Online Supplement

TAC from *Mycobacterium tuberculosis*: a paradigm for toxin-antitoxin systems controlled by SecB-like chaperones

Ambre Sala, Virginie Calderon, Patricia Bordes and Pierre Genevaux

Materials and Methods

Annotation of putative systems

The sequence of the *Mycobacterium tuberculosis* TAC chaperone Rv1957 was used as a query in PSI-BLAST searches among all complete and partial prokaryotic genome sequences available on the NCBI server (http://www.ncbi.nlm.nih.gov) with default parameters. Recursive PSI-BLAST searches were performed using the less conserved retrieved sequences, *i.e.* those with the lowest BLAST score. The procedure was terminated when the results converged to a final stable data set (no new sequences were detected). For each result, the neighborhood was analyzed searching for putative toxin-antitoxin partners on the base of conserved gene organization with the original system (Fig. 1A), except that in some cases we accepted the absence of the toxin gene. Each gene product of the putative partners was used as query in a BLASTP search against the non redundant database (Altschul et al. 1990). Functional annotation was inferred when at least the best hits for the gene located directly upstream of the putative chaperone were annotated as antitoxins (or transcriptional regulator), and eventually as toxins for the more upstream gene. Note that for MTBC and Mpho2 the *Rv1954a* gene of unknown function is conserved.

Taxonomic tree and distribution of solitary SecB versus TAC/AC

The taxonomic tree of bacterial species was generated from the available complete genomes (October 2011) on the basis of the NCBI taxonomic classification (http://www.ncbi.nlm.nih.gov/taxonomy). The tree was then edited with iTol (http://itol.embl.de/index.shtml) and annotated with the solitary SecB chaperones and TAC/AC systems identified. For the sake of clarity, only one strain per species was retained, giving 901 genomes in the tree. Therefore, in some species other strains may not possess the system.

Horizontal Gene Transfer detection

We used Alien Hunter (AH) software (Vernikos and Parkhill 2006) to identify regions that have unusual sequence composition in terms of k-mers for various values of k (called interpolated variable order). An alien region is one whose AH score is above a genome-dependent and automatically calculated threshold based on the sequence composition of the whole genome (termed background composition). AH was run on all the 52 complete genomes possessing a TAC or AC system. Predictions were visualized using Artemis (Rutherford et al. 2000). In some cases, AH predictions could be reinforced when the region encoded other HGT signatures such as tRNA, integrase, transposase, phage genes, plasmid genes or TA genes.

MCL analyses

A graph was built in which nodes correspond to proteins and weighted edges represent the BLASTP log(e-value) obtained between a pair of proteins. Graph partitioning - detection of communities of highly connected nodes - was performed using the Markov clustering program. Sequences used correspond to the proteins whose locus tags are given in table S1.

For SecB chaperones we used the seed of the Pfam02556, corresponding to a subset of 114 representative sequences. Two sequences were removed: one from an archaea, since our analysis focuses on bacteria, and one that our analysis identified as part of a TAC system. The homology relationship was inferred when an e-value less than or equal to 10^{-5} was observed between two chaperone or toxin sequences and 10^{-3} between two antitoxin sequences. Partitioning was performed by using an inflat factor of 1.2.

Determination of antitoxin and toxin families

On the base of MCL results, we used RPS-BLAST (Altschul et al. 1990) to detect conserved domains describing each community. The five MCL antitoxin communities, namely HigA, HicB, MqsA, MqsA-like and HTH-like are respectively described by the conserved domains pfam01381, pfam05534, tigr03830, a partial tigr03830 (no CXXCG N-terminal motifs) and a partial and weakly conserved pfam01381. MCL partition leads to 7 toxin communities from which 3 correspond to the pfam05973 that characterizes the HigB toxin of *M. tuberculosis*, thus these sequences were annotated as HigB toxins. Repartition of HigB toxins in the three communities globally follows taxonomic groups. The MqsR family was named according to the annotation of one of the members of the corresponding community. All toxins associated to HicB antitoxins presented the pfam07927 conserved domain. This domain characterizes the sequences homologues to the YcfA protein from *E. coli*. Since the BLASTP best hits for this protein are annotated as HicA toxins, we named this family HicA. Two toxin communities did not present any conserved domain, but were considered as potential new families. In some cases, putative toxin sequences were excluded from the MCL analysis, but were either described by conserved domains corresponding to the HigB or HicA families, either presented low sequence similarity with proteins annotated as toxins.

