Supplementary data

Understanding specificity of the mycosin proteases in ESX/type VII secretion by structural and functional analysis

Jonathan W. Wagner¹, Timothy J. Evans¹, Jing Chen¹, Haining Zhu¹, Edith N. G. Houben², Wilbert Bitter² and Konstantin V. Korotkov^{1*}

¹Department of Molecular & Cellular Biochemistry and Center for Structural Biology, University of Kentucky, Lexington, Kentucky 40536, USA ²Department of Medical Microbiology and Infection Control, VU University Medical Center, Amsterdam, The Netherlands

*Correspondence to: Konstantin V. Korotkov, Department of Molecular & Cellular Biochemistry, University of Kentucky, 741 South Limestone, Lexington, KY 40536, USA. Phone: 859-323-5493. Fax: 859-323-5504. E-mail: kkorotkov@uky.edu

	signal peptide		ptide –	N-terminal extension		
		1 10	20	30	40	50
МусР1 МусР2 МусР3 МусР4 МусР5	MASPLNRPG	MHRIFLITVAL LRAAAASAALTLVA IRAAFACLAATVVV RTLRLLVVSALATL AGTATIAAVLLASG	ALLTASPASAT LSANVPAAQAI AGWWTPPAWAI SGLGTPVAHAV ALTGLPPAYAI	TPPPIDPGALPP PPSVDPAMVPA GPPVVDAAAQPP SPPPIDERWLPE SPPTIDPGALPP	DVT.GPDQPTE DARPGPDQPMR SGDPGPVAPME SALPAPPRPTV DGPPGPLAPMK	QRVLCASPTTLP RSNSCSTPITVR QRGACSVSGVIP QREVCTEVTAES QNAYCTEVGVLP
МусР1 МусР2 МусР3 МусР4 МусР5	60 70 GSGFHDPPWSNTYLG NPDVAQLAPGFNLVN GTDPGVPTPSQTMLN GRAFGRAERSAQLAD GTDFQLQPKYMEMLN	80 VADAHKFATGAGVT ISKAWQYSTGNGVP LPAAWQFSRGGQL LDQVWRLTRGAGQR LNEAWQFGRGDGVK	90 VAVIDTGVDAS VAVIDTGVSPNI VAIIDTGVQPG VAVIDTGVARHI VAVIDTGVTPH	100 PRVP. AE PGGDF PRLP.VVPGGDY PRLPNVDAGGDF RRLPKVVAGGDY PRLPRLIPGGDY	110, 1 VDQAG <mark>NGLSDC</mark> IMG.EDGLSDC VES.TDGLTDC VFT.GDGTADC VMAGGDGLSDC	20 130 DAHGTLTASIIA DAHGTVVSSIIA GHGTLVAGIVA DAHGTLVAGIA DAHGTLVASMIA
			r	proline-rich insertion		
МусР1 МусР2 МусР3 МусР4 МусР5	GRPAP. AAPLGILPMPR GQPGN. AAPDAQ. AVPANGAVPLPSVPR	AMPATAAFPPPAGP RPVTIPTTETPPPP	PPVTAAPAPPVI	EVPPPMPPPPV	TITQTVAPPPP	PPEDAGAMAPSN EPPPAPGPQPPA
МусР1 МусР2 МусР3 МусР4 МусР5	GP VDRGGGTVTVPSYSG	PDPQTED GRKIAPIDNPRNPH	EPAVPPPPPGA	140 TDGFVGVAPDAR PDGVVGVAPHAT .DGFSGVAPAR SDNFSGVAPDVT PDAFSGIAPGVE	150 LLSIROTSEAF IISIROSSRAF LLSIROSSKF IISIROSSOAF IISIROSSOAF	60 170 EPVGSQANPNDP EPVNPSSAGPNS SPRTSGGDP APVGDPSSTG GLKDPYTGDEDP
	180	190 20	0 210	220	230	240
