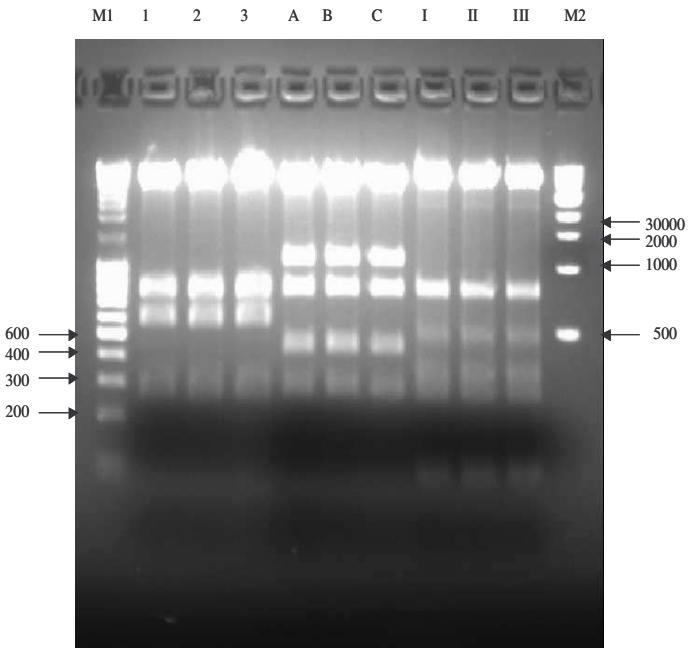

Figure 3.12. Restriction analysis of pSV6-cest, (lanes 1,2,3) pSV6-taq (lanes A, B and C), pSV6-lip (lanes I, II and III). M1 and M2 are standard DNA molecular weight markers in bp. The constructs were digested with a mixture of four restriction enzymes, *Aat*II, *Nde*I, *Mlu*I and *Asc*I. The following DNA fragments were expected and obtained in all the digests, 5631 (corresponding to the shuttle vector backbone fragment), 800 bp (promoter/signal peptide region) and the 270 bp fragment (terminator region). Additional unique bands of 980 and 400 bp for the pSV6 cest, 1600 and 460 bp for the pSV6-Taq and 600 and 80 bp for the pSV-lip plasmid.

Functional expression of the carboxylesterase and lipase genes was firstly tested by patching *Escherichia coli* JM109 cells transformed with pSV6-cest and pSV6-lip on TLB plates. Zones of clearance surrounded pSV6-cest containing colonies after 24 hours of incubation at 37 ^{o}C (Figure 3.13A) and the production of lipolytic activity by *Escherichia coli* clones harbouring the pSV6-lip plasmid could be observed only after extended (at least 48 hours) incubation (Figure 3.13B).

