Type VII Secretion: A Highly Versatile Secretion System

LOUIS S. ATES,1 EDITH N. G. HOUBEN,2 and WILBERT BITTER1,2

1Department of Medical Microbiology and Infection Control, VU University Medical Center, Amsterdam, The Netherlands; 2Section Molecular Microbiology, Amsterdam Institute of Molecules, Medicine and Systems, Vrije Universiteit Amsterdam, 1081 BT Amsterdam, The Netherlands

ABSTRACT Type VII secretion (T7S) systems of mycobacteria secrete substrates over the unusual diderm cell envelope. Furthermore, T7S gene clusters are present throughout the phylum Actinobacteria, and functional T7S-like systems have been identified in Firmicutes. Most of the T7S substrates can be divided into two families: the Esx proteins, which are found in both Firmicutes and Actinobacteria, and the PE and PPE proteins, which are more mycobacterium-specific. Members of both families have been shown to be secreted as folded heterodimers, suggesting that this is a conserved feature of T7S substrates. Most knowledge of the mechanism of T7S and the roles of T7S systems in virulence comes from studies of pathogenic mycobacteria. These bacteria can contain up to five T7S systems, called ESX-1 to ESX-5, each having its own role in bacterial physiology and virulence.

In this article, we discuss the general composition of T7S systems and the role of the individual components in secretion. These conserved components include two membrane proteins with (predicted) enzymatic activities: a predicted ATPase (EccC), likely to be required for energy provision of T7S, and a subtilisin-like protease (MycP) involved in processing of specific substrates. Additionally, we describe the role of a conserved intracellular chaperone in T7S substrate recognition, based on recently published crystal structures and molecular analysis. Finally, we discuss system-specific features of the different T7S systems in mycobacteria and their role in pathogenesis and provide an overview of the role of T7S in virulence of other pathogenic bacteria.

Bacterial secretion systems were initially studied in the Gram-negative bacterium Escherichia coli K-12. When researchers started to explore protein secretion in different Gram-negative bacteria and especially in bacterial pathogens, it was clear that E. coli K-12 was not able to present us with a complete picture of protein secretion systems. Type II, type III, and type IV secretion systems were quickly discovered and revolutionized host-pathogen interaction studies. Gram-negative bacteria need these specialized secretion systems to transport proteins across two membranes (also called a diderm cell envelope). The presence of this complex cell envelope not only means that two membranes have to be crossed, but an additional problem is that there is no energy source at the outer membrane. This means that alternative mechanisms for protein transport need to be present, such as coupling the energy of the inner membrane to protein transport across the outer membrane or crossing the entire cell envelope in a single step. Although the discovery of different secretion systems in Gram-negative bacteria was a major breakthrough, the downside has been that secretion systems in other bacteria have been neglected. It was generally thought that secretion in other bacteria, which are generally monoderm, would completely depend on the universal Sec or Tat system. Only in recent years has this idea begun shifting, and again it started by studying pathogens, i.e., the pathogenic mycobacteria.

The genus Mycobacterium contains a number of important pathogens, including Mycobacterium tuberculosis.
losis and Mycobacterium leprae. These pathogens are generally slow growing and have a distinctive growth pattern, known as cording. Although these bacteria genetically belong to the high-GC Gram-positive bacteria, they produce a second membrane composed of unique and complex lipids (1, 2). The cell envelope of these bacteria is therefore diderm, and protein secretion is as problematic as for Gram-negative bacteria. Among the first identified secreted proteins of M. tuberculosis were two small proteins, known as ESAT-6 and CFP-10 (3). Later they were renamed EsxA and EsxB (4). These proteins lack any obvious canonical secretion signal (Sec or Tat) and therefore seemed to depend on a new secretion system. Detailed analysis of the tuberculosis vaccine strain Mycobacterium bovis BCG indicated that this strain had lost not only both the esxAB genes, but also several surrounding genes responsible for EsxAB secretion (5, 6). Subsequent research showed that many of the genes surrounding esxAB are indeed part of this secretion system (7). This region is now known as the esx-1 locus and contains 20 genes (Fig. 1). With these genes identified, it quickly became clear that mycobacteria have several of these secretion systems. In fact, up to five different esx-loci can be present on the chromosome of mycobacterial species. Recently, additional ESX systems have been identified on conjugative mycobacterial plasmids (8). On these plasmids, the esx clusters are located adjacent to a type IV-like secretion cluster, and together they are required for the conjugation process (8).

To emphasize that these ESX systems are required for secretion over the diderm mycobacterial cell envelope, they were named type VII secretion (T7S) (4, 9). In the literature the terms Wss (for WxG100 secretion system) and ESX have been infrequently used, but in this chapter we will use T7S as a general term for these secretion systems.

**FIGURE 1** Genetic loci of different T7S (and T7S-like) systems. Depicted are the T7S loci, esx-1, esx-4, and esx-5 of M. tuberculosis H37Rv (101), as well the T7S-like systems of S. aureus (strain USA300, annotation based on Anderson et al. [63]) and B. subtilis subsp. subtilis (strain 168, annotation based on Huppert et al. [12]). Color coding represents conserved T7S membrane components (dark blue), (putative) substrates of the systems (green), cytosolic chaperones (yellow), and Firmicutes-specific T7S-like membrane components (light blue).

---

2 ASMscience.org/MicrobiolSpectrum
systems and ESX for the different mycobacterial secretion systems. T7S systems are not specific for pathogenic mycobacteria; they are also present in nonpathogenic mycobacteria, and many other bacteria have homologous systems. Most of these bacteria, such as Rhodococcus, Corynebacterium, and Nocardia species, are closely related to mycobacteria and have a similar dierm cell envelope. Probably, the T7S systems of these species are involved in protein secretion as well, although currently there is no data supporting this. An interesting group of T7S homologs is present in several Firmicutes species, including many Bacillus and Staphylococcus species (10). However, these systems are only distantly related since only a homolog of the EccC membrane component (see below) and homolog(s) of the EsxA substrates are present in these bacteria. These latter systems have recently been shown to function as active secretion systems (11, 12). Clearly, there is an evolutionary link between the dierm T7S systems and the newly identified Firmicutes secretion systems, but nomenclature could be an issue. Previously, we suggested that they might be called type VIIb systems, analogou to the type IV secretion systems where similar heterogeneity occurs (9). We will comprehensively discuss the mycobacterial T7S systems, which have been studied most intensively, and give an overview of the similarities with the Firmicutes T7S-like secretion systems, with a focus on the nature of the substrates and the conserved secretion signal.

COMPOSITION AND FUNCTIONING OF T7S IN MYCOBACTERIA

T7S gene clusters of M. tuberculosis usually share a number of conserved genes that belong to 10 gene families. A general nomenclature for these proteins was agreed on in an early phase (4); the name ESX-conserved component (Ecc) is used for components that are present in most systems, and ESX-1-specific components (Esp) for components that are (mostly) unique for the best-studied secretion system, ESX-1. Finally, for some conserved proteins the old and well-established names were kept, i.e., mycosins (MycP), Esx, and PE and PPE proteins. Within the esx loci we can find genes that encode both substrates and structural components. In addition to the Esx proteins, PE and PPE proteins and most Esp proteins are probably substrates. Thus far, there seem to be seven conserved genes that are required for the secretion process and can therefore be considered as core components of the secretion machinery. Two of these core components (EspG and EccA) are localized to the mycobacterial cytosol, while the other five (EccB, EccC, EccD, EccE, and MycP) contain transmembrane domains (TMDs) and reside in the mycobacterial cell envelope (Fig. 2). All esx gene clusters contain members of these 10 conserved gene families, except for the most archaic cluster, i.e., esx-4, which lacks the pe, ppe, espG, eccA, and eccE genes (Fig. 1). We will first discuss the two cytosolic conserved components of mycobacterial T7S systems and then the membrane components.