МусР1 МусР2 МусР3 МусР4 МусР5	NATPAAGSIRSLARA DEKVKAGTLDSVARA QLAQATLDVAVLAGA VGDVDTMAKA QTAQKIDNVETMARA	VVHAANLGVGVINI IVHAANMGAKVINI IVHAADLGAKVINV VRTAADLGASVINI IVHAANMGASVINI	SEAACYKVSRP SVTACLPAAAP STITCLPADRM SSIACVPAAAA SDVMCMSARNV	IDETSLGASIDY GDQRVLGAALWY VDQAALGAAIWY PDDRALGAAIRY IDQRALGAALAY	AVNVKGVVV AATVKDAVIVA AAVDKDAVIVA AVDVKDAVIVA AVDVKNAVIVA AAVDKDAVIVA	AAGNTGGD AAGNDGEAG AAGNTGASGSVS AAGNTGGA AAGNGGSKK
	250	260 2	70 28	0 290	300	310
МусР1 МусР2 МусР3 МусР4 МусР5	CVQNPAPDPSTPG CGNNPMYDPLDPS ASCDSNPLTDLSRPDD AQCPPQAPGVTR .DCKQNPIFDPLQPDD	DPRGWNN <mark>VQTVVTP</mark> DPRDWHQ <mark>VTVVSS</mark> P DPRNWAGVTSVSIP DS <mark>VTVAVS</mark> P DPRAWNA <mark>VTTVVT</mark> P	AWYAPLVLSVG SWFSDYVLSVG SWWQPYVLSVA AWYDDYVLTVG SWFHDYVLTVG	GIGQTGMP.SSF AVDAYGAA.LDK SLTSAGQP.SKF SVNAQGEP.SAF AVDANGQPLSKM	SMHGPWVDVAA SMSGPWVGVAA SMPGPWVGIAA TLAGPWVDVAA SIAGPWVSISA	PAENIVALGDTG PGTHIMGLSPQG PGENIASVSNSG TGEAVTSLSPFG PGTDVVGLSPRD
	320	330	340 31	50 360	370	380
МусР1 МусР2 МусР3 МусР4 МусР5	EPVNALQGRE(GG.PVNAYPPSRPGE) DGALANGLPDAHQ DG.TVNRLGGQH DG.LINAIDGPDN	GPVPIAGTSFAAAY KNMPFWGTSFSAAY KLVALSGTSYAAGY GSIPISGTSYAAPV SLLVPAGTSFSAAI	VSGLAALLRQR VSGVAALVRAKI VSGVAALVRSR VSGLAALIRAR VSGVAALVRAKI	FPDLTPAQIIHR FPELTAYQVINR YPGLNATEVVRR FPTLTARQVMQR FPELSAYQIINR	ITATARHPGGG IVQSAHNPPAG LTATAHRGARE IESTAHHPPAG LIHTARPPARG	VDDLVGAGVIDA VDNKLGYGLVDP SSNIVGAGNLDA WDPLVGNGTVDA VDNQVGYGVVDP
		linker	t	ransmembran	e helix	
	390	400 41	<u>0</u> 420	430	440	-
MycP1 MycP2 MycP3 MycP4 MycP5	VAALTWD.IPPGP VAALTFN.IPSGD VAALTFN.IPSGD LAAVSSDSIPQAGTA VAALTWD.VPKGP	A S A P Y N V <mark>R R</mark> L P P P V R M A P G A Q <mark>S R</mark> V I T P A G G . A A P <mark>A K</mark> P V A D P I S D P A P V <mark>A V</mark> P V P R R A E P P K Q L <mark>S A</mark> P L V V P	VEPGPDRRPIT APPPPDHRAR PVPAPKDTTPR STPGPSDRRAL QPPAPRDMVPI	AVALVAVGLTLA NIAIG.FVGAVA NVAFAGAAALSV HTAFAGAAICLL WVAAGGLAGALL	LGLGALARRAL TGVLAMAIGAR LVGLTAATVAI ALMATLATASR IGGAVFGTATL	SKR LRRAR ARRRREPTE RLRPGRNGIAGD M <mark>RR</mark> SRKQQ

Fig. S1. Sequence alignment of *M. tuberculosis* mycosin proteases.

The N-terminal extension is indicated by blue line. The proline-rich insertion, which is present in $MycP_2$ and $MycP_5$, is indicated by magenta line. The conserved Cys residues are highlighted in yellow. The catalytic triad residues, as well as the residues coordinating the oxyanion hole, are highlighted in blue. Note that Asp367 is the residue coordinating the oxyanion hole in $MycP_5$.



Fig. S2. Mass spectrometry analysis of the disulfide-linked peptides from MycP1mth.

