ADVANCES IN APPLIED MICROBIOLOGY

VOLUME 73

Advances in
APPLIED MICROBIOLOGY

VOLUME 73

Edited by

ALLEN I. LASKIN
Somerset, New Jersey, USA

SIMA SARIASLANI
Wilmington, Delaware, USA

GEOFFREY M. GADD
Dundee, Scotland, UK

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Academic Press is an imprint of Elsevier

Academic Press is an imprint of Elsevier
525 B Street, Suite 1900, San Diego, CA 92101-4495, USA
30 Corporate Drive, Suite 400, Burlington, MA 01803, USA
32, Jamestown Road, London NW1 7BY, UK

First edition 2010

ISBN: 978-0-12-380987-2
ISSN: 0065-2164

Printed and bound in the USA
10 11 12 13 10 9 8 7 6 5 4 3 2 1

Advances in
APPLIED MICROBIOLOGY

VOLUME 73

CONTRIBUTORS

John F. Cannon
Department of Molecular Microbiology and Immunology, University of Missouri, Columbia, Missouri, USA

Bumsuk Hahm
Department of Surgery, Department of Molecular Microbiology and Immunology, Center for Cellular and Molecular Immunology, Virology Center, University of Missouri-Columbia, Columbia, Missouri, USA

Colin R. Harwood
Centre for Bacterial Cell Biology, Institute of Cell and Molecular Biosciences, Newcastle University, Newcastle upon Tyne, UK

Ralf Hortsch
Institute of Biochemical Engineering, Technische Universität München, Boltzmannstr. 15, Garching, Germany

Susanne Pohl
Centre for Bacterial Cell Biology, Institute of Cell and Molecular Biosciences, Newcastle University, Newcastle upon Tyne, UK

Young-Jin Seo
Department of Surgery, Department of Molecular Microbiology and Immunology, Center for Cellular and Molecular Immunology, Virology Center, University of Missouri-Columbia, Columbia, Missouri, USA

Dirk Weuster-Botz
Institute of Biochemical Engineering, Technische Universität München, Boltzmannstr. 15, Garching, Germany

CONTENTS

CHAPTER 1

Heterologous Protein Secretion by *Bacillus* Species: From the Cradle to the Grave

Susanne Pohl and **Colin R. Harwood**[1]

Contents

Abstract

The Gram-positive bacterium *Bacillus subtilis* and some of its close relatives are widely used for the industrial production of enzymes for the detergents, food, and beverage industries. The choice of these organisms is based almost exclusively on the high capacity of

Centre for Bacterial Cell Biology, Institute of Cell and Molecular Biosciences, Newcastle University, Newcastle upon Tyne, UK
[1] Corresponding author: e-mail address: colin.harwood@ncl.ac.uk

Advances in Applied Microbiology, Volume 73
ISSN 0065-2164, DOI: 10.1016/S0065-2164(10)73001-X

their secretion systems that are, under the right conditions, able to secrete proteins at grams per liter concentrations. In contrast, there are relatively few examples of *Bacillus* species being used for the cytoplasmic production of proteins.

The range of proteins that are capable of high-level production and secretion is limited by a combination of characteristics of both the target protein and the host bacterium. The secretion pathway includes checkpoints that are designed to validate the authenticity of pathway substrates. Although many of these checkpoints are known, only some can be overcome by reengineering the host. As a result, the yield of heterologous protein production is extremely variable. In this review, we consider the *Bacillus* protein secretion pathway from the synthesis of the target protein (cradle) to its emergence at the outer surface of the complex cell wall (grave), and discuss the roles of the various checkpoints both with respect to the target protein and their role on cell homeostasis.

