Crystal structure of an aspartic proteinase domain of the *Mycobacterium tuberculosis* cell surface antigen PE_PGRS16

Deivanayaga V Barathy and Kaza Suguna

Supplementary information

Table S1. Aspartic proteinases present in various species of *Mycobacterium*.

Organism	Total no. of genes coding AP	Gene locus		
M. tuberculosis H37Rv	3	Rv0977	Rv1983*	Rv2519*
M. tuberculosis H37Ra	3	MRA_0984	MRA_1995	MRA_2546
M. tuberculosis F11	3	TBFG_10995.4	TBFG_12014.4	TBFG_12540.4
M.tuberculosis CDC 1551	2	—	MT2036	MT2595
M. tuberculosis Haarlem (draft)	2	—	TBHG_01939.1	TBHG_02455.1
M. tuberculosis C	1	—	TBCG_01932.1	—
M. bovis AF2122/97	3	Mb1002	Mb2005	Mb2548
M. bovis BCG	2	BCG_1031	—	BCG_2540
M. *africanum*	3	MAF_09860	MAF_19940	MAF_25340
M. marinum	3	MMAR_2272[†]	MMAR_2933[††]	MMAR_1538 [†††]

* 43 % identical to Rv0977, [†] 45 % identical to Rv0977, [††] 69 % identical to Rv1983, [†††] 78 % identical to Rv2519

— Absent

All the proteins of mycobacterium species in the same column are 100 % identical to H37Rv proteins except the marinum species.

Table S2. Predicted subsites of Mtb-AP along with the known subsites of eukaryotic aspartic proteinases.

Subsite	Rhizopuspepsin (3APR)	Human pepsin (1PSO)	Mtb-AP
S_4	-	Met12	-
	Thr222	Ser219	Ile193
	-	Gln287	-
S_3	Ile15	Met12	-
	Glu16	-	Glu18
	-	Thr77	
	-	Phe111	Phe106
	Thr221	Gly217	Gly191
	Thr222	Ser219	Ile193
S_2	Gly78	Gly76	Ala65
	Asp79	Thr77	-
	Thr221	Thr218	Gly192
	Ile225	-	-
	-	Gln287	-
	-	Met289	-
S_1	Asp33	Val30	Leu34
	Asp35	Asp32	Asp36
	Tyr77	Tyr75	Tyr64
	-	Gly76	Ala65
	Asp79	Thr77	-
	Ser81	-	-
	Phe114	-	Phe106
	Leu122	Ile120	Val115
	Asp 218	Asp215	Asp189
	Gly220	Gly217	Gly191
S_1'	Gly37	-	Gly38
	Tyr77	-	-
	Gly78	-	Ala65
	Ile216	-	Met187
	Asp218	-	Asp189
	Trp294	-	-
	Ile298	-	Asn247
S_2'	Gly37	Gly34	Gly38
	Ser38	Ser35	Ser39
	Ile75	-	Ala62
	Ser76	-	Gly63
	Ile130	-	-
	Trp194	Tyr189	-

Figure S1 The His-tag peptide residues (cyan) bound to the active site are stabilized by the interactions with the protein residues (green), water molecules (red sphere) and ethylene glycol (labeled as EDO). The dashed lines show the hydrogen bonding interactions. The Zn^{2+} ion interacts with the two catalytic aspartates and two histidine residues from the tag. Protein residues which form van der Waals interactions are shown with 50% transparency.

