

Supporting information

Enhancing Sequence Coverage of Nanodiamond-extracted Early Secretory Proteins from *Mycobacterium tuberculosis* Complex

Po-Chi Soo², Ching-Chieh Lee¹, Meng-Fu Shie¹, Avinash A. Patil¹, Mhikee Janella N. Descanzo¹, Ya-Ching Chin¹, Xi-An Chen¹, Yu-Tze Horng², Chih-Bin Lin³, Jen-Jyh Lee³, Cheng-Kang Chiang⁴, Wen-Ping Peng^{*,1}

¹Department of Physics, National Dong Hwa University, Shoufeng, Hualien, Taiwan

²Department of Laboratory Medicine and Biotechnology, Tzu Chi University, Hualien, Taiwan

³Department of Internal Medicine, Buddhist Tzu Chi General Hospital and Tzu Chi University, Hualien, Taiwan

⁴Department of Chemistry, National Dong Hwa University, Shoufeng, Hualien, Taiwan

FTIR measurement

The functional groups on the surface of the DND was measured using an MB154 FTIR spectrometer interferometer (ABB Inc, Bomem, Zurich, Switzerland) with the sample in ambient environment. Before the measurement, 2 mg of the DND powder was dissolved in 60 μ L of DDW. Then, 20 μ L of the solution was carefully dropped on a silicon wafer and dried in air to form a thin film. The measured spectrum is shown in Figure S1.

Zeta measurements

The zeta potentials of DNDs (50 g/mL) at different pH conditions were measured using a Malvern Zetasizer Nano-ZS (Malvern Instruments Inc., United Kingdom) equipped with a HeNe laser at a wavelength of 632.8 nm (standard: red) at room temperature. Each zeta potential measurement was repeated six times, with the final mean values plotted in supporting Figure S2.

Flowchart for protein analysis

In Figure S5, we developed an algorithm to reduce the interference from other proteins. For those high mass peaks with $m/z > 3000$, the Flex control program cannot automatically select those peptide peaks. Thereby, it is of necessity to pick those peptide peaks manually. Therefore, pre-peak selection is necessary for peptide analysis with CI-CCA matrix (Table 1). Then we removed the baseline and smoothed the mass spectrum. Next for those BSA and ACR[31] peptide peaks were all removed. We noticed that when the ACR peptide peaks were removed, both CFP-10 and ESAT-6 proteins were identified simultaneously by Mascot software with the scores of 55 and 51. The sequence coverage of CFP-10 and ESAT-6 were 94% and 100%.