EspG

EspG was initially not considered a T7S core component (4), because these proteins share relative low sequence identity (20 to 25%). However, recent experiments have clearly shown that these proteins are structural and functional equivalents. To acknowledge this, it would be more appropriate to rename this component Esx conserved component, but we realize that the current name is already established and should be kept. The first indication of a role for EspG in secretion was a study describing that EspG5 is required for the secretion of PE-PPE proteins via the ESX-5 system (13, 14). Subsequently, immunoprecipitation experiments showed that EspG5 specifically interacts with the heterodimeric model substrates PE25 and PPE41 (15) (for dimer formation of T7S substrates see below). Additional biophysical analysis of the EspG5-PE25/PPE41 complex showed a tight interaction. Because EspG5 is located in the cytosol and could not be identified in the culture supernatant, it probably dissociates from the substrate pair upon contacting the secretion machinery. Interestingly, the interaction of EspG with PE-PPE proteins shows considerable system specificity, because EspG5 does not interact with the PE35/PPE68_1 protein pair that is secreted via ESX-1 (15). Conversely, EspG1 of the ESX-1 system does not bind the ESX-5-dependent PE25/PPE41 but does interact with the ESX-1 protein pair PE35/PPE68_1. This suggests that EspG proteins are cytosolic chaperones that specifically recognize their cognate PE-PPE substrates. This component might therefore determine through which T7S system these substrates are transported. The discovery that EspG proteins are PE-PPE-specific chaperones also explains why the esx-4 locus is lacking espG, in addition to pe and ppe genes.

Recently, crystal structures of EspG5 in complex with PE25/PPE41 were obtained (Fig. 3C) (16, 17). The structures reveal that the EspG5 protein has a novel mixed α/β-fold. The crystal structure of EspG3 of the ESX-3 system was solved in one of these studies (16). Despite the low similarity on the sequence level, EspG3 has a highly similar fold, confirming that these proteins
Ates et al.

**FIGURE 2** Model for T7S in mycobacteria. The conserved membrane components (blue) form a complex in which the EccC homolog is the ATPase possibly providing energy for the secretion process. The mycosin (MycP) is not part of the core complex but is essential for successful secretion. The T7S substrates (green) are secreted dependently on the conserved signals YxxxD/E and WxG (red). Secretion of PE-PPE dimers is dependent on the cytosolic chaperones EspG and EccA (yellow). While EspG binds to the substrate pair in the cytosol, EccA might be involved in releasing this chaperone from the PE-PPE dimer upon contact with the membrane complex. In contrast, Esx proteins are not recognized by EspG, and their dependence on the cytosolic chaperones might be indirect due to interdependence of Esx and PE-PPE for secretion. The EspB monomer has a similar fold to PE-PPE dimers and contains the putative secretion signal. Upon translocation, EspB is processed and forms a heptamer with a barrel-like structure. Whether PE-PPE dimers adopt a similar quaternary structure is yet unknown. Secreted substrates can localize to the culture supernatant or remain attached in the capsular layer. Whether the secretion process is a one- or two-step process is not known, so a putative outer membrane component (gray) is indicated by a question mark.
have a similar function. The structure of the EspG5-PE25/PPE41 complex shows that EspG5 binds to the tip of the PPE protein, which is not involved in the interaction with the PE protein. This tip region of PPE contains some hydrophobic residues that are now buried in a hydrophobic groove formed by a central β-sheet and C-terminal α-helical bundle of EspG5. Subsequent mutagenesis in the EspG5-recognition motif of various ESX-5 dependent PPE proteins showed that mutations that abolish EspG5 binding in vitro usually affect substrate secretion (17). Additionally, coexpression of EspG proteins increased the in vitro solubility of PE-PPE pairs, which suggests a role of EspG in keeping these protein pairs soluble. Recently published x-ray diffraction and electron microscopy data of the ESX-1 substrate EspB shows not only that this protein has a highly similar fold as PE-PPE heterodimers (see below, Fig. 3D), but also that it can form pore-like quaternary structures (18). Modeling suggests that EspB could be organized as a heptamer and that the multimerization is mediated by its putative EspG-binding domain. Therefore, the EspG proteins could be responsible for preventing this multimerization from occurring intracellularly. Together, these data suggest that EspG proteins are chaperones involved in substrate recognition and perhaps in preventing their aggregation in the cytosol.

Although the EspG proteins seem to play a central role in recognition of the PE and PPE substrates in the cytosol, they do not interact with the Esx proteins (15). Perhaps the intrinsic solubility of Esx protein pairs makes a role for the EspG chaperone superfluous. EspG might therefore be a dedicated chaperone of PE-PPE proteins (and perhaps several Esp proteins such as EspB) that are more prone to aggregate.

**EccA**

EccA is the second cytosolic protein encoded by most mycobacterial T7S systems. EccA belongs to the AAA+ (ATPases associated with various cellular activities) protein family, members of which are involved in diverse processes, including protein degradation, signal transduction, and (dis)assembly of protein complexes. AAA+ proteins typically form ring-shaped hexamers with a central pore. EccA, also forms oligomers, possibly hexamers, upon overexpression in *E. coli* (19). ATP hydrolysis by these hexamers usually induces a conformational change in the complex that is transferred to the bound substrate(s). All EccA homologs are composed of two

---

**FIGURE 3** Crystal structures of T7S substrates. (A) Structure of the heterodimer EsxB (dark blue) and EsxA (light blue) of *M. tuberculosis* (3FAV). The two proteins form a four-helix bundle. The Tyr of the YxxxD motif and the Gly and Trp residues of the WxG motif that are postulated to together constitute the T7S signal are shown in red (54). (B) The EssA protein of *S. aureus* forms a homodimer that results in two putative secretion signals (YxxxD) on each end of the four-helix bundle (red) (2VRRZ) (139). (C) Crystal structure of PE25 (light blue) and PPE41 (dark blue) of *M. tuberculosis* in complex with the chaperone EspG5 (yellow). EspG interacts with the PPE protein through hydrophobic interactions but not directly with the PE protein. The WxG motif on the PPE and the YxxxD motif on the PE protein together form a putative T7S signal (red residues) (4KXR) (17). (D) Crystal structure of monomeric EspB visualizing an extended secretion signal that includes the YxxxD/E and WxG motif (red residues) (3J83) (18).
conserved domains joined by a linker region. The C-terminal domain of EccA is the ATPase domain and contains all the motifs that are characteristic for AAA+ proteins. In line with this, in vitro ATPase activity of EccA can be pinpointed to this C-terminal domain \((20)\). The N-terminus of EccA contains six tetratricopeptide repeat motifs, which are known for mediating protein-protein interactions in other proteins \((21)\). The structure of this N-terminal domain of \(M.\ tuberculosis\) EccA1 has recently been solved, showing an arrangement of these motifs in a right-handed superhelix \((19)\). Interestingly, the structure of this N-terminal domain of EccA1 resembles the structure of PilF, which is involved in the assembly of the type IV pilus system in \(Pseudomonas aeruginosa\) \((19)\).