I. INTRODUCTION

Members of the genus *Bacillus* are prodigious producers of industrial enzymes such as proteases, α-amylases, and other macromolecular hydrolases (Harwood, 1992). This reflects the fact that, in their natural habitat, the ability to breakdown and utilize soil detritus—particularly that derived from plants (protein, starch, pectin, cellulose, etc.)—provides important sources of nutrients. Because bacteria generally do not take up macromolecules, *Bacillus subtilis* and its close relatives naturally secrete a wide range of hydrolytic enzymes into their environment. The ability of this Gram-positive bacterium to secrete proteins directly into the culture medium at high concentrations is, in part, a reflection of the fact that its cytoplasm is surrounded by a single membrane system, in contrast to the double membrane found in Gram-negative bacteria. From a commercial point of view, the purification of proteins from the culture medium rather than from the cytoplasm is considerably more cost-effective, often leads to improved structural authenticity and reduces the likelihood of the co-purification of endotoxins and other potential contaminants. Given the commercial advantages of their secretory systems, it is surprising that more use has not been made of *Bacillus* species for the production of heterologous proteins. In practice, attempts to use this group of bacteria for the manufacture of heterologous proteins have met with mixed success. The reasons for this are complex, but relate to the intrinsic properties of both the target proteins and the secretion systems themselves. In particular, given the importance of protein secretion to cell growth and integrity, the native secretion systems include a series of quality control checkpoints designed to avoid potentially lethal blockages.

II. EXPRESSION AND SECRETION VECTOR SYSTEMS

A wide variety of expression systems have been described for *Bacillus* species and these have to be combined with targeting sequences for proteins that are required to be secreted. Both self-replicating (mono- and bi-functional) and integrating plasmids have been developed, although the former is necessary for the highest levels of protein production.

Originally, *Bacillus* cloning vectors were based on *Staphylococcus aureus* antibiotic resistance plasmids, such as pUB110 (Km^R), pT181 (Tc^R), and pC194 (Cm^R). These mono-functional plasmids, and the vectors derived from them, replicate using the rolling circle mode of replication and consequently tend to accumulate single-stranded replication intermediates and to suffer from both structural and segregational instability (Bron, 1990; Gruss and Ehrlich, 1989). The identification of theta replicating plasmids, such as pAMβ1 and BS72, led to the development of a newer generation of *Bacillus* vectors and *Escherichia coli–Bacillus* bifunctional vectors that are both segregationally and structurally stable (Bruand *et al.*, 1991; Jannière *et al.*, 1993; Titok *et al.*, 2003). A good example is pMTLBS72, which contains the replication origins of pBR322 for maintenance in *E. coli*, and pBS72 for maintenance in *B. subtilis*. Like all bifunctional vectors, they require antibiotic resistance genes that are selectable in both hosts.

An alternative to autonomously replicating plasmid vectors is to use integration vectors, exploiting the high frequency of recombination between homologous DNA sequences in many widely used strains of *B. subtilis*. For the most part, integrated vectors are stable provided the site of integration is carefully chosen. *B. subtilis* integration vectors are based solely on *E. coli* origins of replication (usually *colE1*-based) since they use regions of homology between the vector and the host chromosome to facilitate integration via single or double crossover recombination events (Harwood *et al.*, 2002).

Industrial strains of *Bacillus* are able to direct the synthesis of extracellular proteins to concentrations in excess of 20 g/l, representing a combination of optimized expression elements, developed strains, media, and growth regimes. Although these systems are not generally available, there is now a range of promoter systems have been developed for the controlled, high-level expression of proteins from *B. subtilis*. Some of these have been adapted from *E. coli* system, others from *Bacillus* species and other Gram-positive bacteria.

The widely used P_{spac} promoter was constructed by fusing the 5′-sequences of a promoter from the *B. subtilis* phage SPO1 and the 3′-sequences of the *E. coli lac* promoter, including the operator (Yansura and Henner, 1984). P_{spac} expression is dependent on the inactivation of a constitutively expressed lactose repressor by IPTG. The P_{spac} promoter functions in plasmid and chromosomal locations and, when present in

multicopy situations, can direct the synthesis of a protein to a significant proportion of total cellular protein. However, this promoter is not sufficiently strong and its inducer is too expensive for large-scale fermentations.