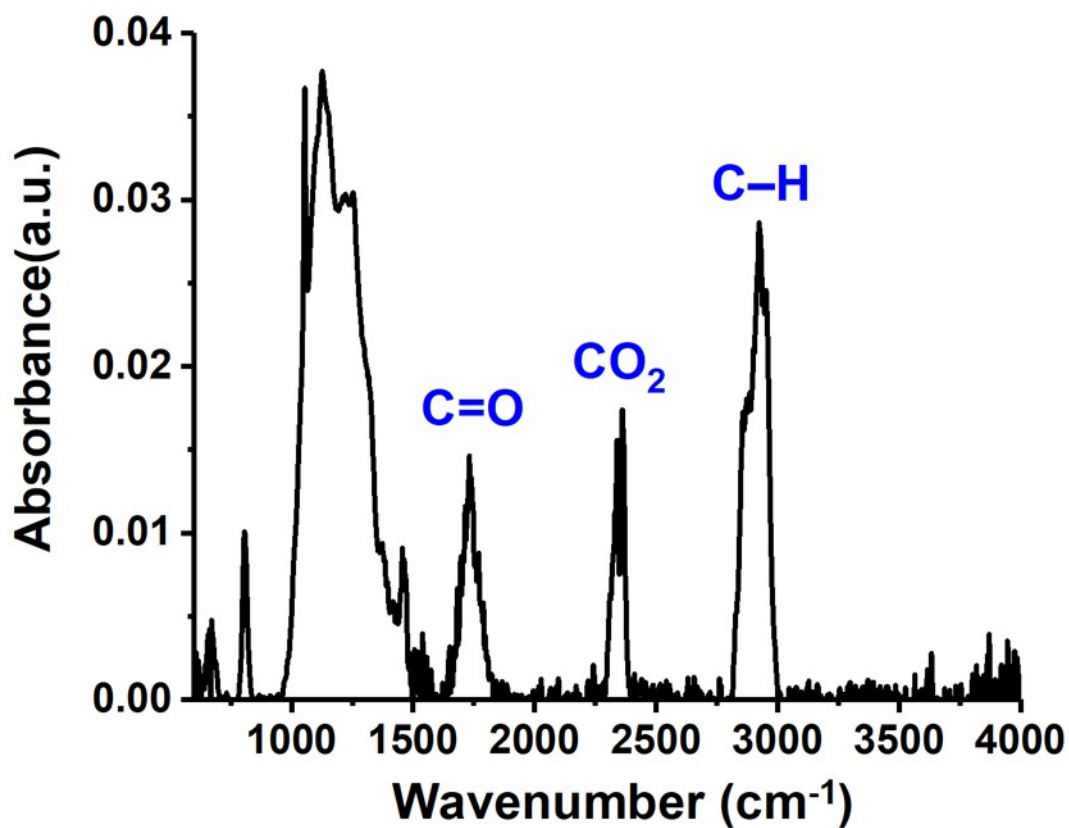


Figure S1 Surface functional groups present on the surface of the 5 nm DND validated using FTIR spectroscopy. It is generally known that NDs are an exceptional platform for protein adsorption. NDs strongly adsorb proteins through the combination of ionic interaction, hydrophobic interaction, hydrogen bonding, and van der Waals interaction. In addition to this, functional groups, such as carboxylic groups, play a significant role in the efficient adsorption of proteins.

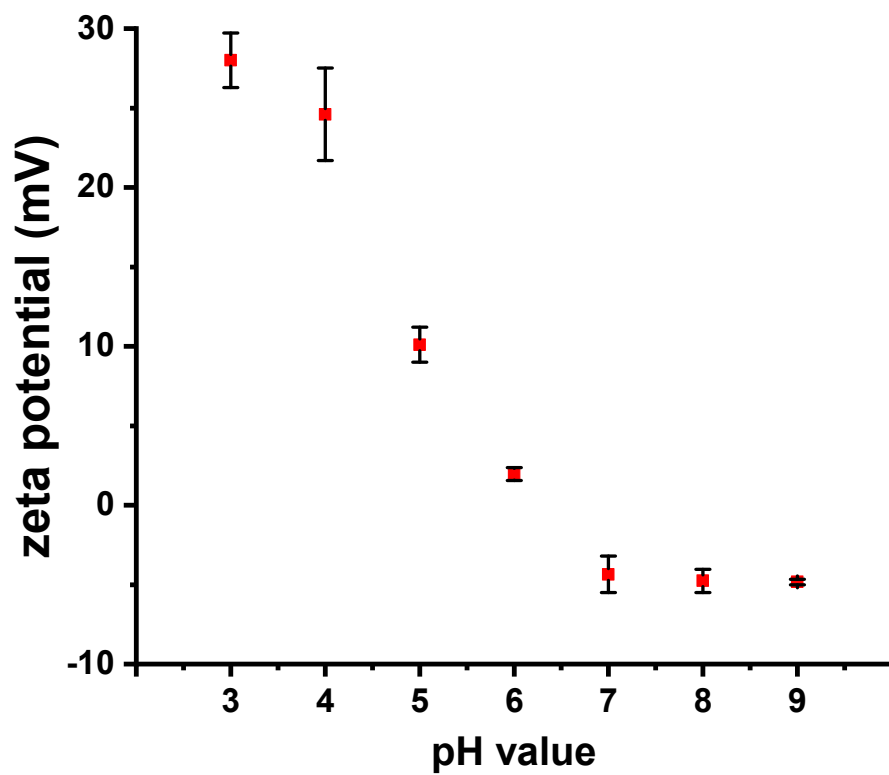


Figure S2 Zeta potential measurement for DND nanoparticles at different pH conditions. The pH was adjusted with 16 mM phosphate buffer.

