

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

Vincent J.C. van Winden<sup>1</sup>, Merel P.M. Damen<sup>2</sup>, Roy Ummels<sup>1</sup>, Wilbert Bitter<sup>1,2</sup> & Edith N.G. Houben<sup>2\*</sup>

### Supporting Information:

**Supporting Table 1.** Identity percentage for the complete protein and the loop regions of MycP<sub>1</sub>, MycP<sub>5</sub> and MycP<sub>P1</sub> 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 *mycP<sub>1</sub>* 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<sub>5</sub>* 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<sub>5</sub>* (P5) and hybrid *mycP<sub>5</sub>* with loop 1 of *mycP<sub>1</sub>* (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<sub>5</sub>* 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<sub>5</sub>* (P5) and an active site mutant of *mycP<sub>5</sub>* (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<sub>1</sub> (P1 or MycP1 WT) or MycP<sub>1</sub> without loop 1 (P1ΔL1 or MycP1dLoop1). Proteins were visualized with Coomassie Brilliant Blue staining.

**Supporting Figure 4.** Multiple sequence alignment of MycP<sub>1</sub> and MycP<sub>5</sub> of *M. marinum* M, and *M. tuberculosis* H37Rv and MycP<sub>P1</sub> 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<sub>5</sub> and MycP<sub>P1</sub>, but different from MycP<sub>1</sub>, are depicted with an open arrowhead (^). Positions where individual domains have been exchanged between MycP<sub>1</sub> and MycP<sub>5</sub> 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<sub>1</sub>, MycP<sub>5</sub> and MycP<sub>P1</sub> of *M. marinum*.

<b>Query Protein</b>	<b>Subject Protein</b>	<b>Entire protein (% identity)</b>	<b>Loop 1 (% identity)</b>	<b>Loop 2 (% identity)</b>	<b>Loop 3 (% identity)</b>
MycP <sub>1</sub> (M)	MycP <sub>5</sub> (M)	46.37	17.65	50	25
MycP <sub>1</sub> (M)	MycP <sub>P1</sub> (E11)	48.77	23.53	58.33	33.33
MycP <sub>5</sub> (M)	MycP <sub>P1</sub> (E11)	59.49	41.18	60	46.15

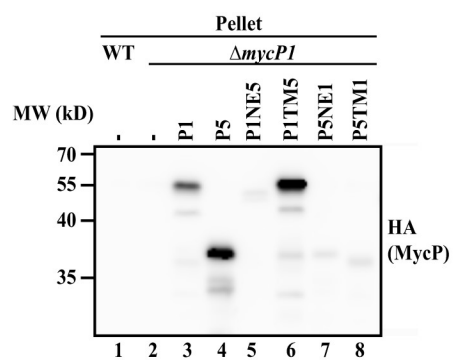
**Supporting Table 2.** Schematic overview of the MycP constructs and their activity in different backgrounds. The composition of the constructs is depicted in Figure 1A and 2A. nd, not determined; -, no expression/secretion/stability; +/-, reduced expression/secretion/stability; +, normal expression/secretion/stability;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

	<b>Mycosin expr.</b>	<b>ESX- 1 secr.</b>	<b>ESX-5 secr.</b>	<b>ESX-5 complex stabil.</b>	<b>Compl. growth for <i>ΔmycP<sub>5</sub></i></b>	<b>Conclusion</b>
<b>P1</b>	+	+	-	nd	nd	
<b>P5</b>	+	-	+	+	yes	
<b>P1ΔNE</b>	-	-	nd	nd	nd	N-terminal extension (NE) is required for protease stability (not system specific)
<b>P5ΔNE</b>	-	nd	-	nd	no	
<b>P1NE5</b>	+/-	+	-	nd	no	
<b>P5NE1</b>	+	-	+	nd	yes	
<b>P1ΔTM</b>	++	-	nd	nd	no	TM domain is required and system specific
<b>P5ΔTM</b>	++	nd	-	nd	no	
<b>P1TM5</b>	+	-	-	-	no	Protease domain is system specific
<b>P5TM1</b>	+/-	-	-	-	yes, but slow growth	
<b>P5ΔL1</b>	+/-	nd	-	nd	no	L1 and L2 are required for protease stability (not system specific for L1)
<b>P5ΔL2</b>	+/-	nd	-	nd	no	
<b>P5ΔL3</b>	+	nd	+	nd	yes	
<b>P5ΔL5</b>	+	nd	+	nd	yes	
<b>P5-L1.P1</b>	+	nd	+	nd	yes	
<b>P.P1</b>	+/-	nd	-	+/-	no	P.P1 protease domain complements ESX-5
<b>P.P1TM 5</b>	+	nd	+	+	yes	