A and B. MycP_{1mth} sample was not pretreated with dithiothreitol (DTT) and iodoacetamide (IAA) prior to trypsin digestion.

C and D. $MycP_{1mth}$ sample was pretreated with DTT and IAA prior to trypsin digestion. The peaks labeled in (A) are the 6+ and 5+ ions of the disulfide-linked peptides:

VC⁵¹AAPTVYPDSNFADR–VPAEPGGDFVDAAGNGMSDC¹²⁰DAHGTLTASVIAGR. The peaks labeled in (B) are the 5+ and 4+ ions of the disulfide-linked peptides: AIVHAANLGAQVINISEAAC²⁰⁶YKVTR–

PIDETGVGAAVNYAVHVKNAVVIAAAGNTGQD**C**²⁴⁴TQNPPPDPAVPSDPR. These peaks disappeared in the corresponding LC-MS analysis of the tryptic digestion samples with DTT pretreatment (C and D).

>MycPl
${\it MQAGLT} {\it RACQSFTAARERSDSGVHRTLLTMVALALLTAPPALA} {\bf IDPPSIDPGAVPPDVTGPDQPTEQRVLCTSPTTLPDS}$
SFHDP PWSNAYMGVGEAHKFATGAGVTVAVIDTGVDASPRVPAEPGGDFVDQAGDGLSDCDAHGTLTASIIGGRPAPTDG
FVGVAPDVRLLSLRQTSEAFEPVGSQPNPNDPNATPAAGSIRSLARAVVHAANLGAGVINISEAACYKVSRPIDEISLGA
AIDYAVNAKNAVVVVAAGNTGGDCSQNPMPDASTPNDPRGWNKVQTVVTPAWYAPLVLTVGGIGQNGVPSSFSMHGPWVG
$VAAPAENIIALGDHGEPVNALQG\underline{REGPVPIAGTSFAAAYVSGLAALV\underline{RQ}\underline{RFPELTPVQVMN}\underline{RITATA\underline{R}HPGGGIDNLVGA}$
GVVNAVAALTWDIPPGPASVPPSV <u>RR</u> LPPP <u>R</u> IEPGPDH <u>R</u> PITMVAVSVLGLTLVLGLGTLAA <u>R</u> AL <u>RR</u>
>MycP3
MKRAGSACLTAILAIFTTGALWAAPPVSAISPPAVDDSVQPPSGTPGPVQTMEQRGACSVSGLIPGTDISVPGASQAMLN
LPAAWQFSRGEGQLVAILDTGVQPGPRLPNVDAGGDFVDSTDGLTDCDGRGTLVAGIVAGQPGNDGFSGVAPAARLLSIR
VTSDKFSARSSGGDPSVERATIDVAALGRAIVHAADRGARVINISSITCLPADRQVDQAALGAAIRYAAVDKDVVIVAAA
GNSGAGGSGSGGTCESNPLTDLSHPDDPRNWAGVTSVSIPSLWQPY <mark>VLSVASLTAGGQPSK</mark> FTMAGPWVGIAAPGENIVS
VSNADGGGLANGLPNQRQQLTPLSGTSYAAGYVSGVAALVRSKYPELTAAQVMHRLTATAHNGARDPSNVVGAGILDPAA
ALTWELPAADGGAPQVAPAKPVSVPPAPQPRNTTPRNVAFAGAAALTVLVAGVAATVAIARRRKDAEGRTAREL
>MycP5
$\underline{MQR} FGTVSGRLRPGRGSTATIAALLLASGALAGLPPAYA \textbf{INPPTIDLGALPPDGTPGPPAPMKQNSYCTEVGVLPGTDFK}$
LQ PKYMEMLNINEAWQFGRGAGVKVAVIDTGVTPHPRFPHLIPGGDYVMGGDGLQDCDAHGTIVASMIGAAPANGALPPP
AVP <u>RR</u> PVTIPTTE <u>K</u> PPPPQTVTLSPVPPQTVTVIPGPPPEEGAPQGEPGPGPVPPPAPGQPPASNHGGGTVTIPSYSGGA
RVTGVDHAGGPRPLDPPPPAPDAFSGIAPEVELISIRQSSQAFGLKDPYTGDEDPQTQQKIDDVETMARAIVHAANMGAS
VINISDVTCMSARNVIDQNALGAAVHYAAVDKNVVIVAAAGDGSKKDCKQNPIFDPLQPDDPRDWNAVTTVVTPSWFSDY
VLTVGAVDTNGQPMTKMSIAGPWVSIAAPGTDVIGLSPRDDGLINAIDGPDNSLLVPAGTSFATAIVSGVVALVRAKYPE
LSAYQIRNRLIHTARPPARGVDNQVGYGVVDPVAALTWDVPEGPVKPPKQLSAPLELPKPPAERNMVPVWVAAGGLTGAL
LIGGAVFGTATLMRRSRKRR

Fig. S3. Mycosin-derived peptides detected by MS/MS analysis.