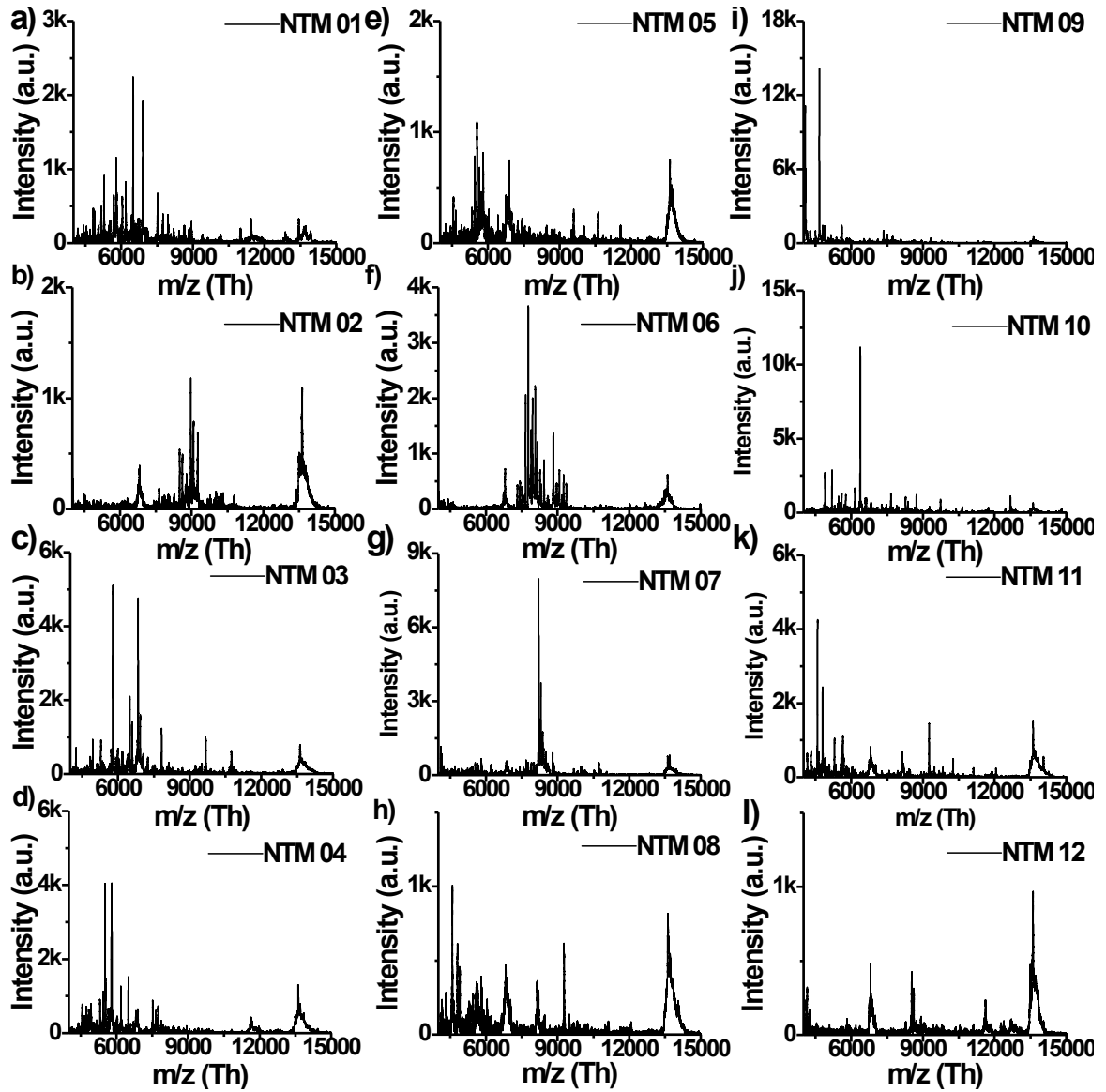


Figure S3 MALDI mass spectra of proteins extracted by DND from NTM isolates, a-l.

TB biomarker protein signals were absent in NTM isolates.