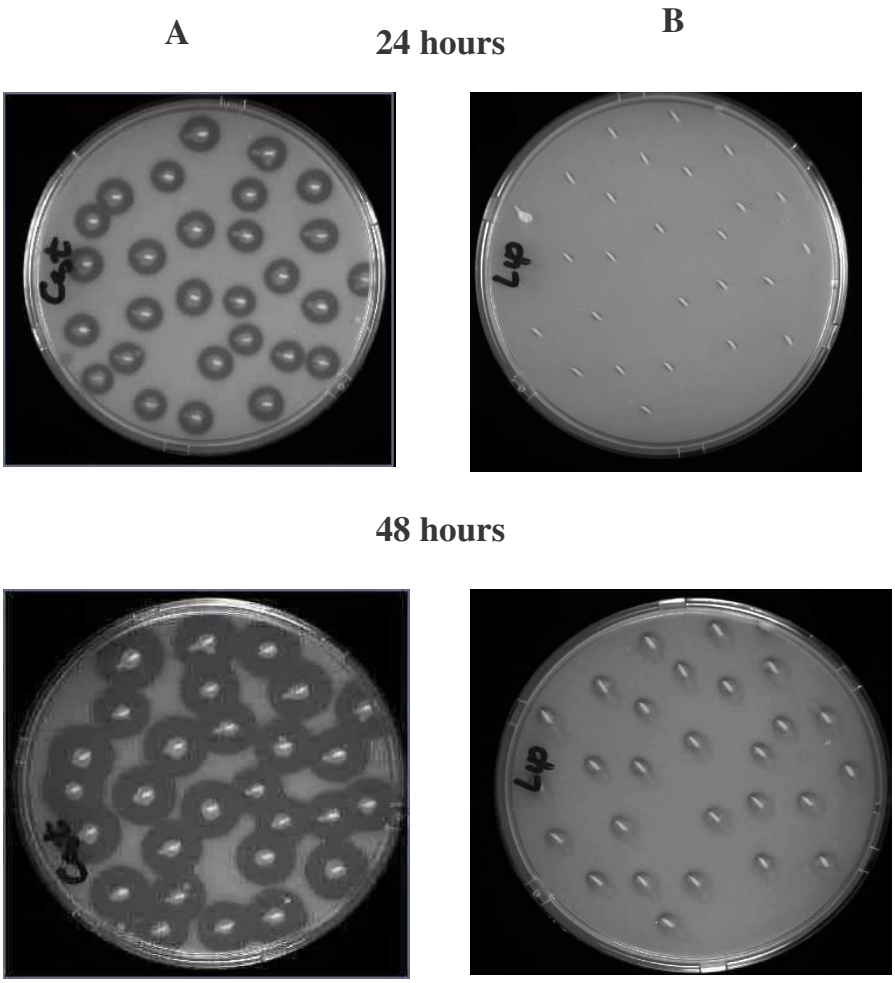

Figure 3.13. A Photograph showing carboxylesterase and lipase production by *Escherichia coli* transformed with pSV6 containing the carboxyesterase and lipase genes. (A) and (B) are carboxylesterase and lipase clones respectively. Pictures were taken after 24and 48 hours of incubation at 37°C. Lipase expression was delayed as clear halos could only be visualized after 48-60 hours of incubation.

3.3.3 Over-expression of model genes in *Bacillus licheniformis* MBB01 and partial purifications of the encoded proteins

The pSV6-cest, pSV6-taq and pSV6-lip plasmids were used to transform *Bacillus licheniformis* MBB01 by the electroporation method with pSV6 as the control. The transformed host was grown on 50 ml nutrient media for 36 hours under conditions that were found to result in high extracellular lipase production. The different supernatants were analysed for proteins of increased intensities at positions corresponding to the sizes of the respective model protein bands. Protein bands of increased intensities at positions of 19-25 kDa, 50-65 kDa and 80-95 kDa were expected from supernatants extracted from *Bacillus licheniformis* cells respectively transformed with pSV6-lip, pSV6-cest and pSV6-taq plasmids. The expected result was obtained only from supernatants of cells transformed with pSV6-cest (Figure3.14). The supernatants were assayed for qualitative presence of their respective activities. The supernatant from cells transformed with pSV6-lip gave activities that were even lower to that of the supernatant from cells transformed with an empty pSV6 plasmid. The assay using the supernatant from cells transformed with pSV6-taq plasmid resulted in successful PCR amplifications, an indication that the sample contained a thermostable DNA polymerase. The assay for the presence of carboxylesterase activity in the supernatant of cells transformed with pSV6-cest could not be done with ease due to the relative instability of *p*-nitrophenyl caproate under conditions of assay. This was further complicated by the presence of endogenous extracellular lipolytic activity of the host strain itself. The supernatants were loaded onto a Nickel affinity chromatography column to capture proteins containing the His tag. The purified samples were analysed on SDS-PAGE (Figure 3.14). Protein bands at expected positions were obtained only with samples that originated from cells transformed with pSV6-cest and pSV6-taq plasmids. The sizes obtained correspond

to the predicted molecular masses of *Bacillus pumilus* carboxylesterase (~56 kDa) and Taq DNA polymerase (~90 kDa) proteins.