Amino acid sequences of $MycP_{1mm}$, $MycP_{3mm}$ and $MycP_{5mm}$. Peptide sequences that were detected by MS/MS analysis are highlighted in grey. Potential trypsin digest sites are underlined. The signal sequences and N-terminal extensions are indicated in italic and bold fonts, respectively. The data were derived from: van der Woude *et al.* (2013) J. Bacteriol. 195, 2050-2059.



Fig. S4. Topology diagram of MycP_{1mth}.

A. MycP_{1mth} consists of a 7 stranded β sheet core (green) surrounded by 8 α helices (red). Three small anti-parallel β sheets are located surrounding the active site pocket. Two of these small β sheets include neighboring disulfide bonds (SHHS). Elements contributed by the Nterminal extension are colored blue. Elements contributed by insertion loops 1-3 are colored purple. Secondary structure diagram made with the help of the Pro-origami server (http://munk.csse.unimelb.edu.au/pro-origami/).

B. Cartoon of MycP_{1mth} with secondary structure elements colored as in (A). The catalytic triad is shown as yellow sticks. Side chains of cysteine residues forming disulfide bonds are shown as orange sticks.

	PE domain
M_tuberculosis_H37Rv	αl α2 <u>000000000000000000000000000000000000</u>
M marinum M M_kansasii_ATCC_12478 M_rhodesiae_NBB3 M_smegmatis_strMC2_155 M_thermoresistibile_ATCC_19527 M_tusciae_JS617 M_chubuense_NBB4 Mycobacterium_spMCS M_vanbaalenii_PYR-1 M_phlei_RIVM601174 M_gilvum_PYR-GCK	MSQPQTVTVDQQEILLNRADEVEAPMATPPTDVPQAPSCLTAANNAAEQLAVSADNVRLYLQAGERER MSQPQTVTVDQQEILTRANDVEAPLPPGKVPPTDVPNPPCALTAAKNAAQQLAUSAENMREFLAAGAKER MTVQRVEPGALTSQAAEMKGQNWHSPAEDAVAPPDAVPSTTDAIANLNENAQSLKEFERWAEVEN .MSEELQYELPGLERKAHECESTRPEGPGDATKPDELATTASVYSKLMASAAKLKATFAAGDREG .MTELRVEPSELQAKATVISAPWPPMPTDPQPACGLELAGIAVSQLYAASAETRAKIAAGEVEA MTVQRVDPTSLTGQAAEMKGQNWHSPSEDAVTPPDALPSTNDAIANLNANAQSLKEFERWAEVEN .MAGDEVRVDPDLNRKAGQIEAISWGDN.PVEMPTVPADGLGTARTAVENLNQNAKTITDFQEYGKKQG .MAGDEVRVEPADLLRAKAGQIEAISWGDN.PVEMPTVPADGLGTARTAVENLNQNAKTITDFQEYGKKQG .MAD.ELKVDPDDLLNRKAGQIEAIPWGED.PSAVAVADPDKLTPTTASVRNLIKNAEALGAEQRWGIAES .MSEQVVRVEPADLRARATLDEEITDEQRRVDVQAPDGLKCSQAAVQNLSANAKAMLGFQRWIETEN .MADGELKVDPNDLTTKAATVEGIPWGSD.PAAVPLSDPDKLSSTTASVQNLVKNATALGAEQKWGQAES
	000000000000000000000000000000000000