EccA is important for T7S, because disruption of eccA genes in \(esx-1\) and \(esx-5\) clusters results in the loss of secretion of Esx and PE-PPE proteins \((22–24)\). In addition, a mutation in the ATPase domain of EccA1 disrupts secretion \((25)\), and EccA2 is, similar to other core components of the ESX-3 system, essential for viability \((26)\). However, EccA-independent secretion through ESX-1 and ESX-5 \((27, 28)\) (Houben and Bitter, unpublished results) has been observed. These results suggest either that EccA plays different roles in different mycobacterial species/strains or that the function of EccA is redundant in specific cases. Interestingly, EccA is, like EspG, not present in ESX-4. This protein therefore could also play a role in the secretion of specific substrates, such as the PE and PPE proteins. Perhaps analogous to the AAA+ protein in the type VI secretion system, EccA could be involved in the disassembly of EspG chaperones from the PE-PPE dimers \((16)\).

**THE T7S MEMBRANE COMPLEX**

After synthesis and folding in the cytosol, the T7S substrates should be guided to the inner membrane, where their transport across the mycobacterial cell envelope is initiated. For the other specialized secretion systems, a complex machinery is involved in the actual secretion process. In line with this, all five conserved T7S membrane components (EccB, EccC, EccD, EccE, and MycP) have been shown to be essential for protein secretion by ESX-1 \((27, 29, 30)\), ESX-5, and ESX-3 \((22, 27, 31)\). Most of these membrane components have large N- or C-terminal hydrophilic domains and only one or two predicted TMDs. EccD is the exception. This protein is highly hydrophobic, usually having 11 predicted TMDs and only a relatively small N-terminal hydrophilic part. Based solely on this characteristic, EccD has been postulated to form the membrane pore through which substrates are transported \((30)\), although there is no functional proof of this. The first evidence of the composition of the transport channel was provided by the observation that four of the conserved components of the ESX-5 system (i.e., EccB, EccC, EccD, and EccE) form a stable membrane complex of \(\sim 1.5\) MDa \((27)\). It is highly likely that this large complex forms the channel through which substrates are transported, although evidence of channel activity has not yet been provided. Because the substrates are probably secreted as (hetero) dimers, the translocation pore should be relatively large to allow passage of these folded structures. Of the four components of the membrane complex, only EccE is not present in all T7S systems; again, ESX-4 lacks this component. This could suggest that this protein is located at the periphery of the complex, which is supported by the observation that EccE is highly sensitive to proteolytic degradation when the membrane complex is treated with trypsin \((27)\).

The presence of classical TMDs with high hydrophobicity predicts that all the Ecc components are inserted in the inner membrane of the mycobacterial cell envelope. Because substrates also need to cross the outer membrane to end up in the extracellular environment, the question remains whether the observed membrane complex also inserts into the outer membrane. The identification of an outer membrane channel is complicated by the fact that our knowledge of mycobacterial outer membrane proteins is limited \((32)\). The only mycobacterial outer membrane protein that has been studied in detail, MspA, does show structural similarity to the outer membrane proteins of Gram-negative bacteria, because MspA spans the outer membrane using short \(\beta\)-sheet transmembrane domains that are organized in a so-called \(\beta\)-barrel. Although the size of the T7S membrane complex would allow it to span both the inner and the outer membrane, none of the subunits have a substantial domain with predicted \(\beta\)-sheets that could form a \(\beta\)-barrel. Perhaps outer membrane transport is mediated by other (unidentified) proteins that more loosely associate with the T7S membrane complex. Of course, other structures, like amphipathic \(\alpha\)-helices, such as those involved in formation of the type IV secretion complex \((33)\), could play a role. However, these latter unusual structures are more difficult to predict in silico.

Another option is that the T7S system is involved only in transport across the inner membrane and that outer membrane transport depends on a separate transport system, which would mean a two-step process.
Although hard evidence for a one-step secretion process across the complex mycobacterial cell envelope is still missing, there is currently also no data showing that the T7S substrates are exposed to the mycobacterial periplasm at any point during the translocation process (34).

**EccC**

Of the four subunits of the T7S membrane complex, only EccC has predicted functional domains; it contains three conserved nucleotide binding domains (NBDs). EccC is one of the most conserved T7S components; it is not only present in all mycobacterial T7S systems, but is also the only protein that is present in T7S-like systems in *Firmicutes*. Apparently, EccC homologs are central players in these secretion systems. While EccC is usually encoded by a single gene, the homolog in the ESX-1 system of *M. tuberculosis* is composed of two distinct proteins. However, it is most likely that these two EccC subunits together form a functional unit (30) and originated from a single gene. All three predicted NBDs of EccC show homology to members of the FtsK/SpoIIIE family of ATPases. This large protein family consists of three conserved nucleotide binding domains (NBDs). EccC is already multimeric as part of a large membrane complex in *Thermomonospora curvata* (38). Possibly, EccC performs similar functions in T7S by recognizing chaperone-substrate complexes at the membrane. Accordingly, EccC of the ESX-1 system was shown to interact with the substrate EsxB in both immunoprecipitation and yeast two-hybrid experiments. This interaction was dependent on the presence of the C-terminal secretion signal of EsxB (37). Although EccC proteins have three NBDs (38), mutational analysis of individual NBDs of EccC molecules of both mycobacteria and *Firmicutes* suggests distinct roles of each domain in secretion; while ATP binding to the first domain is essential for secretion, the other two domains are not strictly required for protein transport (39). In addition, ATP binding to any of the three NBDs of EccC is not sufficient for formation of the ESX-5 membrane complex (39), suggesting that ATP hydrolysis by EccC is not involved in complex assembly, but that it is dedicated to substrate recognition and/or transport.

Recent structural analysis of EccC from the thermophilic actinobacterium *Thermomonospora curvata* (Tc) with and without bound substrate TcEsxB provides important insight to the role of the individual NBDs in substrate recognition and transport (40). In this study, Rosenberg et al. described the structure of TcEccC in complex with a peptide representing the TcEsxB secretion signal, revealing that the signal binds to a hydrophobic pocket of NBD3, while both NBD2 and NBD3 were bound to ATP (Fig. 4). In contrast, NBD1 is visualized in a nucleotide-free state, suggesting that NBD2 and NBD3 are relatively inefficient ATPases. The hydrophobic binding pockets of NBD1 and NBD2 do not bind the secretion signal but are filled by a linker region of the adjacent NBD3. Mutating the linker of NBD2 that binds NBD1 (R543A) subsequently activated ATPase activity by NBD1, and this activity is increased upon binding of TcEsxB to NBD3. These results suggest that substrate binding activates a chain of events leading to ATPase activation. In line with the previously obtained data in mycobacteria and *Bacillus subtilis* (38, 39), the authors showed that mutations in catalytic residues of NBD1 completely abolished ATP hydrolysis, while similar mutations in NBD2 and NBD3 only reduced this activity. Rosenberg et al. provide evidence that TcEccC (R543A) multimerizes *in vitro* upon binding of TcEsxB, indicating that this substrate triggers both ATPase activity and multimerization. However, addition of TcEsxA, most likely the binding partner of TcEsxB, abolishes both multimerization and ATPase activity of this TcEccC construct. How the *in vitro* multimerization of EccC correlates with the settings in the bacterial cell, where EccC is already multimeric as part of a large membrane complex (27), remains to be investigated.

