LOSt and found across the mycobacterial cell envelope
van der Woude, A.D.

2013

document version
Publisher's PDF, also known as Version of record

Link to publication in VU Research Portal

citation for published version (APA)
van der Woude, A. D. (2013). LOSt and found across the mycobacterial cell envelope.
Specific chaperones for the type VII protein secretion pathway

Maria H. Daleke1,2*, Aniek D. van der Woude1,2*, Annabel H.A. Parret3, Roy Ummels1, A. Marit de Groot2, David Watson2, Sander R. Piersma4, Connie R. Jiménez4, Joen Luijink2, Wilbert Bitter1,2, Edith N.G. Houben1,2

1Department of Medical Microbiology and Infection Control, VU University Medical Center, 1081 BT Amsterdam, The Netherlands; 2Section of Molecular Microbiology, Department of Molecular Cell Biology, VU University, 1081 HV Amsterdam, The Netherlands; 3European Molecular Biology Laboratory - Hamburg Outstation, 22603 Hamburg, Germany; 4OncoProteomics Laboratory, Department of Medical Oncology, VU University Medical Center, 1081 HV Amsterdam, The Netherlands

*These authors contributed equally to this work

J Biol Chem. 2012 Sep 14;287(38):31939-47 © the American Society for Biochemistry and Molecular Biology
Specific chaperones for the type VII protein secretion pathway

**Background:** Pathogenic mycobacteria use the type VII secretion systems (T7SS) ESX-1 and ESX-5 to secrete virulence factors, but it is unknown how these systems recognize their cognate substrates.

**Results:** Pulldowns identified specific interactions between cytosolic components of ESX-1 and ESX-5 and subsets of cognate substrates.

**Conclusion:** T7SS substrates interact with associated cytosolic secretion system components.

**Significance:** Cytosolic chaperones contribute to system specificity in T7SS.

Mycobacteria use the dedicated type VII protein secretion systems ESX-1 and ESX-5 to secrete virulence factors across their highly hydrophobic cell envelope. The substrates of these systems include the large mycobacterial PE and PPE protein families, which are named after their characteristic Pro-Glu and Pro-Pro-Glu motifs. Pathogenic mycobacteria secrete large numbers of PE/PPE proteins via the major export pathway, ESX-5. In addition, a few PE/PPE proteins have been shown to be exported by ESX-1. It is not known how ESX-1 and ESX-5 recognize their cognate PE/PPE substrates. In this work, we investigated the function of the cytosolic protein EspG, which is essential for ESX-5-mediated secretion in *Mycobacterium marinum*, but for which the role in secretion is not known. By performing protein co-purifications, we show that EspG interacts with several PE proteins and a PE/PPE complex that is secreted by ESX-5, but not with the unrelated ESX-5 substrate EsxN or with PE/PPE proteins secreted by ESX-1. Conversely, the ESX-1 parologue EspG interacted with a PE/PPE couple secreted by ESX-1, but not with PE/PPE substrates of ESX-5. Furthermore, structural analysis of the complex formed by EspG and PE/PPE indicates that these proteins interact in a 1:1:1 ratio. In conclusion, our study shows that EspG and EspG interact specifically with PE/PPE proteins that are secreted via their own ESX systems, and suggests that EspG proteins are specific chaperones for the type VII pathway.

Bacterial pathogens use dedicated protein secretion systems to export virulence factors that interfere with cellular processes of the host to ensure bacterial survival, multiplication and spread [14]. Also pathogenic mycobacteria, such as *Mycobacterium tuberculosis*, the causal agent of tuberculosis, secrete proteinaceous virulence factors. These bacteria are surrounded by a unique diderm cell envelope that requires a specialized secretion pathway, known as the ESX or type VII secretion pathway, to facilitate protein export [1].

*M. tuberculosis* contains five ESX clusters, designated ESX-1 to ESX-5. The most well-studied locus, ESX-1, is crucial for the virulence of *M. tuberculosis* and secretes two virulence factors of the WXG100 protein family, *i.e.* EsxA and EsxB. In addition, ESX-1 secretes a number of proteins referred to as ESX-1 secretion-associated proteins (Esp) [5, 8, 13, 15, 21, 25] and a few PE/PPE proteins [12, 15, 25]. *Peppe* genes, which are named after conserved Pro-Glu and Pro-Pro-Glu motifs near the N termini of their gene products, are present in high numbers in the genomes of several mycobacterial pathogens and are thought to contribute to mycobacterial virulence [24]. Based on the presence of specific motifs in
their C termini, the PE and PPE proteins are further divided into subfamilies, of which the PE_PGRS subfamily is the largest [18]. Although some PE/PPEs are secreted via ESX-1, ESX-5 is responsible for the secretion of most of these proteins, including many PE_PGRS proteins [2, 3, 6, 11]. The ESX-5 locus is restricted to the genomes of slow-growing mycobacteria. Importantly, all ESX substrates identified to date lack classical signal peptides, and the mechanism of substrate targeting is not yet fully understood [9, 13]. Recently, we demonstrated that the C termini of PE proteins and several other ESX-1 substrates share a conserved YxxxD/E motif that is required for secretion. Intriguingly, this signal does not discriminate between the various ESX pathways [12].

The ESX clusters contain a number of conserved genes, termed ESX conserved components (ecc), which are required for secretion [5]. In addition, there are components that are only present in one or a few ESX systems, one of which is EspG. EspG was originally thought to be specific for ESX-1, -2 and ESX-3, but a homologue with low similarity is also present in the ESX-5 locus. Disruption of this homologue in M. marinum (mmar_2676) blocked secretion via ESX-5 [3] and affected intracellular levels of ESX-5 substrates [28]. Here, we have investigated the role of MMAR_2676/EspG in ESX-5 secretion and we show that this protein selectively interacts with PE/PPE proteins.