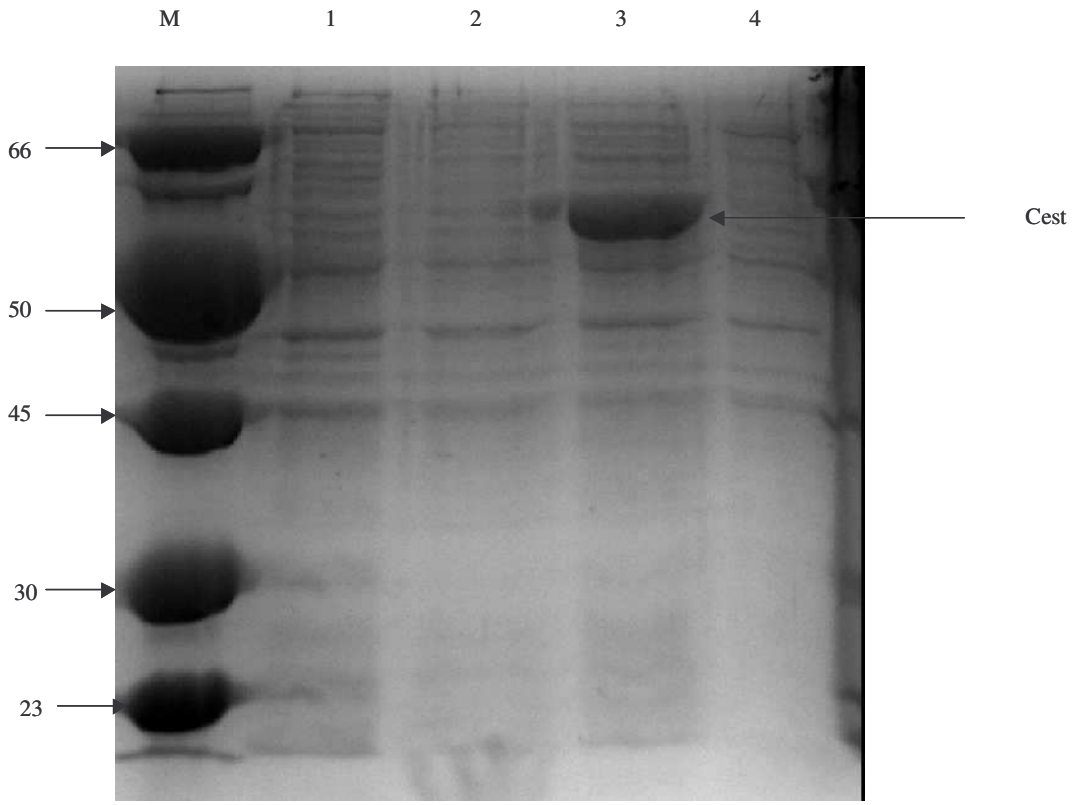

Figure 3.14. Overproduction of *Taq* DNA polymerase, lipase and carboxylesterase. Construction of *Bacillus licheniformis* recombinant strains was done as described. The supernatant fractions of each culture were analysed by SDS-PAGE of a 12.5% gel. Lane M, molecular mass markers in KDa; lane 1, crude extract of the pSV strain (Control); lane 2, crude extract of the pSV6-lip strain; lane 3, crude extract of the pSV6-cest strain; and lane 4, crude extract of the pSV6-taq strain.

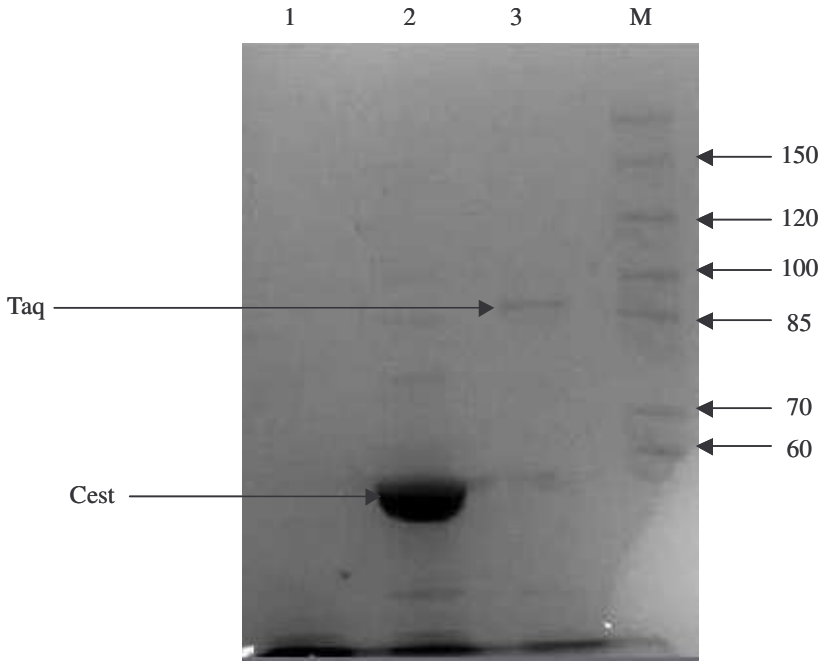

Figure 3.15. Purified recombinant lipase, carboxylesterase and Taq DNA polymerase from *Bacillus licheniformis* were resolved on a 12.5% SDS-PAGE gel. Lanes: 1, lipase; 2, carboxylesterase; 3, Taq DNA polymerase; and 4, molecular weights of standard proteins in kDa.

3.4 Discussion

A physiological study on lipase production by *Bacillus licheniformis* MBB01 revealed that the protein is predominantly expressed during the stationary phase of growth and is secreted into the medium at relatively high levels, particularly in the presence of Tween 80 (Nthangeni *et al*., 2001). It was therefore conceived that the regulatory region (the promoter and the signal peptide region) of this gene could be a useful tool for heterologous expression and secretion of proteins in *Bacillus licheniformis* MBB01. The promoter and signal peptide region of *Bacillus licheniformis* MBB01 lipase gene was therefore cloned, and evaluated for its potential as an expression /secretion tool using the endogenous mature lipase gene, the carboxlesterase gene from *Bacillus pumilus* and a thermophilic DNA polymerase gene from *Thermus aquaticus* as model enzymes. The proteins were selected based on the need to evaluate the secretion capability of the host strain using a known intracellular protein (carboxylesterase) and the efficiency of the system to express, fold properly and secrete to the extracellular a protein from a thermostable protein (Taq DNA polymerase) originating from a thermophile. The endogenous lipase was used as the control protein for expression and extracellular secretion.