**Mycosin**

The only conserved membrane protein of T7S systems that is not part of the ESX-5 membrane complex is mycosin (MycP). Mycosins are subtilisin-like proteases with a classical signal sequence that presumably directs the protease domain to the periplasm. In addition, mycosins have a putative C-terminal TMD to anchor the protein in the membrane (41, 42). Although mycosins are essential for T7S, their role in secretion is still an enigma. The only known substrate for any mycosin is the ESX-1 substrate EspB, which is an ESX-1 secreted protein cleaved by MycP1 (43). In addition, structural analysis of EspB with and without the C-terminus is
cleaved by MycP₁ shows that MycP₁ might be involved in inducing a conformational change of the EspB quaternary structure (18). Although this finding suggests that mycosins are involved in substrate processing, the proteolytic activity of MycP₁ seems to be dispensable for the secretion process (43). Similarly, MycP₅ of the ESX-5 system is essential for ESX-5-dependent secretion, but its protease activity is not (van Winden, Houben, and Bitter, unpublished results). This surprising observation suggests that mycosins have a second role in secretion, besides their function as a protease. Further research is needed to understand this crucial additional role of mycosins in T7S.

Recently, the structures of both Mycobacterium smegmatis and Mycobacterium thermoresistibile MycP₁ were solved, which show a typical subtilisin core domain decorated with several extended loops (44, 45).

The most distinctive structural feature of mycosins is the presence of an N-terminal extension. Many subtilisins are produced with an N-terminal extension, called a propeptide, that prevents premature substrate access to the active site (46). However, classical subtilisin propeptides form a tightly folded structure near the active site, and this propeptide is usually degraded by the subtilisin-like proteases after folding and/or transport. The N-terminal extension of mycosins does not display homology with subtilisin propeptides and is structurally different as well, because it is wrapped around the protease domain and does not block the active site (44, 45). In addition, MycP₁ with an intact N-terminal extension is able to cleave EspB in vitro (44, 45). Therefore, the term “propeptide” for this N-terminal extension of mycosins is probably incorrect. Interestingly, the structure of M. smegmatis MycP₅ was solved, revealing a very
similar fold of the protease domain to the MycP₁ structure, including the N-terminal extension (45). The properties of the active site clefts are quite distinct between these two mycosins, suggesting different substrate specificities.

**T7S SUBSTRATES**

Although the different T7S systems that have been studied in detail secrete different classes of substrates, most of these substrates belong to the EsxAB clan (PFam CL0352) (47). This clan contains six protein families: Esx (WxG100), PE, PPE, LXG, DUF2563, and DUF2580. The best known of these is the Esx family. The first Esx protein that was discovered is the EsxA (ESAT-6) protein of *M. tuberculosis*, which is secreted as a heterodimer together with EsxB (CFP-10). Since then, members of this protein family have been described over a wide range of species, mostly in the phyla Actinobacteria and Firmicutes (11, 48), but have also been found in Verrucomicrobia, Lentisphaerae, Planctomycetes, Chloroflexi, and even some in Proteobacteria (47).

Esx proteins are also called WxG100 proteins (10), based on a short conserved motif, tryptophan-X-glycine (WxG), in the middle of the protein and their typical size of approximately 100 amino acids (Fig. 3A). Esx proteins form two helices separated by a turn, which contains the conserved WxG motif. Although most Esx proteins are small, some contain an extended C-terminal domain. The C-terminal domains of these extended WxG100 proteins are highly variable and, although largely uncharacterized, are predicted to have highly divergent functions. Esx proteins of mycobacteria, such as EsxA and EsxB, are secreted as antiparallel heterodimers (49, 50). These co-secreted dimers are usually encoded by adjacent genes and are part of the same operon (51). In contrast, some Esx proteins from *Streptococcus* (52) and *Staphylococcus* (53) exclusively form homodimers and are encoded by genes that are generally monocistronic (Fig. 3B). It has been suggested that the bicistronic heterodimers evolved after a duplication event, which suggests that the original substrates were homodimers (54).

The structure of the EsxAB complex (55) shows that the N- and C-termini of both EsxA and EsxB are predominantly unstructured. The longest of these unordered stretches, the C-terminus of EsxB, merits special attention since this region has been shown to contain a secretion motif that is required for secretion (see below). The overall structure of Esx dimers is also observed for other Esx proteins, such as EsxGH (56) and EsxRS (57).

**PE and PPE Proteins**

Two other major classes of the T7S substrates belonging to the EsxAB clan are the PE and PPE proteins. Although they belong to different protein families, these proteins form secreted heterodimers and, similar to the mycobacterial Esx substrates, their genes are often located adjacently on the genome. Therefore, they will be discussed together in this section. PE proteins are named for their conserved proline-glutamic acid motifs at positions 8 and 9 of the N-terminus of these proteins, whereas PPE proteins are defined by a proline-proline-glutamic acid motif at positions 7 to 9. PE and PPE proteins have only been described in members of the phylum Actinobacteria and are most widespread in the slow-growing species of mycobacteria (58). The structure of PE25 of *M. tuberculosis* in complex with PPE41 has been solved by Strong and colleagues (59) and shows an antiparallel dimer forming a four-helix bundle, similar to that of the Esx proteins. The recently solved crystal structure of EspB has revealed that this protein is similar to a PE-PPE dimer (Fig. 3D) (18). Interestingly, this protein multiplies as a heptamer, indicating that a similar confirmation could be formed by PE-PPE dimers.

PE proteins are characterized by a conserved N-terminal PE domain of approximately 110 amino acids, which forms a helix-turn-helix structure similar to the Esx proteins, but the turn is less defined compared to the Esx proteins and does not contain the WxG motif. The PE domain interacts with a PPE domain through conserved apolar residues that establish strong hydrophobic interactions (59). A secretion motif in the mostly unstructured C-terminal domain of PE proteins is, similar to EsxB, essential for secretion (54, 60). While the genes coding for the most ancient PE proteins are located within the esx loci and mostly consist of only the PE domain, more recently evolved members of this protein family have (large) extended C-terminal domains. For instance, *M. tuberculosis* contains 99 genes encoding PE proteins, of which 69 belong to the PE_PGRS group, named after polymorphic GC-rich repetitive sequences that code for the C-terminal domains. These C-terminal domains are composed of glycine-rich repeats and can be up to 1,550 residues long. These long PE proteins are secreted to the cell surface by a T7S system (i.e., the ESX-5 system) as well.

The PE domain, with approximately 180 amino acids, is slightly larger than the PE domain and contains five α-helices. The three N-terminal α-helices interact with the PE partner protein. A typical WxG motif is present in the turn between the second and the third α-helix, similar to that of the ESX proteins. The fourth
and fifth α-helices also pair together to form an extension of the PE-PPE dimer. This extended region forms the interaction site for the EspG chaperone (15, 17). Similar to PE proteins, PPE proteins can have (largely) extended C-terminal domains.

As mentioned previously, the interaction of the PPE41 protein with EspG5 is mediated by the conserved hydrophobic tip of the PPE proteins, mostly formed by α-4 and α-5. Bioinformatic analysis suggests that PPE proteins secreted by the different ESX systems in mycobacteria can be grouped on the basis of these conserved residues in the hydrophobic patch (17), and this phenomenon could explain how the EspG chaperones establish system specificity (15). Thus far, the PE and PPE proteins are the only T7S substrates shown to require such a specific chaperone.

