Supporting Information

Knudsen et al. 10.1073/pnas.1314973111

Materials and Methods

Animals. Six- to 8-wk-old female CB6F1 (BALB/c x C57BL/6) [H-$2^{d/b}$] or B6C3F1 (C57BL/6 × C3H) [H-$2^{b/k}$] mice (Harlan) were rested for 1 wk before initiation of experiments. All experiments were done according to Danish Ministry of Justice and Animal Protection Committees and in compliance with European Community Directive 86/609. Mice were housed in animal facilities at Statens Serum Institut, provided with radiation-sterilized food (Harlan) and water ad libitum, and handled in accordance with the Danish Ministry of Justice and Animal Protection Committee regulations by authorized personnel. Infected animals were housed in a biosafety level 3 facility in cages contained within laminar flow safety enclosures (Scantainer).

Recombinant Proteins. All DNA constructs used in this study were codon-optimized for expression in *Escherichia coli* and made by chemical synthesis followed by insertion into the pJexpress 411 vector (DNA2.0). In the five heterodimers, the single Esx proteins are connected by the thrombin-cleavable 9-mer linker GLVPRGSTG, and in the H65 fusion, the three heterodimers are in addition connected by the 20-amino acid linkers LIGAHPRALNVVKFGGAAFL and LGFGAGRLRGLFTNPGSWRI from Rv1986. After transformation into *E. coli* BL21 AI (Invitrogen), protein expression was induced with 0.2% arabinose and the proteins purified from inclusion bodies by a three-step process, as previously described (1), resulting in a very high purity of the final products (>99%). Using the NanoOrange Protein Quantitation Kit (Invitrogen), the protein concentrations were found to be between 0.1 and 0.5 mg/mL and the yield to be between 1 and 13.5 mg purified protein from 3 L of culture. The identity of all purified proteins were confirmed by mass spectrometry analysis (matrix-assisted laser desorption/ionization–time-of-flight).

Immunizations and Infections. Mice were immunized s.c. in the neck or at the base of the tail three times at 2-wk intervals. Cationic adjuvant formulation 01 [CAF01, 75 μg DDA (dimethyldioctadecylammonium)/25 μg TDB (trehalose 6,6′-dibehenate)] was emulsified with 5 μg recombinant antigen to a final volume of 200 μL for each injection. Negative control mice received three equivalent doses of CAF01, and positive control mice received a single dose of 5×10^4 CFU bacillus Calmette–Guérin Danish 1331 (Statens Serum Institut) in the first round of immunization. Ten weeks after the first immunization, the animals were challenged with *Mycobacterium tuberculosis* (*M.tb*) strain Erdman. Using a Biaera exposure system controlled by the AeroMP aerosol management system, virulent mycobacteria suspended in PBS Tween 20 (0.05%) were aerosolized and delivered via the respiratory route at ∼100 CFU per mouse.

Isolation of Cells and CFU Measurements. Blood samples from six mice were pooled within immunization groups before peripheral blood mononuclear cells were isolated by density-gradient centrifugation, using Lympholyte Mammal (Cedar-Lane Laboratories). Splenocytes and lung mononuclear cells were isolated from individual/pooled animals by forcing cells through a 70-μm nylon cell strainer (BD Pharmingen). For CFU measurements, lung homogenates were prepared in PBS Tween 80 (0.05%) from individual mice and plated at threefold serial dilutions on Middlebrook 7H11 Bacto agar. After 3 wk of incubation at 37 °C, the CFUs were enumerated.

Cytokine Secretion Assays. Peripheral blood mononuclear cells, splenocytes, or lung mononuclear cells (2×10^5 per well) were cultured in round-bottomed 96-well plates in 200 μl complete RPMI media [RPMI 1640 supplemented with 1 mM L-glutamine, 50 μM 2-mercaptoehanol, 1% pyruvate, 1% penicillin-streptomycin, 1% Hepes, and 10% (vol/vol) FCS; Gibco Invitrogen], and 2 μg of antigen at 37 °C in a humidified incubator under 95% air, 5% CO_2. Culture supernatants were harvested from lymphocyte cultures after 72 h of in vitro antigen stimulation and tested in triplicates. For detection of secreted IFN-γ, 96-well Maxisorb microtiter plates (Nunc) were coated with 1 μg/mL monoclonal rat anti-murine IFN-γ (clone R4-6A2; BD Pharmingen). Free binding sites were blocked with 2% (wt/vol) milk powder in PBS. IFN-γ was detected with a 0.1 μg/mL biotin-labeled rat anti-murine antibody (clone XMG1.2; BD Pharmingen) and 0.35 μg/mL horseradish peroxidase-conjugated streptavidin (Zymed). The enzyme reaction was developed with 3.3′, 5.5′-tetramethylbenzidine, hydrogen peroxide (TMB plus; Kementec) and stopped with 0.2 M H_2SO_4. rIFNγ (BD Pharmingen) was used as a standard. Plates were read at 450 nm with an ELISA-reader and analyzed with KC4 3.03 Rev 4 software (BioTek).