Also widely used are xylose-inducible promoters based on the *XylR* repressor. Since these promoters originate from *B. subtilis* and related organisms, they have been used to control gene expression without modification (Gartner *et al.*, 1992). When used on high-copy-number expression vectors, an additional copy of the *xylR* gene is usually included to maintain a balance between the number of repressor molecules and operator sites. Although genes in the xylose regulon are usually subject to catabolite repression, the catabolite responsive element (Cre) is not included in the vectors. *XylR*-controlled promoters direct moderately high levels of expression and have the advantage that the inducer, xylose, is relatively cheap.

Constitutive promoters associated with catabolite repressed genes such as the α-amylase genes from *B. licheniformis* or *B. amyloliquefaciens* can be used as non-inducible expression systems by batch-feeding a catabolite repressing carbon source so that the fermenter operates at substrate limiting concentrations. Under these conditions, *B. subtilis* will continue to produce the target protein for several days.

Expression systems continue to be developed for *B. subtilis*. For example, the lactose (*lac*) operator system for controlling gene expression has been combined with the very strong vegetative (σ^A) promoter upstream of the *B. subtilis groESL* operon, encoding the heat-shock protein GroES and GroEL (Phan *et al.*, 2006). More recently, Chen *et al.* (2010) have adopted the T7 expression system for *B. subtilis*. They used an integration vector to insert the gene encoding the T7 RNA polymerase under the control of the P_{spac} promoter, together with the *lacI* gene under the control of the P_{penP} promoter, into the *wprA* gene (encoding a cell wall protease—see later). By flanking the antibiotic and ColE1 *ori* genes of the integration vector with FRT phage integration sites, the inserted DNA was made markerless by the induction of the cognate FLP recombinase gene on a suicide plasmid.

In general, *B. subtilis* has little or no advantage over *E. coli* for the intracellular production of heterologous protein, except for the lack of the highly immunogenic lipopolysaccharides (LPS), traces of which have to be removed from proteins that need to be injected into humans and animals. The main advantage of using *B. subtilis* is its potential for high-level protein secretion with subsequent recovery from the culture medium. Proteins that are targeted for secretion require a targeting signal in the form of a signal peptide and, for proteins that are not naturally secreted, a signal sequence needs to be incorporated into the vector in such a way as to fuse the signal peptide in-frame with N-terminus of the target protein.

Numerous attempts have been made to maximize the secretion of heterologous proteins by identifying optimal *Bacillus* signal peptides. However, while a specific signal peptide may be optimal for the secretion of one particular target protein, it is often found not to be optimal for another, indicating that as yet understood characteristics of both the signal peptide and the mature protein together influence secretion (Brockmeier *et al.*, 2006). For general purposes, the signal peptides from the *B. amyloliquefaciens* α-amylase (AmyQ) and *B. subtilis* alkaline protease (AprE) have been used in many secretion/expression vectors (Olmos-Soto and Contreras-Flores, 2003; Phan *et al.*, 2006).

III. *BACILLUS* SECRETION PATHWAYS

Between 5% and 10% of the proteins encoded by bacteria are secreted across the cytoplasmic membrane using the ubiquitous Sec-dependant (Sec) (Driessen and Nouwen, 2008; Holland, 2004) and twin-arginine translocation (TAT) (Berks *et al.*, 2005; Robinson and Bolhuis, 2004) pathways. The Sec pathway is responsible for the secretion of the majority of these proteins, while the TAT pathway is required for the smaller numbers of proteins that need to be folded prior to translocation. In the case of Gram-negative bacteria such as *E. coli* and *Salmonella enterica*, the need to translocate proteins across a double membrane system has resulted in the evolution of a variety of specialized, substrate-specific, protein secretion pathways (e.g., types I, II, III, IV, V, and VI), some of which require the involvement of the Sec pathway (e.g., types II, IV, and V pathways) (Papaniko *et al.*, 2007). In contrast, Gram-positive bacteria generally lack these specialized pathways, except for homologues of the relatively poorly understood ESAT-6 secretion pathway (ESX; Bitter *et al.*, 2009), substrate-specific Sec pathways associated with homologues of the SecA protein (Rigel and Braunstein, 2008) and phage-associated holin-like proteins for the secretion of endolysins (Borysowski *et al.*, 2006). Only the Sec pathway is currently being exploited for the secretion of heterologous proteins from *B. subtilis* and its relatives, and consequently this review focuses exclusively on this pathway and various attempts that have been made to improve the secretion of foreign protein from these organisms (Brockmeier *et al.*, 2006; Sarvas *et al.*, 2004; Tjalsma *et al.*, 2004).