**EXPERIMENTAL PROCEDURES**

**Bacterial strains and growth conditions**

- M. marinum wt strain E11 [22] and the ESX-5 mutants 7C1 (espG::tn) [3] and Mx2 (eccA::tn) [2] were grown and electroporated as described [11]. *Escherichia coli* BL21(DE3), BL21(DE3) pLysS, DH5α, TOP10F' and XL10-Gold were grown at 37°C in LB medium with shaking at 200 rpm, or on LB agar. When required, ampicillin was used at a concentration of 100 μg/ml, chloramphenicol at 30 μg/ml, kanamycin at 25 μg/ml and hygromycin at 50 μg/ml for mycobacteria and 100 μg/ml for *E. coli."

**Plasmid construction**

PCR reactions were carried out with the Phusion High-Fidelity DNA polymerase (Finnzymes) using primers listed in Table S1. The mmar_2672-2676 operon was amplified by PCR from *M. marinum* E11 genomic DNA using 5′ primer AbMmMiddle-H5-F and 3′ primer 1794-His-SpeI-R, which contains a 6xHis-encoding sequence and a SpeI restriction site. This fragment was cloned in the integrative pUC-Int-cat vector [3], resulting in pUC-Int-cat::Mm2672-76His, which was used to express C-terminally His-tagged EspG under the ag85 promoter in *M. marinum*. Mmar_1460 (ppe33) was amplified from E11 DNA using primers PPE1460-XbaI-F and PPE1460-HindIII-HA-R, which contained a 3′ HA epitope. After digestion with XbaI, this fragment was ligated into the *E. coli*-mycobacterial shuttle vector pSMT3::LipY, which had been digested with NheI and EcoRV to remove the lipY gene, resulting in vector pSMT3::Mm1460-HA.

For expression in *E. coli*, espG and espG were cloned both with and without C-terminal His tag under control of the tet promoter in the pASK-IBA3c vector (IBA Gmbh). The genes were amplified using a 5′ primer containing an *E. coli* ribosomal binding site (RBS) and XbaI digestion site (1794-RBS-XbaI-F or EspG-RBS-XbaI-F) and a 3′ primer with HindIII digestion site (1794-R, 1794-His-HindIII-R or EspG1-His-HindIII-R). After digestion, the fragments were ligated into XbaI and HindIII-digested pASK-IBA3c, resulting in pIBA::EspG5, pIBA::EspG5-His, pIBA::EspG5-His. For expression of PE25 and C-terminally 6xHis-tagged PPE41 in *E. coli* a previously described construct was used in which the rv2431c and rv2430c...
LOSt and found across the mycobacterial cell envelope

genes were each placed behind a RBS, i.e. pET29b(+):Rv2430-31c-His [26]. To create an unlabeled version of PPE41, the rv2430c gene was amplified with primers pET41_F and pET41noHis_R containing NcoI and HindIII sites from pSMT3::PE25-HA-PPE41 [12]. This fragment was subsequently used to replace the rv2430c-His sequence in pET29b(+):Rv2430-31c-His, thereby generating pET29b(+):Rv2430-31c. Similarly, a fragment coding for C-terminally His-tagged PPE41 without its last 20 aa was amplified from pSMT3::PE25-HA-PPE41 with primers pET41_F and pET41d20C_R. Subsequently, replacement of the rv2430c-His fragment in pET29b(+):Rv2430-31c-His as described above resulted in pET29b(+):Rv2430-31cΔ20C-His. Plasmids for E. coli expression of PE35 and PPE68_1, with and without a C-terminal His tag, were constructed by first amplifying mmar_0185 from pSMT3::MmPE35-HA-MmPPE68_1 [12] with primers pET35_F and pET35_R containing an NdeI and a KpnI site, respectively. This fragment was ligated into pET29b(+)::Rv2430-31c-His after NdeI/KpnI digestion, thereby replacing the rv2431c gene. Subsequently, the rv2430c-His fragment of this vector was substituted for either wt or His labeled PPE68_1-encoding fragments, amplified with the 5’ NcoI site-containing primer pET-31cF, which contained a PstI restriction site, and either pET68noHis_R or pET68_R, both with HindIII sites, from pSMT3::PE35-HA-PPE68_1. To express EsxM and EsxN in E. coli, the encoding genomic region was amplified by PCR and introduced into cloning vector pJet1 (Fermentas). After digestion by XbaI and XhoI, the resulting fragment was introduced in pET15b(+), which was cut with the same enzymes, resulting in pET15b(+):esxMN.

**Secretion analysis and fractionation of M. marinum** – Preparation of whole cell lysates, extraction of surface proteins by treatment with Genapol X-080 and precipitation of proteins secreted into the medium of M. marinum were carried out as described [11].

Mycobacterial cultures grown to an OD$_{600}$ of ~1.0 were incubated for 1 h with 100 μg/ml ciprofloxacin (Sigma), washed twice with PBS, and subsequently resuspended in PBS supplemented with Complete protease inhibitor cocktail (Roche), 1 mM EDTA and optionally 10 mM imidazole. Cells were broken by two-times passage through a One Shot cell disrupter (Constant Systems Ltd) at 0.83 kbar, adding 1 mM DTT after the first press. Unbroken cells were spun down by repeated centrifugation at 3000 g, and subsequently the cell envelope and soluble fractions were separated by ultracentrifugation at 100,000 g.