M_tuberculosis_H37Rv M_marinum_M M_kansasii_ATCC_12478 M_rhodesiae_NBB3 M_smegmatis_strMC2_155 M_thermoresistibile_ATCC_19527 M_tusciae_JS617 M_chubuense_NBB4 Mycobacterium_spMCS M_vanbalenii_PYR-1 M_phlei_RIVM601174 M_gilvum_PYR-GCK	QRLATSLRNAAAYGEVDEEAATALDNDGEGTVQAËSAGAVGDSSAELTDTRVATAGEPN QRLATSLRNAAAYGEVDEEAATALDNDGNGEVDAOSAGGAGGAGGAGQTESLEETPKVAAAGESD QKLATSLRNAAAYGEVDESSATALNSDGSGTVEAOSAGGAGGAGQTESLEETPKVAAAGESD QKLATSLRNAAKAYGEVDDYGTAIDNPERITAVDATAVPAPQTQTPEIPGPASTPRLLDASG ERIAAAIRAAAGAYQKVDDDYGTAIDNPERITAVDATAVPAPQTQTPEIPGPASTPRLLDASG ERIAAAIRAAAGAYQKIEEQKAAELSRQMNGSDAPPPAAEAVVYDDMSGIPGPLAIPSMEYPSAAAA.AD RRLSQSLVAAANAYTEFDARAKAALDAGEPIPSEPVFPRYDIEPPAPETGLPIGQCPGPPG QRIAEMLEIAAQAYQKVDDDYGKAIDDPERAAVAAITTIASPQTQPPEIPGPPSTRLLDASG QRLAETLRSVAAAYTVDQNAMDNINKT.IPG.ETPAPPAPVIPAANAIPAPTPPQPLPTQRGMTDE ERLAAVLSEAADVYQOIDNRAAQALEYPDGPRSVEPVPVNPTLPPSVPIEFPATVP.AVPGGQT.DAG RRLAETLRLVAQAYRAVDEASAGNIDST.MPGGTSSPAPAPVPIPPNEFPATVP.AVPGGQT.ADG DYLAQLFANTATAYEAVDAACGKALEAPGRAAAVEQIQVVPPQVEYPPPVPVGAQETPAAVE RRLAETLRLVAAAYQDVD
	PPE domain
1	<u>000000000</u> <u>00000000000000000000000000</u>
M_tuberculosis_H37Rv M_marinum_M M_kansasii_ATCC_12478 M_rhodesiae_NBB3 M_smegmatis_strMC2_155 M_thermoresistibile_ATCC_19527 M_tusciae_JS617 M_chubuense_NBB4 Mycobacterium_spMCS M_vanbaalenii_PYR-1 M_phlei_RIVM601174 M_gilvum_PYR-GCK	FMDLKEAARKLETGD. QGASLAHFADGWNTFNLTLQGDVKRFRGFDN. WEGDAATACEASLDQORQWILH FTDLKTAATKLESGD. QGTSMNNFADGWNNFNLSLQRDIKRFRGFDN. WEGDAATACEASMDQOKEWILH FTDLKTAATKLESGD. QGTSMNNFADGWNNFNLSLQRDIKRFRAFD. WEGDAATACEASMDQOKEWILH YDLKTAATKLESGD. QGRSLVDFANAWNDYNFALQGDVKRFRAFD. WEGDAATACEASMDQOKEWILH YSNVKQTQAELTAPD. TGTSLKTAMLHWGIASKRVENNRP. KPPPGDWEGDAADAAYARMTTFGSWLTQ EMDWEAAARIIHSGDTQALSMKYFRDQWRDYQSTLEGHGRHFANPAEGWAGAAAETCAEAQRRLSTWWAD YLDVKTAAAQINAPD. QTASLQEFAANWRAYSTQLSERANSFNLDDVNWGGTGAESAQDAMTEHRQWLQA YSNVIQTQAELSAPD. TGTSLKTAMLWGIASNRVENNRP. KPPPGDWEGEAADAAYARMTTFGSWLTQ YMDVKQAQGSLLSAPD. TGTSLKTAMLQWGIASNRVENNRP. KPPPGDWEGEAADAAYARMTTFGNWLTQ YMDVKQAQGSLLSAPD. TGTSLKTAMLQWGIASNRVENNRP. KPPPGDWEGEAADAAYARMTTFGNWLTQ YMDVKQAQGSLLSGD. QAASLQAAAAKWAQNGATLAAAAANFETSDVNWEGEAADAAYARMTTFGNWLTQ YMDVKAAAQIIHSG. DSGPMRTYGEQAKKFAGTLRDRAASAFEILDGINWDGTAAEKAGDAVROHKEWLEK ILDPIETDNKLLEGD. QAAALRAAAARWSANAVRLADASAPFEIKIQNWEGEAADAAYIKFKSFGGWLQU YYIVPETEKALQQGD. DGASLRDAWFQWSHASRLADDAQV. EPDLRNWEGEAADAAQORLHRFNEQLVQ SLDPIQTDKKLAEGD. QGASLRAAATEWTANAARLTEAAIPFEVRMQNWEGVAAEAAYTKLKSFGGWLQA