Flow Cytometry. Splenocytes or lung mononuclear cells (2×10^6 cells per well) were stimulated in vitro in V-bottom 96-well plates at 37 °C in 200 μl complete media containing anti-CD49d (1 μg/mL) and anti-CD28 (1 μg/mL) antibodies in the presence of recombinant antigen (2 μg/mL) for 1 h, and subsequently incubated for 5–6 h in the presence of 10 μg/mL brefeldin A (Sigma-Aldrich). After overnight storage at 4 °C, cells were washed in FACS buffer (PBS containing 0.1% sodium azide and 1% FCS) and subsequently stained for 30 min at 4 °C for surface markers with mAbs, as indicated, using 1/100 dilutions of anti-CD4-allophycocyanin-Cy7 (clone GK1.5) and anti-CD44-FITC (clone IM7) (all BD Pharmingen). Cells were then washed in FACS buffer, permeabilized using the Cytofix/Cytoperm kit (BD Pharmingen) according to the manufacturer's instructions, and stained intracellularly for 30 min at 4 °C in dilutions of 1/100, using anti-IFN-γ-PE-Cy7 (clone XMG1.2; eBioscience), anti-TNF-PE (MP6-XT22; BD Pharmingen), or anti-IL-2-allophycocyanin (clone JES6-5h4; BD Pharmingen) mAbs. Cells were subsequently washed with BD Perm/Wash buffer (BD Pharmingen), resuspended in FACS buffer, and analyzed using a FACSCanto flow cytometer (BD Pharmingen) and FlowJo software v.8.8.7 (Tree Star). The relative proportions of cells producing different combinations were determined using PESTLE and SPICE v.5.22 software.

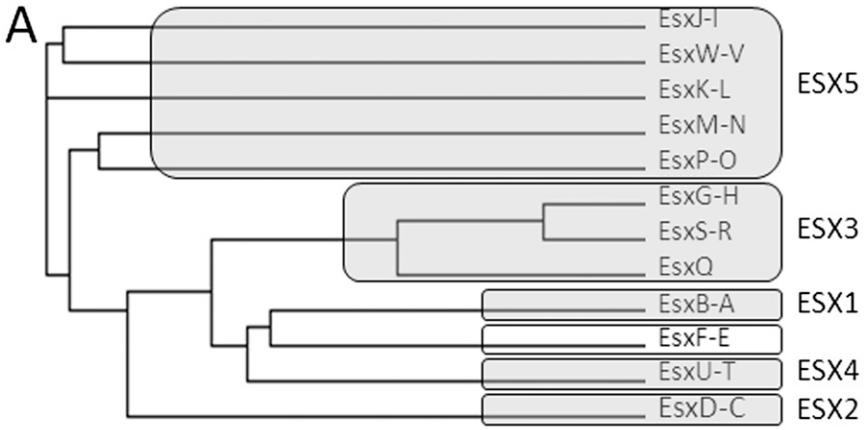
In Vitro and in Vivo Expression of Selected *M.tb* Genes. We determined the gene expression profile of 13 selected *M.tb* genes in in vitro cultures and followed their in vivo expression in lungs of CB6F1 mice infected via the aerosol route with *M.tb* strain Erdman. The expression profile was determined by reverse-transcribing isolated total RNA and amplifying cDNA as previously described (2), followed by quantification in individual real-time PCR reactions. The 13 genes represented Esx substrates for all five ESX secretion systems plus antigens 85A, 85B, and 85C.