### Other T7S Substrates

The EsxAB clan also contains the DUF2563, DUF2580, and LxG families. The first of these is a small Mycobacterium-specific protein family about which not much is known. Members of the second family, DUF2580, are restricted to the mycobacteria as well, but some members of this family are T7S substrates: both the ESX-1 substrates EspC and EspF belong to this family. Interestingly, the genes encoding these substrates are located in an operon with genes encoding other T7S substrates, i.e., EspA and EspE, respectively. Although EspA and EspE do not officially belong to the EsxAB clan, the structure prediction program (Phyre2) (61) predicts with high confidence a helix-turn-helix domain at the N-terminus of these proteins, with Esx proteins as the best template. Therefore, these proteins possibly form heterodimers with EspC and EspF. Another known ESX-1 substrate is EspB. Recently, the structure of this protein was elucidated, which showed that it has an N-terminal helix-turn-helix motif followed by a T7S signal motif (see below and Fig. 3D) (60). The major surprise was that the adjacent region of this domain showed structural homology to PPE proteins, which means that this protein forms a four-helix bundle by itself and therefore could be secreted as a monomer. The study by Solomonson et al. (18) further shows that these monomeric structures multimerize as a heptamer with a barrel-shaped structure. This suggests that PE-PPE dimers could form similar quaternary structures, which might have important implications to predict their functions.

**Firmicutes** have different T7S substrates as well, both EsxAB-like proteins and proteins that do not seem to belong to the EsxAB clan. Recently, in *Staphylococcus aureus* two new substrates were identified with the somewhat misleading names EsxC and EsxD. Although they officially do not belong to the EsxAB clan (and therefore also not to the Esx family), again structure prediction indicates a helix-turn-helix motif.

The final and perhaps most intriguing protein family within the EsxAB clan is LxG. This recently described extended protein family contains diverse proteins, including (putative) endonucleases and toxins (62). These members have an N-terminal LxG domain and a nucleosome domain (SUKH family) at the C-terminus. Members of this family are mainly found in the Firmicutes. Unfortunately, secretion of these proteins has not been studied. However, because they belong to the EsxAB clan, secretion via T7S seems likely. Supporting this prediction is the observation that homologs of SUKH endonucleases in other bacteria have different N-terminal domains that are linked to secretion; in Proteobacteria they are predicted to be secreted via the two-step secretion pathway, and in Actinobacteria, via a classical (Sec/Tat) secretion pathway.

### T7S Signal

How are the T7S substrates recognized, and what determines system specificity? The first indication of a secretion signal was identified for the EsxAB dimer. Deletion of the unstructured C-terminal tail of EsxB completely blocked secretion (37). Furthermore, in the same study it was shown that in yeast two-hybrid experiments the C-terminal tail of EsxB interacts with the EccC protein of ESX-1. This first study could not identify crucial residues within this C-terminal tail, but later it was shown that there was indeed a conserved consensus, albeit slightly more upstream (60). The presence of a C-terminal secretion signal was also shown for other T7S substrates, including EspC and PE25. Detailed analysis showed that two residues within the C-terminal tail of PE25 are crucial and that the spacing between these residues is important as well. The identified signal, YxxxxD/E, is crucial for the secretion of Esx, PE, and Esp proteins in mycobacteria (60). If we look at all T7S substrates from different organisms, a broader consensus must be used, as described by Poulsen et al. (54). Interestingly, although the secretion signal was first identified in the unstructured C-terminal tail of T7S substrates, more recent structural studies of Esx, PE, and EspB proteins showed that this signal is part of the elongated second helix. In these structures the two crucial residues, i.e., tyrosine and the acidic amino acid, are located on the same side of the helix (54). Furthermore, the tyrosine is positioned close to the conserved tryptophan residue of the WxG motif in the turn region.
of the dimer partner. It has been suggested that this tryptophan residue is therefore also part of the secretion signal (54). Because the deletion of several C-terminal residues beyond the YxxxD/E motif is already enough to block secretion, the signal probably extends at least 10 amino acids further than the consensus sequence, but without any clear conserved features.

Recent data (40) indicate that this unstructured tail is recognized by the C-terminal domain of EccC and is responsible for the multimerization of EccC. Probably, a similar secretion signal is present for T7S substrates in Firmicutes, but the role in secretion has not been studied in the same detail. Although the C-terminus was shown to be important for these Firmicutes substrates, the wrong amino acids were examined (i.e., an erroneous YxxxD/E motif), without any effect (54, 63).

One puzzling observation is that exchange of the secretion signal between an ESX-1 and an ESX-5 substrate restores secretion but does not redirect the substrate to another secretion system. Therefore, this secretion signal does not seem to determine system specificity. Apparently, a second signal is required for this. For the PE-PPE heterodimeric complex this second signal is probably provided by the chaperone EspG, which was shown to specifically recognize substrates of the cognate secretion system. However, it is still unknown what characteristic of the Esx dimers determines their systems’ specificity.

Other major questions concerning substrates and substrate specificity remain. For instance, the large majority of PE and PPE proteins in mycobacteria seem to be without an obvious partner protein. Are they secreted as single proteins, or are they also secreted as a heterodimer with an unknown partner? Furthermore, many of these PE and PPE proteins contain large C-terminal extensions. A good example is LipY, one of the few PE-PPE proteins with a known function. This protein contains a PE domain in Mycobacterium marinum, a PPE domain in Mycobacterium gilvum (13). Interestingly, whereas both the PE and PPE domain were shown to be interchangeable for secretion to the cell surface via T7S, exchange for a classical signal sequence did not result in surface localization. What is interesting to note is that many of the WxG (and LxG) proteins in Actinobacteria and Firmicutes have C-terminal protein domains with described functions as well, indicating that there might be overlapping structural and mechanistic similarities between these proteins and the PE-PPE proteins. Unfortunately, no protein structures are known for these putative substrates, and therefore many questions about the folding, chaperones, secretion partners, and functions of these proteins remain.

THE ROLE OF T7S SYSTEMS IN VIRULENCE

ESX-1 System of Pathogenic Mycobacteria

The specific roles of T7S systems and their substrates have been best described in mycobacteria. The ESX-1 system of M. tuberculosis was the first T7S system identified; this region was lost when M. bovis was cultured for 11 years by Calmette and Guerin to create the live attenuated-vaccine strain M. bovis BCG (51, 64). It was named region of difference 1 (RD1), and Pym et al. showed that this region is (mainly) responsible for the attenuation of M. bovis BCG and the vole bacillus Mycobacterium microti (65).

Since then, the importance of the ESX-1 system for the virulence of mycobacteria has been strongly established (29, 66–68), but the role of the individual secreted substrates has been more difficult to investigate. The first ESX-1 substrates that were identified were EsxA and EsxB (3), which are major secreted proteins and dominant T-cell antigens. As such, they were regarded as the principle substrates and responsible for the attenuation of M. bovis BCG. The EsxA protein seemed to be the prime candidate for mediating phagosomal escape (see below). Later it was shown that more proteins are secreted via this system; these proteins are known as ESX-1-associated proteins (Esp). EspA and EspB, in particular, seem to be involved in virulence as well (18, 69–71). A complicating factor in determining the role of individual components in virulence is their codependence for secretion. Secretion of EspA and possibly EspB is shown to be linked to functional secretion of EsxAB and vice versa (68). Redirection of ESX-1 substrates through other T7S systems could help to solve this conundrum.