	α5 α6 α7
M_tuberculosis_H37Rv M_marinum_M M_kansasii_ATCC_12478 M_rhodesiae_NBB3 M_smegmatis_strMC2_155 M_thermoresistibile_ATCC_19527 M_tusciae_JS617 M_cubuense_NBB4 Mycobacterium_spMCS M_vanbaalenii_PYR-1 M_phlei_RIVM601174 M_gilvum_PYR-GCK	200 210 220 230 240 250 MAKLSAAMAKOAQYVAQLHVWARREHPTYEDIVGLERLYAENPSARDQ. ILPVYAEYQ MAKLSASLAKOANFMAQLQLWARRCHPTLADIVVELERLAKD.PDYQEQ. AIKLYAEYQ MAKLSAALAKOAQYIAQLUVWARRCHPTLADIVVELERLAKD.PDYQEQ. AIKLYAEYQ MAKLSAALAKOAQYIAQLUVWARRCHPTLADIVVELERLAKD.PDYQEQ. AIKLYAEYQ MAKLSAALAKOAQYIAQLUVWARRCHPTLADIVVELERLAKD.PDYQEQ. AIKLYAEYQ MAKLSAALAKQAQYIAQLUVWARRCHPTLADIVVELERLAKD.PDYQEQ. AIKLYAEYQ MAKLSAALAKQAQYIAQLUVWARRCHPTLADIVYELERAMMQLASLTGPHGGVRAQQEIEKIRRMEELQ AIKLYAEYQ MAKDAECGRLAQEATTFVDAHDKLSVANHPTLENVREFEETEWASEWDR. QNAWAMLQ MAGVAENLADHAEKLADAHKKAVQEHP.TEEEVERVENTIVRQN. FRWMDWHQLQ LSEAWYDLAEAASKILSAHNEAKSAHDPITKEYTELEARMRELANVTGPHGGVRVQREMERIRKRMEELQ LSGAWTRLAAEAEKVAMAHATTLGNHTPIAEEYAQLEAQLPAALANGGS. AARVIQLKMEKLQ LAESYEFLGTEATIDMADAHDKWSGOHP.TVAEVEVEAVEARVAQAGQAGKD. ALARAMMUYLALQ LAEAYEFLGTEATIDMADAHDKWSGUP.TURVAAUVAAQYEALQTQLMTTSMDGG. AQRAVQLQMEQLY LSEGWRELAQAADKIRQAHEGAOSAHTPVAEEYEQLEEQLKALPQRITPGNYEWVMKEQRIHQRMEELQ LAGKWNQLAAEAEKLAAAHDQAKAANAPIRIEFEALQSEFVKQGMA. ASQASRDRLQQLY
M_tuberculosis_H37Rv M_marinum_M M_kansasii_ATCC_12478 M_rhodesiae_NBB3 M_smegmatis_strMC2_155 M_thermoresistibile_ATCC_19527 M_tusciae_JS617 M_chubuense_NBB4 Mycobacterium_spMCS M_vanbaalenii_PYR-1 M_phlei_RIVM601174 M_gilvum_PYR-GCK	260 270 280 QRSEKVITE YNNKAALEPVNPKPPAIKIDPP ETSEKVISEVNTKADLEPVNPPKPPAIKIDPP QKSEQVISEVNTKATLEPVNPPKPPAIKIDPP AQSDQVRQEYAGDATFAPVRPADP AQSDQVRQEYAGDATFAPVRPADP AQSDQVRQEYAGDATFAPVRPADP AQSDQVRQEYAGDATFAPVRPADP AQSDQVRQEYAGDATFAPVRPADP AQSDQVRQEYAGDATFAPVRPADP AQSDQVRQEYAGDATFAPVRPADP AQSDQVRQEYAGDATFAPVRPADP AQSDQVRQEYAGDATFAPVRPADP AQSDEVREYASATADSALSGIEVPPPPTGAANVPVTTNGDPRGRYSRDADLKAKYIGEGAPGGGGAR AKSDEVREYASSATFAPVRPADP ARSDDVRKSALSALSGIEVPPPPRGAPTPVRSNGDPR. PAGKPPQEPRGGSSG. QESETATRETYAKAGQPTPVRPAPVPPAPAVPYPISSNGDPR. AQSOCRVREYARAGQPTPVQPPQPRPDAVAPAPAVTGNGDPNRRGMPAHTEESRRGGSSAGQPSSGG KESEATRETYAKAGQPTPVQPPQPRPDAVAPATTRSGGSSS. QESEATRETYAKAGQCPTVQPPQPRPDAVAPATTRSGGSSGSS. RESEAVREITYANGGRPTPVVQPPQPRPDSVAPTSNGDPRDRCVPARTEDEERRDFWGAGEOPNAGG