IV. SUBSTRATE RECOGNITION

A crucial early event in the Sec secretion pathway is the identification, by cytoplasmic components, of substrates that are destined for secretion. Proteins that are targeted for translocation across the cytoplasmic

membrane are identified by their possession of an N-terminal extension, the signal peptide, that is removed during the latter stages of secretion (Bendtsen *et al.*, 2005; Nielsen *et al.*, 1997). The signal peptides of bacterial Sec pathway substrates exhibit a similar structural organization: they are usually between 20 and 30 amino acids in length, and have a positively charged amino terminal (N) region, followed by a hydrophobic (H) central region and a short cleavage (C) region containing the target site for signal peptidase (Fig. 1.1). The signal peptides fall into two distinct types, as defined by the class of signal peptidase responsible for the cleavage event that releases the mature protein and their final location. Type I signal peptidases cleave the most abundant class of secretory substrates, which include proteins associated with the cell wall or which are released into the culture medium. *B. subtilis* encodes five chromosomally encoded Type 1 signal peptidases, namely SipS, SipT, SipU, SipV, and SipW. None

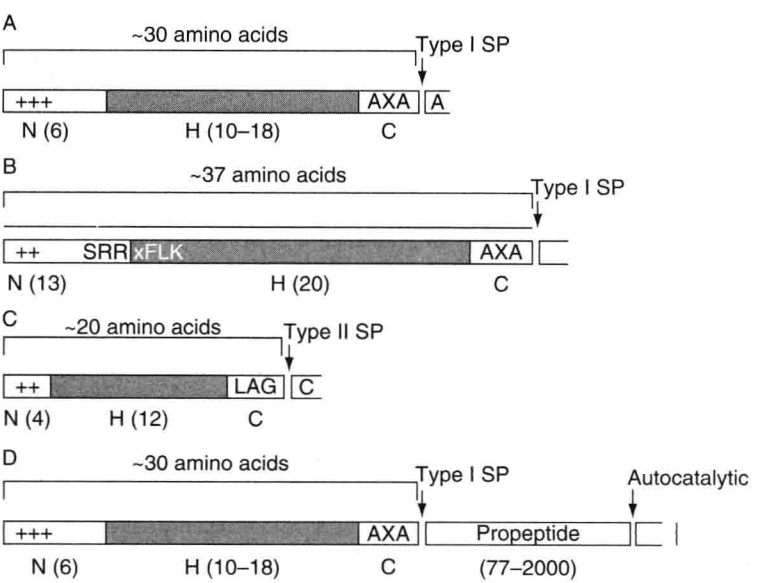

FIGURE 1.1 The main features of *Bacillus* signal peptides and propeptides. The N-terminal (N), hydrophobic (H), and cleavage (C) regions are identified by contrasting shading with their average lengths indicated in brackets. A. Sec-dependent signal peptide cleaved by a Type I signal peptidase (SP); B. TAT-dependent signal peptide, cleaved by a Type I signal peptidase (SP); C. Lipoprotein signal peptide cleaved by the Type II signal peptidase (SP); D. The signal peptide and propeptide (prepropeptide) at the N-terminal end of a secretory protein requiring the propeptide for folding on the *trans* side of the cytoplasmic membrane. The signal peptide is removed by a Type I enzyme and the propeptide either autocatalytically or by a coexisting protease.

are essential, although either SipS or SipT must be present to maintain viability (Tjalsma *et al.*, 1999) and are therefore regarded as paralogues of the single Type I signal peptidase (Lep) of *E. coli*. SipU, SipV, and SipW are minor signal peptidases that appear to be involved in the secretion of specific substrates. The type I signal peptides (Fig. 1.1) of Gram-positive bacteria are, on average, longer (~30 amino acid) and more hydrophobic than those of Gram-negative bacteria (~25 amino acids). They have similar consensus cleavage sequences (AXA↓) and the signal peptides of *B. subtilis* tend to be functional in *E. coli* and vice versa, albeit with differing efficiencies (Zanen *et al.*, 2005).