**Expression of mycobacterial proteins in E. coli** – Overnight cultures of E. coli TOP10F’ harboring pASK-IBA3-plasmids were diluted to an OD$_{660}$ of 0.05 and incubated until an OD$_{660}$ of 0.3, at which expression of recombinant proteins was induced by the
addition of 200 μg/ml anhydrotetracycline for 2 hours. Proteins were expressed from pET29b(+) vectors in E. coli BL21(DE3) pLysS essentially as described in [26]. Briefly, overnight cultures were diluted to an OD$_{660}$ of 0.05 in medium containing 0.2% dextrose to repress premature expression of recombinant proteins. At an OD$_{660}$ of 0.6, expression of the recombinant proteins was induced for 3 hours in the presence of 0.4 mM isopropyl β-D-thiogalactoside (IPTG). EsxM/EsxN were expressed from pET15b(+) using the same strain and conditions, but for induction 0.1 mM IPTG was used. PE25$_{Δ15C}$ and C-terminally His-tagged PPE41 were expressed constitutively in E. coli XL10-Gold from the P$_{lac}$ promoter in the pSMT3::PE25$_{Δ15C}$-PPE41-His vector.

Cells were harvested by centrifugation, resuspended to a concentration of 50 OD units/ml in a buffer containing 50 mM NaH$_2$PO$_4$ and 300 mM NaCl (pH 8.0), and lysed in a One Shot cell disrupter (Constant Systems Ltd) at 1.72 kbar. Unbroken cells were removed by low-speed centrifugation, and soluble proteins were obtained by ultracentrifugation. Subsequently, the E. coli lysates were diluted in 50 mM NaH$_2$PO$_4$ and 300 mM NaCl (pH 8.0) in such a way that the concentration of soluble recombinant protein was similar in all samples.

Protein co-purification assays – Ni$^{2+}$-purifications of His-labeled proteins expressed in M. marinum were performed by loading filtered soluble lysates on washed HisTrap HP Columns (GE Healthcare) at 0.5 ml/min. Optionally, the filtered M. marinum lysates were precleared on HisTrap sepharose columns that had been stripped from Ni$^{2+}$, prior to loading on a Ni$^{2+}$-loaded column. After extensive washing of the column with phosphate buffer (20 mM phosphate, 150 mM NaCl, pH 8.0) containing 50 mM imidazole, bound proteins were eluted with 5 ml phosphate buffer containing 150 mM imidazole, and collected in 500 μl fractions. Wash and elution fractions were precipitated with 10% TCA. Alternatively, lysates of M. marinum expressing proteins of interest were incubated with Ni-NTA agarose beads (Qiagen) for 1 hour at room temperature with head-over-head rotation. After washing the beads five times with phosphate buffer containing 20 mM imidazole, bound proteins were eluted by incubation with 100 mM, 200 mM and 500 mM imidazole, respectively. Immunoprecipitation of HA-tagged proteins were performed using the HA Tag IP/Co-IP kit (Pierce).

For the in vitro pulldowns, soluble lysates of E. coli containing similar amounts of recombinant proteins were mixed with Ni-NTA agarose beads (Qiagen) for 1 hour at room temperature. Washing was performed as described above using a buffer containing 50 mM NaH$_2$PO$_4$, 300 mM NaCl and 20 mM imidazole (pH 8.0), and proteins were eluted using this buffer supplemented with 100 mM or 250 mM imidazole.

Cloning, protein expression and purification for molecular mass measurements – espG$_5$ was cloned with an N-terminal His tag under control of the T7 promoter in the pETM-11 vector (EMBL). The gene was amplified using a 5’ primer containing a NcoI digestion site (AP-303) and a 3’ primer with XhoI digestion site (AP-304). After digestion, the fragments were ligated into NcoI and XhoI-digested pETM-11, resulting in pAP200. For expression of PE25 and C-terminally His-tagged PPE41 in E. coli a previously described construct was used in which the rv2431c and rv2430c genes were each placed behind a RBS [26]. All proteins were produced in E. coli BL21(DE3) after induction with 0.2 mM IPTG and overnight incubation at 20°C. Cells were lysed by sonication in 25 mM HEPES (pH
7.5), 150 mM NaCl supplemented with protease inhibitor mix (Serva). Lysates were cleared by centrifugation and subjected to standard immobilized metal ion affinity chromatography (IMAC) using HisTrap HP columns (GE Healthcare). After washing with 20 mM imidazole, recombinant proteins were eluted with a linear gradient of 20 to 250 mM imidazole in 25 mM HEPES (pH 7.5), 150 mM NaCl. Relevant protein fractions were pooled and concentrated and subjected to protease treatment for tag removal. His-tags were cleaved from EspG5 by overnight incubation with tobacco etch virus protease at 4°C. Protease and uncleaved protein were removed by a second IMAC step. In case of PE25/PPE41-His, biotinylated thrombin (Novagen) was used which was subsequently removed by streptavidin beads (Invitrogen) according to manufacturer’s recommendations. All proteins were further purified to homogeneity by preparative size-exclusion chromatography (HiLoad 16/600 Superdex 200 pg, GE Healthcare) in 20 mM HEPES (pH 7.5), 150 mM NaCl. EspG5 as well as PE25/PPE41 were concentrated using Vivaspin centrifugal concentrators (Corning) to a final concentration of 8.6 and 5.3 mg/ml, respectively. Protein concentrations were estimated by measurements of absorbance at 280 nm using theoretical molar absorption coefficients calculated by ProtParam.

Size Exclusion Chromatography–Tridetector Analysis (SEC-TDA) – Proteins were analyzed with a Viscotek 305 TDA (Malvern Instruments, Malvern, UK) which monitors light scattering, refractive index, and UV absorbance. This setup was connected to an analytical size exclusion column (Superdex 200 10/300 GL, GE Healthcare) equilibrated at 20°C in 20 mM HEPES (pH 7.5), 150 mM NaCl as running buffer. Flow rate was set to 0.5 ml/min, and the sample volume was 100 μl. Proteins were diluted to 1 mg/ml in running buffer prior to analysis. To detect heterotrimeric complex formation, EspG5 was mixed with PE25/PPE41 at a 3-fold molar excess and incubated 30 min at 20°C before injection. The provided OmniSEC software was used to acquire and evaluate all data. Molecular masses were estimated using refractive index combined with light scattering data using BSA as an internal control using a refractive index increment with protein concentration (dn/dc) of 0.185 ml/g.