Strains and media

In vivo assays were performed in the strain *Escherichia coli* W3110 Δ*secB*::Cm^R (Ullers et al. 2007). Bacteria were grown at 37°C in LB medium supplemented when necessary with ampicillin (100 µg/mL) or kanamycin (100 µg/mL).

Plasmid constructs

Plasmids pSE380Δ*Nco*I (Genevaux et al. 2004), pSE380-SecB (Ullers et al. 2004), pSE380-Rv1957, pMPMK6 and pMPMK6-HigBA (Bordes et al. 2011) were described previously. To construct plasmid pSE380-SmegB, the gene *Msmeg_2143* was amplified by PCR from *Mycobacterium smegmatis* mc^2155 genomic DNA using primers smegB-for (5'-gacaattgcatatgattgagcgggacggcgcgcccac-3') and smegB-rev (5'-gaaagcttggatcctcaggcgtcctcggccgagtccatag-3'). PCR fragment was digested with *Mfe*I/*Hind*III and ligated into pSE380Δ*Nco*I digested with *Eco*RI/*Hind*III.

In vivo assays

The *in vivo* assays were performed as described (Bordes et al. 2011) with minor modifications. For complementation of Rv1957 function in controlling HigBA, the *E. coli* W3110 Δ*secB* strain co-transformed with plasmids pMPMK6-HigBA and pSE380Δ*NcoI*, pSE380-Rv1957 or pSE380-SmegB, were grown to mid-log phase in LB containing 0.2% glucose, kanamycin and ampicillin. Cultures were serially diluted and spotted on LB-ampicillin-kanamycin agar plates with or without arabinose and IPTG inducers as indicated. Plates were incubated at 37°C overnight. For complementation of the SecB chaperone activity at low temperature, fresh transformants of *E. coli* W3110 Δ*secB* containing pSE380Δ*NcoI*, pSE380-SecB, pSE380-Rv1957 or pSE380-SmegB were grown to mid-log phase at 37°C in LB containing 0.2% glucose and ampicillin. Cultures were serially diluted and spotted on LB-ampicillin

plates with or without IPTG as indicated and incubated at 37°C overnight or at 16°C for 5 days.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403-410

Bordes P, Cirinesi AM, Ummels R, Sala A, Sakr S, Bitter W, Genevaux P (2011) SecB-like chaperone controls a toxin-antitoxin stress-responsive system in *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A 108: 8438-8443

Genevaux P, Keppel F, Schwager F, Langendijk-Genevaux PS, Hartl FU, Georgopoulos C (2004) In vivo analysis of the overlapping functions of DnaK and trigger factor. EMBO Rep 5: 195-200

Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B (2000) Artemis: sequence visualization and annotation. Bioinformatics 16: 944-945

Ullers RS, Ang D, Schwager F, Georgopoulos C, Genevaux P (2007) Trigger Factor can antagonize both SecB and DnaK/DnaJ chaperone functions in *Escherichia coli*. Proc Natl Acad Sci U S A 104: 3101-3106

Ullers RS, Luirink J, Harms N, Schwager F, Georgopoulos C, Genevaux P (2004) SecB is a bona fide generalized chaperone in *Escherichia coli*. Proc Natl Acad Sci U S A 101: 7583-7588

Vernikos GS, Parkhill J (2006) Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the *Salmonella* pathogenicity islands. Bioinformatics 22: 2196-2203

Fig. S1. Conserved universal SecB-like motif designed with Web Logo (http://weblogo.berkeley.edu/logo.cgi) for the 60 TAC/AC chaperone sequences and the 114 SecB sequences of the seed sample of the pfam02556 conserved domain. This motif is located at positions 106-114 on the *E. coli* SecB sequence, and 138-146 on the *M. tuberculosis* Rv1957 sequence. The size of the amino-acid letters, measured in bits, indicates the level of conservation at the position.