The absence of a membrane-enclosed periplasm means that *Bacillus* species have a higher proportion of lipoproteins than their Gram-negative counterparts (Tjalsma *et al.*, 2004). Because lipoproteins have to be diacylglycerol-modified prior to attached to the outer surface of the membrane, they are targeted by distinct signal peptides that are recognized and cleaved by the single Type II signal peptidases (LspA). Consequently, while topologically similar to Type I signal peptides, Type II signal peptides share discrete characteristics that include shorter N and H regions and a different consensus cleavage site, referred to as a Lipobox: [LITAGMV]-[ASGTIMVF]-[AG]-↓C-[SGENTAQR] (Fig. 1.1; Sutcliffe and Harrington, 2002; Tjalsma and van Dijl, 2005). The amino acid at the N-terminus of the mature lipoprotein is invariably a Cys residue that, when lipo-modified, serves to tether the protein to the outer leaflet of the cytoplasmic membrane (Juncker *et al.*, 2003).

V. INTRACELLULAR CHAPERONING AND PILOTING TO THE SEC TRANSLOCASE

Although bacterial signal peptides have been well characterized, surprisingly little is known about the intracellular events associated with targeting and the subsequent piloting of their cargo proteins to the Sec translocase, particularly in Gram-positive bacteria. The process requires three key elements; (a) the identification of the target secretory protein, preferably as it emerges from the ribosome; (b) its interaction with chaperone proteins that prevent it folding into a secretion incompetent state; (c) its piloting to the membrane-bound translocase.

There is considerable uncertainty about the intracellular events associated with protein secretion in bacteria. Since there are few relevant studies in *B. subtilis*, and for the most part both these bacteria share the same components, we discuss recent progress in *E. coli* (Luirink and Sinning, 2004; Zhang *et al.*, 2010). The three key intracellular players are the Signal Recognition Particle (SRP), SecA, and SecB (Fig. 1.2). The current view is that the SRP is required for the cotranslational targeting

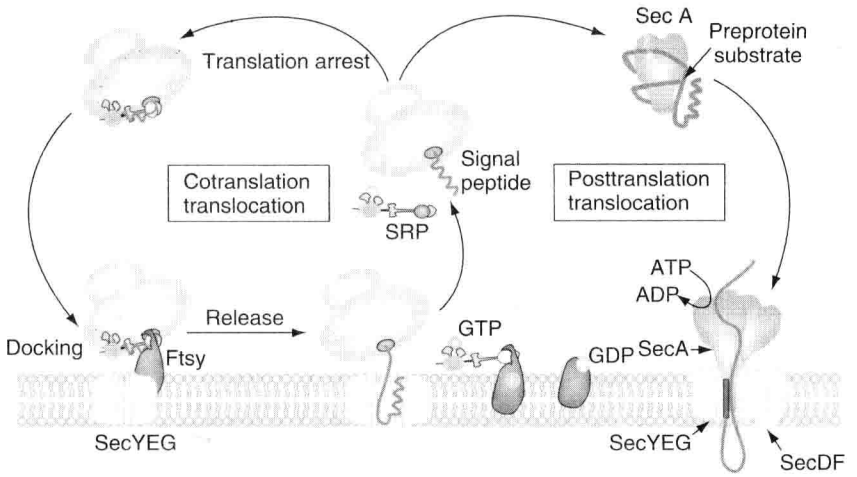

FIGURE 1.2 Diagrammatic representation of the cytoplasmic chaperoning and targeting pathways of *Bacillus subtilis* based on the Signal Recognition Particle (SRP) and SecA cycles.

of integral membrane proteins to the inner membrane via the Sec translocase. SRP, a ribonucleoprotein complex consisting of 4.5S RNA and Ffh, interacts specifically with signal sequences of nascent membrane proteins emerging from the ribosome (Neher *et al.*, 2008). The resulting complex then docks at the membrane and ultimately the Sec translocase via a membrane-bound receptor, FtsY. The result is a switching of the translocase to a transversal opening mode, and the lateral release of target proteins into the membrane. Ffh and FtsY are members of the SRP-GTPase protein family that are essential for viability (Chen *et al.*, 2008).

