

RESEARCH ARTICLE

The putative propeptide of MycP1 in mycobacterial type VII secretion system does not inhibit protease activity but improves protein stability

Demeng Sun¹, Qing Liu¹, Yao He¹, Chengliang Wang¹, Fangming Wu², Changlin Tian^{1,2}✉, Jianye Zang¹✉

¹ National Laboratory for Physical Science at the Microscale and School of Life Sciences, University of Science and Technology of China, Hefei 230026, China

² High Magnetic Field Laboratory, Chinese Academy of Sciences, Hefei 230031, China

✉ Correspondence: cltian@ustc.edu.cn (C. Tian), zangjy@ustc.edu.cn (J. Zang)

Received September 18, 2013 Accepted October 21, 2013

ABSTRACT

Mycosin-1 protease (MycP1) is a serine protease anchored to the inner membrane of *Mycobacterium tuberculosis*, and is essential in virulence factor secretion through the ESX-1 type VII secretion system (T7SS). Bacterial physiology studies demonstrated that MycP1 plays a dual role in the regulation of ESX-1 secretion and virulence, primarily through cleavage of its secretion substrate EspB. MycP1 contains a putative N-terminal inhibitory propeptide and a catalytic triad of Asp-His-Ser, classic hallmarks of a subtilase family serine protease. The MycP1 propeptide was previously reported to be initially inactive and activated after prolonged incubation. In this study, we have determined crystal structures of MycP1 with (MycP1^{24–422}) and without (MycP1^{63–422}) the propeptide, and conducted EspB cleavage assays using the two proteins. Very high structural similarity was observed in the two crystal structures. Interestingly, protease assays demonstrated positive EspB cleavage for both proteins, indicating that the putative propeptide does not inhibit protease activity. Molecular dynamic simulations showed higher rigidity in regions guarding the entrance to the catalytic site in MycP1^{24–422} than in MycP1^{63–422}, suggesting that the putative propeptide might contribute to the conformational stability of the active site cleft and surrounding regions.

KEYWORDS type VII ESX-1 secretion system, serine protease, propeptide, crystal structure, EspB cleavage, molecular dynamic simulations

INTRODUCTION

Mycobacterium tuberculosis is one of the most infectious pathogenic bacteria, and has caused huge numbers of fatalities in humans for millennia. Mycobacteria are able to evade host immune responses by delivering virulence molecules into host cells, and they have developed a specialized secretion system, the type VII secretion system (T7SS), to excrete virulence factors across their highly impermeable cell envelope (Stanley et al., 2003; Abdallah et al., 2007; DiGiuseppe Champion and Cox, 2007). ESX-1 secretion system is the intensely studied T7SS in *M. tuberculosis*. ESX-1 secretion is required for early replication and full virulence in macrophages, and has multiple effects on host cells (Brodin et al., 2004; McLaughlin et al., 2007; Raghavan et al., 2008). Studies in *M. tuberculosis* and related mycobacteria have identified components necessary for ESX-1 export activity, including predicted ATPases (EccA1 and EccCb1) and several membrane proteins (MycP1, EccB1, EccCa1, EccD1, and EccE1) (Stanley et al., 2003; Ligon et al., 2012). In addition to conventional substrates (EAST-6 and CFP-10), several other substrates of ESX-1 export system were also identified: EspA, EspB, EspC, and EspR (Fortune et al., 2005; McLaughlin et al., 2007; Xu et al., 2007). It was also known that secretion of all substrates was mutually dependent with components of the T7SS (Fortune et al., 2005).

MycP1 is an essential component of the ESX-1 T7SS, and gene deletion abolished secretion in the complex lacking MycP1 (Ohol et al., 2010). At the same time, introduction of catalytically inactive MycP1 constructs (protease activity abolished by mutation of the catalytic residues) into mycobacteria enhanced ESX1-mediated secretion (Ohol et al., 2010). This

counter-intuitive finding led to the proposal of a dual role for MycP1 in the regulation of ESX-1 secretion (Ohol et al., 2010).

Primary sequence analysis identified MycP1 as member of the subtilase class of serine proteases, with an additional N-terminal Sec signal sequence and a C-terminal transmembrane domain that anchors the protein to the inner membrane, leaving the catalytic peptidase domain in the extra-cytoplasmic space (Brown et al., 2000; Dave et al., 2002). Classical subtilase family proteases have a propeptide and an Asp-His-Ser catalytic triad. The propeptide may facilitate the folding of the protease domain by acting as an inter-molecular chaperone (Zhu et al., 1989). After self-cleavage during the maturation process, the severed propeptide can act as a potent inhibitor (Zhu et al., 1989; Shinde and Inouye, 1993), and removal or degradation of the propeptide may be required to generate the active protease. Several crystal structures of subtilisins have been obtained for inactive auto-processed (propeptide-subtilisin complex) (Gallagher et al., 1995; Jain et al., 1998; Tanaka et al., 2007), matured (Bott et al., 1988; Tanaka et al., 2007), and unautoprocessed forms (Tanaka et al., 2007). These structures showed that the propeptide was separated from the protease domain but remained locating in the substrate entrance to the active site (Gallagher et al., 1995; Jain et al., 1998). Interactions between residues in the propeptide and protease domains therefore play an essential role in manipulating the catalytic activity of subtilases.

Previous studies indicated that the secreted protein EspB, a substrate of MycP1, mediates a feedback mechanism in the maintenance of virulence in mycobacteria (Ohol et al., 2010). EspB was cleaved at two different sites by activated MycP1 protease (Ohol et al., 2010; Solomonson et al., 2013). The EspB assay can be used to study the maturation process of the MycP1 protein, including propeptide cleavage and its role in regulating protease activity.

To answer the above questions, we have determined the crystal structures of MycP1 with (MycP1^{24–422}) and without (MycP1^{63–422}) the putative propeptide. The structures revealed that the putative propeptide wraps around the protease domain in a way that is very different from classical subtilases. The overall similarity of the protease domains in the two structures was high, indicating minimal conformational interference from the propeptide. EspB cleavage assays verified that there was no direct inhibition of proteolytic activity by the propeptide. Molecular dynamic simulations suggested a possible role of the putative propeptide in maintaining the conformational stability of regions guarding entrance to the catalytic site of MycP1.

RESULTS AND DISCUSSION

Crystal structure of MycP1^{24–422}

MycP1 from *M. smegmatis* shares about 80% sequence identity to its homolog in *M. tuberculosis* (Fig. S1), making it a suitable model for studying mycosins of pathogenic mycobacteria. Previous studies have indicated that MycP1 from *M. smegmatis* contained a Sec secretion signal sequence (1–23) and a

transmembrane helix (423–446) (Ohol et al., 2010) (Fig. S2A). In this study a construct encoding MycP1^{24–422} was engineered for structural and functional studies. MycP1^{24–422} was over-expressed in *E. coli*, purified, crystallized and the crystal structure was determined.