Fig. S5. Sequence alignment of EspB proteins (continues on the next page).

l

	290	300	310	320	330
M_tuberculosis_H37Rv M_marinum_M M_kansasii_ATCC_12478 M_rhodesiae_NBB3 M_smegmatis_strMC2_155 M_thermoresistible_ATCC_19527 M_tusciae_JS617 M_chubuense_NBB4 Mycobacterium_spMCS M_vanbaalenii_PYR-1 M_phlei_RIVM601174 M_gilvum_PYR-GCK		QGTIPGFLMPF QGLIPSFLMPF QKIASLGGAQ PSMSPMSANSA MGSPVASPFPV QKMGESMGSPP AQMSPMSAADQ QAAAPMGGS APVSPMSAAEQ DRMGDSLGGPT APVSPP.VAEQ	SDGSGVTPGTGM GDGSTG.LASGM SDGSGGTPVTGM SQVGSQQGGGGS GEEQSSGSPSSG PASPEPVTAAAE PQSGSQQGGG AAQAAQQGGGSPG PAGQGGQGAQPAGA AAQGAQQAAPQG PQGAG PPQATQQSGQQE	IPAAPMVPPTG ITPP.MIPPTG IPA.PMAPAG IPSGGSSSGG IGSPSGSPSGG ICGGAPAGGC I.AGGSPSGG ISSCGGSPSG ISSCGGSPSG ISSCGGSPGG ISSPAGGQQGG	SPGGLPADTA GAGG.TPDVNT GAGGGMPAGTS GAGGGMPAGTS GASPSAGGSPGGAPG SPGGAPSGGGAPGGGMPGGL GSPPAGGGSPGGGMPGGL GSSPAGGGAPAGAPAGGGL SQGDGSQGGSPAGGSQGGAPG GAPAGGSPGGSPGGSPG GQGAPGGSPGGSPGGSPG
	prima	ary MycP ₁			secondary MycP ₁
	cleav 340 350	age site		370	cleavage site
M_tuberculosis_H37Rv M_marinum_M M_kansasii_ATCC_12478 M_rhodesiae_NBB3 M_smegmatis_strMC2_155 M_thermoresistibile_ATCC_19527 M_tusciae_JS617 M_chubuense_NBB4 Mycobacterium_spMCS M_vanbaalenii_PYR-1 M_phlei_RIVM601174 M_gilvum_PYR-GCK	AQLTSAGREAAA.LSG AELTSAGREAASNLSK GGPAPSAPKLPT SDLPGGPDIPGLD PGGAPGGKPEMPKMPT GETPDGTPKLPT AGMPGLGKG.EPKVPT GGLPGGGPDGLPPLD GGMPGLGKG.EPKLPT GGMPTGTPGSELPT	DVAVKAASLGC GLGVKPMSLGC DPGMKPMSLGC DPSLRPAGSC DPSLKPAGGC DPSLKPAGGC DPSLKPAGGC GPLLKPAAAGG DPALKPAAASC DPALKPAAASC DPSLRPAAAGC	G G G G G	VPSAPLGS LGGMPMGD GGAGGAGAAP GGAGGGAPAP GGAGGGAPAP GGAGGGAPAP GGGGGVPAAP GGGGGVPAGP GGGGGVPAGP GGGGGVPAGP	. A I GGAE SVRPAGAGD I AGL AALAGE SVRPAAAGD I AGA PAGLAGTD SVRPAAAGD I AGA LSPAVTPE TVAPGPP MPVA LGPAVGAD SVSPSPSSTRGGG LOPAVGAE TVAPGPP VPVG LSPAVTPE TVAPGPP VPVG LOPAVGAE TVAPTPVAAASAH LOPAVGAE TVAPTPVAAAA MSAPVAAE TVAPAPTSTTPVS LOPAVGAE TVAPAPTSTPVS
					hemelesuresien
					nomology region
	400	410	420	430	440 450
M_tuberculosis_H37Rv M_marinum M M_kansasii_ATCC_12478 M_rhodesiae_NBB3 M_smegmatis_strMC2_155 M_thermoresistibile_ATCC_19527 M_tusciae_JS617 M_chubuense_NBB4 Mycobacterium_spMCS M_vanbaalenii_PYR-1 M_phlei_RIVM601174 M_gilvum_PYR-GCK	GQGGAAG GQGGAAG AQGKAAAAA AGPAAAAAPAG VGVPGGPGGGAGGAGC AAPVGGPMAGGGMG AGPASATGTPGG GAAPGGAAGAACAPGG AAAAGPAGAMG GPGVSSGSGVAGG APASTQAGTGG APASAGSGGGGAAG	AALGGGCMGME SGMGGGCMGME SGMGGGMGMGM MAG.GCMGGM MMGGMGGMGM MAGGMGGMAPM GGG.MGGMAPM AMGGMGGMAPM AMGGMGGMAPM MMG.GGMAPM	MGAAHQQQQGG. MGGAGQQQGG. .HGAQGQQGS. .GAGHGQCQQGKE GAGHGQCQQKE I.HGAQGQQCKE I.HGAQGAQQGKE IHG.AQSGGSG.E GHGAGRGQQGGE MGGAHGGGAG.E AHGAGAGAQGKE HG.AHGCGAG.E	AKSKGSQQDD AKSKGQQDE SKSKGAQQDD KRRDPRLAPD KRRDPRLAPD KRRDPLTAPD KKRNPQLSQD KKRTPGLAPD KKRNPALSPD KKRNPALSPD KKRDPRLAPD	SALY TEDRAWTEAVIGNRRRQ SALY TEDRAWTEAVIGNRRRQ SDLY TEDRAWTEAVIGNRRRQ DLY EDRAWTESVVGNRRRK DLY EDRAHTEGVIGHRARR SDLY EDRAHTEGIIGNRRRK SDLY EDRPWTEAVIGNRRRR SDLY UEDRPHTEAVIGNRVRR SDLY UEDRPHTEVIGVRPRR SDLY EDRPHTEQVIGVRPRR SDLY EDRPHTEVIGNRRRS SDLY UEDRPHTEVIGVRPR
M_tuberculosis_H37Rv M_marinum_M	460 DSKESK DNK				