Fig. S2. MCL analysis of the TAC/AC antitoxin sequences. The MCL communities are depicted with brackets and the corresponding antitoxin family names are indicated. For each system, the color assigned for the chaperones communities (Fig. 2A) was conserved. Note that the sequences from the MqsA and MqsA-like families formed one community when increasing the e-value threshold at 10^{-2}.

Fig. S3. (A) Steady-state expression of SmegB and Rv1957. Cultures of W3110 Δ*secB* containing plasmids pMPMK6-HigBA and pSE380Δ*NcoI*, pSE380-SmegB or pSE380-Rv1957 were grown to mid-log phase at 37°C. Protein expression was induced with 50µM (left) or 500µM (right) IPTG, and 20min later with 0.1% arabinose for 1h. Whole cell extracts were separated on SDS-PAGE. (B) Chaperone protection of the HigA antitoxin. The extracts from (A) induced with 50µM IPTG were analyzed by western blot using a rabbit anti-HigA antibody (Bordes et al. 2011). (C) SmegB/SecB and SmegB/Rv1957 amino acid sequence alignments realized with MAFFT and edited with Jalview. The residues were colored by conservation with a threshold of 8.5. Positions are colored with a grayscale for similarity, and in black for identity.

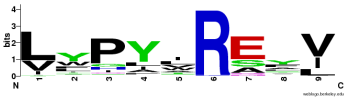
Table S1. [A] Strain abbreviations used in the figure 2 legend.*Pathogenic strains