The crystal structure of MycP1^{24–422} was refined to a resolution of 2.15 Å with an *R* and *R*_{free} of 18.82% and 21.41%, respectively. Crystallographic statistics are summarized in Table 1. Each asymmetric unit contained a single MycP1^{24–422} molecule. The C-terminal residues 409–422 and the hexahistidine tag could not be modeled due to weak electron density.

Sequence alignment showed around 30% sequence identity between MycP1^{24–422} and subtilisin-like protease family members, even though some regions of MycP1^{24–422} were not found in other subtilases (Fig. 1). The overall structure of MycP1^{24–422} includes a central parallel β-sheet sandwiched between α-helices, plus several extended loops (Figs. 2A and S3). A total of four groups of β-sheets (a highly twisted six-stranded parallel β-sheet, two two-stranded and one three-stranded anti-parallel β-sheets) and ten α-helices are visible. A β-α-β unit consisting of a central six-stranded parallel β-sheet (β3, β2, β5, β7, β9, and β10) with five α-helices (α3, α4, α5, α8, and α9) packed on both sides of the central sheet formed the core of the protease domain. Recently, the crystal structure of Mycosin-1 (MycP1^{24–407}) was reported by Solomonson, M. et al. (2013). Structure alignment of MycP1^{24–422} protein in our study with the MycP1^{24–407} gave a root mean square deviation (RMSD) value of 0.193 Å over 322 Cα atoms (Fig. S2B). The overall structure of MycP1^{24–422} is similar to MycP1^{24–407}. Structure of the MycP1^{24–422} core domain was also aligned with other subtilases. Structure alignment showed high similarity with the *Bacillus* AK.1 serine protease (Smith et al., 1999), subtilisin BL (Goddette et al., 1992), and subtilisin Carlsberg (Cianci et al., 2010), with RMSD values of 0.884, 0.940, and 0.829 Å over 173, 187, and 168 Cα atoms, respectively (Fig. 2B). However, it was found the predicted MycP1 catalytic cleft was unusually large in comparison to typical subtilisin-like enzymes of known structure (Solomonson et al., 2013). This is in part due to a 22-residue insertion after the β3 substrate-binding strand (Arg154–Thr175). Further, the β12/β13 strand is extended (Pro322–Pro327) to form a second mycosin-specific insertion that sits over the cleft to form a “lid” over the catalytic site (Fig. 3A).

Sequence alignment of MycP1 with other subtilases identified conserved residues involved in catalysis, including the catalytic triad Asp-His-Ser (Fig. 1). It was previously predicted that the residues involved in hydrolysis are Asp90, His121, and Ser332 in *M. tuberculosis* MycP1 (Ohol et al., 2010), corresponding to Asp92, His123, and Ser334 in the *M. smegmatis* enzyme (Figs. S1 and S2). In the MycP1 structure, the catalytic triad residues are located at the edge of the central β-sheet, with Asp92 being the last residue of strand β2 and His123 and Ser334 positioned at the beginning of α3 and α8, respectively (Fig. 2A). The spatial arrangement of the catalytic triad in MycP1 aligned well with those in subtilisin Carlsberg, subtilisin

Table 1. Data collection, phasing, and refinement statistics

	Se-MycP1 ^{24–422}	Native MycP1 ^{24–422}	MycP1 ^{63–422}
Data collection			
Wavelength (Å)	0.9791	0.9791	0.9791
Space group	<i>P</i> ₄ ₁ ₂ ₁	<i>P</i> ₄ ₁ ₂ ₁	<i>P</i> ₂ ₁ ₂ ₁
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	<i>a</i> = <i>b</i> = 149.44, <i>c</i> = 230.64	<i>a</i> = <i>b</i> = 85.11, <i>c</i> = 193.14	<i>a</i> = 83.31, <i>b</i> = 83.61, <i>c</i> = 98.25
α , β , γ (°)	$\alpha = \beta = \gamma = 90$	$\alpha = \beta = \gamma = 90$	$\alpha = \beta = \gamma = 90$
Resolution (Å)	50.00–3.20 (3.26–3.20) ^a	30.00–2.15 (2.19–2.15)	50.00–2.25 (2.29–2.25)
No. of unique reflections	43770	37215	28885
R_{merge}^b (%)	11.6 (77.3)	7.1 (50.7)	10.2 (44.4)
$\ I\sigma\ $	28.5 (3.8)	25.9 (4.1)	12.9 (2.8)
Completeness (%)	99.6 (99.7)	99.9 (99.8)	99.5 (99.9)
Multiplicity	9.2 (9.4)	7.6 (7.7)	4.0 (4.1)
Wilson B-factor (Å ²)	80.1	34.6	30.2
Refinement			
R_{work}^c (%)		18.82	16.60
R_{free}^d (%)		21.41	22.18
No. of atoms			
Protein		2813	5032
Water		177	291
GOL		60	
Average B-factors (Å ²)			
Protein		43.31	33.39
Water		41.15	37.97
GOL		56.36	
R.M.S.Deviations			
Bond lengths (Å)		0.0084	0.0100
Bond angles (°)		1.2240	1.3615
Protein Geometry ^e			
Poor rotamers (%)		1.01	0.80
Ramachandran favored (%)		98.18	97.98
Ramachandran outliers (%)		0	0

^a The values in parentheses refer to statistics in the highest shell.

^b $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$ where $I_i(hkl)$ is the intensity of the *i*th measurement, and $\langle I(hkl) \rangle$ is the mean intensity for that reflection.

^c $R_{\text{work}} = \sum_{hkl} ||F_o| - |F_c|| / \sum_{hkl} |F_o|$, where F_o and F_c are the observed and calculated structure factor amplitudes, respectively.

^d R_{free} was calculated with 5.0% of the reflections in the test set.

^e Statistics for the Protein Geometry from an analysis using MolProbity.

BL, and *Bacillus* AK.1 (Fig. 2C). The combined sequence and structural homology strongly indicated that MycP1^{24–422} is a member of the subtilisin serine protease family.

The putative propeptide wrapped around the protease domain of MycP1

With MycP1 predicted to be a member of the subtilase family, the N-terminal segment (residues 21–63) of the *M. tuberculosis* enzyme was assumed to be a putative propeptide (Ohol et al.,

2010) that would inhibit the activity of protease, but activate the protease function after its self-cleavage and degradation. The putative propeptide of the *M. smegmatis* protein (Ile24–Arg65; Fig. S2A) was observed to adopt a disordered structure in the crystal structure of MycP1^{24–422} (Fig. 3A). In the middle of the disordered region, a short β -strand (β_1 ; residues T45, E46, and Q47) was observed to form an antiparallel β -sheet with β_4 of the core protease domain (Fig. 3A). In addition, a disulfide bond connecting the putative propeptide and core protease domain was found between Cys51 and Cys120 (Fig. 3A).