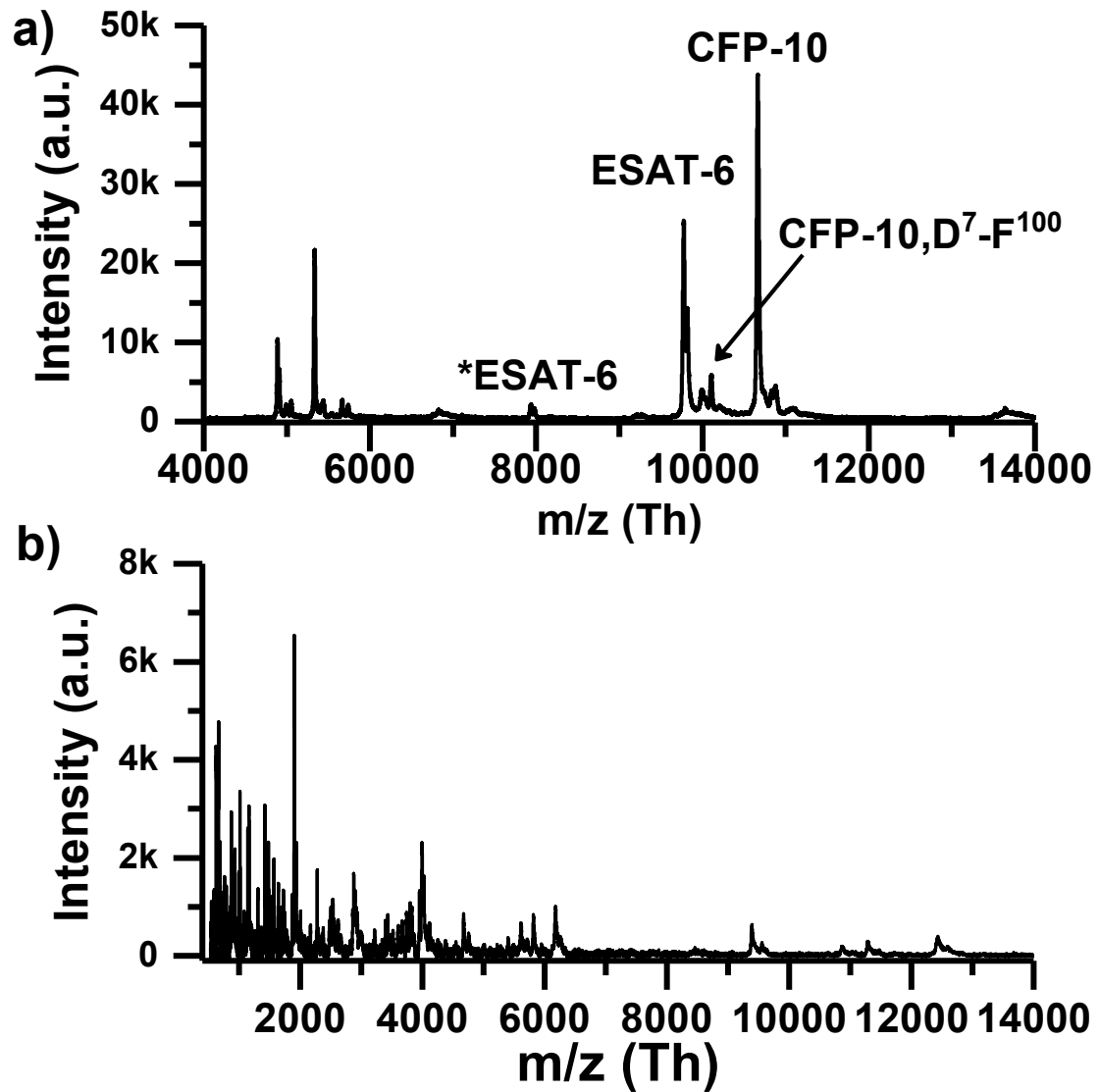


Figure S4 a) MALDI mass spectrum of CFP-10 and ESAT-6 proteins extracted by DND from MTC isolates. b) Tryptic peptide mass spectra of MTC proteins (Figure S3a) after on-diamond digestion. The CFP-10 and ESAT-6 protein peaks were almost absent after trypsin digestion.