Epitope-Binding Predictions. The Immune Epitope Database Analysis Resources server (3), one of the most accurate prediction servers available (4), was used to perform MHC class 2

binding predictions for H65, each of its six protein components, and the two linkers included in H65. This was done for 34 HLA-DBR1 alleles representing the high-frequency HLA-DRB1 alleles among tuberculosis (TB) high-burden populations (5). The Immune Epitope Database documents a large number of peptides tested for binding to various MHC class 2 allelic variants, which are used to predict binding affinities for related peptides (6). Binding affinities are given as IC_{50} values (half maximal inhibitory concentration) in units of nanomoles. Experimental data have classified peptides with $IC_{50} < 1{,}000$ nM as binders and >1,000 nM as nonbinders (7). For conservative reasons, 500 nM was selected as cutoff (7). Epitope binding predictions were generated using the prediction algorithms CombLib (6), SMM (8), NN (9), or NetMHCIIpan (10). Where possible, a consensus approach of multiple algorithms was used to identify the predicted epitopes (Table S2). All of the algorithms provide prediction output in the form of IC_{50} values. The generated prediction results were compiled in Excel 2010 (Microsoft). The various algorithms often generated multiple positive hits of various lengths, centered on the same binding core. In cases like this, the smallest core fragment was identified and counted only once to avoid overestimation.

Statistical Analysis. Prism 5 software (GraphPad) was used for all statistical analyses. CFU data were log-transformed before analyses. One-way ANOVA combined with Tukeys multiple comparison test was used for comparing between multiple groups. Statistical significant differences are marked by asterisks in figures and explained in the figure legends.

1. Aagaard C, et al. (2009) Protection and polyfunctional T cells induced by Ag85B-TB10.4/IC31 against Mycobacterium tuberculosis is highly dependent on the antigen dose. *PLoS One* 4(6):e5930.
2. Aagaard C, et al. (2011) A multistage tuberculosis vaccine that confers efficient protection before and after exposure. *Nat Med* 17(2):189–194.
3. Zhang Q, et al. (2008) Immune epitope database analysis resource (IEDB-AR). *Nucleic Acids Res* 36(Web Server issue, suppl 2):W513–W518.
4. Lin HH, Zhang GL, Tongchusak S, Reinherz EL, Brusic V (2008) Evaluation of MHC-II peptide binding prediction servers: Applications for vaccine research. *BMC Bioinformatics* 9 (Suppl 12):S22.
5. Davila J, McNamara LA, Yang Z (2012) Comparison of the predicted population coverage of tuberculosis vaccine candidates Ag85B-ESAT-6, Ag85B-TB10.4, and Mtb72f via a bioinformatics approach. *PLoS ONE* 7(7):e40882.
6. Wang P, et al. (2010) Peptide binding predictions for HLA DR, DP and DQ molecules. *BMC Bioinformatics* 11:568.
7. Wang P, et al. (2008) A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. *PLOS Comput Biol* 4(4): e1000048.
8. Nielsen M, Lundegaard C, Lund O (2007) Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC Bioinformatics* 8(1):238.
9. Nielsen M, Lund O (2009) NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. *BMC Bioinformatics* 10:296.
10. Nielsen M, et al. (2008) Quantitative predictions of peptide binding to any HLA-DR molecule of known sequence: NetMHCIIpan. *PLOS Comput Biol* 4(7):e1000107.



B

	EsxW-V (192)	*EsxJ-I* (192)	*EsxK-L* (192)	*EsxM-N* (191)	*EsxP-O* (192)
EsxW-V	1,00	0,99	0,98	0,96	**0,95**
EsxJ-I		1,00	0,98	0,96	0,95
EsxK-L			1,00	0,95	0,97
EsxM-N				1,00	0,97
EsxP-O					1,00

C

	EsxG-H (193)	*EsxS-R* (193)	*EsxQ* * 120
EsxG-H	1,00	0,88	0,43
EsxS-R		1,00	**0,44**
EsxQ			1,00

* homology over 120 amino acids

Fig. S1. Phylogenic relationship among substrates for the five ESX secretion systems in *Mycobacterium tuberculosis*. (*A*) Cladogram based on multiple sequence alignment of ESX substrate dimers using Clustaw W2. (*B*) Homology among five closely related ESX5 substrates. Lowest homology is in bold. One equals 100% identity. Total amino acid number for each substrate dimer is given in parenthesis. (*C*) Homologs to the esxG-H genes and possible substrates for the ESX3 secretion system. One locus encodes only one of the proteins in the dimer substrate (esxQ) with modest homology to the other substrates.