To achieve this, a shuttle vector denoted pSV6 was constructed using the pUC18/pUB110 replicative and selective elements. The vector contained the promoter and the signal peptide region of *Bacillus licheniformis* a multiple cloning site a sequence encoding a 6 X His tag and the rrnT1T2 terminator from *Escherichia coli*. The His tag is for possible incorporation to the C-termini of expressed proteins to facilitate purifications by Ni-NTA affinity chromatography. The terminator has been proved to be functional and effective in stabilizing gene transcriptions in expression systems aimed for *Bacillus* species (Doi *et al*., 1996). The

shuttle vectors, containing the respective genes were transformed into *Bacillus licheniformis* MBB01 strain. Transformation of *Bacillus licheniformis* MBB01 was found to be difficult without any published protoplast regeneration methods reported for *Bacillus* species (Bakhiet and Stahly, 1985; Wu and Welker, 1989; Bourne and Dancer, 1986, Akamatsu and Taguchi, 2000) being successful. Transformation by electroporation provided relatively the best results, although significantly poor, with transformation efficiencies of less than 50 transformants per/μg DNA being obtained. This efficiency was in many experiments not consistent, with reasons for failures not known.

Bacillus licheniformis MBB01 transformed with plasmids containing genes for the model enzymes were grown for 36 hours in 50 ml nutrient broth media containing Tween 80, the conditions that coincided with the highest production of the endogenous extracellular lipase (Nthangeni 2001, PhD thesis). Analysis of protein expression using the supernatants on SDS-PAGE did not reveal any over-production of the lipase or Taq DNA polymerase proteins, but a comparatively very high level over-production of the carboxylesterase was observed as judged by protein staining techniques. A very distinct band at visual intensities comparison of more than 50% of all the proteins was observed at the position corresponding to that of the carboxylesterase protein. The very same size of the protein band was also captured by Ni-NTA affinity chromatography, an indication that the protein contained a His-tag. It was concluded that this intense protein band is the carboxylesterase protein and this was further confirmed by enzyme assays where the fraction resulted in pronounced hydrolysis of the substrate as compared to the control enzyme assays. Although there was no over-produced protein band at the position corresponding to the Taq DNA polymerase, passing the supernatant through the Ni-NTA chromatography column resulted in the recovery of three protein bands, one of which had a size of about 90 kDa corresponding to the predicted

molecular mass of Taq DNA polymerase protein (Desai and Pfaffle, 1995). Enzyme assay of the fraction using the PCR technique resulted in successful PCR amplifications concluding that the Taq DNA polmerase was indeed expressed in *Bacillus licheniformis*, correctly folded and secreted to the extracellular medium. The expression level is however not good, and the failure to express it at levels equal to that of the carboxylesterase could be speculated to be the heterologous nature of the protein, proteolytic degradation, non-optimal folding or codon bias. The expression of the endogenous lipase gene surprisingly failed, with no relative increase of the lipase activity in the experimental supernatants and no His-tagged protein was recovered after Ni-NTA chromatography. The reason for this failure is not known. It was however observed that even in *Escherichia coli*, delayed production of the lipase activity was observed. This is in contrast to the results obtained in Chapter 2 (Section 2.3.6) where a plasmid carrying the complete lipase gene together with the promoter and the signal peptide resulted in pronounced hydrolysis of the substrate on TLB plate. The probability that the reading frame was incorrect was ruled out by complete nucleotide sequencing of the expression cassette contained within the pSV6-lip plasmid. The expression cassette as contained within the pSV6-lip plasmid would in *Escherichia coli*, have the lipase protein containing (after the predicted signal peptidase cleavage site) an Ala, His and Met amino acid residues preceding the mature lipase. This is probably contributing to a distortion in the conformation of the lipase protein and thereby resulting in delayed and reduced lipase activity production in *Escherichia coli* as no posttranslational processing of the pre-protein is expected. The pre-lipase protein would theoretically, be efficiently processed in *Bacillus licheniformis*, resulting in properly folded and functional lipase.

In conclusion, the study demonstrated that the promoter and signal peptide region of *Bacillus licheniformis* MBB01 could be used as tools for expression and secretion of heterologous proteins. It is noteworthy that an intracellular protein was successfully secreted to the extracellular by the incorporation of the signal peptide to the N-terminal region of the intracellular carboxylesterase and the protein was expressed to very high levels. A thermostable protein was also expressed and secreted to the extracellular in a functional conformation. It should however be the subject of future researches, the post-translational processing status of the secreted proteins, which could be done by expressing and purification of the secreted proteins followed by N-terminal sequencing to assess the processing of the signal peptide. This study has however proved that the expression system described here could be considered for full development into a commercial tool for heterologous production of industrial proteins. Another related study was done during the course of this study to construct a *Bacillus subtilis* lipase gene-based expression system to overproduce lipase in *Bacillus subtilis* (Ma *et al*., 2006), indicating the feasibility of the lipase promoter for expression.

3.5 References.

Akamatsu T, Taguchi H. (2000). Interspecific transformation of *Bacillus subtilis* competent cells by chromosomal DNA in lysates of protoplasts of *Bacillus amyloliquefaciens*. *Biosci Biotechnol Biochem.* **64**:275-279.

Archibald, A. R., I. C. Hancock, and C. R. Harwood. (1993). Cell wall structure, synthesis, and turnover, p. 381-410. *In* A. L. Sonenshein, J. A. Hoch, and R. Losick (ed.), *Bacillus subtilis* and other gram-positive bacteria. American Society for Microbiology, Washington, D.C.