M_marinum_M	DNK
M_kansasii_ATCC_12478	DSKESK
M_rhodesiae_NBB3	DVQDGRQGGQDDK
M_smegmatis_strMC2_155	EKDSGKQQ
M thermoresistibile ATCC 19527	ESQEGKESK
M_tusciae_JS617	DVQDVKQSEQEDQ
M chubuense NBB4	RGAPDDMKKES
Mycobacterium_spMCS	.PRDRKDSK
M_vanbaalenii_PYR-1	RGAAEDTKKGDT.
M_phlei_RIVM601174	SAGKSEGTENK
M_gilvum_PYR-GCK	RGTPEDGKRGETS

Fig. S5. Sequence alignment of EspB proteins.

The amino acid sequence alignment of EspB proteins from the ESX-1 clusters of mycobacteria was prepared using ClustalW2 (<u>http://www.ebi.ac.uk/Tools/msa/clustalw2/</u>). The PE and PPE domains, and the secondary structure elements are indicated as predicted by the HHpred server (<u>http://toolkit.tuebingen.mpg.de/hhpred</u>). The MycP₁ cleavage sites are indicated by vertical arrows. The YxxxD/E type VII (ESX) secretion motif in PE domain and the WxG motif in PPE domain are highlighted in blue. EspB proteins have a homology region of unknown function, indicated by orange rectangle.