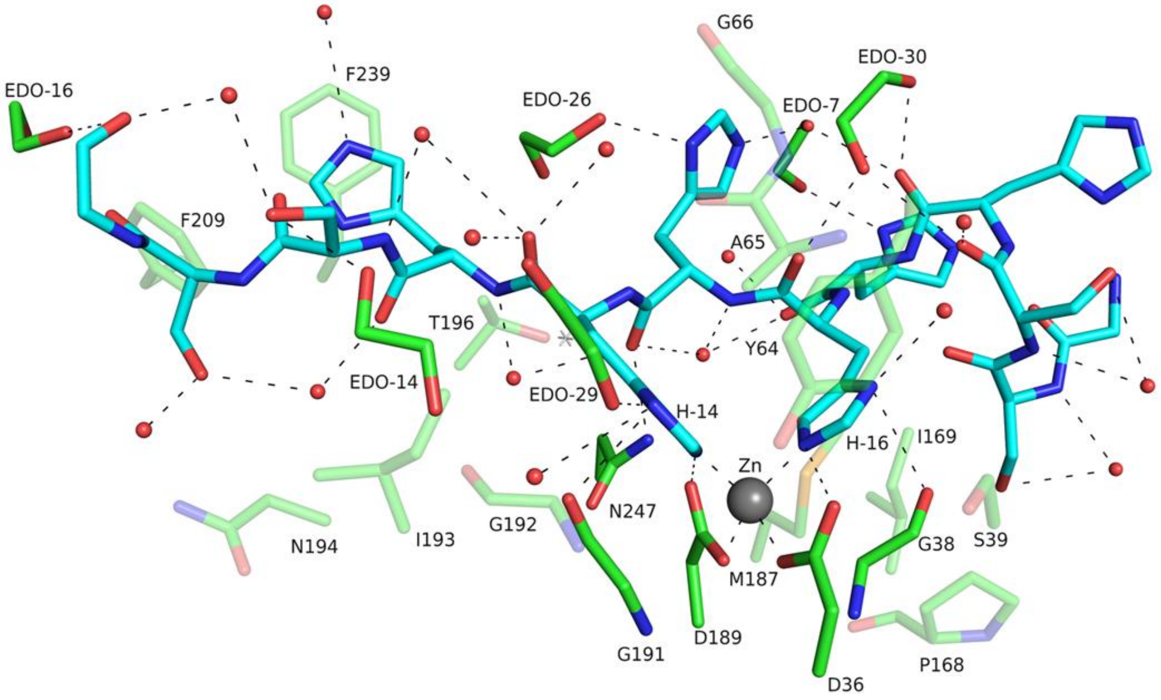

Figure S2 CD spectra of different constructs of Mtb-AP.