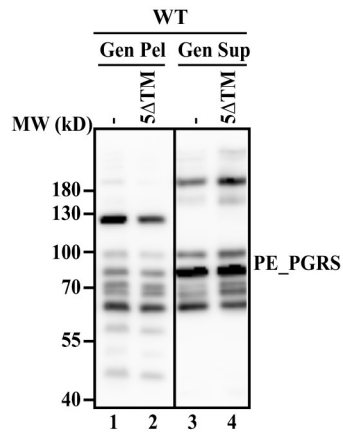
**Supporting Table 3.** List of primers used in this study.

<b>Name</b>	<b>Sequence (5' → 3')</b>
MYCP5ECORI FW	ccggaattccatatgcagcgattcggaccgtt
MYCP5HINDIII HA REV	gccgaagcttctacgcgtagtccggcacgtcgtacgggtatcgccgcttccgtga
MYCP5ΔNE FW	cgcaatcaatcctccgaccagccgaagtacatggagat
MYCP5ΔNE REV	atctccatgtacttcggctgggtcggaggattgattgcg
MYCP5ΔTM HA REV	gccgaagcttctacgcgtagtccggcacgtcgtacgggtaattgcgctctcggggcg
MYCP5NE1 REV	gcattcctcatgtacttcggcggatcgtgaaagctggaa
MYCP5NE1 FW	ttccagctttcacgatccgccgaagtacatggagatgc
MYCP5TM1 REV	acaccgccaccattgtgatattgcgctctcggggcg
MYCP5TM1 FW	cgcccgcagagcgcaatatacaatgggtggcggtgt
MYCP5ΔLOOP1 FW	gccaatcgagccaggcgaagatcgacgacgttgagac
MYCP5ΔLOOP1 REV	gtctcaacgtcgtcgtatcttcgctggctcgattggc
MYCP5ΔLOOP2 FW	gcggccggcgacggcgtcaccacgggtgacg
MYCP5ΔLOOP2 REV	cgtcaccaccgtggtgacgccgtcggccggc
MYCP5ΔLOOP3 FW	cgtgacgacggcttgatcgtggcagagttttgcg
MYCP5ΔLOOP3 REV	cgcaaaactcgtgccagcgatcaagccgtcgtcag
MYCP5ΔLOOP5 FW	gtcggcgccggccaacgatgattcagcgggattg
MYCP5ΔLOOP5 REV	caatcccgtgaatgcatcgttggccggcgccgcac
MYCP5-LOOP1-P1 FW	gccaatcgagccaggcgttcgaaccggttggtcgcg
MYCP5-LOOP1-P1 REV	gtctcaacgtcgtcgtatcttcggcgtcgcgttgggat
MYCP1ECORI FW	ccggaattccatatgcagcgaggactgacac
MYCP1HINDIII HA REV	gccgaagcttctacgcgtagtccggcacgtcgtacgggtatcgccgcctcagcg
MYCP1ΔNE FW	gcaatcgatcctcctcggcggcgggaggatcgattgc
MYCP1ΔNE REV	gcgttactccacggcggcgaggaggatcgattgc
MYCP1ΔTM HA REV	gccgaagcttctacgcgtagtccggcacgtcgtacgggtacggcggtgatccggg
MYCP1NE5 REV	tataggcgttactccacggctgcaactgaagtcggtg
MYCP1NE5 FW	caccgacttcaagttgcagccgtggagtaacccctata
MYCP1TM5 REV	ccaccatactggcaccatcgggcgggtgatccggg
MYCP1TM5 FW	cccggatcaccgcccgatggtgccagtatgggtgg
MYCP1MTHNDEI FW	ccgcatatgatcgaacctccggtga
MYCP1MTHXHOI REV	gccgctcgagctacacatcccagggtcagcgc
MYCP1MTHΔL1 FW	ccacacctcggcggcgaccgccggctcgtcgc
MYCP1MTHΔL1 REV	gcagcgagccggcggctcggccggagggtgtgg
MYCP1.P1 FW	ccggaattccatatggtgaggactgcgcaaaaag
MYCP1.P1 HA REV	gccgaagcttctacgcgtagtccggcacgtcgtacgggtatgaccccgccccctc
MYCP1.P1-TM5 REV	cccttcgggcacgtcccacgtcaaggcggccaccg
MYCP1.P1-TM5 FW	cggtggccgcttgacgtgggacgtgccgaaggg

