Protease and transmembrane domain of the type VII secretion mycosin protease determine system-specific functioning in mycobacteria

Vincent J.C. van Winden[1], Merel P.M. Damen[2], Roy Ummels[1], Wilbert Bitter[1,2] & Edith N.G. Houben[2*]

Supporting Information:

Supporting Table 1. Identity percentage for the complete protein and the loop regions of $MycP_1$, $MycP_5$ and $MycP_{P1}$ of *M. marinum*.

Supporting Table 2. Schematic overview of the MycP constructs and their activity in different backgrounds. The composition of the constructs is depicted in Figure 1A and 2A. nd, not determined; -, no expression/secretion/stability; +/-, reduced expression/secretion/stability; +, normal expression/secretion/stability; ++, increased expression.

Supporting Table 3. List of primers used in this study.

Supporting Figure 1. Immunoblot detection of the hybrid HA-tagged constructs (see Figure 2A), expressed in the *mycP1* knockout of *M. marinum*.

Supporting Figure 2. Immunoblot analysis of Genapol pellets (Gen Pel, cellular) and Genapol supernatant (Gen Sup, surface localized proteins) of wild-type (WT) *M. marinum*, complemented with a truncated *mycP5* without its TM region (5ΔTM). Proteins were visualized with anti-PE_PGRS (ESX-5 substrates).

Supporting Figure 3. (A) Immunoblot detection of the HA-tagged constructs of wild-type *mycP5* (P5) and hybrid *mycP5* with loop 1 of *mycP1* (P5-L1.P1). (B) Immunoblot analysis of Genapol pellets (Gen Pel, cellular) and Genapol Supernatant (Gen Sup, surface localized proteins) of wild-type (WT) *M. marinum* and a *mycP5* knockout, complemented with P5 or P5-L1.P1. Proteins were visualized with anti-PE_PGRS (ESX-5 substrates). (C) Immunoblot detection of HA-tagged wild-type *mycP5* (P5) and an active site mutant of *mycP5* (P5SA) in *M. smegmatis*. (D) SDS-PAGE analysis of purified wild-type EspB, both full-length and processed after co-incubation with either purified $MycP_1$ (P1 or MycP1 WT) or $MycP_1$ without loop 1 (P1ΔL1 or MycP1dLoop1). Proteins were visualized with Coomassie Brilliant Blue staining.

Supporting Figure 4. Multiple sequence alignment of $MycP_1$ and $MycP_5$ of *M. marinum* M, and *M. tuberculosis* H37Rv and $MycP_{P1}$ of *M. marinum* E11. The N-terminal extension and the extended loops are highlighted in grey. The linker domain (linker) and predicted transmembrane domain are indicated by grey boxes. Amino acids constituting the predicted signal sequence are shown in grey. Amino acids of the mycosins, excluding the N-terminal extension and the four loops, are colored based on characteristics (*i.e.* basic, acidic, polar or nonpolar). Residues in the protease and linker domain, excluding the loops, that have conserved characteristics between $MycP_5$ and $MycP_{P1}$, but different from $MycP_1$, are depicted with an open arrowhead (^). Positions where individual domains have been exchanged between $MycP_1$ and $MycP_5$ are indicated by closed arrowheads (▼); positions where specific domains have been deleted are indicated by an asterisk (*).

Supporting Table 1. Identity percentage for the complete protein and the loop regions of $MycP_1$, $MycP_5$ and $MycP_{P1}$ of *M. marinum*.