Figure 1. Sequence alignment of *M. smegmatis* MycP1. The protein sequence was aligned with three subtilisin-like protease members (AK.1 serine protease, subtilisin BL, and subtilisin Carlsberg). The conserved catalytic triad residues (Asp, His, and Ser) are marked with blue stars. The sequence alignment was generated using ClustalW2 (Larkin et al., 2007) and decorated using ESPript (Gouet et al., 1999).

The structure of the putative propeptide in MycP1 is significantly different from propeptide structures of typical subtilases

such as Tk-subtilisin, in which the propeptide forms a discrete globular domain comprising a four-stranded antiparallel

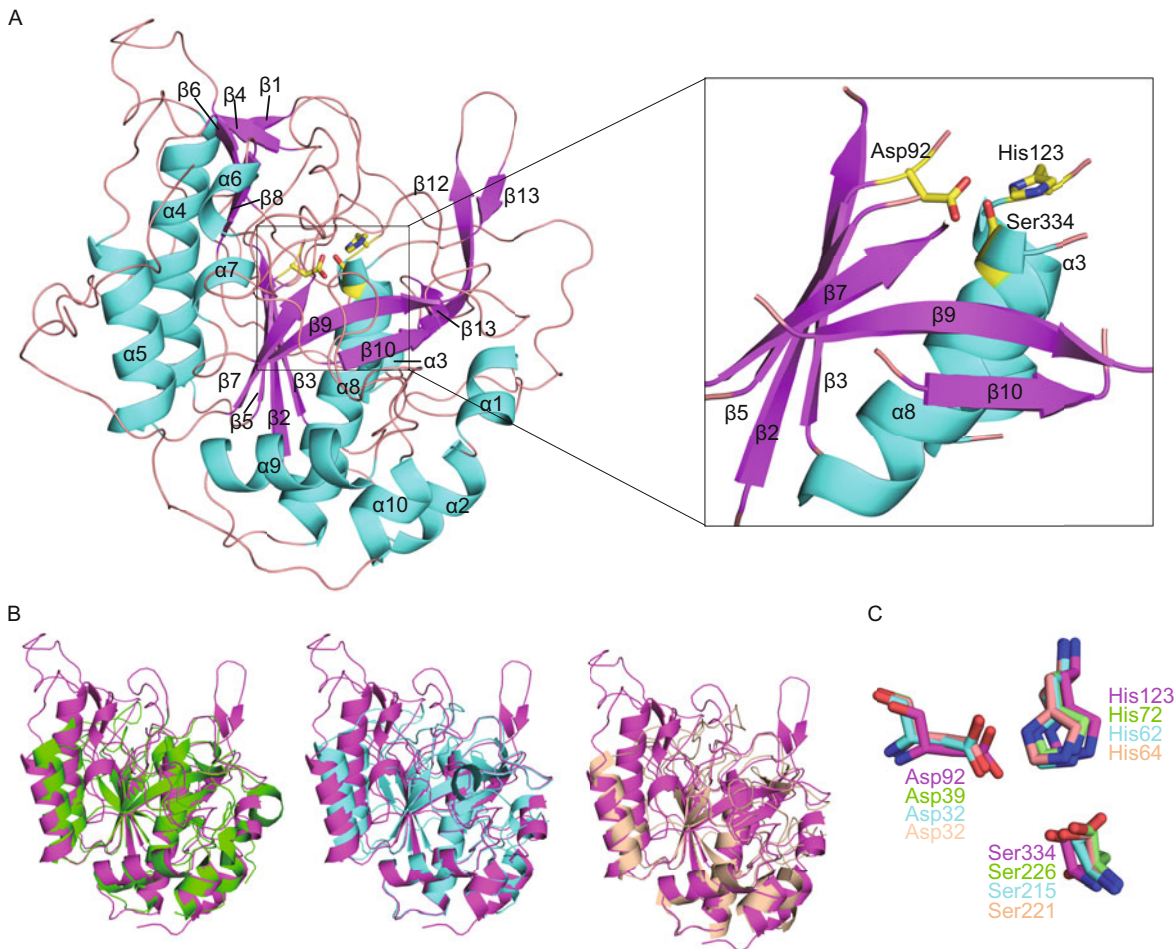


Figure 2. Crystal structure of MycP1^{24–422} and comparison with subtilisin-like serine proteases. (A) Overall ribbon representation of MycP1^{24–422}. Catalytic triad in the active site is shown in detail. (B) Structure superimposition of MycP1^{24–422} (magenta) with AK.1 serine protease (PDB, 1DBI; green), subtilisin BL (PDB, 1ST3; cyan), and subtilisin Carlsberg (PDB, 2WUW; wheat). (C) Structure superimposition of the conserved catalytic triads of MycP1^{24–422} and the three subtilisin-like serine proteases.

β -sheet and two α -helices (Fig. 3B) (Tanaka et al., 2007). In the Tk-subtilisin structure, the propeptide domain blocked the entrance of the protease domain active site (Fig. 3B). During the maturation process of classical subtilases, cleavage of the propeptide not only exposes the serine protease active site, but also abolishes interactions between the propeptide and the core protease domain, leading to conformational changes of the protease domains (Fu et al., 2000; Tanaka et al., 2007). Therefore, the propeptide was considered as a potential inhibitor, and subtilases are only active after cleavage and degradation of the propeptide. However, in this study, the putative propeptide region (Ile24–Arg65) of MycP1 was not observed to form a standalone globular structure, but instead wraps around the core protease domain in a largely unstructured mode, with specific antiparallel β -sheet and disulfide bridge interactions (Fig. 3A).