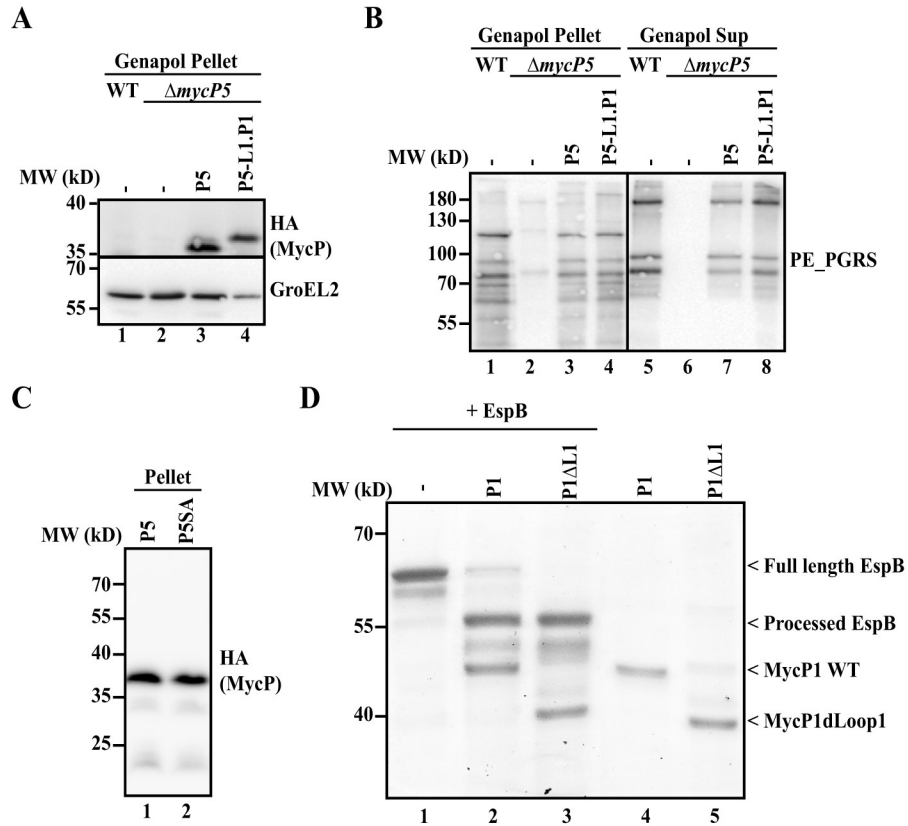
**Supporting Figures:**



**Supporting Figure 1.** Immunoblot detection of the hybrid HA-tagged constructs (see Figure 2A), expressed in the *mycP<sub>1</sub>* 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<sub>5</sub>* 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<sub>1</sub> (P1 or MycP1 WT) or MycP<sub>1</sub> without loop 1 (P1 $\Delta$ L1 or MycP1dLoop1). Proteins were visualized with Coomassie Brilliant Blue staining.

signal sequence \*  
 MycP1mmar MQAGLTRACQSFTAARERSDSGV---HRTLLTMVALALLTAPPALAIIDPPSIDPGAVP 55  
 MycP1mtb -----M-----HRIFLITVALALLTASPASAITPPPIDPGALP 33  
 MycPp1 -----MRTAQKACAASAVAVVGLAGLWGAAPPAAAVERPQIDVGATP 42  
 MycP5mmar -----MQRFGTVSGRLRPG-RGSTATAAALLASGALAGLPPAYAINPPTIDLGALP 51  
 MycP5mtb -----MQRFGTGSSRSWCG-RAGTATIAAVLLASGALTGLPPAYAISPTIDPGALP 51