Query Protein	**Subject Protein**	**Entire protein (% identity)**	**Loop 1 (% identity)**	**Loop 2 (% identity)**	**Loop 3 (% identity)**
$MycP_1$ (M)	$MycP_5$ (M)	46.37	17.65	50	25
$MycP_1$ (M)	$MycP_{P1}$ (E11)	48.77	23.53	58.33	33.33
$MycP_5$ (M)	$MycP_{P1}$ (E11)	59.49	41.18	60	46.15

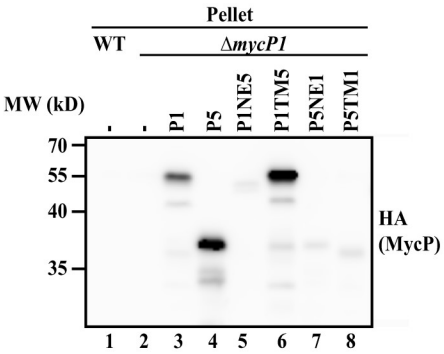
Supporting Table 2. Schematic overview of the MycP constructs and their activity in different backgrounds. The composition of the constructs is depicted in Figure 1A and 2A. nd, not determined; -, no expression/secretion/stability; +/-, reduced expression/secretion/stability; +, normal expression/secretion/stability; ++, increased expression.

	Mycosin expr.	ESX-1 secr.	ESX-5 secr.	ESX-5 complex stabil.	Compl. growth for *ΔmycP5*	Conclusion
P1	+	+	-	nd	nd	
P5	+	-	+	+	yes	
P1ΔNE	-	-	nd	nd	nd	N-terminal extension (NE) is required for protease stability (not system specific)
P5ΔNE	-	nd	-	nd	no	
P1NE5	+/-	+	-	nd	no	
P5NE1	+	-	+	nd	yes	
P1ΔTM	++	-	nd	nd	no	TM domain is required and system specific
P5ΔTM	++	nd	-	nd	no	
P1TM5	+	-	-	-	no	Protease domain is system specific
P5TM1	+/-	-	-	-	yes, but slow growth	
P5ΔL1	+/-	nd	-	nd	no	L1 and L2 are required for protease stability (not system specific for L1) L3 and L5 are dispensable
P5ΔL2	+/-	nd	-	nd	no	
P5ΔL3	+	nd	+	nd	yes	
P5ΔL5	+	nd	+	nd	yes	
P5-L1.P1	+	nd	+	nd	yes	
P.P1	+/-	nd	-	+/-	no	P.P1 protease domain complements ESX-5
P.P1TM5	+	nd	+	+	yes	

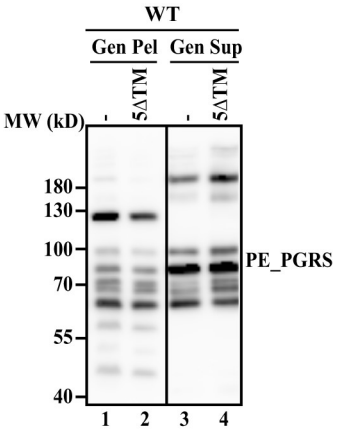
Supporting Table 3. List of primers used in this study.