SDS-PAGE, immunoblotting and sera – Proteins were separated by SDS-PAGE and stained with Coomassie Brilliant Blue G-250 (CBB; BioRad), or transferred to nitrocellulose membranes by western blotting. The membranes were incubated with mouse monoclonal antibodies directed against the conserved repeat of PE_PGRS proteins (mAb 7C4.1F7) [3], the influenza hemagglutinin epitope (HA.11; Covance), the C-terminal 6xHis epitope (11922416001; Roche Applied Science), EsxA (Hyb76-8; Statens Serum Institut, Copenhagen, Denmark) or GroEL2 (CS44; John Belisle, National Institutes of Health, Bethesda, MD, Contract AI-75320); or with rabbit polyclonal serum reactive against PPE41 [2], PPE68 [23], EspG5, FtsH/Mmar_0752 [27] or EsxN (Mtba.9A) [4]. Secondary horseradish peroxidase-conjugated goat anti-mouse IgGs (A106PS, American Qualex) or goat anti-rabbit IgGs (611-1302, Rockland) were detected with ECL (Roche Applied Science or Pierce).

Mass spectrometric analysis – For analysis by nanoLC-MS/MS, protein lanes from CBB-stained gels were excised, prepared and analyzed as described [25]. Identified proteins were annotated using the Mycobrowser website [20]. For MS/MS analysis, samples were prepared essentially as described [3] and analyzed by MALDI-TOF/TOF MS (AB Sciex TOF/TOF 5800). MS/MS spectra
Specific chaperones for the type VII pathway

Figure 1. Complementation of the \( espG_5::tn \) mutant and pulldown experiment in \( M. marinum \) (A and B) Immunoblots of equal amounts of whole cells pellets (P), whole cells treated with Genapol (Gp), Genapol surface extracts (Gs), and two-fold excess of culture supernatants (S) from \( M. marinum \) wt strain E11, the isogenic \( espG_5::tn \) mutant 7C1, and the mutant expressing \( espG_5::tn \)-His, all expressing PE25-HA/PPE41, were probed with antibodies against EspG5, PE_PGRS, PPE41, HA-tag, EsxA and the lysis control GroEL2. (C) Immunoblot analysis of total (T), soluble (Sol) and cell envelope (CE) fractions of \( M. marinum \) Mx2 (\( eccA_5::tn \)) expressing EspG5-His. EspG5 was detected using mAb directed against the 6xHis epitope, and cross-reactive polyclonal antiserum was used to detect the cytosolic protein MMAR_0752 and the inner membrane protein FtsH as controls for the fractionation. (D) Purification of EspG5-His from soluble lysates of Mx2 expressing EspG5-His from a plasmid (+) or the empty strain (-) on Ni-NTA beads. Unbound proteins (FT), the last washing step (W) and proteins eluted with 100 mM (E1), 200 mM (E2) and 500 mM (E3) imidazole were separated by SDS-PAGE in a ratio of 1:20:10:10:10, and stained with CBB or probed with antiserum directed against the His-epitope in immunoblots. Eluted EspG5-His is marked by an asterisk, and specific co-eluted proteins are marked by an arrowhead.
search was performed against the *M. marinum* FASTA database using Mascot software (Matrix Science). Searches were performed with a peptide mass tolerance of 0.15 Da and a fragment mass tolerance of 0.1 Da, while allowing a single site of miscleavage and oxidation of methionine as a variable modification. Only proteins with a score of >50 were considered significant.

**RESULTS**

**EspG$_5$ interacts with PPE proteins.**

To study the function of EspG$_5$, we expressed C-terminally His-tagged EspG$_5$ in the *M. marinum* espG$_5$::tn mutant. Immunostaining with an antibody recognizing multiple PE_PGRS proteins showed that expression and (partially) secretion of these ESX-5 substrates were restored (Fig. 1A). Moreover, treatment of cells with the mild detergent Genapol to extract surface proteins [25] showed that surface localization of PE_PGRS proteins did not differ from the wild type strain (Fig. 1B). We also investigated secretion of heterologously expressed PE25 and PPE41. As previously shown, secretion was markedly reduced in the espG$_5$::tn mutant strain [3, 12]. Importantly, secretion was restored to wild-type levels upon complementation with EspG$_5$-His (Fig. 1A). As controls we used the intracellular protein GroEL2 and the secreted ESX-1 substrate EsxA. Together, these results show that EspG$_5$-His is functional. Notably, EspG$_5$ is not a secreted protein, as both EspG$_5$ and EspG$_5$-His were only detected in the cell pellet and not extracted with Genapol (Fig. 1A/B). The subcellular localization of EspG$_5$ was examined by a fractionation procedure, which showed that EspG$_5$-His was exclusively present in the soluble fraction of the cells (Fig. 1C). As a control for successful fractionation an antiserum that recognizes both the membrane protein FtsH and cytosolic MMAR_0752 was used.