N-terminal extension \* protease domain  
 MycP1mmar PDV-TGPDQPTQRVLCSTPTLLPDSSFDHPPWSNAYMGVGEAHKFATGAGVTVAVIDTG 114  
 MycP1mtb PDV-TGPDQPTQRVLCASPTLLPGSGFDHPPWSNTYLGVADAHKFATGAGVTVAVIDTG 92  
 MycPp1 PDGPPGPDQPMRQGTFCCTKVTLPGTDIRVQPHYMDMLDLAEAWRFRGAGQKVAVIDTG 102  
 MycP5mmar PDGTPGPPAPMKQNSYCTEVGLVPGTDFKLQPKYMEMLNINEAWQFGRGAGVKVAVIDTG 111  
 MycP5mtb PDGPPGPLAPMKQNAICTEVGLVPGTDFQLQPKYMEMLNINEAWQFGRGAGVKVAVIDTG 111  
^ ^ ^ ^ ^ ^

\* loop 5  
 MycP1mmar VDASPRVPA-EPGGDFVDQAGDGLSDCDAHGTLTASIIIGRP----- 155  
 MycP1mtb VDASPRVPA-EPGGDFVDQAGNGLSDCDAHGTLTASIIAGRP----- 133  
 MycPp1 VSPHRLRLTLVAGGDYVAVAGDGLSDCDAHGTIVAALIAAGQSDGKTPLPVPPRQTRRDP 162  
 MycP5mmar VTPHPRFPHLIPGGDYVM-GDGLQDCDAHGTIVASMI GAAPANGALP--PPAVPRRPVT 168  
 MycP5mtb VTPHPRRLRLLIPGGDYVMAGDGLSDCDAHGTLVASMIAAVPANGAVP--LPSVPRRPVT 169  
^ ^ ^ ^ ^ ^

\*  
 MycP1mmar ----- 155  
 MycP1mtb ----- 133  
 MycPp1 VPTTEA PPPPP-----PPQTVTVQVPPPPPEGAG-----WRPRPAPQVVV--PA 205  
 MycP5mmar IPTTEKPPPPQTVTTLSPVPPQTVTVI-PGPPPEEGAPQGEPPGPPVPPAPG-QPPASNH 226  
 MycP5mtb IPTTETPPPPPQTVTTLSPVPPQTVTVI-PAPPPPEEGVPPGAPVPGPEPPPPAPGPPAVDR 228

\*  
 MycP1mmar -----APTDFGVGVAPDVRLLSLRQTS EAF 180  
 MycP1mtb -----APTDFGVGVAPDARLLSLRQTS EAF 158  
 MycPp1 GHGTLASEG--PGP--GSDDPAGSGS--PESPLSARGADGFSGVAPDVQVIAIRQSSKAF 259  
 MycP5mmar GGGTVTIPSYSGGARVTVGDHAGGPPRLDPPP--PAPDAFSGIAPVEVELISIRQSSQAF 283  
 MycP5mtb GGGTVTVPSYSGGRKIAPIDNPRNPHPSAPSPALGPPDAFSGIAPGVEIISIRQSSQAF 288  
^ ^

\* loop 1  
 MycP1mmar EPVGSQPNPNPNATPAAGSIRSLARAVVHAANLGAGVINISEAACYKVS RPID EISLGA 240  
 MycP1mtb EPVGSQANPNPNATPAAGSIRSLARAVVHAANLGAGVINISEAACYKVS RPID EISLGA 218  
 MycPp1 SPKDAFAGNQDPATRRKAGDIRTMARAI VHAANLGATVINISEVSCMSVTDIIDQRDLGA 319  
 MycP5mmar GLKDPYTDEDPQTQKIDNVE TMAR-----DVTCMSARNVIDQNALGA 327  
 MycP5mtb GLKDPYTDEDPQTQKIDNVE TMARAI VHAANMGASVINISDVCMCSARNVIDQRALGA 348  
^ ^ ^