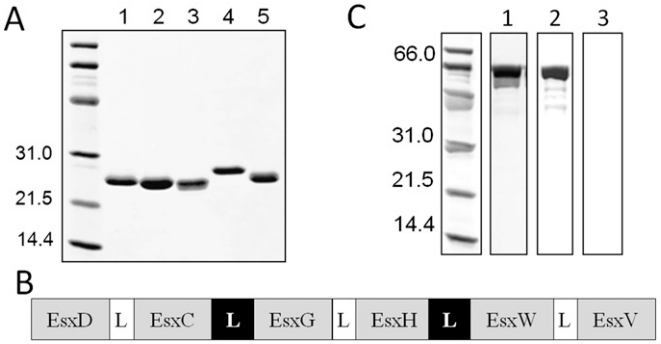


Fig. S2. Dimer fusions of Esx proteins. (*A*) Quality control of the purified protein dimers. Coomassie stained SDS gel. Lane 1, EsxB-EsxA; lane 2, EsxD-EsxC; lane 3, EsxG-EsxH; lane 4, EsxU-EsxT; lane 5, EsxW-EsxV. A molecular weight standard was included for size confirmation. All protein identities were confirmed by mass spectroscopy. (*B*) Schematic representation of H65 showing the six subunits separated by linkers. The individual proteins in the secretion dimers are linked (L, white box) with the same nine-amino acid linker (GLVPRGSTG). The three heterodimers are connected by two nonidentical 20-amino acid-long linkers (L, black box). (*C*) Quality control of the purified H65 fusion protein. H65 fusion protein in SDS-PAGE gels that have been stained with Coomassie blue (1) or transferred to membranes and developed using anti-His (2) or anti-*Escherichia coli* (2) antibodies.

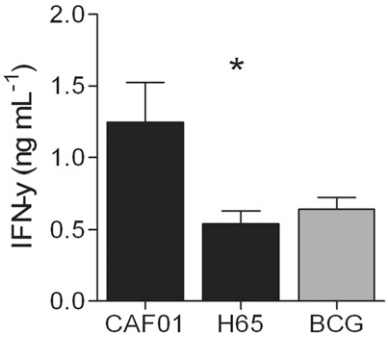


Fig. S3. In vivo INF-γ in lungs of infected mice. Groups of mice were vaccinated with H65 or bacillus Calmette–Guérin or injected with CAF01 adjuvant. Six weeks after the third vaccination, all animals were infected with *M.tb*, and after 6 wk of infection, IFN-γ was measured by ELISA in supernatants from homogenized lungs.

Table S1. Characteristics of the 23 ESAT-6 family proteins

T7S system	H37Rv*	Protein	Common name	kDa	pI	Motif†	Identified in culture filtrate‡
ESX1	**Rv3874**	EsxB	CFP10	10.8	4.31	YSRAD	Yes
ESX1	**Rv3875**	EsxA	ESAT-6	9.9	4.19	—	Yes
ESX2	**Rv3890c**	EsxC	ES6_11	9.9	4.17	No	No
ESX2	**Rv3891c**	EsxD	—	11.2	4.43	No	No
ESX3	**Rv0287**	EsxG	TB9.8	9.8	6.51	YVAAD	Yes
ESX3	**Rv0288**	EsxH	TB10.4	10.4	4.35	—	Yes
ESX3	Rv3017c	EsxQ	TB12.9	12.9	8.06	No	No
ESX3	Rv3019c	EsxR	TB10.3	10.3	4.13	—	No
ESX3	Rv3020c	EsxS	PE28	9.8	6.68	YVAAD	No
ESX4	**Rv3444c**	EsxT	—	11.1	6.34	No	No
ESX4	**Rv3445c**	EsxU	—	11.4	6.68	No	No
ESX5	*Rv1037c*	EsxI	Mtb9.9D	9.8	4.48	—	Yes
ESX5	*Rv1038c*	EsxJ	ES6_2	11.0	5.02	YEQQE	Yes
ESX5	*Rv1197*	EsxK	ES6_3	11.0	5.02	YEQQE	Yes
ESX5	*Rv1198*	EsxL	Mtb9.9C	9.9	4.83	—	Yes
ESX5	*Rv1792*	EsxM	TB11.0	10.9	5.02	YEQQE	Yes
ESX5	*Rv1793*	EsxN	Mtb9.9A	9.9	4.56	—	Yes
ESX5	*Rv2346c*	EsxO	Mtb9.9E	10.0	4.56	—	Yes
ESX5	*Rv2347c*	EsxP	ES6_7	11.0	5.02	YEQQE	No
ESX5	***Rv3619c***	EsxV	ES6_1	9.8	4.48	—	No
ESX5	***Rv3620c***	EsxW	ES6_10	11.0	5.02	YEQQE	Yes
None	Rv3904c	EsxE	ES6_12	9.6	5.45	—	No
None	Rv3905c	EsxF	ES6_13	10.5	4.47	YQHNE	No

*Proteins used in this study are in bold.
†The amino acid pattern YXXXD/E has been identified as an ESX secretion marker present in one of the dimer partners, directing secretion of both proteins. Dash indicated a motif found in a predicted secretion partner.
‡For references, see the TubercuList Webpage (http://tuberculist.epfl.ch/).