Taxonomic group	Abb.[A]	Species	Toxin	Antitoxin	Chaperone	Ref. Genome	
γ-proteobacteria	Acal	*Acidithiobacillus caldus* SM-1	x	Atc_1144	Atc_1145	CP002573.1	
	Ajoh	*Acinetobacter johnsonii* SH046	HMPREF0016_03309	HMPREF0016_03310	HMPREF0016_03311	GG704981.1	
	Aehr	*Alkalilimnicola ehrlichii* MLHE-1	x	Mlg_2307	Mlg_2308	CP000453.1	
	Mmet	*Methylomonas methanica* MC09	Metme_2813	Metme_2812	Metme_2811	CP002738.1	
	Nhal	*Nitrosococcus halophilus* Nc4	Nhal_3521	Nhal_3522	Nhal_3523	CP001798.1	
	Pana	*Pantoea ananatis* AJ13355	x	PAJ_0757	PAJ_0758	AP012032.1	
	Pflu	*Pseudomonas fluorescens* SBW25	PFLU_4126A	PFLU_4126B	PFLU_4126C	AM181176.4	
	VchoA	*Vibrio cholerae* RC385	VCRC385_03678	VCRC385_03679	VCRC385_03680	GG774561.1	
	VchoB	*Vibrio cholerae* HE-09	VCHE09_0202	VCHE09_0203	VCHE09_0204	AFOP01000111.1	*
	Taue	*Tolumonas auensis* DSM 9187	x	Tola_2854	Tola_2853	CP001616.1	
β-proteobacteria	Axyl	*Achromobacter xylosoxidans* C54	HMPREF0005_02907	HMPREF0005_02908	HMPREF0005_02909	ACRC01000563.1	*
	Ctai	*Cupriavidus taiwanensis* LMG19424	pRALTA_0477	pRALTA_0478	pRALTA_0479	**CU633751.1**	
	Gcap	*Gallionella capsiferriformans* ES-2	Galf_0727	Galf_0728	Galf_0729	CP002159.1	
	Rsol	*Ralstonia solanacearum* CFBP2957	RCFBP_mp30545	RCFBP_mp30544	RCFBP_mp30543	**FP885907.1**	*
δ-proteobacteria	Dalk	*Desulfatibacillum alkenivorans* AK-01	Dalk_1596	Dalk_1597	Dalk_1598	CP001322.1	
	Dace	*Desulfobacca acetoxidans* DSM 11109	Desac_2259	Desac_2260	Desac_2261	CP002629.1	
	Dret	*Desulfohalobium retbaense* DSM 5692	Dret_0389	Dret_0390	Dret_0391	CP001734.1	
	Glov	*Geobacter lovleyi ATCC* BAA-1151	Glov_3492	Glov_3491	Glov_3490	CP001089.1	
	Gura	*Geobacter uraniireducens* Rf4	Gura_1411	Gura_1410	Gura_1409	CP000698.1	
	Pcar	*Pelobacter carbinolicus* DSM 2380	Pcar_0788	Pcar_0789	Pcar_0790	CP000142.2	
	Saci	*Syntrophus aciditrophicus* SB	SYN_00016	SYN_00015	SYN_00014	CP000252.1	
	Udes	*uncultured Desulfobacterium sp. env.*	N47_E41910	N47_E41920	N47_E41930	FR695877.1	
Firmicute	Amet	*Alkaliphilus metalliredigens* QYMF	Amet_3134	Amet_3133	Amet_3132	CP000724.1	
	Cbes	*Caldicellulosiruptor bescii* DSM 6725	Athe_2320	Athe_2319	Athe_2318	CP001393.1	
	Chyd	*Caldicellulosiruptor hydrothermalis D*SM 18901	Calhy_0207	Calhy_0208	Calhy_0209	CP002219.1	
	Ckro	*Caldicellulosiruptor kronotskyensis* DSM 18902	Calkro_0234	Calkro_0235	Calkro_0236	CP002330.1	
	Csac	*Caldicellulosiruptor saccharolyticus DSM 8903*	Csac_0625	Csac_0626	Csac_0627	CP000679.1	
	LplaA	*Lactobacillus plantarum plantarum* ATCC 14917	HMPREF0531_10367	HMPREF0531_10368	HMPREF0531_10369	ACGZ02000005.1	
	LplaB	*Lactobacillus plantarum plantarum* ST-III	x	LPST_C0260	LPST_C0261	CP002222.1	