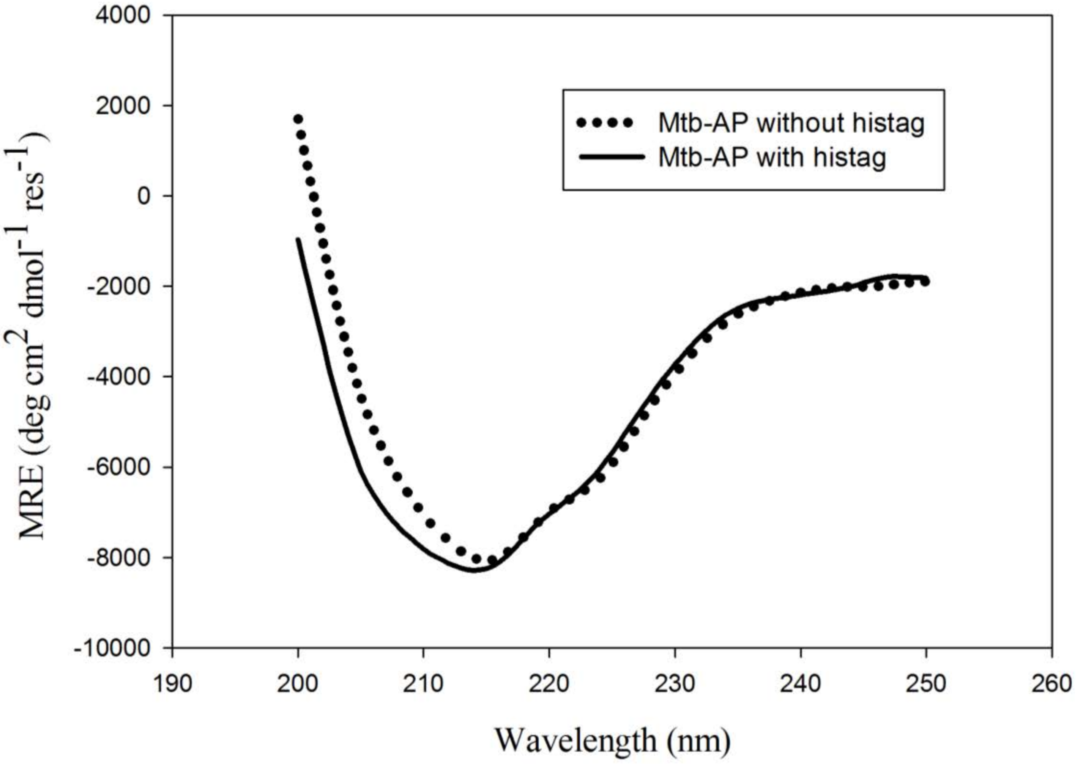

Figure S3 Multiple sequence alignment of aspartic proteinases from Mycobacterium species. Rv denotes *M. tuberculosis* and MMAR denotes *M. marinum*. The DT/SG motif is shown in red. The N-terminal HHG motif and the sequence NTG which replaces the C-terminal HHG motif are shown in blue.

```
Rv0977          STTLTNATVPLQLVNTTEPVVFISLNGGQMVPVLLDTGSTGLVMDSQFLT-----QNFGP 55
Rv1983          ----DGRTVPLEIIHVTEPTVHANVNGGPTSTILVDTGSAGLVVSPEDVGGILGVLHMGL 56
Rv2519          --DPVNVAVPLRVENN-FPLVNLLVNRGPTVPILLDTGSSSLVIPFWKIG----WQNLGL 53
MMAR_2272       ------------------PVVTISVGGGPGIAVTVDTGASGLLVRPQDVN----LQSLGT 38
MMAR_2933       -------TVPLEVVNVTEPVVNVNVNGGHSTPVLIDTGSAGLVMQVKDVGGPLGLLRMGL 53
MMAR_1538       -----NASVPLYLDNN-FPAVNVSINGGPSVPVLLDTGSAGLVVPIWDIG----LQNLGV 50
                                  * *   :. *   .: :***::.*::    :        :*

Rv0977          VIGTGTAGYAGG----LTYNYNTYSTTVDFGNGLLTLPTSVN----VVTSS--SPGTLGN 105
Rv1983          PTGLSISGYSGG----LYYIFATYTTTVDFGNGIVTAPTAVNVVLLSIPTSPFAISTYFS 112
Rv2519          PTGFDVVHYGNG----VSIVYADVPTTVDFGGGAATTPTSVHVGILPYPRNLDSLVLIAS 109
MMAR_2272       PTGSGAVTYGNSSYAFNTVQYQTYQTTVNFGNGIVTNPTNV-----AVATS--ATQTING 91
MMAR_2933       PSGISMSAYSGG----LTYLFATYPTTVDFGNGIVTSTTGVDVVLFSIPTSPYALTTWLN 109
MMAR_1538       PTGFDVIRYGNG----VNILYADFNTTVDFGSGAVTAPTSVQVGILPFPTSLQGLTLIAM 106
                  * .   *...        :    ***:**.*  * .* *       . .  .

Rv0977          FLSRSGAVG-------------VLGIGPNNG---FPGT-SSIVTAMPGLLNNGVLIDESA 148
Rv1983          ALLADPTTTPFEAYFGAVGVDGVLGVGPNAV---GPGP-SIPTMALPGDLNQGVLIDAPA 168
Rv2519          GGAFGPNGN------------GILGIGPNVGSYAVSGPGNVVTTDLPGQLNEGTLIDIPG 157
MMAR_2272       VTTSIPLSS----------LPLYLGIGPNND---FPLP-DQVTAALPGDLNQGVLINTNL 137
MMAR_2933       ALWSNPLTTPFDAYFQSAGVDGVLGVGPNAV---GPGP-SIPTQALGGGLGQGLLIDMKG 165
MMAR_1538       GNGFGPTGH------------GVLGIGPNIN-AATGGHGNVVTTALPGQLNEGELINIPQ 153
                                       **:***          .  .  : * *.:* **:

Rv0977          GILQFGPNTL-TGGITISGAPISTVAVQID----NGPLQQAP-VMFDSGGINGTIPSALA 202
Rv1983          GELVFGPNPLPAPNVEVVGSPITTLYVKID----GGTPIPVP-SIIDSGGVTGTIPSYVI 223
Rv2519          GYMQFGPNTG-TPITSVTGAPITVLNVQIGGYDPNGGYWSLP-SIFDSGGNHGTLPAVIL 215
MMAR_2272       GYLQFGANPL-TPVASVTGSPVTDLQIQIN----NGPLQPATGSFIDSGGLYGTIPSSLI 192
MMAR_2933       GELVFGPNPL-TPEFSISGAPIATLWVSVN----GGAPVAVP-SIIDSGGVMGTIPSSVI 219
MMAR_1538       GYMQFGPNTG-TPITSVSGVPITTLDVQFGGYDPLGTYYPVT-SIVDSGGNHGTIPGIIL 211
                * : **.*.  :    : * *:: : :...     *     .  :.****  **:*. :

Rv0977          S-LPSGGFVPAGTTISVYTSDGQTLLYSYTTTATNTPFVTSGG-VMNTGHVPFAQQPIYV 260
Rv1983          G----SGTLPANTNIEVYTS-PGGDRLYAFNTNDYRPTVISSG-LMNTGFLPFRFQPVYI 277
Rv2519          GTGQTTGYAPPGTVISISIHDNQTLLYQYTTTASNSP-VVTADPRLNTGLTPFLLGPVYI 274
MMAR_2272       PGVPVGYSVPVGTTITVYTTDGVQLYSQTVTGSTNAPLVVPSNNPFNTGNYPFLLGPIYI 252
MMAR_2933       G----GSTLPANTNITVYTDNTMTTEVYHYSTNDYQPTVISSG-LMNTGFLPFWNQPVYI 274
MMAR_1538       GTGQTSGVVPPGTVISISTNDNQTLLYSYTTTGTDSP-VVTGNIPMNTGLLPFALGPVYI 270
                         * .* * :             .     * * ...  :***  **   *:*:

Rv0977          SYSPTAIGTTTFN---- 273
Rv1983          DYSPSGIGTTVFDHPA- 293
Rv2519          SNNPSGVGTVVFNYPPP 291
MMAR_2272       SNSPTGGGQTIFDF--- 266
MMAR_2933       DYSPAGTGTTVFDMP-- 289
MMAR_1538       SNSPSGVGTVVFNYPPP 287
                . .*:. * . *:
```