Name	Sequence (5' → 3')
MYCP5ECORI FW	ccggaattccatatgcagcgattcggtaccgtt
MYCP5HINDIII HA REV	gccgaagcttctacgcgtagtccggcacgtcgtacgggtatcgccgcttccgtga
MYCP5ΔNE FW	cgcaatcaatcctccgacccagccgaagtacatggagat
MYCP5ΔNE REV	atctccatgtacttcggctgggtcggaggattgattgcg
MYCP5ΔTM HA REV	gccgaagcttctacgcgtagtccggcacgtcgtacgggtaattgcgctctgcgggcg
MYCP5NE1 REV	gcatctccatgtacttcggcggatcgtgaaagctggaa
MYCP5NE1 FW	ttccagctttcacgatccgccgaagtacatggagatgc
MYCP5TM1 REV	acaccgccaccattgtgatattgcgctctgcgggcg
MYCP5TM1 FW	cgcccgcagagcgcaatatcacaatggtggcggtgt
MYCP5ΔLOOP1 FW	gccaatcgagccaggcgaagatcgacgacgttgagac
MYCP5ΔLOOP1 REV	gtctcaacgtcgtcgatcttcgcctggctcgattggc
MYCP5ΔLOOP2 FW	gcggccggcgacggcgtcaccacggtggtgacg
MYCP5ΔLOOP2 REV	cgtcaccaccgtggtgacgccgtcgccggccgc
MYCP5ΔLOOP3 FW	cgtgacgacggcttgatcgctggcacgagttttgcg
MYCP5ΔLOOP3 REV	cgcaaaactcgtgccagcgatcaagccgtcgtcacg
MYCP5ΔLOOP5 FW	gtgcggcgccggccaacgatgcattcagcgggattg
MYCP5ΔLOOP5 REV	caatcccgctgaatgcatcgttggccggcgccgcac
MYCP5-LOOP1-P1 FW	gccaatcgagccaggcgttcgaaccggttggctcg
MYCP5-LOOP1-P1 REV	gtctcaacgtcgtcgatcttcggcgtcgcgttgggat
MYCP1ECORI FW	ccggaattccatatgcaggcaggactgacac
MYCP1HINDIII HA REV	gccgaagcttctacgcgtagtccggcacgtcgtacgggtatcggcgcctcagcg
MYCP1ΔNE FW	gcaatcgatcctccctcgccgccgtggagtaacgc
MYCP1ΔNE REV	gcgttactccacggcggcgagggaggatcgattgc
MYCP1ΔTM HA REV	gccgaagcttctacgcgtagtccggcacgtcgtacgggtacgggcggtgatccggg
MYCP1NE5 REV	tataggcgttactccacggctgcaacttgaagtcggtg
MYCP1NE5 FW	caccgacttcaagttgcagccgtggagtaacgcctata
MYCP1TM5 REV	ccacccatactggcaccatcgggcggtgatccggg
MYCP1TM5 FW	cccggatcaccgcccgatggtgccagtatgggtgg
MYCP1MTHNDEI FW	ccgcatatgatcgaacctccggtga
MYCP1MTHXHOI REV	gccgctcgagctacacatcccaggtcagcgc
MYCP1MTHΔL1 FW	ccacacctcggcggcgaccgccggctcgctgc
MYCP1MTHΔL1 REV	gcagcgagccggcggtcgccgccgaggtgtgg
MYCP1.P1 FW	ccggaattccatatggtgaggactgcgcaaaaag
MYCP1.P1 HA REV	gccgaagcttctacgcgtagtccggcacgtcgtacgggtatgaccccgccccccgtc
MYCP1.P1-TM5 REV	cccttcgggcacgtcccacgtcaaggcggccaccg
MYCP1.P1-TM5 FW	cggtggccgccttgacgtgggacgtgcccgaaggg

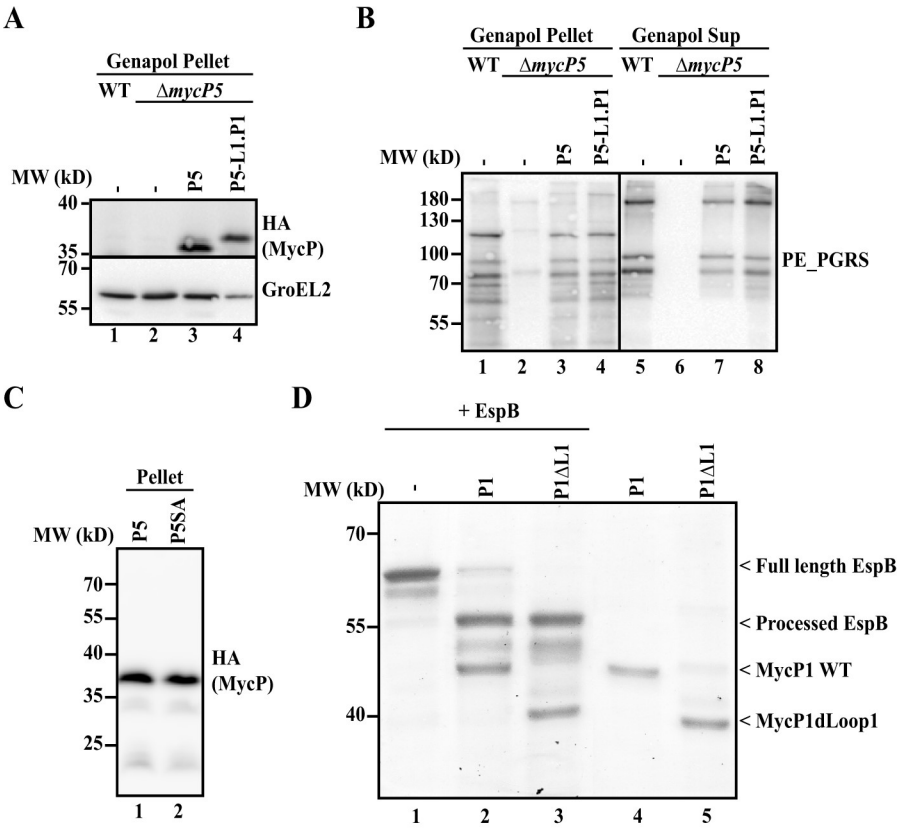
Supporting Figures:



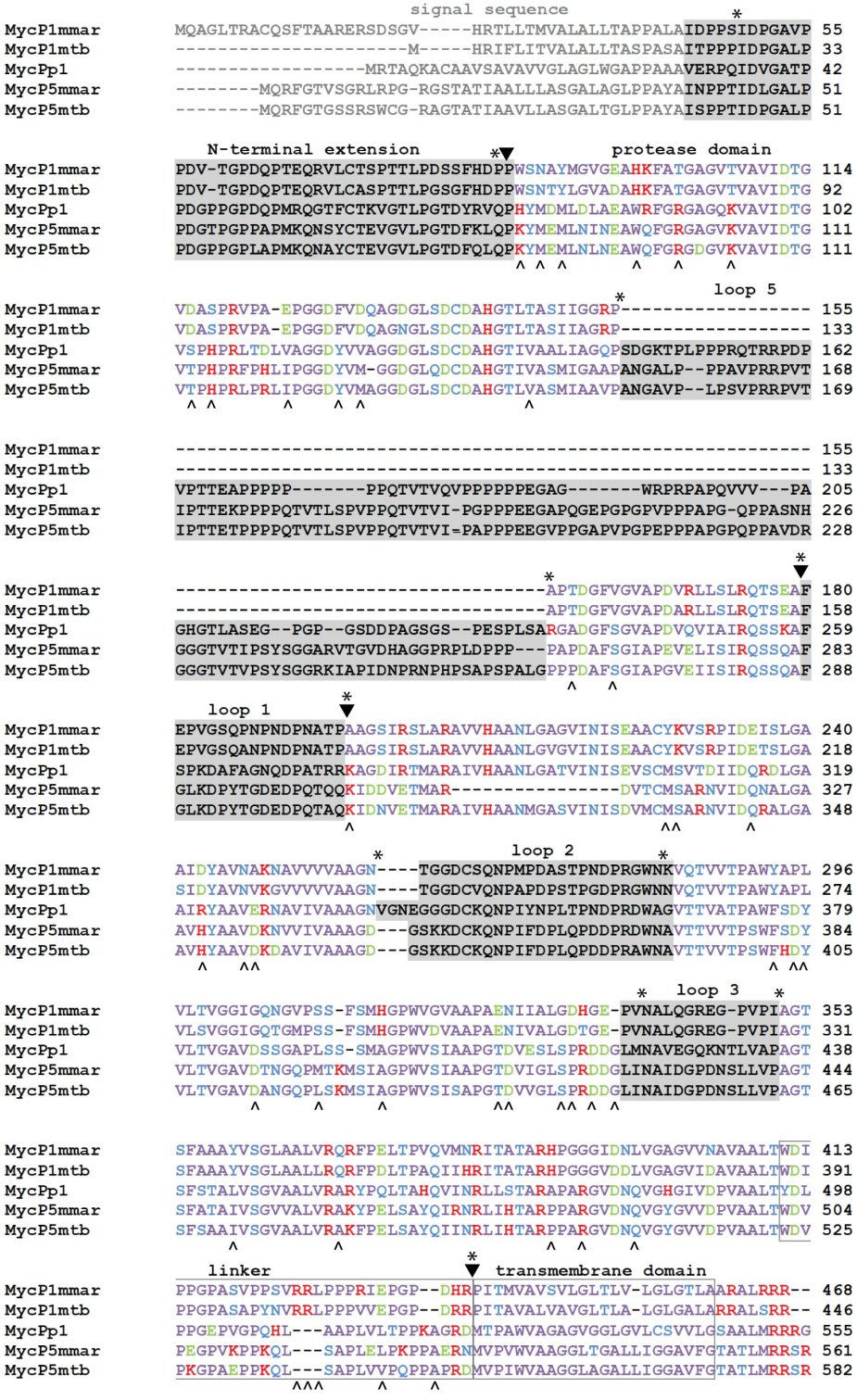
Supporting Figure 1. Immunoblot detection of the hybrid HA-tagged constructs (see Figure 2A), expressed in the *mycP*$_1$ knockout of *M. marinum*.