In conventional subtilases, the propeptide was also reported to act as an intra-molecular chaperone, aiding the correct fold-

ing of the core protease domain (Zhu et al., 1989; Shinde et al., 1993; Shinde and Inouye, 1995; Fu et al., 2000). However, in this study, cleavage of the propeptide from MycP1^{24–422} did not influence the structure of the core domain. Crystal structure of MycP1^{63–422}, the MycP1 protein with its propeptide truncated, was determined to a resolution of 2.25 Å with an R and R_{free} of 16.60% and 22.18%, respectively. Superposition of the MycP1^{24–422} and MycP1^{63–422} structures (RMSD of 0.281 Å for 268 C α atoms) revealed almost identical core protease domains (Fig. 4A). This finding does not support an intra-molecular chaperone function for the propeptide of MycP1, unlike that of Tk-subtilisin (Tanaka et al., 2007). Both the unstructured and intimately wrapped features of the MycP1 propeptide were in stark contrast to propeptide structures of classical subtilases, and suggest that the N-terminal segment (Ile24–Arg65) of MycP1 does not function as a conventional propeptide in the regulation and maturation of MycP1. To verify this hypothesis, functional assays were conducted to analyze the potential in-

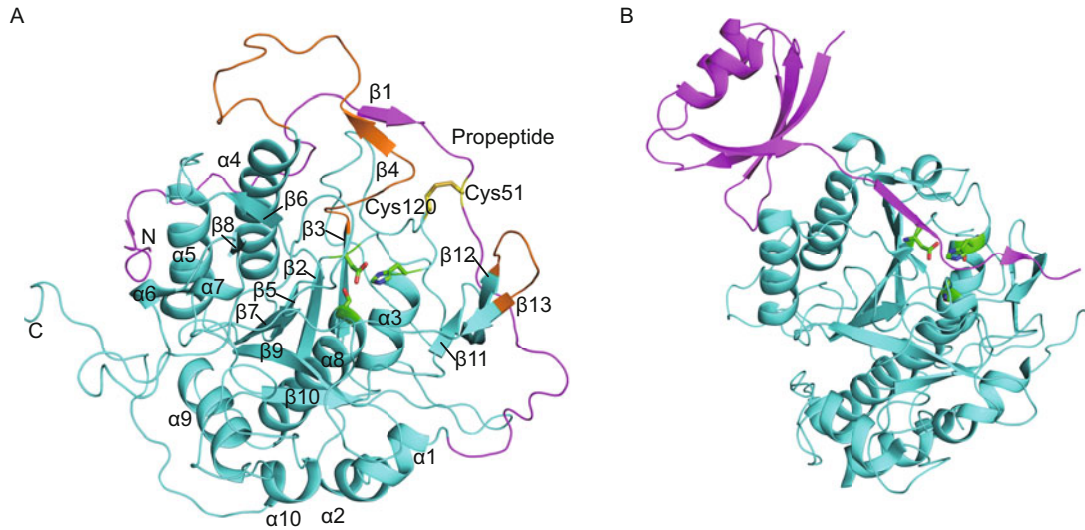


Figure 3. Structural comparison of MycP1²⁴⁻⁴²² and Tk-subtilisin. (A) Structure of MycP1²⁴⁻⁴²². The N-terminal propeptide is colored magenta, and the disulfide bond Cys51–Cys120 is yellow. The catalytic triad is shown in stick conformation and the mycosin-specific insertion forming the lid over the catalytic triad is shown in orange. (B) The structure of Tk-subtilisin (PDB: 2Z2Z). The N-terminal propeptide is colored magenta.

hibitory role of the putative propeptide.

EspB cleavage assay

In mycobacteria, EspB is an important substrate of the ESX-1 T7SS. Previous studies reported that the catalytically inactive MycP1 mutant did not cleave EspB, while active MycP1 processed EspB efficiently (Ohol et al., 2010). Two cleavage sites for MycP1 in EspB were identified, although it was not clear whether the active MycP1 contained the putative propeptide or not (Ohol et al., 2010). To verify the functional role of the putative propeptide of MycP1, the EspB cleavage assay was conducted using MycP1²⁴⁻⁴²² and MycP1⁶³⁻⁴²².

Following the previously established EspB cleavage protocol (Ohol et al., 2010; Solomonson et al., 2013), protease activity of MycP1²⁴⁻⁴²² and MycP1⁶³⁻⁴²² against purified EspB was conducted *in vitro*. EspB was mixed with MycP1²⁴⁻⁴²² or MycP1⁶³⁻⁴²² at a molar ratio of 5:1 (Fig. 4B), and the proteolytic activity was analyzed using SDS-PAGE. The catalytically inactive MycP1-S334A mutant was used as negative control.

After overnight incubation with MycP1 at 37°C, EspB migrated at a lower molecular weight (~45 kDa) compared with the intact EspB (~60 kDa; Fig. 4B). An additional lower molecular weight band at ~14 kDa corresponding to a truncated EspB fragment was observed for EspB cleavage using both MycP1²⁴⁻⁴²² and MycP1⁶³⁻⁴²² (Fig. 4B). The truncated bands were not observed when EspB was treated with the proteolytically inactive mutants MycP1²⁴⁻⁴²²-S334A and MycP1⁶³⁻⁴²²-S334A (Fig. 4B). The observed protease activity of both MycP1²⁴⁻⁴²² and MycP1⁶³⁻⁴²² strongly indicated that the putative propeptide did not inhibit the protease activity. In a prolonged assay under the same conditions, EspB cleavage by MycP1²⁴⁻⁴²² was found to be more efficient than MycP1⁶³⁻⁴²² (Fig. 4C).

Notably, MycP1²⁴⁻⁴²² cleaved more than 90% of EspB after a 24 h incubation, while MycP1⁶³⁻⁴²² cleaved about 50% (Fig. 4C). This observation strongly indicated the non-inhibitory effect of the putative propeptide to MycP1 protease, but mechanism of higher cleavage efficiency of MycP1 with the presence of the propeptide region is interesting and should be explored in the future using other methods.

Flexibility analysis of MycP1²⁴⁻⁴²² or MycP1⁶³⁻⁴²²

Following observation of similar structures for MycP1 with or without the propeptide, and establishing the non-inhibitory nature, the possible role of the propeptide was investigated using 15 ns molecular dynamics simulations with the crystal structures of MycP1²⁴⁻⁴²² and MycP1⁶³⁻⁴²² determined in this study. The RMSD for all backbone C α atoms as a function of simulation time was calculated (Fig. S4A). The RMSD values remained at 1.2 Å for MycP1²⁴⁻⁴²² and 1.7 Å for MycP1⁶³⁻⁴²² after 320 ps, demonstrating the stable conformations of these structures within the 15 ns simulations. Structure calculations were analyzed using the B-factors from the MD trajectory, and a strong correlation was observed with the experimental B-factors from the crystallographic data for MycP1²⁴⁻⁴²² (Fig. S4B) and MycP1⁶³⁻⁴²² (Fig. S4C), indicating a robust MD simulation for both structures.