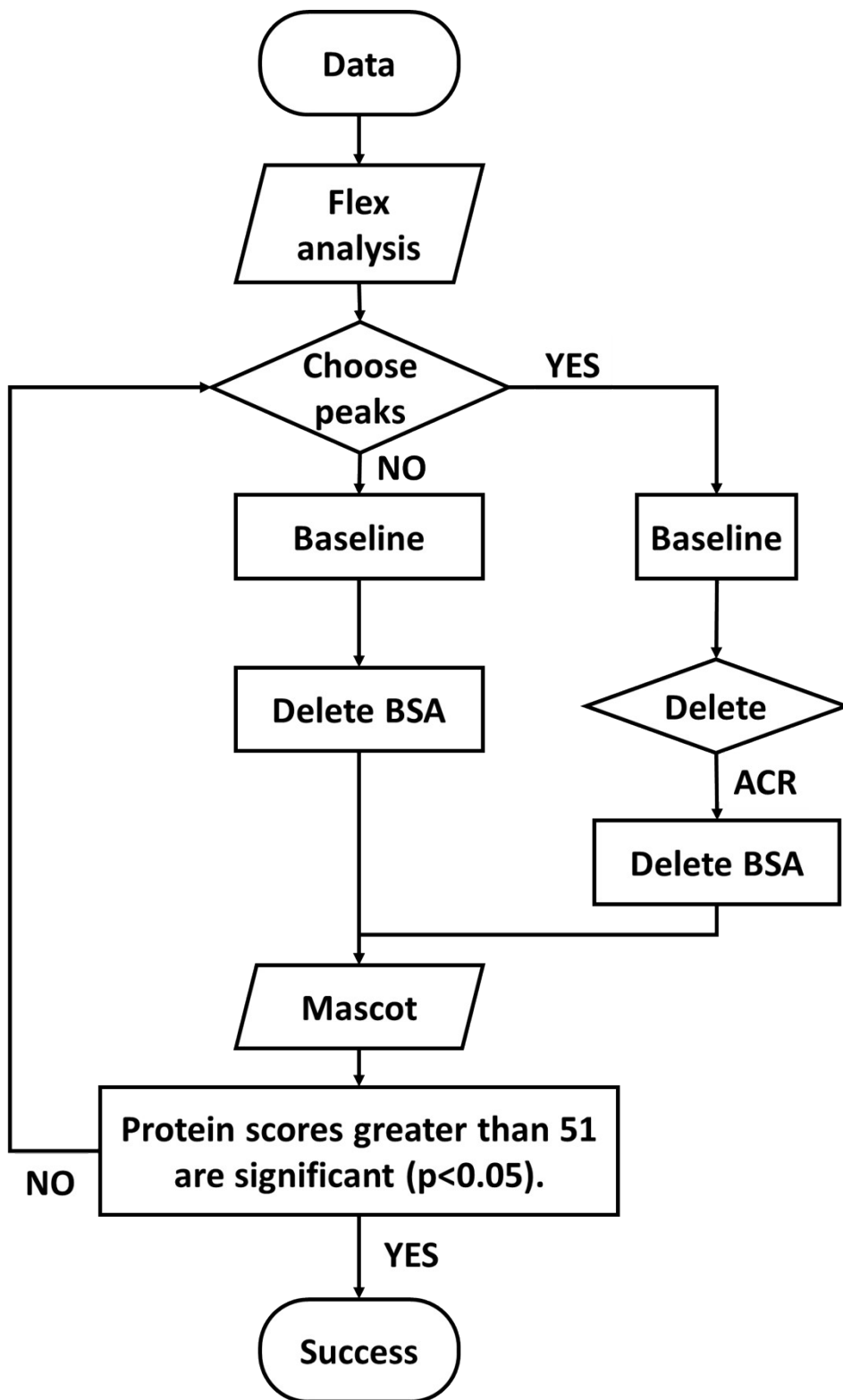


Figure S5 Flowchart for peptide sequence analysis

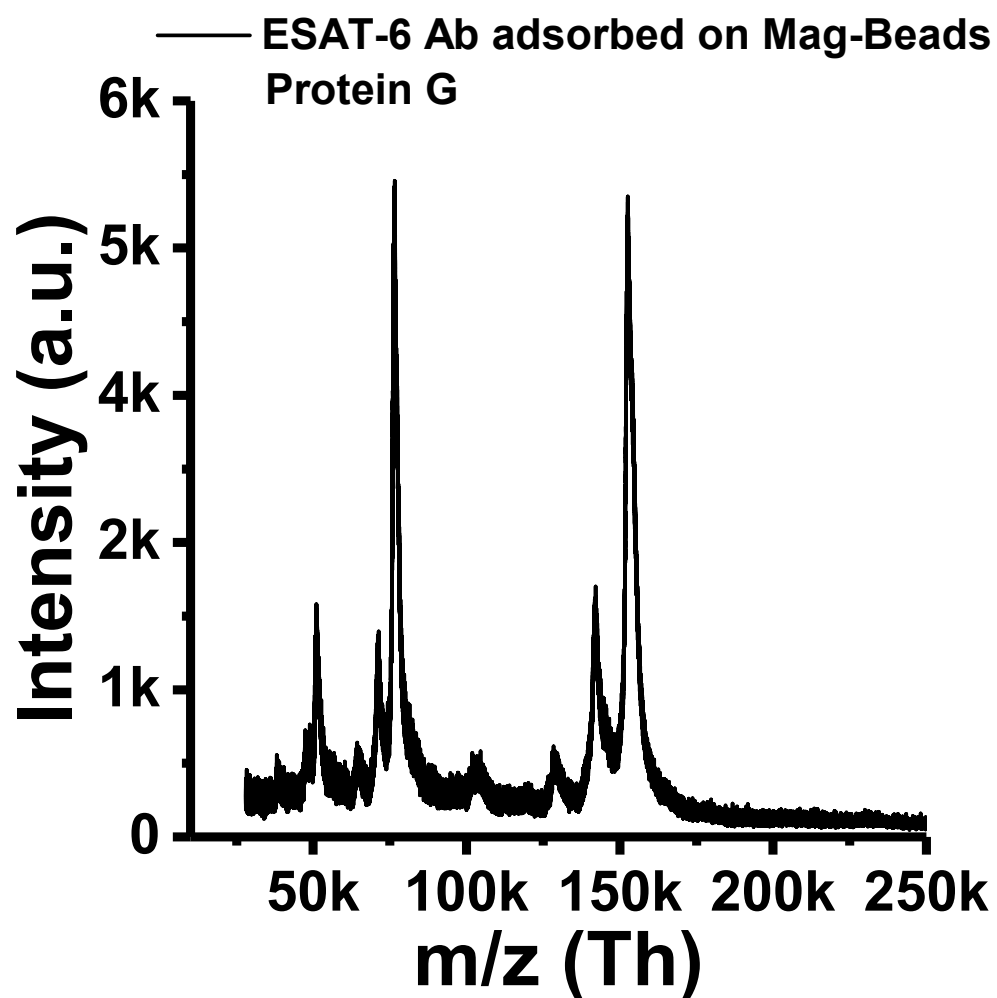


Figure S6 MALDI mass spectrum of ESAT-6 Ab attached to the Mag-Beads Protein G beads.