Bakhiet N, Stahly D.P. (1985). Studies on transfection and transformation of protoplasts of Bacillus larvae, *Bacillus subtilis* and *Bacillus popilliae*. *Appl Environ Microbiol.* **49**:577-581.

Bourne N, Dancer B.N. (1986). Regeneration of protoplasts of *Bacillus subtilis* 168 and closely related strains. *J Gen Microbiol.* **132**:251-255.

Desai U. J. and P. K. Pfaffle (1995). Single-Step Purification of a Thermostable DNA Polymerase Expressed in *Escherichia coli*. *BioTech.* **19**:780-784

Doi R.H., Wang L-F, Brucker R (1996). Extracellular serine protease and a *Bacillus subtilis* alkaline neutral serine protease mutant strain. United States Patent No. 5 585 0253.

Furutani; Y., Shimada H., Mita I, Manabe K., Honjo M., Tomioka N., (1997). DNA base sequence containing regions involved in the production and secretion of a protein, recombinant DNA including the whole or a part of the DNA base sequence, and method of producing proteins by use of the recombinant DNA. United States Patent No. 5 612 192.

Kontinen V.P., Sarvas M. (1993). The PrsA lipoprotein is essential for protein secretion in *Bacillus subtilis* and sets a limit for high-level secretion. *Mol Microbiol.* **8:**727-737.

Laemmli, U.K. (1970). Most commonly used discontinous buffer system for SDS electrophoresis. *Nature* **227**: 680-688.

Li W., Zhou X., Lu P. (2004). Bottlenecks in the expression and secretion of heterologous proteins in *Bacillus subtilis*. *Res. Microbiol.* **8**:605-610.

Ma J, Zhang Z, Wang B, Kong X, Wang Y, Cao S, Feng Y. (2006). Overexpression and characterization of a lipase from *Bacillus subtilis*. *Protein Expr Purif.* **45**:22-29.

Nthangeni M.B., H. Patterton, A. van Tonder, W.P. Vergeer and D. Litthauer, (2001). Over-expression and properties of a purified recombinant *Bacillus licheniformis* lipase: a comparative report on *Bacillus* lipases. *Enzyme Microbiol. Technol.* **28**: 705–712.

Nthangeni M.B. (2001). Molecular and kinetic properties of recombinant *Bacillus* lipase. PhD thesis.

Nthangeni M.B., Ramagoma F, Tlou M.G, Litthauer D. (2005). Development of a versatile cassette for directional genome walking using cassette ligation-mediated PCR and its application in the cloning of complete lipolytic genes from *Bacillus* species. *J. Microbiol. Methods.* **61**: 225-234.

Petit-Glatron M.F, Grajcar L, Munz A, Chambert R. (1993). The contribution of the cell wall to a transmembrane calcium gradient could play a key role in *Bacillus subtilis* protein secretion. *Mol Microbiol.* **9**:1097-1106.

Rhee JK, Ahn DG, Kim YG, Oh JW. (2005). New thermophilic and thermostable esterase with sequence similarity to the hormone-sensitive lipase family, cloned from a metagenomic library. *Appl Environ Microbiol.* **71**:817-825.

Sambrook J., E.F. Fritsch and T. Maniatis, (1989). Molecular cloning: a laboratory manual, Cold Spring Harbor Laboratory Press, New York.

Shyamala V. and G.F. Ames, (1993). Genome walking by single specific primer–polymerase chain reaction. *Methods Enzymol.* **217**: 436-446.

Sloma A., Hannett; N. M., Stephens; M. A., Rudolph C. F., Rufo, Jr., Gerald A., Pero; J. (1992). Expression of heterologous DNA using the *Bacillus coagulans* amylase gene. United States Patent No. 5,171,673.

Tjalsma, H., M.A. Noback, S. Bron, G. Venema, K. Yamane and J.M. van Dijl, (1997). *Bacillus subtilis* contains four closely related type I signal peptidases with overlapping substrate specificities. Constitutive and temporally controlled expression of different *sip* genes. *J. Biol. chem.* **272**: 25983–25992.

van Dijl, J.M., Bolhuis, A., Tjalsma, H., Jongbloed, J.D.H., de Jong, A. and Bron, S., (2001). Protein transport pathways in *Bacillus subtilis*: a genome-based road map. In: Sonenshein, A.L., Hoch, J.A. and Losick, R., Editors, 2001. *Bacillus subtilis* and its Closest Relatives: from Genes to Cells, American Society for Microbiology, Washington, DC, pp. 337–355.

Wilson C., Ron T.M.R., Christianson T., Maurer K-H, Weiss A. (1996). Expression systems for commercial production of cellulase and xylanase in *Bacillus subtilis* and *Bacillus licheniformis*. United States Patent No. 5 888 800.

Winkler U.K. and Stuckman M (1979). Glycogen, Hyaluronate, and some other polysaccharides greatly enhance the formation of exolipase by *Serratia marcescens.* *J. Bacteriol.* **138**: 663-670.