The structural flexibility of MycP1²⁴⁻⁴²² and MycP1⁶³⁻⁴²² was assessed using the root mean square fluctuation (RMSF) of C α atoms after the entire MD calculations. The RMSF for residues in five regions (R1, residues 155–180; R2, 205–220; R3, 238–265; R4, 315–330; R5, 368–377) in the MycP1⁶³⁻⁴²² MD simulation was observed to be generally higher than those of the corresponding MycP1²⁴⁻⁴²² regions (Fig. 5A). These five regions were mapped to the MycP1 structure. The R1–R4 regions

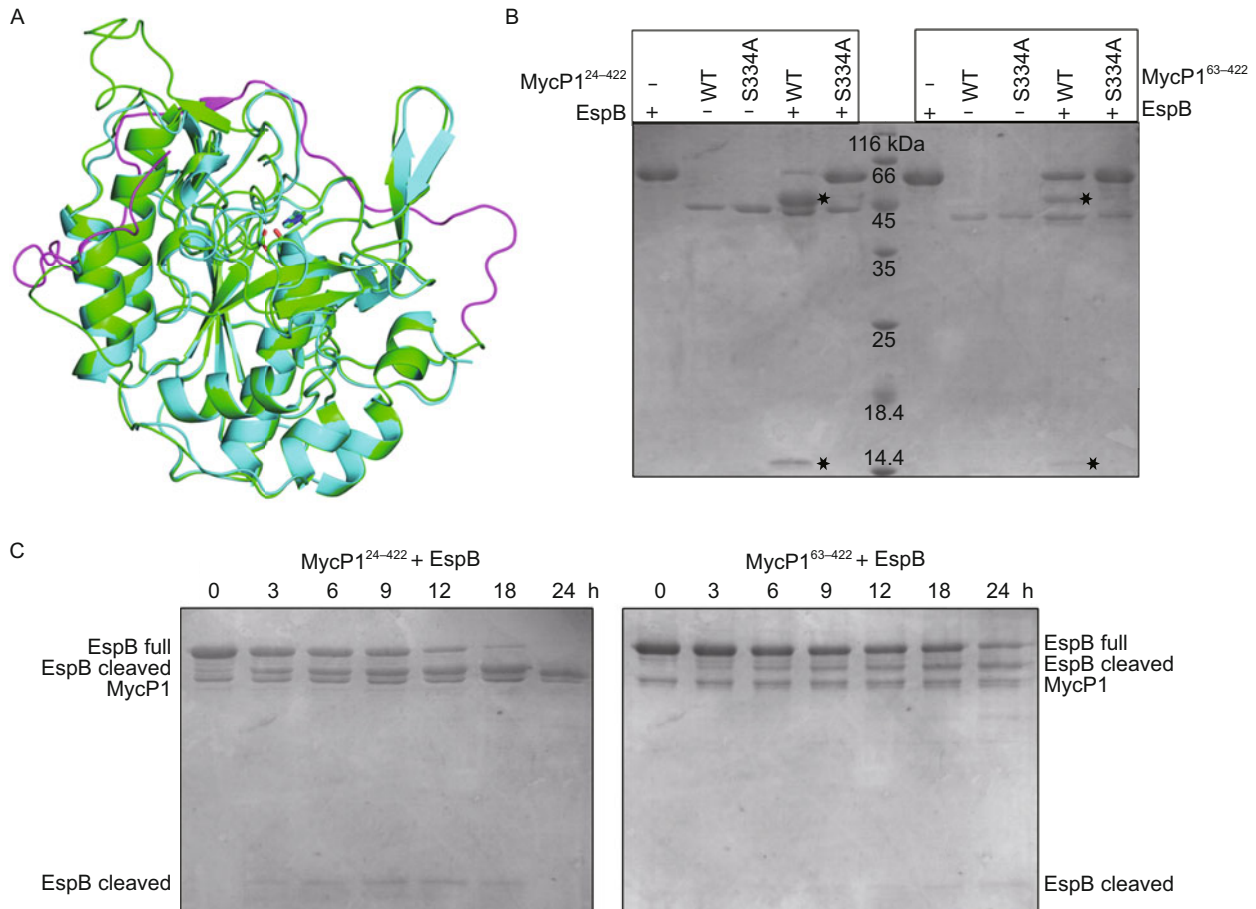


Figure 4. Structure and protease activity comparison of MycP1²⁴⁻⁴²² and MycP1⁶³⁻⁴²². (A) Structure superposition of MycP1²⁴⁻⁴²² (green) and MycP1⁶³⁻⁴²² (cyan). The putative propeptide of MycP1²⁴⁻⁴²² is colored in magenta. (B) SDS-PAGE of the *in vitro* protease activity assay of MycP1²⁴⁻⁴²², MycP1⁶³⁻⁴²², MycP1²⁴⁻⁴²²-S334A, and MycP1⁶³⁻⁴²²-S334A using EspB as substrate. Cleavage products are indicated by asterisks. Molecular weight markers are shown. (C) SDS-PAGE gel showing MycP1²⁴⁻⁴²² and MycP1⁶³⁻⁴²² cleavage of EspB at the indicated time points.

were found to be in close proximity to the putative propeptide region and also surround the entrance to the catalytic site of the protease, while the R5 region was observed to be close to R3 (Fig. 5B). The more rigid conformation of MycP1²⁴⁻⁴²² may be due to the stabilizing interactions between the putative propeptide and residues in the R1–R4 regions of the protease domain. The higher stability of MycP1²⁴⁻⁴²² was consistent with the protease assay, in which higher protease efficiency was also observed for MycP1²⁴⁻⁴²² (Fig. 4C). In combination, the EspB cleavage efficiency data and molecular dynamics flexibility analysis suggest the putative propeptide stabilizes the conformation of the MycP1 protease domain.

Classical subtilases have been shown to use calcium ions and disulfide bonds as two strategies for stabilizing their catalytic domains (Siezen and Leunissen, 1997). Since no Ca²⁺ ions or appropriate chelating amino acid residues were present in the MycP1 structure, calcium ions were not utilized by MycP1 (Solomonson et al., 2013). Alternatively, two disulfide bonds (Cys51–Cys120 and Cys206–Cys244) were observed

in the MycP1²⁴⁻⁴²² structure, which may be the only possible stabilization strategy used by MycP1. Disruption of the first disulfide bond through truncation of the putative propeptide region could significantly reduce the overall stability of the protein. Together with the antiparallel β -sheet comprised of β 1 (Thr45–Gln47) and β 4, the disulfide bridges and the rest of the putative propeptide appear to contribute to the stability of the protease domain.

In a recent report, similar results were observed on the structure and function of MycP1²⁴⁻⁴⁰⁷ (Solomonson et al., 2013). However, Solomonson, M. et al. only described the structure of MycP1²⁴⁻⁴⁰⁷. They have tried to over-express the truncated construct without the putative propeptide, but failed in not only structural studies, but also related functional EspB cleavage studies. Structural study of MycP1⁶³⁻⁴²² reported here strongly indicated the majority of the putative propeptide is not required for folding. The reduced protease activity of MycP1⁶³⁻⁴²² suggested that the N-terminal sequence does not serve as a typical inhibitory propeptide.