To identify putative interacting partners of EspG$_5$, a Ni-NTA pulldown was performed using membrane-free lysates of the espG$_5$::tn mutant complemented with EspG$_5$-His. However, although purification of EspG$_5$ was successful, no specific interacting partners were detected (results not shown), perhaps because interactions are transient in the presence of a functional ESX-5 system. Therefore, this pulldown experiment was repeated using a different ESX-5 secretion mutant strain, *i.e.* eccA$_5$::tn. Importantly, this strain is deficient in ESX-5-mediated secretion also after introduction of the EspG$_5$-His construct (not shown) and some of the ESX-5 substrates do accumulate in the cytosol [2]. We therefore hypothesized that putative interactions between EspG$_5$ and other proteins might be prolonged in this strain. Indeed, in two pulldowns, of which the second was optimized to reduce background binding to the beads (see Experimental procedures), we could now observe protein bands that were exclusively present in the eluted fractions of the EspG$_5$-His expressing strain (see Fig. 1D for the optimized pulldown). To identify these interacting proteins, total protein isolations with and without EspG$_5$-His were subjected to nanoLC-MS/MS analysis. As also observed after SDS-PAGE analysis (Fig. 1D), proteins with the highest spectral counts were present in both samples and could therefore be considered as proteins that aspecifically bind to the Ni$^{2+}$ sepharose beads. Interestingly, when focusing on proteins that were detected in the second, optimized experiment and that had more than two spectral counts in at least one of the experiments, we identified only three proteins specifically co-eluting with EspG$_5$-His (Table 1), all belonging to the SVP subfamily of PPE proteins [18]. Of note, PE/PPE proteins often have a low number of trypsin recognition sites and are therefore notoriously underrepresented in mass spectrometry experiments. These
Specific chaperones for the type VII pathway

results indicate that EspG<sub>5</sub> specifically interacts with PPE proteins.

To confirm this interaction a co-purification assay was performed using lysates of the eccA<sub>5</sub>::tn strain expressing only EspG<sub>5</sub>-His or both EspG<sub>5</sub>-His and a C-terminally HA-tagged version of one of the identified PPEs, i.e. PPE33 (Fig. 2A). Although a minor amount of PPE33-HA bound to the beads in the absence of EspG<sub>5</sub>-His, significantly more was co-purified with EspG<sub>5</sub>-His (Fig. 2A). The unrelated cytosolic control MMAR_0752 was only present in the flow-through fraction (Fig. 2A). Furthermore, in the reciprocal pulldown using beads coated with HA mAb, both EspG<sub>5</sub>-His and endogenous EspG<sub>5</sub> were specifically co-purified, whereas in control lysates both forms of EspG<sub>5</sub> ended up in the flow-through (Fig. 2B). In conclusion, these data confirm that EspG<sub>5</sub> interacts with PPE33.

Table 1. Interacting partners for EspG<sub>5</sub> in M. marinum Mx2 identified by nanoLC-MS/MS analysis

<table>
<thead>
<tr>
<th>Accession number</th>
<th>Annotation</th>
<th>M&lt;sub&gt;W&lt;/sub&gt;</th>
<th>Spectral counts (% amino acid coverage)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Experiment 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- + EspG&lt;sub&gt;5&lt;/sub&gt;-His</td>
</tr>
<tr>
<td>MMAR_3661</td>
<td>PPE38&lt;sup&gt;a&lt;/sup&gt;</td>
<td>37 kDa</td>
<td>0</td>
</tr>
<tr>
<td>MMAR_1460</td>
<td>PPE33&lt;sup&gt;a&lt;/sup&gt;</td>
<td>47 kDa</td>
<td>5 (8%)</td>
</tr>
<tr>
<td>MMAR_1514</td>
<td>PPE51_1</td>
<td>38 kDa</td>
<td>4 (7%)</td>
</tr>
</tbody>
</table>

<sup>a</sup> Annotated as their orthologues in M. tuberculosis H37Rv [1].

EspG<sub>5</sub> interacts with PE/PPEs secreted by ESX-5

The PPE proteins that we identified as interacting partners of EspG<sub>5</sub> all belong to the SVP subfamily. To test whether EspG<sub>5</sub> also interacts with other PPE proteins secreted via ESX-5 we selected the well-known heterodimeric substrate complex PE25/PPE41 [2, 6, 12, 26]. The PE25/PPE41 complex was expressed in combination with EspG<sub>5</sub>-His in the M. marinum eccA<sub>5</sub>::tn strain Mx2 (eccA<sub>5</sub>::tn) expressing (A) PPE33-HA or (C) PE25-HA/PPE41, alone (-) or in combination with EspG<sub>5</sub>-His (+), or (B) Mx2 expressing EspG<sub>5</sub>-His, with (+) or without (-) co-expression of PPE33-HA. Proteins were detected with antisera directed against PPE41, the cytosolic ATPase MMAR_0752, EspG<sub>5</sub> or the HA- or His-tags.

Figure 2. EspG<sub>5</sub> interacts with PPE33 and PPE41 in M. marinum. Immunoblots of total input material (T), unbound proteins (FT), the final washing step (W) and eluted proteins (E) in a 1:1:10:10 ratio from pulldown assays using (A, C) Ni-NTA beads and (B) beads coated with anti-HA mAb, in lysates of M. marinum strain Mx2 (eccA<sub>5</sub>::tn) expressing (A) PPE33-HA or (C) PE25-HA/PPE41, alone (-) or in combination with EspG<sub>5</sub>-His (+), or (B) Mx2 expressing EspG<sub>5</sub>-His, with (+) or without (-) co-expression of PPE33-HA. Proteins were detected with antisera directed against PPE41, the cytosolic ATPase MMAR_0752, EspG<sub>5</sub> or the HA- or His-tags.
and found across the mycobacterial cell envelope

mutant and a pulldown was performed. Immunoblot analysis of obtained fractions showed that PPE41 indeed specifically co-eluted with EspG₅-His (Fig. 2C), indicating that EspG₅ also interacts with PPE41.

To substantiate these results in vitro interaction studies were performed, using lysates of E. coli cells expressing either the PE25/PPE41 complex or C-terminally His-tagged EspG₅. The soluble fractions were mixed and the His-containing EspG₅ protein was isolated using Ni-NTA agarose beads. Interestingly, two specific bands co-eluted with EspG₅-His (~32 kDa), with an apparent molecular weight of ~24 kDa and ~15 kDa, respectively (Fig. 3A, lanes 13-16). Immunoblot analysis identified the larger band as PPE41, whereas the apparent molecular weight of the smaller band corresponds to that of PE25, indicating that EspG₅ interacts with the PE25/PPE41 complex. In a control purification using only the PE25/PPE41-containing lysate the majority of PPE41 was detected in the flow-through (Fig. 3A, lanes 9-12). The interaction between EspG₅ and PPE41 was confirmed in a reciprocal pulldown assay using C-terminally His-tagged PPE41 (Fig. 3A, lanes 1-8).