Wu S.C, Wong S.L. (1999). Development of improved pUB110-based vectors for expression and secretion studies in *Bacillus subtilis*. *J. Biotech* **72**:185-95.

Wu L.J, Welker N.E. (1989). Protoplast transformation of *Bacillus stearothermophilus* NUB36 by plasmid DNA. *J Gen Microbiol.* **135**:1315-1324.

Xue, G.-P., J. S. Johnson, and B. P. Dalrymple. (1999). High osmolarity improves the electrotransformation efficiency of the gram-positive bacteria *Bacillus subtilis* and *Bacillus licheniformis. J. Microbiol. Methods* **34**:183-191.

Chapter 4

Summary and concluding remarks

Bacillus represents a genus of Gram-positive bacteria, which are ubiquitous in nature being found in soil, water, and airborne dust. They are distinguished from other Gram- positive bacteria by their ability to produce endospores when environmental conditions become unfavourable. Certain Bacilli species have been implicated as causative agents of diseases such as food poisoning and anthrax while others have been applied in the homologous production of products such as riboflavin, ribose, poly-γ-glutamic acid and industrially important enzymes like alpha-amylases, lipases and proteases. Due to their ability to express and secrete large quantities of proteins, coupled with the Generally Regarded As Safe (GRAS) status of some of the species including *Bacillus subtilis*, *Bacillus licheniformis* and *Bacillus amyloliquefaciens*; *Bacillus* species have been employed as industrial workhorses in the production of homologous proteins. This has attracted Bacilli for development as expression systems for expression of heterologous proteins. The promoters that have been used to drive heterologous expression of proteins are predominantly of genes expressing extracellular proteins. These promoters are even more attractive due to the fact that their genes also contain sequences encoding signal peptides that direct the translocation of proteins to the extracellular, which make the downstream processing of the expressed proteins easy. A physiological study on the production of extracellular lipase from *Bacillus licheniformis* revealed that the protein is expressed at relatively high levels particularly when grown on nutrient broth in the presence of the detergent Tween 80. The DNA fragment encoding mature lipase from this strain

has been cloned previously. The objective of the study was to clone the promoter and the signal peptide region of the *Bacillus licheniformis* MBB01 extracellular lipase and to evaluate its potential as a tool for expression and secretion of endogenous and heterologous proteins.

The study described the improved method for genome walking based on the cassette ligation-mediated PCR principle which was used to clone the promoter and signal peptide region of the lipase gene from *Bacillus licheniformis* MBB01. A 200 bp DNA cassette flanked by various restriction enzymes was introduced within the multiple cloning site of pUC18 plasmid to yield the pLigCas plasmid. Excision of the cassette with restriction enzymes located at opposite ends of the cassette resulted in the release of an efficiently annealed cassette with ends that are ligatable to a compatibly enzyme-restricted genomic DNA sample. Treatment of the excised cassette with alkaline phosphatase prevented self ligation between cassette DNA molecules and ensured preferential ligation with targeted enzyme restricted genomic DNA fragments. The PCR technique referred to as Single-Strand Amplification PCR which involves an initial amplification using a lone primer designed based on the known region of the target region and the cassette-target DNA ligation mixture as the template was employed. The single stranded DNA product obtained during the initial SSA-PCR is used as a template in the second PCR by employing a nested locus specific primer paired with a cassette specific primer in a conventional PCR which results in increased selectivity and specificity of the PCR product. The SSA-PCR technique was used to amplify DNA fragments of 800 and 2200 base pairs respectively corresponding to the regions upstream and downstream to the mature lipase gene fragment of *Bacillus licheniformis* MBB01. Nucleotide sequence analysis

revealed the presence of the open reading frame encoding the isochorismatase gene located downstream to the lipase gene. Isochorismatase is a family of hydrolase enzymes, that is also referred to as dihydro-2,3 dihydroxybenzoate synthase. Some enzymes belonging to this family have been found to be capable of catalyzing the bioconversion of isochorismate in the presence of water to produce dihydro-2,3-dihydroxybenzoate and pyruvate, which are used as important chiral starting materials in the manufacture of bioactive substances such as the siderophore enterobactin and carbasugars. The other ORF encoding a hypothetical conserved protein of unknown function in *Bacillus* was located on the complementary strand upstream to the promoter region of the lipase gene.

The promoter and signal sequence of the extracellular lipase from *Bacillus licheniformis* was incorporated into an *Escherichia coli* / *Bacillus* shuttle vector comprising of replicative elements from pUB110 and pUC18. The shuttle vector denoted pSV6 contained the multiple cloning site, a sequence encoding 6 His tag and the rrnB T1T2 terminator from *Escherichia coli*. The genes encoding the carboxylesterase from Bacillus pumilus, Taq DNA polymerase from *Thermus aquaticus* and the mature lipase from *Bacillus licheniformis* were subcloned into the shuttle and the enzymes were expressed as C-terminally His tagged proteins in *Bacillus licheniformis* MBB01 strain. The endogenous mature lipase gene could not be expressed while the Taq DNA polymerase and the carboxylesterase genes were successfully expressed and processed to the extracellular medium. The level of expression was however different, with the carboxylesterase being expressed at levels that according to visual estimations on SDS-PAGE, were 50 % more than that of the background proteins.