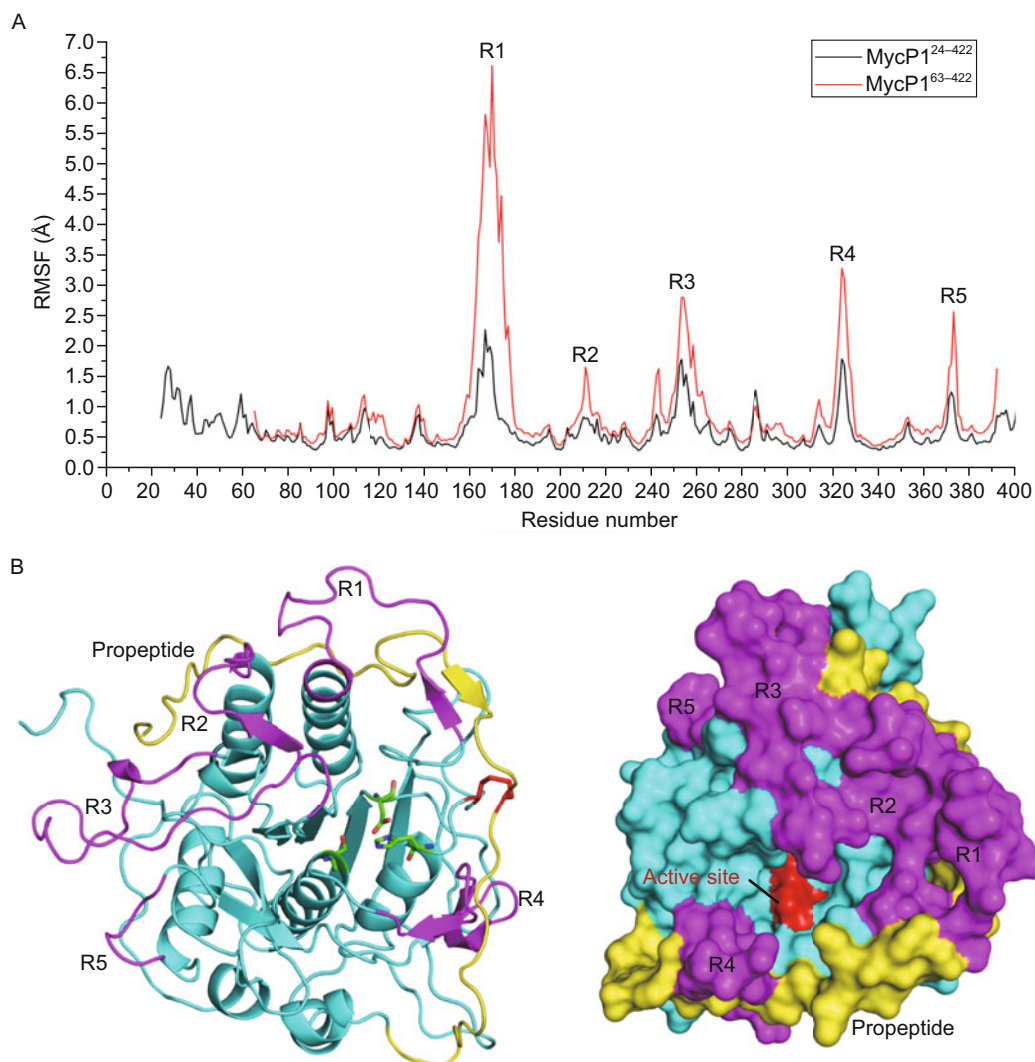


Figure 5. Molecular dynamic simulations showing the differences in flexibility between MycP1²⁴⁻⁴²² and MycP1⁶³⁻⁴²². (A) RMSF values of Ca atoms in MycP1²⁴⁻⁴²² and MycP1⁶³⁻⁴²². The regions with high RMSF deviations between the two proteins are indicated. (B) Ribbon (left panel) and surface (right panel) representations of the loop regions with high flexibility (magenta) in MycP1, showing their relative locations to the core domain (cyan) and the putative propeptide (yellow).

In summary, structural and biochemical studies demonstrated that the putative propeptide of MycP1 neither serves as an intramolecular chaperone, nor as an inhibitor of proteolytic activity. Rather, truncation of the propeptide reduced protease efficiency, and impaired the stability of the protease domain. Together, the results in this study indicate that the N-terminal sequence of MycP1 (Ile24–Arg65) does not serve as a typical inhibitory propeptide, but instead enhance the stability of the protease domain.

MATERIALS AND METHODS

Cloning, expression, and purification of MycP1

The DNA sequence encoding *M. smegmatis* MycP1²⁴⁻⁴²² (including the propeptide and the core protease domain, Fig. S2A) was synthesized with codon optimization for *E. coli* (Sangon, Shanghai) and

inserted into the pET21b vector (Novagen, CA) to include a C-terminal hexahistidine tag. The DNA sequence encoding MycP1⁶³⁻⁴²² was amplified using PCR and inserted into the pET-22b vector (Novagen, CA) to include an N-terminal hexahistidine tag. The constructs were transformed into *E. coli* BL21 (DE3) (Stratagene) for protein expression. Cells were grown in M9 medium at 37°C until OD₆₀₀ reached 0.8, then protein production was induced using 0.5 mmol/L IPTG (12 h at 25°C for MycP1²⁴⁻⁴²²; 48 h at 12°C for MycP1⁶³⁻⁴²²). MycP1 proteins were purified by affinity chromatography using a Ni-NTA column (QiaGen) followed by size-exclusion chromatography using a HiLoad 16/60 Superdex 200 column (GE Healthcare). The purified proteins were analyzed using SDS-PAGE.

Crystallization, data collection, and structure determination

MycP1²⁴⁻⁴²² was purified in buffer containing 20 mmol/L Tris-HCl (pH 8.0), 200 mmol/L NaCl, and 2 mmol/L DTT, and initially crystallized

using the sitting-drop vapor-diffusion method at 10°C. Microcrystals appeared in a number of conditions in 1 week. After optimization, the best MycP1^{24–422} crystals used for data collection were grown in 1.6 mol/L MgSO₄, 0.1 mol/L MES (pH 5.5), with a protein concentration of 20 mg/mL. Crystals of the SeMet-derivative of MycP1^{24–422} were also screened and obtained in a crystallization condition containing 0.1 mol/L Mg(Ac)₂, 0.1 mol/L sodium cacodylate (pH 6.5), and 15% (w/v) PEG 6000.

MycP1^{63–422} was purified in buffer containing 50 mmol/L Tris-HCl (pH 8.0), 500 mmol/L NaCl, 5% (v/v) glycerol, and 2 mmol/L DTT, and initially crystallized using the sitting-drop vapor-diffusion method at 10°C. Microcrystals appeared in 1 month. The best MycP1^{63–422} crystals used for data collection were grown in 0.1 mol/L sodium citrate (pH 5.6), 20% (v/v) 2-propanol, and 20% (w/v) PEG 4000, with a protein concentration of 5 mg/mL.