Proteins that are targeted beyond the membrane (including membrane-anchored lipoproteins) associate with SecB (Bechtluft *et al.*, 2009), a secretion-specific cytoplasmic chaperone, as they emerge from the ribosome, rather than the SRP. The role of the tetrameric SecB is to maintain the secretory proteins in the essential unfolded (i.e., secretion competent) state required for translocation through the Sec translocase. SecB binds to the mature region of its secretory substrates, in a groove that forms between the interacting dimers (Dekker *et al.*, 2003). How SecB recognizes its cargo is still unclear. The SecB tetramer is organized at a dimer of dimers, with inwardly facing α-helices and outwardly facing β-sheets. Each dimer binds one molecule of the SecA–ADP dimer that interacts with the translocase, via the highly conserved C-terminal SecB binding domain. Once the ternary complex (preprotein/SecB/SecA) interacts with the translocase, the replacement of ADP with ATP leads to the release and recycling of SecB (Luirink and Sinning, 2004).

In the early stages of secretion, it is clear that the SRP and SecB pathways converge and are in competition for substrates. SRP appears to have a preference for signal peptides with more hydrophobic "H" regions, while the remaining substrates appear to be directed to the SecB via the ribosome-bound trigger factor.

The intracellular processing of secretory proteins in the Gram-positive bacterium *B. subtilis* is similar except that its SPR RNA is significantly longer (271 nucleotides) than that of its *E. coli* counterpart (114 nucleotides) and contains an *Alu* domain to which an additional histone-like protein, HBsu, is attached (Eichler, 2003; Nakamura *et al.*, 1999). Since *Alu* domains are present in Eukaryal and Archael SRP RNA, both of which undergo translational arrest and subsequent cotranslation translocation, it would be interesting to know if these processes also take place in *B. subtilis*. The second major difference is the absence, in all Gram-positive bacteria, of SecB (or identifiable functional homologue), even though their SecA proteins still encode the highly conserved 22 amino acid C-terminal SecB binding domain. Attempts to identify a direct homologue of SecB, or to show that other intracellular chaperones such as GroEL/ES and DnaK play a major role in secretion, have been unsuccessful. However, one potential protein chaperone has been identified that might fulfill the role of SecB, namely CsaA (Müller *et al.*, 2000; Shapova and Paetzel, 2001). Although the evidence for its role in secretion needs to be strengthened, CsaA has been shown to interact with SecA, to bind to peptides and is upregulated under secretion stress (Linde *et al.*, 2003; Müller *et al.*, 1992; Vitikainen *et al.*, 2005).

The absence of SecB, together with the more hydrophobic nature of the signal peptides of Gram-positive bacteria, has led some to suggest that SRP provides the chaperone activity for all integral membrane and secreted proteins and others to suggest that SecA alone can perform this function for secretory proteins (Zanen *et al.*, 2005, 2006). Depletion of *B. subtilis* Ffh reduces Sec-dependent secretion, although experiments designed to elucidate the precise role and substrate specificity of SRP pathway have provided ambiguous results. This is presumably due to the pleiotropic effects of the depletion of Ffh on the insertion of the translocase itself into the cytoplasmic membrane. Until the molecular processes associated with these early cytoplasmic events are better understood, currently there are few opportunities to improve this stage of secretion by rational intervention.

Potentially, some of these issues can be circumvented by the use of the Twin Arginine Transporter (Tat) pathway (Berks *et al.*, 2005). In this case target proteins are characterized by a signal peptide that includes R–R at the junction of the N- and H-regions. The Tat pathway is adapted for the secretion of folded proteins, which potentially obviates the need for cytoplasmic chaperoning. However, although there are examples of the