Virulence

Although the interdependence of ESX-1 substrates makes the study of individual substrates difficult, major steps have been taken to elucidate the function of ESX-1 and its substrates in virulence. The ESX-1 system plays an important role in the macrophage infection cycle of pathogenic mycobacteria. Several species of Mycobacteria, including M. tuberculosis, M. marinum, and M. leprae, have been shown to translocate from the phagosome to the cytosol in late stages of the macrophage infection cycle (72). Upon translocation to the cytosol, bacteria start to replicate and ultimately induce a
necrosis-like cell death. After necrosis of the host cell the bacteria can spread to neighboring macrophages within the host (73). Therefore, escape into the cytosol is essential for full virulence. This crucial step is dependent on a functional ESX-1 system (72), since phagosomal escape is not observed for M. bovis BCG or esp-1 mutants of M. tuberculosis and M. marinum (72, 74). Furthermore, it has also been shown that mycobacteria can lyse erythrocytes in an ESX-1- and contact-dependent manner (24, 75). This phenomenon is seen as a direct effect of the membrane-disrupting potential of ESX-1 substrates (75, 76). Several studies have implicated EsxA as the main membrane-disrupting protein of the ESX-1 system (76, 77), although some questions remain. Most importantly, EsxA and EsxB are present in nonpathogenic mycobacteria as well as pathogenic mycobacteria and seem to have similar functions (78). Furthermore, Mycobacterium kansasii has several subtypes, of which only one shows phagosomal escape and a concomitant increased virulence. Avirulent subtypes of M. kansasii do efficiently secrete EsxA (74). Perhaps different ESX-1-dependent substrates are together required for phagosome escape. EspA would be an important candidate for this, since this protein is highly upregulated within the phagosomal environment. Another important substrate is EspB, which was shown to be crucial for the virulence of M. marinum. Interestingly, this protein forms ring-shaped heptamers with a hydrophobic domain (18). As such, it could be involved in perturbation of host cell membranes.

In M. marinum, ESX-1 proteins have been shown to localize in the capsular layer at the cell surface (2). This surface localization of ESX-1 substrates would be in line with the membrane-disrupting capacities of mycobacteria (79). Whether this phenomenon is the same for M. tuberculosis remains to be established. Furthermore, the components of the ESX-1 machinery have been shown to be localized to the cell poles of multiple species of mycobacteria (80, 81). The ESX-1 system and its substrates seem to be enriched in new cell poles with active peptidoglycan synthesis, indicating that there might be a role for the ESX-1 system in cell wall growth (81).

ESX-1 Regulation

The ESX-1 system is active in culture medium but is upregulated when mycobacteria encounter host cells. This process is regulated by multiple transcriptional regulators. The best-described transcriptional regulator is the two-component sensor kinase PhoPR system. The PhoPR system is activated by a decrease in pH (82), as well as increased Cl⁻ concentration (83), which are conditions that occur in maturing phagosomes. The avirulent M. tuberculosis strain H37Ra contains a single point mutation in the DNA binding domain of PhoP, which abrogates DNA binding and results in reduced virulence (84). PhoP was later shown to regulate secretion of ESX-1 indirectly through regulation of the espACD locus (85).

Another gene that is part of the PhoP regulon is whiB6 (Rv3862c), which is a transcriptional regulator situated in close genetic proximity to the esp-1 gene cluster (86). In clinical M. tuberculosis strains, PhoP binds to the promoter region of whiB6, which induces transcription. Increased production of WhiB6 positively regulates the expression of several esp-1 genes and in that way is able to increase ESX-1 secretion. Point mutations in the promoter region of whiB6 in laboratory strains H37Rv and H37Ra result in an inverted regulation of whiB6 by PhoP, making PhoP a negative regulator of ESX-1 secretion (86).

The espACD operon is also regulated by other factors, one of which is EspR. This protein was discovered as a regulator of ESX-1 secretion and was first hypothesized to be an ESX-1 substrate itself (87) but was later confirmed to be a more general nucleoid-associated protein that is not secreted (88). EspR is a transcriptional regulator with a unique helix-turn-helix structure and a C-terminal domain that is essential for dimerization. In its dimerized form it binds two specific operator sites situated 177 bp apart, thereby probably creating a loop in the promoter region of espACD (89).

Finally, the espACD operon is also regulated by the two-component regulator MprAB (69) and the regulator Lsr2 (90). Altogether, a picture is emerging of a secretion system that is tightly regulated by multiple layers of regulation adapted to different conditions. These networks can be highly relevant for the virulence capacities of clinical strains, since small mutations in regulators or promoter regions can have significant effects on virulence (84, 91).

Role of ESX-1 and T7S in Horizontal Gene Transfer

Although the role of the ESX-1 system in pathogenic mycobacteria is tightly linked with intracellular routing and host cell death, a completely different role has been described in M. smegmatis. This bacterium is capable of horizontal gene transfer, in which large genomic fragments are exchanged between different strains (reviewed in reference 92). This process is an atypical form of conjugation and is called distributive conjugal transfer. Interestingly, this process is dependent on
ESX-1 in different and opposing ways: esx-1 mutants of donor strains are shown to be hyperconjugative (78), whereas esx-1 mutants in the recipient strains are hypoconjugative (92). This paradoxical dual role of ESX-1 in mycobacteria has not been mechanistically explained yet, but it will be interesting to discover if DNA is transported through this T7S system or if secreted proteins located at the cell surface play a role in cell-cell adhesion. An interesting observation is that the hyperconjugative phenotype of the M. smegmatis esx-1 mutants can be complemented with its M. smegmatis counterparts. However, distributive conjugal transfer has not been described in M. tuberculosis or other slow-growing mycobacteria (94, 95).

The ESX-1 system of M. smegmatis is not the only ESX system that has been linked to DNA transport. Recently, a conjugative plasmid (pRAW) was discovered in M. marinum (8). This plasmid efficiently conjugated between different slow-growing species of Mycobacteria, including M. tuberculosis. No conjugation to fast-growing mycobacteria was observed, indicating specific genetic requirements for recipient strains. The pRAW plasmid was shown to contain a newly identified T7S system called ESX-P1. This ESX system was shown to be essential for conjugal transfer of pRAW. ESX-P1 has the highest homology with the ESX-5 system (8).

ESX-3

The ESX-3 system is essential for the growth of M. tuberculosis, as was shown by both directed mutagenesis and saturated transposon mutagenesis (26, 96). ESX-3 is regulated by iron- and zinc-dependent transcription regulators ideR (97) and Zur (FurB) (98). Siegrist et al. have shown that expression of the esx-3 system is highly upregulated when iron chelators are added to the growth medium (99). Additionally, several groups have reported that ESX-3 is essential for growth only under iron-limiting conditions (99, 100). It is postulated that a functional ESX-3 system is crucial for the uptake of iron via mycobactin, a siderophore produced by mycobacteria, but a mechanism for this phenomenon remains to be elucidated. In conditions of high zinc availability, the essentiality of ESX-3 can be rescued as well. The most efficient complementation of ESX-3 essentiality is achieved by supplementing both iron and zinc, which suggests that mycobactin uptake is not the only function of the ESX-3 system (100). There seems to be only a partial correlation of functional ESX-3 secretion and mycobactin uptake, indicating that the role of ESX-3 is probably broader than metal ion acquisition. It is somewhat counterintuitive that a protein secretion system should be involved in the uptake of metal ions, but perhaps ESX-3 substrates play a role in the uncoupling of the Fe-mycobactin complex or in the actual binding of this siderophore at the cell surface.

The only known substrates of the ESX-3 system are EssG and EssH, which are both encoded by the esx3 locus. These two proteins are secreted as a heterodimer, just like their ESX-1 homologs EssAB. Two other putative ESX-3 substrates are PE5 and PPE4, which are encoded by the esx-3 cluster and which are also essential. However, experimental evidence for the secretion of these proteins via ESX-3 is lacking.