	Txyl	*Thermoanaerobacterium xylanolyticum* LX-11	Thexy_1763	Thexy_1762	Thexy_1761	CP002739.1	
Actinobacteria	Cmic1	*Clavibacter michiganensis subsp. Sepedonicus* ATCC33113	CMS2954	CMS2955	CMS2956	AM849034.1	*
	Cmic2		pCS0043	pCS0044	pCS0045	**AM849035.1**	*
	Cglu	*Corynebacterium glutamicum* ATCC13032	Cgl1746	Cgl1745	Cgl1744	BA000036.3	
	Mpho1	*Microlunatus phosphovorus* NM-1	x	MLP_51410	MLP_51420	AP012204.1	
	Mpho2		MLP_49450	MLP_49460	MLP_49470	AP012204.1	
	Mafr	*Mycobacterium africanum* GM041182	MAF_19780	MAF_19790	MAF_19800	FR878060.1	*
	MbovA	*Mycobacterium bovis* AF2122/97	Mb1990	Mb1991	Mb1992	BX248340.1	*
	MbovB	*Mycobacterium bovis* BCG Pasteur 1173P2	BCG_1994	BCG_1995	BCG_1996	AM408590.1	
	MbovC	*Mycobacterium bovis* Tokyo 172	JTY_1978	JTY_1979	JTY_1980	AP010918.1	
	Mcan	*Mycobacterium canettii* CIPT 140010059	MCAN_19711	MCAN_19721	MCAN_19731	HE572590.1	*
	Mgil	*Mycobacterium gilvum* PYR-GCK	Mflv_3208	Mflv_3207	Mflv_3206	CP000656.1	
	Msme	*Mycobacterium smegmatis* MC2 155	x	MSMEG_2144	MSMEG_2143	CP000480.1	
	MtubA	*Mycobacterium tuberculosis* H37Rv	Rv1955	Rv1956	Rv1957	BX842578.1	*
	MtubB	*Mycobacterium tuberculosis* CDC1551	MT2004	MT2005	MT2006	AE000516.2	*
	MtubC	*Mycobacterium tuberculosis* KZN 1435	TBMG_02037	TBMG_02036	TBMG_02035	CP001658.1	*
	MtubD	*Mycobacterium tuberculosis* H37Ra	MRA_1964	MRA_1965	MRA_1966	CP000611.1	*
	MtubE	*Mycobacterium tuberculosis* F11	TBFG_11983	TBFG_11984	TBFG_11985	CP000717.1	*
	Pfre	*Propionibacterium freudenreichii subsp. shermanii* CIRM-BIA1	PFREUD_02980	PFREUD_02970	PFREUD_02960	FN806773.1	
	Ropa	*Rhodococcus opacus* B4	x	ROP_pROB01-01180	ROP_pROB01-01170	**AP011116.1**	
	Xcel	*Xylanimonas cellulosilytica* DSM 15894	Xcel_2923	Xcel_2922	Xcel_2921	CP001821.1	
Thermotogae	Tlet	*Thermotoga lettingae* TMO	Tlet_0094	Tlet_0093	Tlet_0092	CP000812.1	
	Tmar	*Thermotoga maritima* MSB8	TM1329	TM1330	TM1331	AE000512.1	
	Tnap	*Thermotoga naphthophila* RKU-10	Tnap_1472	Tnap_1471	Tnap_1470	CP001839.1	
	Tnea	*Thermotoga neapolitana* DSM 4359	CTN_1255	CTN_1254	CTN_1253 + 52	CP000916.1	
	Tpet	*Thermotoga petrophila* RKU-1	Tpet_1452	Tpet_1451	Tpet_1450	CP000702.1	
	Tbac1	*Thermotogales bacterium* mesG1.Ag.4.2 ctg90	ThebaDRAFT_1153	ThebaDRAFT_1154	ThebaDRAFT_1155	AEDC01000002.1	
	Tbac2		x	ThebaDRAFT_1157	ThebaDRAFT_1158	AEDC01000002.1	
Deinococcus-Thermus	Drad	*Deinococcus radiodurans* R1	x	DR_B0141	DR_B0142	**AE001826.1**	
Verrumicrobia	Oter	*Opitutus terrae* DSM 11246	x	Oter_0530	Oter_0531	CP001032.1	
Synergistete	Acol	*Aminobacterium colombiense* DSM 12261	Amico_0944	Amico_0945	Amico_0946	CP001997.1	