Figure **S4** Structure-based sequence alignment of Mtb-AP with Bla g 2 (1YG9), bovine chymosin (4CMS) and porcine pepsin (3PEP) shows the insertion of alanine residue in Mtb-AP analogous to phenylalanine in Bla g 2 in the flap region.

```
                                            ↓
MTAP.pdb    46   --------FLTQNF-G-PVIGTGTAGYAGGLTYNYNTYSTTVDFGNGLLTLPTSVNVVTS    95
1YG9.pdb    55   CPNLQKY-EK--LKPKYISDGNVQVKFFDTGSAVGRGIEDSLTISQ-LTTSQQDIVLADE   110
4CMS.pdb    54   N--HQRFDPRKSSTFQ-NLGKPLSIHY-GTGSMQGILGYDTVTVSN-IVDIQQTVGLSTQ   108
3pep.pdb    52   D--HNQFNPDDSSTFE-ATSQELSITY-GTGSMTGILGYDTVQVGG-ISDTNQIFGLSET   106
```

Figure S5 Conformation of the flap residue tyrosine. (A) Tyrosine of Mtb-AP (green: Y64) is in a different conformation from that of other aspartic proteinases like porcine pepsin (brown; Y75) and rhizopuspepsin (grey; Y77). (B) The tyrosine of Mtb-AP has the same conformation as in the self-inhibited states of chymosin (pink; Y75) and saccharopepsin (cyan; Y75).