ESX-5

The ESX-5 system is probably the most recently evolved T7S system in mycobacteria and is only present in slow-growing species (58). The slow-growing mycobacteria include most pathogenic species, such as M. tuberculosis, M. leprae, M. marinum, Mycobacterium ulcerans, and Mycobacterium avium. The evolution of the ESX-5 system, probably through a duplication of the esx-2 gene cluster, was followed by a large expansion of the pe and ppe genes in species such as M. tuberculosis (101) and M. marinum. Genomic analysis suggests that ESX-5 could be responsible for the secretion of most PE and PPE proteins and that the ESX-5 system plays an important role in virulence and/or the slow-growing phenotype of mycobacteria. Different biochemical methods have experimentally validated the first hypothesis; indeed, many PE and PPE proteins of M. marinum and M. tuberculosis are secreted via ESX-5, especially the recently evolved subclasses of PE and PPE proteins, such as the PE-PGRS and PPE-MPTR proteins (14).

After secretion these proteins can be identified mainly at the cell surface but also in the culture filtrate (14, 22, 102). There are indications that a number of these surface-localized PE-PPE proteins play a direct role in virulence by interacting with host immune receptors, such as Toll-like receptor (TLR) 2 and the inflammasome (reviewed in reference 103), but recent data indicate that there could be an important second role for ESX-5 substrates. Just like for ESX-3, the membrane components of the ESX-5 system are essential for the in vitro growth of M. bovis BCG, M. marinum (39), and M. tuberculosis H37Rv on culture plate or broth (96, 104). However, this essentiality is not observed in all M. tuberculosis strains; for instance, in CDC1551 esx-5 mutations can be readily identified (22, 27). Recent data could explain this apparent discrepancy; the essentiality of esx-5 is linked to the permeability of the outer membrane. The ESX-5 system is no longer essential when
outer membrane permeability is increased, either by mutations in lipid biosynthesis or by the introduction of the pore-forming protein MspA from *M. smegmatis*. Slow-growing mycobacteria do not produce MspA-like pores and as such probably need alternative mechanisms for nutrient transport over the outer membrane. The ESX-5 system and its substrates are responsible for the uptake of fatty acids and probably other nutrients (39). The expression of MspA-like pores makes mycobacteria more vulnerable to defense systems of the host (105), which suggests that the slow-growing mycobacteria utilize ESX-5 substrates to acquire nutrients while maintaining resistance against bactericidal host factors.

In *M. marinum*, esx-5 mutants with a strongly reduced secretion phenotype (recall that knock-out mutants are not viable) are attenuated in cell-infection experiments and show delayed phagosomal escape. The ESX-5 system also seems to be involved in inducing cell death through inflammasome activation via an unknown mechanism (106). An espG5 transposon mutant of *M. marinum* is attenuated in zebra fish embryo experiments, but surprisingly, this mutant is hypervirulent in adult zebra fish (107). This hypervirulence is characterized by increased bacterial numbers in the organs, an altered pro-inflammatory cytokine response, and more severe immunopathology (107). This suggests that the ESX-5 system might be responsible for manipulation of the adaptive immune response, which is not developed in zebra fish embryos. However, the same effect was observed in rag⁻²⁻ zebra fish that lacked B-cells and T-cells, indicating that antigen-specific adaptive immune responses are probably not responsible for the hypervirulence of the esx-5 mutant. Additionally, expression of dormancy-related genes was not clearly affected in the esx-5 mutant strain, indicating that hypervirulence is not due to an inability to enter the dormant phase of infection (107). In contrast to these results, Bottai et al. (22) showed that an eccD₅ mutant of *M. tuberculosis* has reduced cell wall integrity and is significantly attenuated in a SCID mouse model.

One of the few ESX-5 substrates with a well-studied and-defined function is the LipY protein. This protein is a PE protein in *M. tuberculosis* but a PPE protein in *M. marinum*, illustrating that the PE-PPE domains are probably not involved in the function of these proteins but are responsible for the secretion (13, 102). LipY is a lipase, which can efficiently degrade long chain triacylglycerols that are present in the host cell or stored inside the mycobacteria (108, 109). Utilization of triacylglycerols is important during latent infection, and expression of LipY is upregulated in *in vitro* models of dormancy (109). *M. bovis* BCG overexpressing *lipY* was shown to lose its capacity to induce protection against *M. tuberculosis* (110), and an *M. tuberculosis* strain which overexpressed LipY was hypervirulent in a mouse model (111). These data indicate that LipY is an important virulence factor in *M. tuberculosis*.

### PE_PGRS

As mentioned previously, the major group of PE proteins is the so-called PE_PGRS proteins, which are characterized by a long C-terminal region with glycine-rich repeats. Notably, PE_PGRS proteins are not present in all ESX-5-containing mycobacteria. *M. avium* and *Mycobacterium xenopi*, for instance, do not contain any PE_PGRS protein (58). PE_PGRS proteins have been hypothesized to be involved in several pathogenesis-related or immunological mechanisms. The PE_PGRS wag22 was proposed to contain a C-terminal fragment with fibronectin binding properties (112). In a subsequent study the wag22 gene was shown to be a vaccine candidate that prevents reactivation of *M. tuberculosis* infection in a mouse model (113). PE-PGRS proteins have also been proposed to protect against proteasomal degradation and therefore reduced activation of CD8⁺ T-cells (114, 115). The glycine-alanine-rich repeat regions of PE_PGRS proteins are similar in composition to the EBNA1 protein of the human Epstein-Barr virus. The EBNA1 protein was shown to be a virulence factor of Epstein-Barr virus that is able to block its own proteasomal processing (116), and the same phenomenon was shown for PE_PGRS proteins, suggesting a role in immune evasion for these proteins (114, 115). However, in recent publications it has been shown that repetitive protein sequences of EBNA1 are not involved in preventing antigenic presentation. Instead, expression of these proteins seems to be self-limited by the mRNA sequence coding for these proteins, which is very rich in purines. These unstructured, purine-rich mRNA sequences make translation efficiencies high enough to create sufficient protein levels for infection but low enough to avoid immune recognition (117, 118). These results question the function of PE_PGRS proteins in blocking epitope processing.

The PE_PGRS proteins have also been hypothesized to be a source of antigenic variation in mycobacteria (119). However, a conserved hierarchy of immune recognition of various PE-PPE proteins suggests that differential immune responses are driven not by antigenic variation but by infection-stage-specific expression of these proteins (120). Additionally, sequencing of multiple clinical strains and their *pe_pgrs* sequences has led to
the observation that *pe_pgrs* diversity as a whole is not driven by antigenic pressure (121, 122). In fact, bio-informatic analysis indicates that the PE_PGRS proteins contain very few epitopes. Of the 1,649 known epitopes of *M. tuberculosis*, only 3 are situated in the PGRS domain of PE_PGRS proteins (121). Nonetheless, there seems to be antigenic pressure on some of the *pe_pgrs* genes, while others seem to evolve neutrally. This suggests that the group of PE_PGRS proteins may not be as homogenous as sometimes thought and that individual genes or subgroups of these genes and proteins of these groups need to be examined for their biological roles.

One of the PE_PGRS proteins that has been studied in more detail is PE_PGRS33 (Rv1818c). Polymorphisms in *pe_pgrs33*, including large insertions, deletions, or truncations, correlate with an absence of lung cavities in patients (123). As shown for other PE proteins, the PE domain of PE_PGRS33 is necessary for the surface localization of this protein (124, 125). The PGRS domain of PE_PGRS33 is shown to localize to the mitochondria of the host cell and is able to induce apoptosis (126). Surprisingly, the full-length PE_PGRS33 was also able to induce necrosis, showing that the PE and linker domains can perform additional functions in this mechanism. Another study showed that PE_PGRS33 was able to induce host-cell apoptosis by specifically binding to the TLR2 (127).