Table S2.

Abb.	Alien position (bp)	Score	Threshold	HGT signatures
Acal	1147000-1205500	27,013	17,294	tRNA, integrase, transposase, TA
Acol	967500-992500	38,741	19,367	tRNAs, integrase
Aehr	2615000-2625000	18,559	14,79	TA
CgluA	1775000-1997500	44,318	21,483	Transposases, integrases, tRNAs
Cmic1	3112500-3120000	18,976	13,934	
Dace	2508000-2514000	20,695	9,881	
Dalka	2092500-2100500	13,515	10,854	
Gcap	770000-790000	35,573	23,483	Transposases, phage genes, TA
Glov	3742500-3750500	14,562	13,312	Integrases, phage genes
Mafr	2197500-2212500	19,589	16,268	TA
MbovA	2185000-2209000	20,593	16,123	TA
MbovB	2200000-2215000	17,529	15,043	TA
MbovC	2182500-2207500	19,734	14,473	TA
Mcan	2242500-2250000	16,687	13,091	TA
Mgil	3402500-3412500	21,549	16,209	
Mmet	3095500-3112500	20,64	15,037	Transposases, phage genes
Mpho1	5470000-5479500	14,316	9,19	
Mpho2	5225000-5255000	26,982	9,19	Integrases, transposases, tRNA
Msme	2212500-2232500	17,852	11,478	Phage genes, transposase
MtubA	2192500-2203500	15,566	13,415	TA
MtubB	2190000-2201000	18,18	15,778	TA
MtubC	2215000-2225000	19,425	15,433	TA
MtubD	2202500-2213500	15,956	13,857	TA
MtubE	2205000-2212500	15,361	13,239	TA
Oter	630000-675000	44,881	15,171	Integrase
Pana	885000-905000	25,061	12,298	tRNA, phage genes, integrase
Pcar	925000-957500	17,491	11,921	Integrases
Pflu	4565000-4585000	32,909	11,461	PROPHAGE
Pfre	345000-371500	21,123	10,393	ARNt, transposases
Taue	3128000-3140000	58,532	14,546	TA, phage genes
Tmar	1317500-1359000	26,068	14,564	tRNA
Tnap	1452500-1470000	21,482	13,493	tRNA
Tnea	1213000-1236500	24,765	15,194	tRNA
Tpet	1430000-1445000	27,148	18,904	tRNA
Amet	Not detected			Transposase, tRNA
Cbes	Not detected			Transposases
Ckro	Not detected			Transposase
Csac	Not detected			Transposase
Nhal	Not detected			Plasmid genes
Tlet	Not detected			tRNA