To analyze the stoichiometry of the complex formed by EspG₅ and PE25/PPE41, we performed size exclusion chromatography in line with light-scattering and refractive analysis. This revealed that EspG₅ migrated as a single species with an estimated molecular mass of 32 kDa, indicating that the protein exists as a monomer in solution (Fig. 3B). Although the molecular mass of the PE25/PPE41 complex is similar to that of monomeric EspG₅, PE25/PPE41 elutes slightly earlier from the gel filtration column. This is due to its tightly folded helical structure [26], resulting in an increased hydrodynamic radius in contrast to the apparent more globular nature of EspG₅. Upon mixing of EspG₅ with PE25/PPE41 complex, a peak was detected with a molecular mass of approximately 60 kDa, corresponding to a 1:1:1 heterotrimeric complex (Fig. 3B).

In addition to PE/PPE proteins, the ESX-5 system also secretes EsxN. However, our nanoLC-MS/MS analysis of proteins co-purified with EspG₅-His did not reveal an interaction with EsxN. To determine if this other class of ESX-5 substrates also interacts with EspG₅, another in vitro pulldown assay was performed. EsxN forms a heterodimer with EsxM and therefore both proteins were co-expressed from an E. coli expression vector. Although EspG₅-His was efficiently pulled down, no co-purification of EsxN was observed (Fig. 3C), indicating that EspG₅ specifically interacts with PE/PPE proteins.

EspG₅ does not interact with the type VII secretion signal.

We have previously shown that the flexible C terminus of PE25 contains a YxxxD/E motif that is required for secretion [12]. To investigate whether EspG₅ interacts with this domain, in vitro pulldowns were carried out with E. coli lysates containing PPE41 and PE25 lacking the C-terminal 15 aa. As shown by immunoblotting, EspG₅ does not interact with the type VII secretion signal.
Specific chaperones for the type VII pathway

A. Peptide analysis of EspG5 and PE25/PPE41 with and without His tags.

B. Molecular Weight (Da) comparison:

<table>
<thead>
<tr>
<th>EspG5</th>
<th>PE25/PPE41</th>
<th>PE25/PPE41-His</th>
</tr>
</thead>
<tbody>
<tr>
<td>32,480</td>
<td>32,004</td>
<td></td>
</tr>
<tr>
<td>32,501</td>
<td>31,706</td>
<td></td>
</tr>
<tr>
<td>64,981</td>
<td>59,737</td>
<td></td>
</tr>
</tbody>
</table>

C. Peptide analysis of EspG5 and PE25/PPE41 with and without His tags.


UV absorbance (280 nm)

Molecular weight (Da)
still co-purified with PPE41, indicating that binding of EspG₅ does not depend on the YxxxD/E motif (Fig. 3D). Using a similar approach we showed that also the flexible C-terminal domain of PPE41 is not essential for the interaction with EspG₅ (Fig. 3D). Together, these experiments indicate that EspG₅ interacts with the structured core-domain of the PE/PPE complex.

**EspG homologues specifically bind cognate PPEs**

We have recently shown that PE35 and PPE68_1 are secreted by ESX-1 in *M. marinum* [12, 25]. To determine whether EspG₅ only recognizes PE/PPE substrates of ESX-5, we performed an *in vitro* pulldown using lysates of *E. coli* strains expressing EspG₅ and C-terminally His-labeled PPE68_1. In parallel, EspG₅ was

![Figure 4. EspG homologues interact with cognate PE/PPE substrates. SDS-PAGE/CBB and immunoblot analyses of fractions from protein co-purifications. (A) Eluates of Ni-NTA pulldowns in mixed lysates of *E. coli* strains expressing recombinant PE25/PPE41, PE35/PPE68_1, EspG₅-His and EspG₁-His are shown in a 1:1 ratio. (B-C) Total input material (T), unbound proteins (FT), final washing step (W) and eluted proteins (E) from (B) HA immunoprecipitations using lysates of *M. marinum* strain Mx2 (*eccA5::tn*) expressing EspG₅-His, with (+) or without (-) co-expression of PE35-HA/PPE68_1 are shown in a 1:1:10:10 ratio and (C) a Ni²⁺ purification using a lysate of Mx2 expressing EspG₅-His and PE35-HA/PPE68_1 are shown a 1:1:4:4 ratio. In immunoblots, proteins were detected with sera directed against EspG₅, PPE68, PE41, EspN, the cytosolic protein MMAR_0752 or the His- and HA-epitopes. (A and C) Symbols mark bands representing EspG₅ (*), PPE41 (*), EspG₁ (*) and PPE68_1 (*), and the unidentified *M. marinum* protein that reacts with the PPE68 antiserum (*). (D) CBB stained SDS-gel shown the eluates from panel (B), where specific protein bands that were analyzed by MALDI-TOF/TOF are indicated (<).
incubated with soluble fractions of \textit{E. coli} expressing PE25/PPE41-His or empty \textit{E. coli} strains, as positive and negative controls, respectively. Interestingly, EspG$_5$ did not co-elute with PPE68$_1$-His (Fig. 3D). This was again confirmed in a reciprocal \textit{in vitro} pulldown assay using EspG$_5$-His (Fig. 4A). Together, these results show that EspG$_5$ does not interact with PPE proteins secreted via the ESX-1 pathway.