Together, these data show that the PE_PGRS proteins are important manipulators of the host immune response. However, the role of the PE_PGRS proteins remains to be fully elucidated. Understanding the role of these proteins will be a major step in increasing our knowledge of *M. tuberculosis* and will most likely contribute to our understanding of disease pathogenesis as well as vaccine design.

**TYPE VII(-LIKE) SECRETION IN MONODERM BACTERIAL SPECIES**

T7S systems are widely present in mycobacteria and closely related bacteria. Mycobacteria are characterized by a diderm cell envelope, the outer membrane of which is characterized by the presence of unusual lipids known as mycolic acids. These mycolic acids are partially linked to the arabinogalactan layer and partially coupled to trehalose molecules. Mycolic acids are not unique for mycobacteria; they can also be found in several close relatives, including Corynebacteria and Rhodococcus species. These bacteria therefore have a diderm envelope similar to mycobacteria and usually have a locus potentially coding for a T7S system (128). Unfortunately, T7S systems have not been analyzed in these bacteria.

Interestingly, a number of monoderm bacteria also have T7S(-like) systems, and some of these have been studied in more detail. The first one we will discuss is the T7S system present in streptomycetes. Although these bacteria are monoderm and do not produce mycolic acids, they do belong to the same taxonomic order (*Actinomycetales*) as mycobacteria and have a well-conserved T7S system; in *Streptomyces* species all genes coding for the five conserved membrane proteins, i.e., homologs of EccBCDE and MycP, are usually present (128, 129). Several substrates seem to be present in these loci, of which the two small Esx proteins have indeed been shown to be secreted by *Streptomyces coelicolor* (129). In two species of *Streptomyces*, i.e., *S. coelicolor* and *Streptomyces scabies* (129, 130), deletion of genes coding for the Esx-like T7S substrates resulted in abnormal spore formation and/or altered timing of spore formation. Surprisingly, absence of the T7S components did not show the same effect as the substrates, which seems to suggest that these Esx proteins have an intracellular effect on spore formation.

A number of *Firmicutes* species have a T7S-like system that contains only two of the original genes, i.e., an *eccC* homolog and one or more of the *esx* genes. In addition, there are usually a variable number of other (membrane) proteins required for secretion (discussed below). In contrast to the T7S systems in mycolic-acid-containing bacteria, the T7S-like systems in *Firmicutes* were probably acquired relatively recently, because their composition and distribution are highly variable. For instance, *Streptococcus agalactiae* is the only species in the *Streptococcaceae* family that has an intact T7S-like system, and the T7S-like system of *Bacillus cereus* is more similar in composition and homology to that of *Listeria* species than those of other *Bacillus* species. Most data on T7S-like systems in *Firmicutes* has been obtained for *B. subtilis* and *S. aureus*, so these systems will be discussed in more detail.

Although protein secretion in *B. subtilis* has been studied for decades, only recently, a T7S-like system (Fig. 1) was shown to be functional, which seems to be due to defective regulation in domesticated *Bacillus* strains (53). In undomesticated strains secretion could be readily observed in late log phase. The secreted substrate is an Esx-like protein known as YukE (53). Proteomic analysis showed that, under normal laboratory conditions, this is the only T7S-like substrate (12). YukE is secreted as a dimer and, in analogy to the mycobac-
terial T7S substrates, the C-terminal tail is required for secretion. Unfortunately, the extracellular function of YukE is currently unknown. Detailed mutational analysis has shown that, in addition to the EccC-homolog (called YukB or YukBA), YukCD and YueBC are also required (12, 53). Apart from YukD, these proteins are all membrane proteins. Of these membrane proteins, YueB is the one with most potential TMDs (6). Because this protein probably forms a dimer (131), this complex could be somewhat similar to EccD of mycobacteria. YueB-like proteins therefore could form a central component in T7S-like systems. In contrast to EccD, the YueB protein has a very substantial extracellular domain of more than 800 amino acids. This extracellular domain is a receptor for the bacteriophage SPP1 (131). The only cytosolic protein that is essential for secretion, YukD, has a ubiquitin-like fold (132). Despite this structural similarity, YukD does not seem to be conjugated, either to itself or to other proteins (132). For the YukD homolog of S. aureus (SAUSA300_0281, also described as EsaB [11] [Fig. 1]), it has been shown that this protein is involved in intracellular posttranscriptional regulation of a specific substrate of the T7S-like system (133). Therefore, this ubiquitin-like protein perhaps functions as a specific chaperone or regulator.

The T7S-like system of S. aureus is called Ess and contains the same elements as the T7S-like system in B. subtilis, but the gene order is different (Fig. 1). The only difference is that in S. aureus an additional gene is identified that might be required for secretion, i.e., esaD (134). However, the effect of an esaD deletion on secretion is not absolute; a major reduction of protein secretion is observed when this gene is deleted, but not a complete blockade. Moreover, in other species esaD homologs are not always linked to this secretion system, and the C-terminal domain of ESAd is predicted to form a nuclease/hydrolase. Therefore, more research is needed to substantiate this component. Another feature of T7S-like systems that is conserved between B. subtilis and S. aureus is the variability in secretion between strains. Both the amount and timing of secretion is highly strain dependent for S. aureus (135), which indicates that this system is controlled by a complex and variable regulation system. Some of these regulators have been identified; the production of the main substrate EsxA, in particular, seems to be controlled by several regulators (63, 136).

The most important difference between the T7S-like systems in S. aureus and B. subtilis is the number of substrates that are secreted. Whereas B. subtilis secretes only a single substrate, four substrates have been identified thus far for S. aureus. Two Esx proteins are secreted, as well as two proteins that do not officially belong to the EsxAB clan (63) but are predicted to form an N-terminal helix-turn-helix motif. These new substrates are unfortunately called EsxC and EsxD. All these substrates are encoded by genes within the T7S-like locus.

Ess secretion is important for both colonization and persistence of S. aureus (11, 135). Disruption of this T7S-like system or deletion of the substrates results in significantly reduced bacterial numbers and reduced pathology in a mouse model. Interestingly, a recent report indicates that this secretion system is involved in the intracellular survival of S. aureus in epithelial cells by blocking apoptosis and promoting escape (137). If substantiated, this would be highly similar to the function of the ESX-1 system in pathogenic mycobacteria. In line with this function, close homologs of the S. aureus Ess system are present in other pathogens, such as Listeria monocytogenes and B. cereus.

CONCLUDING REMARKS

In approximately one decade of research on T7S, major advances have been made to elucidate the mechanism of T7S and its role in pathogenesis. However, many questions remain about the functioning and structure of T7S systems. Is the T7S membrane channel in mycolic-acid-containing bacteria a double membrane-spanning complex, mediating transport in a one-step mechanism, or are there unidentified outer membrane components of T7S systems that are responsible for transport over the specific outer membrane? Solving the structure of the mycobacterial T7S membrane complex will answer many of these questions and would be a major breakthrough in this field. Another topic that needs to be addressed is the function of individual T7S substrates. Unfortunately, the interdependence of substrates for secretion and complex regulatory and posttranslational networks complicates research on this topic. The observed roles of T7S systems and their substrates in virulence and in nutrient acquisition show that these systems are promising targets for future drug development. Recently, the first T7S inhibitors have been identified and could be the front-runners of new classes of drugs against pathogenic mycobacteria (138). A more detailed understanding of T7S systems and their substrates could lead to new concepts for treatment of mycobacterial diseases but might also shed light on novel potential virulence factors secreted by other Actinobacteria and Firmicutes.
REFERENCES