TAC chaperones motif



Solitary SecB motif

Fig. S1

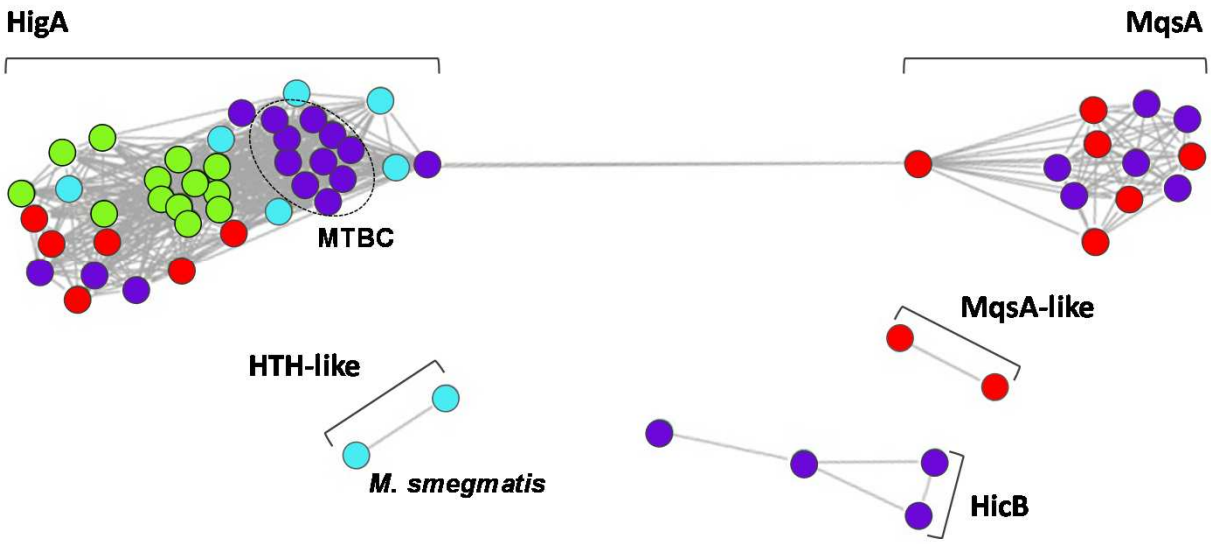


Fig. S2

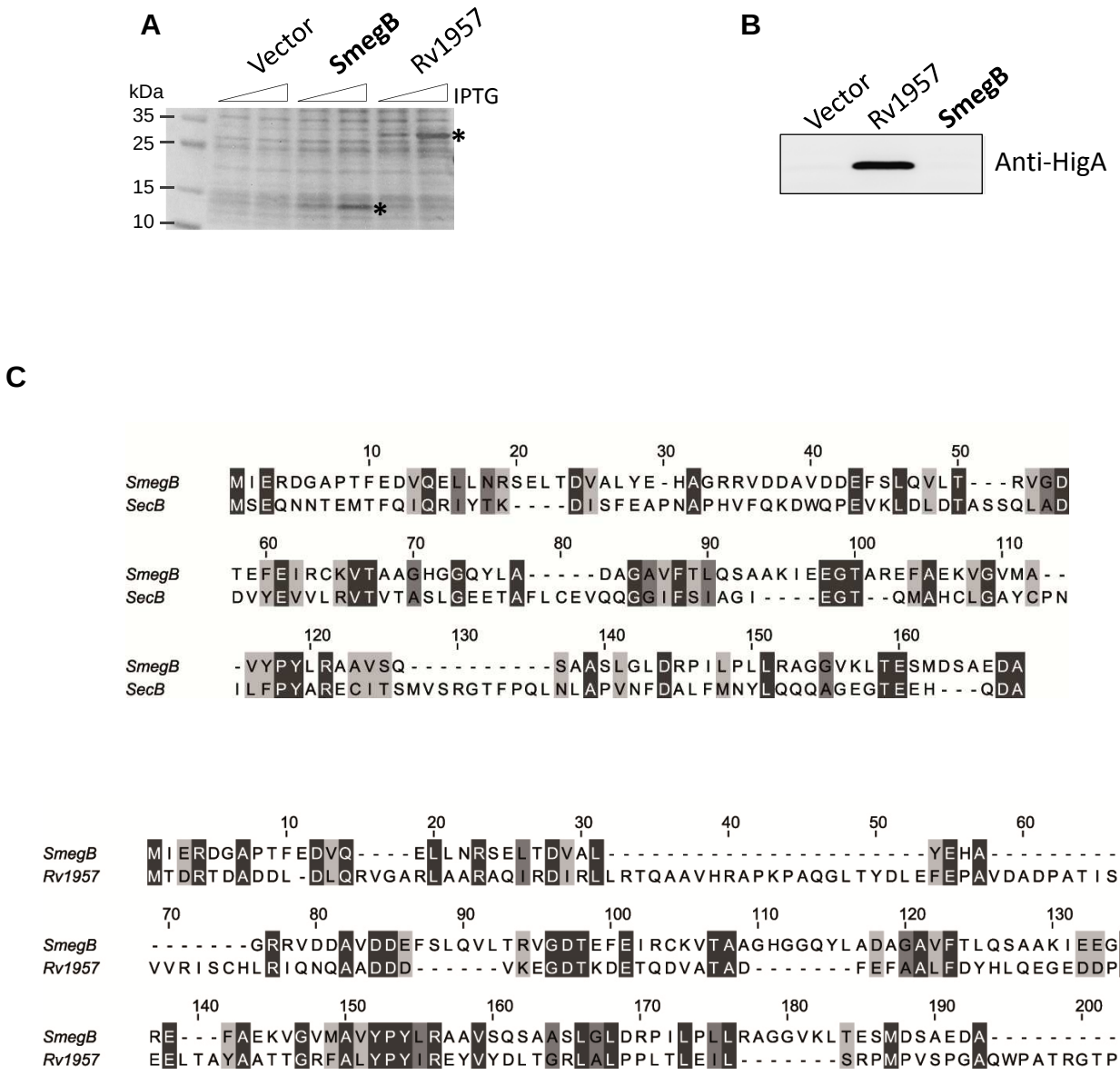

A
Vector
SmegB
Rv1957
kDa
IPTG
35
25
15
10
B
Vector
Rv1957
SmegB
Anti-HigA
C
SmegB
SecB
SmegB
SecB
SmegB
SecB
SmegB
Rv1957
SmegB
Rv1957
SmegB
Rv1957

Fig. S3