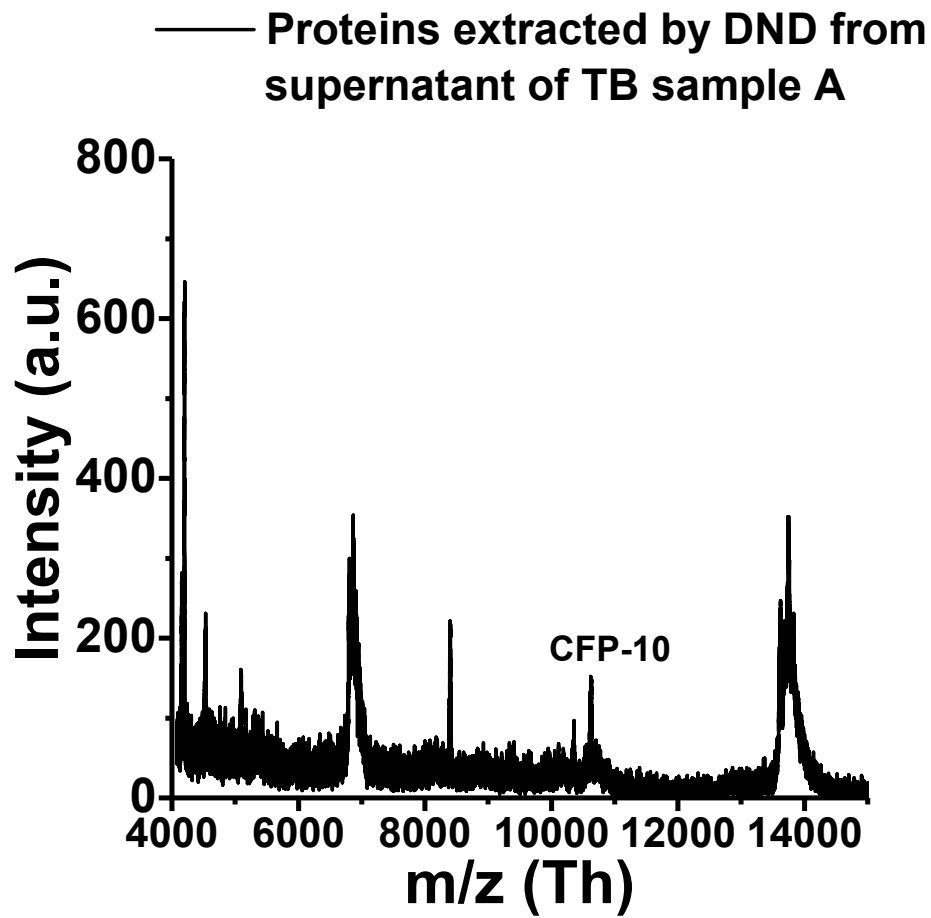


Figure S7 MALDI mass spectrum of the protein extracted by DND from leftover supernatant of the TB sample A. Only little CFP-10 ion signal was found.