EspG$_5$ is 22\% identical at the amino acid level to the ESX-1 encoded protein EspG$_1$, which has been suggested to affect the stability of PPE68 in \textit{M. tuberculosis} [7]. These facts led us to hypothesize that EspG$_1$ interacts with PE/PPE proteins secreted via ESX-1. To investigate this, we performed \textit{in vitro} pulldowns using lysates of \textit{E. coli} expressing C-terminally His-tagged EspG$_1$, mixed with lysates containing PE25/PPE41 or PE35/PPE68$_1$. In both reactions, a band of \textasciitilde27 kDa was visible in Coomassie Brilliant Blue (CBB)$^\text{-}$stained gels, which was identified by immunoblotting as EspG$_1$ (Fig 4A). In the pulldown with PE25/PPE41 no additional specific bands were visible on protein gels and only a weak band was detected with antibodies directed against PPE41. In contrast, when using the PE35/PPE68$_1$-containing lysate a major specific band of \textasciitilde37 kDa was detected, which was recognized by antibodies directed against \textit{M. tuberculosis} PPE68 (Fig 4A). These results show that while EspG$_5$ specifically interacts with the ESX-5 substrates PE25/PPE41, the ESX-1-encoded EspG$_1$ fails to associate with PPE41, but does interact with the ESX-1-substrate PPE68$_1$.

To confirm these observations, we introduced a plasmid encoding C-terminally HA-labeled PE35 and PPE68$_1$ in the \textit{M. marinum} eccA$_5$::tn mutant expressing EspG$_5$-His. Notably, this strain is also affected in ESX-1 secretion, and ESX-1 substrates accumulate in the cytosol [3]. In an HA immunoprecipitation assay, PPE68$_1$ co-purified with PE35-HA (Fig. 4B), showing that this PE/PPE couple indeed forms a complex. Of note, an additional band was visible just below the expected molecular weight of PPE68$_1$ in the anti-PPE68$_1$ immunoblot, which could represent endogenous PPE68$_1$. This immunoprecipitation also showed that PE35-HA does not interact with EspG$_5$, which was also confirmed in a reciprocal pulldown (Fig. 4B/C). However, CBB-staining of the pulldown for PE35-HA did reveal specific co-purification of two proteins with an apparent molecular mass of \textasciitilde36 and \textasciitilde30 kDa (Fig. 4D). MS/MS analysis showed that the \textasciitilde36 kDa product was indeed PPE68$_1$, whereas the \textasciitilde30 kDa product was identified as (endogenous) EspG$_1$. These results show that also in \textit{M. marinum} the ESX-1 substrates PE35 and PPE68$_1$ interact specifically with EspG$_1$. Taken together, we conclude that EspG$_5$ and EspG$_1$ are both specific chaperones for cognate PE/PPE proteins.

\textbf{DISCUSSION}

In this study, we show that the ESX-5-encoded MMAR$_{2676}$, which is required for ESX-5-mediated protein secretion [3, 11], interacts with cognate PE/PPE proteins. The similarity of MMAR$_{2676}$ to the EspG proteins encoded by other ESX clusters was previously considered too low for this protein to be a true EspG homologue [5]. Here, we provide compelling experimental evidence that despite this modest similarity, MMAR$_{2676}$ and EspG$_1$ are functionally conserved. Therefore, we propose that MMAR$_{2676}$ should be annotated as EspG$_5$.

Pulldown of His-labeled EspG$_5$ resulted in co-purification of PE and PPE proteins, but interestingly not of the ESX-5-encoded WXG100 proteins EsxM and EsxN. These results suggest that EspG$_5$ specifically interacts with members of the PE/PPE families. This finding seems to correlate with the composition of the different ESX loci, as all ESX loci that contain \textit{pep/ppe} genes also have an \textit{espG} homologue.
Consistently, ESX gene clusters in close relatives of mycobacteria that lack pel/ppe genes, e.g. in Corynebacterium glutamicum and Streptomyces coelicolor, also lack an espG homologue [17]. However, Nocardia farcinica has an additional ESX locus that contains both a pel/ppe gene couple and an adjacent espG-like gene, as well as a small pel-ppe-espG gene cluster at a distinct genomic location (Fig. 5). These findings together confirm the link between the espG and the pel/ppe families.

What is the function of the EspG proteins? Previous studies have indicated that inactivation of espG₂ or espG₅ abolishes secretion and/or affects the intracellular levels of some ESX substrates [3, 7, 11, 16]. This suggests that EspG proteins play a role in expression, stability and/or secretion of the substrates. As intracellular levels of constitutively expressed PE and PPE proteins were also markedly affected in an espG₅ mutant [11], it is unlikely that EspG₅ directly regulates the mRNA expression of the ESX-5 substrates. Rather, EspG₅ might act as a chaperone that maintains the substrates in a translocation-competent form, protects them from degradation prior to secretion, and/or serves as a targeting factor for ESX-5 secretion. We have recently shown that the PE/PPE proteins secreted via different type VII secretion systems share a conserved secretion signal, located in the C terminus of the PE proteins, but that this signal does not determine through which ESX system they are secreted [12]. System specificity might be mediated through targeting of secreted effectors by specific chaperones, similar to the type III secretion pathway in Gram-negative bacteria. In type III secretion, several different classes of chaperones are distinguished, which have one or a few cognate substrates and function independently of ATP [29, 30]. EspG proteins could function similarly. Consistently, deletion of the secretion signal-containing C terminus of PE25 did not affect the interaction with EspG₅. Inspection of the sequences of PE/PPE proteins and their cognate EspG chaperones does not immediately reveal residues or motifs that could explain the specificity of the interaction. Notably, EspG₁ and EspG₅ only share low homology. It is plausible that the basis for the specificity is a structural motif in the EspG protein and/or structural core-domain of the PE/PPE complex.