Crystals were harvested and soaked in a cryo-protectant solution consisting of the crystallization condition supplemented with 20% (v/v) glycerol for several seconds, and then flash-cooled in liquid nitrogen. X-ray diffraction data were collected on beamline BL17U at the Shanghai Synchrotron Radiation Facility (SSRF). Data were processed using HKL-2000 (Otwinowski and Minor, 1997).

The structure of MycP1^{24–422} was determined using the single-wavelength anomalous dispersion (SAD) method (Hendrickson, 1991). The SeMet sites were located and the initial phases were calculated using the Phenix software package (Adams et al., 2002). Residues in one asymmetric unit were automatically built using the Autobuild program. After tracing the initial model manually using the COOT program (Emsley and Cowtan, 2004), the model was refined against the native data using REFMAC5 (Vagin et al., 2004), then rebuilt interactively using the σ_A -weighted electron density maps with coefficients $2mF_o - DF_c$ and $mF_o - DF_c$ using COOT (Emsley and Cowtan, 2004).

Using the determined MycP1^{24–422} structure as a search model, the structure of MycP1^{63–422} was solved by molecular replacement using the Phaser program (McCoy et al., 2005). The model was refined to 2.25 Å using alternate cycles of REFMAC5 (Vagin et al., 2004) and COOT (Emsley and Cowtan, 2004). The quality of the resultant models was analyzed using MolProbity (Chen et al., 2010). The crystal structure coordinates of MycP1^{24–422} and MycP1^{63–422} have been deposited in the PDB with accession codes 4KB5 and 4M1Z, respectively.

Expression, purification, and cleavage assay of recombinant EspB

The DNA sequence encoding full-length EspB was amplified using PCR from a gene library of *M. tuberculosis* and inserted into the pET-22b vector (Novagen) to include an N-terminally fused hexahistidine-MBP tag. The construct was transformed into *E. coli* BL21 (DE3) (Stratagene) for protein expression. Cells were grown in LB medium at 37°C until the OD₆₀₀ reached 0.8, then protein production was induced using 0.5 mmol/L IPTG at 37°C for 4 h. The hexahistidine-MBP-tagged protein was purified by affinity chromatography using a Ni-NTA column (Qiagen). The hexahistidine-MBP tag was cleaved by incubation with PreScission protease, and the tag-free protein was further purified by size-exclusion chromatography using a HiLoad 16/60 Superdex 200 GL (GE Healthcare). The purified EspB protein was analyzed using SDS-PAGE.

The EspB cleavage assay was performed in reaction buffer containing 20 mmol/L MES pH 6.0 and 100 mmol/L NaCl. EspB (10 μmol/L)

was digested with wild-type MycP1^{24–422}, MycP1^{63–422} or proteolytically inactive MycP1 at an enzyme/substrate molar ratio of 1:5 at 37°C. Samples were taken at the indicated time points, and the reactions were stopped by the addition of SDS-PAGE loading buffer and subsequent boiling. Degraded EspB products were analyzed using SDS-PAGE and visualized using Coomassie Blue staining.

Molecular dynamics simulations

The crystal structures of MycP1^{24–422} and MycP1^{63–422} were used to prepare molecular dynamic simulations. The two proteins were simulated under the same conditions using GROMACS-4.5.4 (Van Der Spoel et al., 2005) and the amber99SB-ILDN force field. The model was initially energetically minimized to remove geometric strain. The energy-minimized model was hydrated using the TIP3P water model and a dodecahedron box. Energy minimization was performed using a steepest descent approach, with the convergence criterion set as 1 kJ/mol-nm and a step size of 0.01 nm. Position restrained MD simulation was then carried out, with the protein position fixed and solvent equilibrated freely. During this process, temperature was increased to 303 K by a reverse annealing process and the pressure was set to 1 bar. The equilibration was finished in 320 ps. MD simulations (15 ns) without position restraints were implemented, using the MD integrator and a step size of 2 fs. Mass translation of the whole system was removed every 10 steps. The Van der Waals interaction cutoff was set as 1.4 nm. RMSD, RMSF, and B-factors were calculated according to backbone C α atoms. The whole trajectory was fitted to the start frame.

ACKNOWLEDGMENTS

This work was supported by funds from the National Basic Research Program (973 Program) (Nos. 2011CB911104 and 2012CB917202), and the National Natural Science Foundation of China (Grant No. 31100538) to F. W., (Grant No. 31170817) to C. T. We thank staff at the beamline BL17U at the Shanghai Synchrotron Radiation Facility (SSRF) for assistance with data collection, and graduate students in Prof. Haiyan Liu's group for help with molecular dynamic simulations.

D. S. carried out the structural and functional studies. Q. L. performed the molecular dynamic simulations and data analyses. Y. H. participated in protein expression, purification and functional assays. C. W. participated in crystal diffraction data collection. J. Z. and C. T. conceived the study, participated in the experimental design, data analyses and discussion, and wrote the paper, with help from all of the authors.

ABBREVIATIONS

IPTG, isopropylthiogalactoside; MD, molecular dynamics; MycP1, mycosin-1 protease; PAGE, polyacrylamide gel electrophoresis; RMSD, root mean square deviation; RMSF, root mean square fluctuation; SAD, single-wavelength anomalous dispersion; SDS, sodium dodecyl sulfate; SeMet, seleno-methionine; T7SS, type VII secretion system

COMPLIANCE WITH ETHICS GUIDELINES

Demeng Sun, Qing Liu, Yao He, Chengliang Wang, Fangming Wu, Changlin Tian and Jianye Zang declare that they have no conflict of interest.

This article does not contain any studies with human or animal subjects performed by any of the authors.