Table S1 Mascot analysis of TB samples extracted by DND with CHCA, DHB, and CI-CCA matrices

● **CHCA**

(a-1) CFP10 sequence coverage: 87%

**MAEMKTDAAT LAQEAGNFER ISGDLKTQID QVESTAGSLQ GQWRGAAGTA
AQAAVVRFQE AANKQKQELD EISTNIRQAG VQYSRADEEQ QQALSSQMGF**

(a-2) Mascot analysis result of CFP10, Mascot score is 66

Start-end	Observe	Mr(expt)	Miss	Peptide
1 – 5	609.3460	608.3387	0	-.MAEMK.T
2 – 20	2054.0080	2053.0007	1	M.AEMKTDAATLAQEAGNFER.I
6 – 20	1593.5180	1592.5107	0	K.TDAATLAQEAGNFER.I
27 – 44	2003.7050	2002.6977	0	K.TQIDQVESTAGSLQGQWR.G
45 – 57	1142.6130	1141.6057	0	R.GAAGTAAQAAVVR.F
65 – 77	1573.6060	1572.5987	1	K.QKQELDEISTNIR.Q
67 – 77	1317.5530	1316.5457	0	K.QELDEISTNIR.Q
78 – 85	908.5180	907.5107	0	R.QAGVQYSR.A
78 – 100	2558.4290	2557.4217	1	R.QAGVQYSRADEEQQQALSSQMGF.-

● **DHB**

(b-1) CFP10 sequence coverage: 87%

**MAEMKTDAAT LAQEAGNFER ISGDLKTQID QVESTAGSLQ GQWRGAAGTA
AQAAVVRFQE AANKQKQELD EISTNIRQAG VQYSRADEEQ QQALSSQMGF**

(b-2) Mascot analysis result of CFP10, Mascot score is 65

Start-end	Observe	Mr(expt)	Miss	Peptide
1 – 5	625.4120	624.4047	0	-.MAEMK.T+ Oxidation (M)
2 – 20	2053.7400	2052.7327	1	M.AEMKTDAATLAQEAGNFER.I
6 – 20	1594.1020	1593.0947	0	K.TDAATLAQEAGNFER.I

27 – 44	2004.6430	2003.6357	0	K.TQIDQVESTAGSLQGQWR.G
45 – 57	1142.8300	1141.8227	0	R.GAAGTAAQAAVVR.F
65 – 77	1574.0720	1573.0647	1	K.QKQELDEISTNIR.Q
78 – 85	908.5950	907.5877	0	R.QAGVQYSR.A
78 – 100	2559.5270	2558.5197	1	R.QAGVQYSRADEEQQALSSQMGF.-
86 – 100	1668.1340	1667.1267	0	R.ADEEQQALSSQMGF.-

● CI-CCA (without the deletion of ACR)

(c-1) ACR sequence coverage: 59%

MATTLPVQRH PRSLFPEFSE LFAAFPSFAG LRPTFDTRLM RLEDEMKEGR
YEVRAELPGV DPKDVDIMV RDGQLTIKAE RTEQKDFDGR SEFAYGSFVR
TVSLPVGAD EDDIKATYDKG ILTVSVAVSE GKPTEKHIQI RSTN

(c-2) Mascot analysis result of ACR, Mascot score is 66

Start-end	Observe	Mr(expt)	Miss	Peptide
2 – 9	885.4510	884.4437	0	M.ATTLPVQR.H
42 – 50	1106.6370	1105.6297	1	R.LEDEMKEGR.Y
42 – 50	1121.6200	1120.6127	1	R.LEDEMKEGR.Y+ Oxidation (M)
48 – 54	908.2080	907.2007	1	K.EGRYEV.R
51 – 54	566.3540	565.3467	0	R.YEV.R
51 – 4	1588.7760	1587.7687	1	R.YEVRAELPGVDPDK.D
55 – 71	1869.2750	1868.2677	1	R.AELPGVDPDKDVDIMVR.D
65 – 71	847.5490	846.5417	0	K.DVDIMVR.D
72 – 78	774.5140	773.5067	0	R.DGQLTIK.A
82 – 90	1095.6310	1094.6237	1	R.TEQKDFDGR.S
86 – 90	609.3620	608.3547	0	K.DFDGR.S
91 – 100	1162.7530	1161.7457	0	R.SEFAYGSFVR.T
120 – 141	2364.2300	2363.2227	1	K.GILTVSVAVSEGKPTEKHIQIR.S