Our results are in contrast to the recent observation that knockout of espG₅ in M. tuberculosis does not affect PPE41 secretion [6]. However, these differences are not unique as functional differences between species have also been shown for EspG₅. This protein is necessary for ESX-1-mediated secretion of WXG100 proteins in M. marinum and M. smegmatis but not in M. tuberculosis [7, 10, 16]. In the latter species, disruption of espG₁ did affect both secretion and stability of PPE68 [7]. These observations suggest that EspG proteins can be (partly) redundant.

In conclusion, we propose that the
EspG homologue of each ESX system recognizes its cognate PE/PPE proteins, maintains them in a stable conformation and possibly also directs them to the membrane-embedded secretion machinery (Fig. 6). Subsequently, recognition of the YxxxD/E secretion motif by a component of the secretion machinery, potentially the FtsK/SpoIIIIE-like ATPase EccC [9], initiates translocation through the secretion channel. Future experiments will determine whether EspG\textsubscript{5} is able to deliver PE/PPE proteins to ESX-5 membrane components.

ACKNOWLEDGEMENTS
We thank Gregory Koningstein, Melissa Graewert and the SPC facility at EMBL Hamburg for technical support, Michael Strong for plasmid DNA, and Ida Rosenkrands and Roland Brosch for providing sera.

FOOTNOTES
*These authors contributed equally to this work.
#This work was supported by funding from an ALW Open Program grant to M.H.D and an ECHO grant to A.D.vdW, both from the Netherlands Organization for Scientific Research, and from the European Community Seventh Framework Programme (FP7/2007–2013) under grant 201762 to W.B.
$The abbreviations used are: CBB, Coomassie Brilliant Blue; Ecc, ESX conserved components; Esp, ESX-1 secretion-associated proteins; T7SS, type VII secretion system.


Specific chaperones for the type VII pathway


SUPPLEMENTAL DATA

Table S1. Primers used in this study.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence (5’ → 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AbMmMiddle-H5-F</td>
<td>aag g g cac cc g gaaatg ttaaa</td>
</tr>
<tr>
<td>1794-His-SpeI-R</td>
<td>cag c ac tag ttc ag tgg atg gtg atg aac tc tt c tat gc g tttt cc</td>
</tr>
<tr>
<td>AW0001-F</td>
<td>ac g tgg ttc c gc g tag ttc ag</td>
</tr>
<tr>
<td>1794-His-HindIII-R</td>
<td>ccc aag ctttc ag tgg atg g tg atg aac tc tt c tat gc g tttt cc</td>
</tr>
<tr>
<td>Novel plBA fw</td>
<td>ttttg tc atgc tc gc ag g</td>
</tr>
<tr>
<td>EspG1-RBS-Xbal-F</td>
<td>ccc tt c tag atac c gc g gca aaaaatg c gc g tc gc ct gc ta</td>
</tr>
<tr>
<td>EspG1HIS-HindIII-R</td>
<td>ccc aag ctttc ag tgg atg gtg atg aac cct gc g gc g gc g gc g gc g</td>
</tr>
<tr>
<td>PPE1460-Xbal-F</td>
<td>ccc tt c tag aat gc ttc tag c gc g gtttc c</td>
</tr>
<tr>
<td>PPE1460-HindIII-HA-R</td>
<td>ccc aag ctttc ac gc g tag tc gc g ac g g c gc g tc gac g g tag gc c c aat gt gc g tc gc a</td>
</tr>
<tr>
<td>1794-RBS-Xbal-F</td>
<td>ccc tt c tag atac c gc g gca aaaaatg g atc aac aag tac gc g</td>
</tr>
<tr>
<td>1794-R</td>
<td>ag ttc gc g ttg cttt gc ag t</td>
</tr>
<tr>
<td>pET-31cF</td>
<td>aaaaatgc g c gc g gta aacc cttt cc ttt t</td>
</tr>
<tr>
<td>pET-30cHisR</td>
<td>ccc g gatat cct ttc gg ggc ttg ttag cag</td>
</tr>
<tr>
<td>pET-d15cF</td>
<td>gcc ttg ac ac tg g gc g c taa gc a c aaaaatttt c tttaac</td>
</tr>
<tr>
<td>pET-d15cR</td>
<td>gttaac aaaaatttt gc g tag c cttg c ac c gc c ggc c gtc aag gc</td>
</tr>
<tr>
<td>pET41_F</td>
<td>cat gc c atgc gct ttc gc gag gc</td>
</tr>
<tr>
<td>pET41 noHis R</td>
<td>ccc aag ctttc tag tc gc gac g g tag ac g</td>
</tr>
<tr>
<td>pET41d02C_R</td>
<td>ccc aag cttag aac gc g c gc c c g c c gat gc g gc gc gc gc</td>
</tr>
<tr>
<td>pET35_F</td>
<td>gcgc catag cg at cc ttttta</td>
</tr>
<tr>
<td>pET35_R</td>
<td>gg gg g tac ctttag cgc gc g tag tc gc g gac gc</td>
</tr>
<tr>
<td>pET68_F</td>
<td>catgc c atgc gct tgg gac gc cgc t gccc c</td>
</tr>
<tr>
<td>pET68_R</td>
<td>ccc aag ctttag aac gc g cgc g cgc c c g c c g c c g c c c g c</td>
</tr>
<tr>
<td>pET68noHis R</td>
<td>ccc aag ctttc ac gc gc gc gc gc gc gc gc</td>
</tr>
<tr>
<td>esxM-F</td>
<td>ggg gc c ttc gtaa cttt</td>
</tr>
<tr>
<td>esxN-R</td>
<td>gata tgc ggc gc gtc ttt</td>
</tr>
<tr>
<td>AP-303</td>
<td>atata c atgc g atc aac aag tac gc</td>
</tr>
<tr>
<td>AP-304</td>
<td>atata tc gc gat tc aac tc tc tag c g tttt cc</td>
</tr>
</tbody>
</table>