REFERENCES

- Abdallah, A.M., Gey van Pittius, N.C., Champion, P.A., Cox, J., Luirink, J., Vandenbroucke-Grauls, C.M., Appelmek, B.J., and Bitter, W. (2007). Type VII secretion—mycobacteria show the way. *Nat Rev Microbiol* 5, 883–891.
- Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K., and Terwilliger, T.C. (2002). PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* 58, 1948–1954.
- Bott, R., Ultsch, M., Kossiakoff, A., Graycar, T., Katz, B., and Power, S. (1988). The three-dimensional structure of *Bacillus amyloliquefaciens* subtilisin at 1.8 Å and an analysis of the structural consequences of peroxide inactivation. *J Biol Chem* 263, 7895–7906.
- Brodin, P., Rosenkrands, I., Andersen, P., Cole, S.T., and Brosch, R. (2004). ESAT-6 proteins: protective antigens and virulence factors? *Trends Microbiol* 12, 500–508.
- Brown, G.D., Dave, J.A., Gey van Pittius, N.C., Stevens, L., Ehlers, M.R., and Beyers, A.D. (2000). The mycosins of *Mycobacterium tuberculosis* H37Rv: a family of subtilisin-like serine proteases. *Gene* 254, 147–155.
- Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66, 12–21.
- Cianci, M., Tomaszewski, B., Helliwell, J.R., and Halling, P.J. (2010). Crystallographic analysis of counterion effects on subtilisin enzymatic action in acetonitrile. *J Am Chem Soc* 132, 2293–2300.
- Dave, J.A., Gey van Pittius, N.C., Beyers, A.D., Ehlers, M.R., and Brown, G.D. (2002). Mycosin-1, a subtilisin-like serine protease of *Mycobacterium tuberculosis*, is cell wall-associated and expressed during infection of macrophages. *BMC microbiology* 2, 30.
- DiGiuseppe, P., Champion, P.A., and Cox, J.S. (2007). Protein secretion systems in *Mycobacteria*. *Cell Microbiol* 9, 1376–1384.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60, 2126–2132.
- Fortune, S.M., Jaeger, A., Sarracino, D.A., Chase, M.R., Sassetti, C.M., Sherman, D.R., Bloom, B.R., and Rubin, E.J. (2005). Mutually dependent secretion of proteases required for mycobacterial virulence. *Proc Natl Acad Sci U S A* 102, 10676–10681.
- Fu, X., Inouye, M., and Shinde, U. (2000). Folding pathway mediated by an intramolecular chaperone. The inhibitory and chaperone functions of the subtilisin propeptide are not obligatorily linked. *J Biol Chem* 275, 16871–16878.
- Gallagher, T., Gilliland, G., Wang, L., and Bryan, P. (1995). The prosegment-subtilisin BPN' complex: crystal structure of a specific 'foldase'. *Structure* 3, 907–914.
- Goddette, D.W., Paech, C., Yang, S.S., Mielenz, J.R., Byströf, C., Wilke, M.E., and Fletterick, R.J. (1992). The crystal structure of the *Bacillus lentus* alkaline protease, subtilisin BL, at 1.4 Å resolution. *J Mol Biol* 228, 580–595.
- Gouet, P., Courcelle, E., Stuart, D.I., and Metz, F. (1999). ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* 15, 305–308.
- Hendrickson, W.A. (1991). Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* 254, 51–58.
- Jain, S.C., Shinde, U., Li, Y., Inouye, M., and Berman, H.M. (1998). The crystal structure of an autoprocessed Ser221Cys-subtilisin E-propeptide complex at 2.0 Å resolution. *J Mol Biol* 284, 137–144.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948.
- Ligon, L.S., Hayden, J.D., and Braunstein, M. (2012). The ins and outs of *Mycobacterium tuberculosis* protein export. *Tuberculosis (Edinb)* 92, 121–132.
- McCoy, A.J., Grosse-Kunstleve, R.W., Storoni, L.C., and Read, R.J. (2005). Likelihood-enhanced fast translation functions. *Acta Crystallogr D Biol Crystallogr* 61, 458–464.
- McLaughlin, B., Chon, J.S., MacGurn, J.A., Carlsson, F., Cheng, T.L., Cox, J.S., and Brown, E.J. (2007). A mycobacterium ESX-1-secreted virulence factor with unique requirements for export. *PLoS Pathog* 3, e105.
- Ohl, Y.M., Goetz, D.H., Chan, K., Shiloh, M.U., Craik, C.S., and Cox, J.S. (2010). *Mycobacterium tuberculosis* MycP1 protease plays a dual role in regulation of ESX-1 secretion and virulence. *Cell Host Microbe* 7, 210–220.
- Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol* 276, 307–326.
- Raghavan, S., Manzanillo, P., Chan, K., Dovey, C., and Cox, J.S. (2008). Secreted transcription factor controls *Mycobacterium tuberculosis* virulence. *Nature* 454, 717–721.
- Shinde, U., and Inouye, M. (1993). Intramolecular chaperones and protein folding. *Trends Biochem Sci* 18, 442–446.
- Shinde, U., and Inouye, M. (1995). Folding mediated by an intramolecular chaperone: autoprocessing pathway of the precursor resolved via a substrate assisted catalysis mechanism. *J Mol Biol* 247, 390–395.
- Shinde, U., Li, Y., Chatterjee, S., and Inouye, M. (1993). Folding pathway mediated by an intramolecular chaperone. *Proc Natl Acad Sci U S A* 90, 6924–6928.
- Siezen, R.J., and Leunissen, J.A. (1997). Subtilases: the superfamily of subtilisin-like serine proteases. *Protein Sci* 6, 501–523.
- Smith, C.A., Toogood, H.S., Baker, H.M., Daniel, R.M., and Baker, E.N. (1999). Calcium-mediated thermostability in the subtilisin superfamily: the crystal structure of *Bacillus Ak.1* protease at 1.8 Å resolution. *J Mol Biol* 294, 1027–1040.
- Solomonson, M., Huesgen, P.F., Wasney, G.A., Watanabe, N., Gruninger, R.J., Prehna, G., Overall, C.M., and Strynadka, N.C. (2013). Structure of the Mycosin-1 Protease from the *Mycobacterium* ESX-1 Protein Type VII Secretion System. *J Biol Chem* 288, 17782–17790.
- Stanley, S.A., Raghavan, S., Hwang, W.W., and Cox, J.S. (2003). Acute infection and macrophage subversion by *Mycobacterium tuberculosis* require a specialized secretion system. *Proc Natl Acad Sci U S A* 100, 13001–13006.
- Tanaka, S., Matsumura, H., Koga, Y., Takano, K., and Kanaya, S. (2007). Four new crystal structures of Tk-subtilisin in unautoprocessed, autoprocessed and mature forms: insight into structural changes during maturation. *J Mol Biol* 372, 1055–1069.
- Vagin, A.A., Steiner, R.A., Lebedev, A.A., Potterton, L., McNicholas,

- S., Long, F., and Murshudov, G.N. (2004). REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D Biol Crystallogr* 60, 2184–2195.
- Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E., and Berendsen, H.J. (2005). GROMACS: fast, flexible, and free. *J Comput Chem* 26, 1701–1718.
- Xu, J., Laine, O., Masciocchi, M., Manoranjan, J., Smith, J., Du, S.J., Edwards, N., Zhu, X., Fenselau, C., and Gao, L.Y. (2007). A unique Mycobacterium ESX-1 protein co-secretes with CFP-10/ESAT-6 and is necessary for inhibiting phagosome maturation. *Mol Microbiol* 66, 787–800.
- Zhu, X.L., Ohta, Y., Jordan, F., and Inouye, M. (1989). Pro-sequence of subtilisin can guide the refolding of denatured subtilisin in an intermolecular process. *Nature* 339, 483–484.