The fact that the activities of the carboxylesterase and the Taq DNA polymerase could be detected in the supernatant is a preliminary indication that the host strain is capable of translocating the otherwise intracellular proteins to the extracellular medium. Further studies, are however required to confirm the nature and status of the N-terminal sequences of the secreted proteins. A provision has been made to introduce within the pSV6 vector at a later stage the gene encoding type 1 Signal peptidases that could be co-expressed to facilitate cleavage of the signal peptide and prevent the probable bottleneck of signal peptide processing. The signal peptidase gene together with a functional promoter could be introduced by subcloning using the *Cla*1/*Nhe*1/*Asc*I polycloning sites within the pSV6 expression vector. The *Nhe*1 site is compatible with *Spe*1, *Xba*1 and *Avr*II, and the *Asc*1 is a rare cutter (as it is an 8 nucleotide recognizing restriction enzyme) which is also compatible with sticky ends generated by *Mlu*1. The size of the pSV6 plasmid is also large and this could restrict the cloning of genes containing large open reading frames, and the bigger the size the higher the instability of the recombinant plasmid. The size could be effectively reduced by the replacing the kanamycin and ampicillin resistance genes in pSV6 by a single gene encoding the chloramphenical acetyltransferase gene from pNW33N plasmid (available from *Bacillus* Genetic Stock Center, Ohio State University, Ohio, USA), which is reportedly capable of serving as a selection marker in both *Escherichia coli* and *Bacillus* hosts.

Chapter 4

Opsomming en Samevattende Opmerkings

Bacillus is een van die verteenwoordigers van die genus van Gram-positiewe bakterieë wat alomteenwoordig is en kan gevind word in grond, water en lugversreide stofpartikels. Hierdie genus word onderskei van ander Gram-positiewe bakterieë deur oor die vermoë te beskik om endospore te produseer gedurende ongunstigde omgewingstoestande. Sekere Bacilli spesies word verbind met siektes soos voedselvergiftiging en antraks, terwyl andere gebruik word vir die produksies van ribo flavien, ribose, poli-γ-glutamaat en industriële ensieme soos alfa-amilases, lipases en proteases. Die feit dat *Bacillus* spesies groot hoeveelhede proteïne kan uitdruk en uitskei en dat die gasheer GRAS-status (algemeen aanvaar as veilig) beskik, het gelei tot die gebruik van sekere van die spesies in die genus (*Bacillus subtilis, Bacillus licheniformis* en *Bacillus amyloliquefaciens*) as industriële gashere vir die produksie van homoloë proteïne. Hierdie aantreklike Bacilli eienskappe het gelei tot die ontwikkeling van uitdrukkingsisteme vir die uitdrukking van heteroloë proteïne. Die promoters is gebruik vir die uitdrukking van meestal gene wat proteïne kodeer wat geteiken word vir ekstrasellulêre uitdrukking. Hierdie promoters het ook die voordeel dat hulle gene seinpeptiede besit wat proteïene uitvoer na die ekstrasellulêre omgewing om sodoende stroomaf hantering van die proteïene te vergemaklik. ‘n Fisiologiese studie van *B. liceniformis* het aangedui dat die ekstrasellulêre lipase op ‘n hoë vlak uitgedruk word indien dit in nutriënt boeljoen en in die teenwoordigheid van Tween 20 gegroei word. Die DNA fragment wat hierdie

volwasse lipase kodeer, is alreeds voorheen gekloneer. Die doel van die studie was om die promoter en die seinpeptied van *B. liciniformis* MBB01, ‘n ekstrasellulêre lipase, te kloneer en sodoende ook te evalueer vir die potensiële uitdrukking en sekresie van homoloë en heteroloë proteïene in die gasheer.

Hierdie studie beskryf ‘n vebeterde kassetligeringsbemiddelde PKR metode om in onbekende gedeelte van die genoom in te beweeg om sodoende die promoter en die seinpeptied gebied van die lipase geen van *B. liciniformis* MBB01 te kloneer. ‘n Tweehonderd basispaar (bp) DNA kasset met verskeie beperkingsensiem herkenningsvolgordes op die punte is in die veeldoelige kloneringsgebied van pUC18 gekloneer om sodoende plasmied pLigCas te lewer. Verwydering van hierdie kasset met beperkingsensieme aan die beide eindes het gelei tot ‘n kasset met eindes wat aanpasbaar was met dienooreenkomstig gesnyde genoom DNA monsters. Behandeling van hierdie kasset met alkaliese fosfatase het sel-ligering van kaset DNA molekules verhoed en sodoende gelei to die meer effektiewe ligering van die kasset met die aanpasbare gesnyde genoom DNA fragmente. Die PKR tegniek, verwys as enkelstring amplifiserings PKR (SSA-PKR), wat die gebruik van ‘ n enkel inleier, gebasseer op bekende inligting van die teiken gebied en die kasset teiken behels, is gebruik met die DNA ligeringsmengsel as templaat. Die enkelstring DNA produk wat verkry is gedurende die SSA-PKR is tydens ‘n tweede PKR rondte gebruik as templaat deur gebruik te maak van ‘n interne inleier (gebasseer op die bekende gedeelte van die teiken) en ‘n inleier gebasseer op die kasset, om sodoende deur konvensionele PKR ‘n produk met verhoogde spesifisiteit en selektiwiteit te lewer. Die SSA-PKR tegniek is gebruik om fragmente van 800 en 2200