Table S2. Gene expression of selected *M. tuberculosis* genes

Gene	H37RV identity	Protein size[†]	Gene expression, RGCN*		$RGCN_{gene1}$+ $RGCN_{gene2}$		Percentage of total	
			In vivo, $t = 17$	In vitro	In vivo, $t = 17$	In vitro	In vivo, $t = 17$	In vitro
esxW[‡]	Rv3620c	98	31975	39502				
esxV[‡]	Rv3619c	94	48802	53961	80777[§]	93463	4,6[¶]	8,9
esxB	Rv3874	100	102456	274157				
esxA	Rv3875	95	915779	581589	1018235	855746	58,3	81,3
esxG	Rv0287	97	254912	25261				
esxH	Rv0288	96	232947	23816	487859	49077	27,9	4,7
esxD	Rv3891c	107	18301	22531				
esxC	Rv3890c	95	35725	14063	54026	36594	3,1	3,5
esxT	Rv3444c	100	39	322				
esxU	Rv3445c	105	261	223	300	545	0,0	0,1
fbpB	Rv1886c	325	24656	1770				
fbpC	Rv0129c	340	50348	2418				
fbpA	Rv3804c	338	29833	12543	104837	16731	6,0	1,6
			$RGCN_{total}$:		1746034	1052156		

*RNA gene copy number (RGCN) measured by RT-PCR.
[†]Number of amino acids.
[‡]Due to high sequence homology the five potential ESX-5 substrates cannot be distinguished by gene expression analysis.
[§]Calculated as 31975 + 48802.
[¶]Calculated as (80777/1746034) × 100.

Table S3. Epitope-binding predictions for 34 HLA-DRB1 alleles common in TB high-burden populations

	Esx proteins								
DRB1	D	C	G	H	W	V	L1*	L2†	H65
*01:01	14	7	17	12	5	7	1	4	67
*01:02	10	7	11	8	8	7	1	4	56
*03:01	1	1	0	1	2	2	1	4	12
*03:02	1	2	3	3	3	0	0	2	14
*04:01	3	3	7	6	5	3	0	0	27
*04:03	0	0	1	4	1	0	0	1	7
*04:04	4	3	6	6	5	1	0	0	25
*04:05	2	1	1	3	2	2	0	1	12
*04:11	3	1	2	3	4	5	0	2	20
*07:01	5	1	5	2	2	1	0	3	19
*08:01	7	1	8	4	3	5	1	3	32
*08:02	0	1	3	1	1	1	0	0	7
*08:03	8	2	7	9	6	5	1	3	41
*08:04	7	3	8	5	4	5	1	3	36
*08:07	0	0	5	2	2	0	0	2	11
*09:01	1	0	3	3	1	1	0	1	10
*10:01	8	6	9	7	5	5	1	3	44
*11:01	1	1	2	1	1	1	0	0	7
*12:01	1	0	2	0	0	1	1	4	9
*12:02	8	2	7	6	4	5	1	3	36
*13:01	5	3	7	4	4	5	1	3	32
*13:02	3	1	2	1	2	1	0	0	10
*13:03	11	9	12	10	10	8	2	4	66
*14:01	2	0	2	2	3	1	0	1	11
*14:02	5	3	6	4	5	3	1	3	30
*14:03	1	0	4	2	2	2	1	2	14
*14:04	4	0	4	2	4	3	0	3	20
*14:05	5	1	6	5	5	4	1	3	30
*14:13	10	8	10	9	8	7	1	4	57
*15:01	0	0	2	2	1	1	0	1	7
*15:02	1	0	4	2	1	1	0	2	11
*15:03	3	0	4	2	3	3	1	3	19
*15:04	2	0	4	2	3	3	0	4	18
*16:02	5	0	8	4	6	6	1	3	33
Sum:	141	67	182	137	121	105	18	79	850

The prediction algorithms CombLib, SMM-align, NN-align, and/or NetMHCIIpan were used.

*Combined results for the three nine-amino acid linkers including 5 amino acids upstream and downstream.

†Combined results for the two 20-amino acid linkers including 5 amino acid upstream and downstream.