\* loop 2  
 MycP1mmar AIDYAVNAKNAV VVAAGN---TGGDCSQNPMPDASTPNDPRGWNK VQT VVTPAWYAPL 296  
 MycP1mtb SIDYAVNVKGV VVAAGN---TGGDCVQNPAPDPSTPGDPRGWNVQTVVTPAWYAPL 274  
 MycPp1 AIRYAAVERNAVIVAAAGNVGNEG GDCQNPIYNPLTPNDPRDWAGVTTVATPAWFS DY 379  
 MycP5mmar AVHYAAVDKNVIVAAAGD---GSKKDCQNPIFDPLQDDPRDWNVTTVVTPSWFS DY 384  
 MycP5mtb AVHYAAVDKDAVIVAAAGD---GSKKDCQNPIFDPLQDDPRAWNVTTVVTPSWFHDY 405  
^ ^ ^ ^ ^

\* loop 3 \*  
 MycP1mmar VLTVGGIGQNGVPSS-FSMHGPVWVGAAPAENI IALGDHGE-PVNALQGREG-PVPIAGT 353  
 MycP1mtb VLSVGGIGQTMGPSS-FSMHGPWVVAAPAENIVALGDTGE-PVNALQGREG-PVPIAGT 331  
 MycPp1 VLTVGAVDSSGAPLSS-SMAGPWVSAAPGTDV ELSPRDDGLMNAVEGQKNTLVAPAGT 438  
 MycP5mmar VLTVGAVD TNGQPMTKMSIAGPWVSAAPGTDVIGLSPRDDGLINAIDGPDNSLLVPAGT 444  
 MycP5mtb VLTVGAVDANGQPLSKMSIAGPWVSAAPGTDVIGLSPRDDGLINAIDGPDNSLLVPAGT 465  
^ ^ ^ ^ ^ ^ ^

\*  
 MycP1mmar SFAAAVYVSGLAALVRQRFP ELPVQVMNRITATARHPGGGIDNLVAGV VNAVAALTWDI 413  
 MycP1mtb SFAAAVYVSGLAALLRQRFPDLTPAQI IHRITATARHPGGGVDDL VGAGVIDAVAALTWDI 391  
 MycPp1 SFSTALVSGVAALVRARYPQLTAHQVINRLLSTARAPARGVNDQVGHGIVDPVAALTVDL 498  
 MycP5mmar SFATAIVSGVVALVRAKYPELSAYQIRNRLIHTARPPARGVNDQVGVVDPVAALTVDV 504  
 MycP5mtb SFSAAIVSGVAALVRAKFPELSAYQIINRLIHTARPPARGVNDQVGVVDPVAALTVDV 525  
^ ^ ^ ^ ^

\* linker transmembrane domain  
 MycP1mmar PPGPASVPPSVRRLPPRI EPGP--DHRPITMVAVSVLGLTLV-LGLGTLAARALRRR-- 468  
 MycP1mtb PPGPASAPYVNRRLPPPV EPGP--DRRPI TAVLVAVGLTLA-LGLGALARRALSRR-- 446  
 MycPp1 PPGPEVPGQHL---AAPLVLP PKAGRDMTPAWVAGAGVGLGVLCVSVLGSAAALMRRR 555  
 MycP5mmar PEGVPKPKQL---SAPLELPKPPAERNMVPVWVAAGGLTGALLIGGAVFGTATLMRRSR 561  
 MycP5mtb PKGPAEPPKQL---SAPLVVPPAPPRDMVPIWVAAGGLAGALLIGGAVFGTATLMRRSR 582  
^ ^ ^ ^ ^

MycP1mmar --- 468 Red = basic (KRH)  
 MycP1mtb --- 446 Green = acidic (DE)  
 MycPp1 AGS 558 Blue = polar (STCYNQ)  
 MycP5mmar KRR 564 Purple = nonpolar (GAVLIMFWP)  
 MycP5mtb KQQ 585



**Supporting Figure 4.** Multiple sequence alignment of MycP<sub>1</sub> and MycP<sub>5</sub> of *M. marinum* M, and *M. tuberculosis* H37Rv and MycP<sub>P1</sub> 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<sub>5</sub> and MycP<sub>P1</sub>, but different from MycP<sub>1</sub>, are depicted with an open arrowhead (^). Positions where individual domains have been exchanged between MycP<sub>1</sub> and MycP<sub>5</sub> are indicated by closed arrowheads (▼); positions where specific domains have been deleted are indicated by an asterisk (\*).