bp te lewer wat onderskeidelik ooreenstem met die stroomop en die stroomaf gedeeltes van die teikengebied. Basispaaropeenvolgingbepalings het die oopleesraam van ‘n isochorismatase geen onthul wat stroomaf van die lipase geen voorgekom het. Isochorismase behoort tot ‘n hidrolase ensiem familie, elders ook verwys as die dihidro-2,3-dihidroksiebensoaat sintase. Sekere ensieme wat tot hierdie familie behoort het die vermoë om, in die teenwoordigheid van water, die bio-omskakeling van isochorismate na dihidro-2,3-dihidroksiebensoaat en pirovaat te kataliseer, produkte wat belangrike beginmateriale is in die produksie van kirale boublokke vir die produksie van bio-aktiewe molekule soos “siderophore enterobactin” en koolstofsuikers. Die ander oopleesraam, wat ‘n gekonserveerde proteïen van onbekende funksie in *Bacillus* kodeer, is aangetref op die komplimentêre string stroomop van die promoter gebied van die betrokke lipase geen.

Die promoter en seinpeptied van die ekstrasellulêre lipase van *B. liciniformis* MBB01 is in die *Escherichia coli* / *Bacillus* pendel vektor, wat repliseringselemente van pUB110 en pUC18 bevat, gekloneer. Die vektor, pSV6, bestaande uit die bogenoemde elemente het addisioneel die veeldoelige kloneringsgebied, ‘n volgorde wat vir 6xHis aminosure kodeer en die rrnB T1T2 termineerder vanuit *E. coli* bevat. Gene wat die karboksielesterase vanuit *Bacillus pumilis*, die DNA polymerase vanuit *Thermus aquaticus* en die volwasse lipase vanuit *B. liciniformis* kodeer, is gekloneer in pSV6 en sodoende uitgedruk as C-terminale histidien bevattende proteïene in *B. liciniformis* MBB01. Uitdrukking van die homoloë volwasse lipase was onsuksesvol, maar aktiewe uidrukking en sekresie van beide die DNA polimerase (vanuit *T. aquaticus*) en die karboksielesterase (vanuit *B. pumilis*) het suksesvol plaasgevind. Die vlakke van

uitdrukking het egter gevarieer, met die karboksielesterase wat sowat 50% meer was as die agtergrond proteïene in die sel (gebasseer op visuele skattings van SDS-PAGE studies).

Die teenwoordigheid van beide karboksielesterase en DNA polimerase aktiwiteit in die ekstrasellulêre omgewing is ‘n aanduiding dat eens intrasellulêre proteïene effektief uitgeskei kan word na die ekstrasellulêre omgewing. Verdere studies moet wel gedoen word om die N-terminale gedeeltes van die proteïene te ondersoek. ‘n Verdere voorsorgmaantreël om ‘n tipe 1 seinpeptidase in die pSV6 vektor te plaas om dit sodoende gelyktydig uit te druk saam met die teiken proteïen en moontlik sal kan fasiliteer in die verwydering van die seinpeptied en hopelik ook ‘n bottelnek in die verwerking van die seinpeptiede te verhoed. Hierdie seinpeptidase geen saam met ‘n funksionele promoter ingevoeg word in die pSV6 plasmied deur gebruik te maak van die *Cla*I/*Nhe*I/*Asc*I veeldoelige kloneringsgebied. Die *Nhe*I volgorde is aanpasbaar met die volgordes van *Spe*I, *Xba*I en *Avr*II. Die *Asc*I ensiem is ‘n 8 bp beperkingsensiem wat die teenwoordigheid van die herkenningsvolgorde baie raar maak en sodoende dan ook klonering vergemaklik en gesnyde DNA is aanpasbaar met *Mlu*I gesnyde DNA. Die pSV6 plasmied se grootte kan ook moontlike probleme veroorsaak indien daar gewerk word met groot oopleesrame, wat dan kan lei tot ‘n onstabiele rekombinante plasmied as gevolg van die grootte. Die grootte van die pSV6 plasmied kan effektief verminder word deur die ampisillien en kanamisien weerstandbiedende merkers te vervang met ‘n chloramfenikol asetieltransferase geen vanaf plasmied pNW33N (beskikbaar by *Bacillus* Genetick Stock Centre, Ohio State University, Ohio, USA). Daar is bewys dat hierdie merkers kan dien as ‘n seleksiemerker in beide *E. coli* en die *Bacillus* gashere.