137 – 141 666.4880 665.4807 0 K.HIQIR.S

(c-3) CFP10 sequence coverage : 94%

MAEMKTDAAT LAQEAGNFER ISGDLKTQID QVESTAGSLQ GQWRGAAGTA
AQAAVVRFQE AANKQKQELD EISTNIRQAG VQYSRADEEQ QQALSSQMGF

(c-4) Mascot analysis result of CFP10, Mascot score is 54

Start-end	Observe	Mr(expt)	Miss	Peptide
1 – 5	609.3620	608.3547	0	-.MAEMK.T
6 – 20	1594.1860	1593.1787	0	K.TDAATLAQEAGNFER.I
27 – 57	3127.0220	3126.0147	1	K.TQIDQVESTAGSLQGQWRGAAGTAAQAAVVR.F
45 – 57	1142.8420	1141.8347	0	R.GAAGTAAQAAVVR.F
58 – 64	807.4920	806.4847	0	R.FQEAANK.Q
65 – 77	1574.2020	1573.1947	1	K.QKQELDEISTNIR.Q
78 – 85	908.2080	907.2007	0	R.QAGVQYSR.A
78 – 100	2559.5940	2558.5867	1	R.QAGVQYSRADEEQQALSSQMGF.-
86 – 100	1669.4120	1668.4047	0	R.ADEEQQALSSQMGF.-

● CI-CCA (deletion of ACR)

(d-1) CFP10 sequence coverage: 94%

MAEMKTDAAT LAQEAGNFER ISGDLKTQID QVESTAGSLQ GQWRGAAGTA
AQAAVVRFQE AANKQKQELD EISTNIRQAG VQYSRADEEQ QQALSSQMGF

(d-2) Mascot analysis result of CFP10, Mascot score is 55

Start-end	Observe	Mr(expt)	Miss	Peptide
1 – 5	609.3450	608.3377	0	-.MAEMK.T
6 – 20	1594.1300	1593.1227	0	K.TDAATLAQEAGNFER.I
27 – 44	2004.7380	2003.7307	0	K.TQIDQVESTAGSLQGQWR.G
27 – 57	3127.2820	3126.2747	1	K.TQIDQVESTAGSLQGQWRGAAGTAAQAAVVR.F
45 – 57	1142.8230	1141.8157	0	R.GAAGTAAQAAVVR.F

58 – 64	807.4630	806.4557	0	R.FQEAANK.Q
65 – 77	1574.1550	1573.1477	1	K.QKQELDEISTNIR.Q
78 – 85	908.3740	907.3667	0	R.QAGVQYSR.A
78 – 100	2559.9670	2558.9597	1	R.QAGVQYSRADEEQQALSSQMGF.-
86 – 100	1669.1380	1668.1307	0	R.ADEEQQALSSQMGF.-

(d-3) ESAT6 sequence coverage: 100%

**MTEQQWNFAG IEAAASAIQG NVTSIHSLLD EGKQSLTKLA AAWGGSGSEA
YQGVQKQWDA TATELNNALQ NLARTISEAG QAMASTEGNV TGMFA**

(d-4) Mascot analysis result of ESAT6, Mascot score is 51

Start-end	Observe	Mr(expt)	Miss	Peptide
1 – 33	3514.7730	3513.7657	0	-.MTEQQWNFAGIEAAASAIQGNVTSIHSLLDDEGK.Q
2 – 33	3387.6210	3386.6137	0	M.TEQQWNFAGIEAAASAIQGNVTSIHSLLDDEGK.Q
2 – 38	3943.2920	3942.2847	1	M.TEQQWNFAGIEAAASAIQGNVTSIHSLLDDEGKQSLTK.L
34 – 38	576.3780	575.3707	0	K.QSLTK.L
39 – 57	1909.2270	1908.2197	0	K.LAAAWGGSGSEAYQGVQK.W
39 – 74	3792.0670	3791.0597	1	K.LAAAWGGSGSEAYQGVQKWDATATELNNALQNLAR.T
58 – 74	1901.7320	1900.7247	0	K.WDATATELNNALQNLAR.T
58 – 95	3988.9560	3987.9487	1	K.WDATATELNNALQNLARTISEAGQAMASTEGNV TGMFA.-

Oxidation (M)