Supporting Figure 2. Immunoblot analysis of Genapol pellets (Gen Pel, cellular) and Genapol supernatant (Gen Sup, surface localized proteins) of wild-type (WT) *M. marinum*, complemented with a truncated *mycP*$_5$ without its TM region (5ΔTM). Proteins were visualized with anti-PE_PGRS (ESX-5 substrates).



Supporting Figure 3. (A) Immunoblot detection of the HA-tagged constructs of wild-type *mycP*$_5$ (P5) and hybrid *mycP*$_5$ with loop 1 of *mycP*$_1$ (P5-L1.P1). (B) Immunoblot analysis of Genapol pellets (Gen Pel, cellular) and Genapol Supernatant (Gen Sup, surface localized proteins) of wild-type (WT) *M. marinum* and a *mycP*$_5$ knockout, complemented with P5 or P5-L1.P1. Proteins were visualized with anti-PE_PGRS (ESX-5 substrates). (C) Immunoblot detection of HA-tagged wild-type *mycP*$_5$ (P5) and an active site mutant of *mycP*$_5$ (P5SA) in *M. smegmatis*. (D) SDS-PAGE analysis of purified wild-type EspB, both full-length and processed after co-incubation with either purified MycP$_1$ (P1 or MycP1 WT) or MycP$_1$ without loop 1 (P1ΔL1 or MycP1dLoop1). Proteins were visualized with Coomassie Brilliant Blue staining.


signal sequence
N-terminal extension
protease domain
loop 5
loop 1
loop 2
loop 3
linker
transmembrane domain
MycP1mmar
MycP1mtb
MycPp1
MycP5mmar
MycP5mtb
--- 468 Red = basic (KRH)
--- 446 Green = acidic (DE)
AGS 558 Blue = polar (STCYNQ)
KRR 564 Purple = nonpolar (GAVLIMFWP)
KQQ 585

Supporting Figure 4. Multiple sequence alignment of $MycP_1$ and $MycP_5$ of *M. marinum* M, and *M. tuberculosis* H37Rv and $MycP_{P1}$ of *M. marinum* E11. The N-terminal extension and the extended loops are highlighted in grey. The linker domain (linker) and predicted transmembrane domain are indicated by grey boxes. Amino acids constituting the predicted signal sequence are shown in grey. Amino acids of the mycosins, excluding the N-terminal extension and the four loops, are colored based on characteristics (*i.e.* basic, acidic, polar or nonpolar). Residues in the protease and linker domain, excluding the loops, that have conserved characteristics between $MycP_5$ and $MycP_{P1}$, but different from $MycP_1$, are depicted with an open arrowhead (^). Positions where individual domains have been exchanged between $MycP_1$ and $MycP_5$ are indicated by closed arrowheads (▼); positions where specific domains have been deleted are indicated by an asterisk (*).