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 ΔsecB::CmR (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 pSE380ΔNcoI (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* mc²155 genomic DNA using primers smegB-for (5’-gacaattgcatatgattgagcgggacggcgcgcccac-3’) and smegB-rev (5’-gaaagcttggatcctcaggcgtcctcggccgagtccatag-3’). PCR fragment was digested with MfeI/HindIII and ligated into pSE380ΔNcoI digested with EcoRI/HindIII.

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 ΔsecB strain co-transformed with plasmids pMPMK6-HigBA and pSE380ΔNcoI, 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 ΔsecB containing pSE380ΔNcoI, 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 ΔsecB containing plasmids pMPMK6-HigBA and pSE380ΔN*col*, 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.
| Taxonomic group | Abb. | Species | Toxin | Antibiotic | Chaperone | Ref. Genome |
|-----------------|------|---------|-------|------------|-----------|-------------|
| **γ-proteobacteria** | AacI | Acidithiobacillus caldus SM-1 | x | MGG_1101 | MGG_1100 | CP002573.1 |
| | AcoI | Acinetobacter johnsonii SH046 | x | MHPREF0006_0330 | MHPREF0006_0331 | | |
| | Aehr | Alkalimicrobium ehrlichii MLHE-1 | x | Mgl_2300 | Mgl_2308 | CP000453.1 |
| | Mmet | Methylobacterium methanicum MC09 | x | Metme_2813 | Metme_2812 | | |
| | Nhal | Nitrospira halophila Nl04 | x | Nhal_3521 | Nhal_3522 | CP002738.1 |
| | Pana | Planotalea anamatis AJ13255 | x | PAJ_0167 | PAJ_0158 | AP013292.1 |
| | Flr | Pseudomonas fluorescens SBW25 | x | PFLU_0406 | PFLU_0428 | AM811764.1 |
| | VchoA | Vibrio cholerae RC385 | x | VCRCS85_0367 | VCRCS85_0369 | CP0074551 |
| | VchoB | Vibrio cholerae HE-09 | x | VCHB09_0202 | VCHB09_0204 | AFO01000011.1 |
| | Taue | Tolypolum aurensc DSM 9187 | x | Tola_2654 | Tola_2653 | CP000161.1 |
| **β-proteobacteria** | Avyl | Acinetobacter xylosidans C54 | x | MHPREF0006_0297 | MHPREF0006_0298 | ACRC0001056.1 |
| | CtaI | Carnobacterium lauisenii LM21424 | x | GLALTA_0477 | GLALTA_0478 | CUA31751.1 |
| | Geap | Gallionella capsuliformans ES-2 | x | Gil_0727 | Gil_0728 | CP002159.1 |
| | Rsol | Ralstonia solanaceae CPB2357 | x | RCPBP_mp0354 | RCPBP_mp0355 | FP8SR907.1 |
| **δ-proteobacteria** | Dalk | Desulfobacterium alkalovorans AK-01 | x | Dak_1586 | Dak_1597 | CP002122.1 |
| | Dace | Desulfohalobus acidovorans DSM 11108 | x | Desac_2259 | Desac_2260 | CP002629.1 |
| | Dret | Desulfohalobus rebmae DSM 5692 | x | Dret_0389 | Dret_0390 | CP002741.1 |
| | Glov | Gobobacter livelyi ATCC 8181 | x | Glov_3492 | Glov_3493 | CP000688.1 |
| | Gura | Gobobacter uniradicicola R14 | x | Gura_1411 | Gura_1410 | CP000698.1 |
| | Pcar | Pelobacter carbinolicus DSM 2380 | x | Pcar_0789 | Pcar_0789 | CP000142.2 |
| | Satl | Sulfuritalea acidiphila S10 | x | SYN_0016 | SYN_0015 | CP000252.1 |
| | Udes | uncultured Desulfobacterium sp. env. | x | UDD_41915 | UDD_41916 | EE666777.1 |
| **Firmicutes** | Amet | Alkaliphilus metallospecificus GEYM | x | Amet_3123 | Amet_3123 | CP000724.1 |
| | Cbes | Cloacibacterium bresii DSM 6725 | x | Cbres_2318 | Cbres_2319 | CP000391.1 |
| | Chyd | Cloacibacterium hydrothermalis DSM 18901 | x | Chhy_0257 | Chhy_0258 | CP000219.1 |
| | Ckro | Cloacibacterium koreensis DSM 18902 | x | Ckro_0234 | Ckro_0235 | CP000230.1 |
| | Coac | Cloacibacterium saccharoliicus DSM 8803 | x | Coac_0625 | Coac_0626 | CP000679.1 |
| | LplA | Lactobacillus plantarum plantarum ATCC 14917 | x | LPTST_0260 | LPTST_0261 | CP002222.1 |
| | LplB | Lactobacillus plantarum plantarum ST-III | x | LPLB_0260 | LPLB_0261 | CP002222.1 |
| **Acinetobacteria** | Cmci1 | Clavibacter michiganensis subsp. Scedopedicus ATCC33113 | x | CSM2954 | CSM2955 | AAS40904.1 |
| | Cmci2 | Clavibacter michiganensis subsp. Scedopedicus ATCC331 | x | CSM2954 | CSM2955 | AAS40904.1 |
| | Cgu | Clostridium acetobutylicum ATCC8253 | x | Cgl_1742 | Cgl_1743 | BA00036.1 |
| | Mpho1 | Microbacterium phosphovorus NM-1 | x | MPP_0450 | MPP_0451 | AP012204.1 |
| | Mpho2 | Microbacterium phosphovorus NM-2 | x | MPP_0450 | MPP_0451 | AP012204.1 |
| | MaFr | Mycobacterium africanaum G6041182 | x | MAFR_19780 | MAFR_19780 | FP8HR894.1 |
| | MbloK | Microbacterium bosri AP1222-67 | x | MBloK_1991 | MBloK_1992 | CP004340.1 |
| | MbloV | Microbacterium boxii BCG Pasteur 1175P2 | x | BCG_1994 | BCG_1995 | CP004340.1 |
| | MbloW | Microbacterium boxii BCG Pasteur 1175P2 | x | BCG_1994 | BCG_1995 | CP004340.1 |
| | Mcan | Mycobacterium canettii CPT14001059 | x | MCan_19712 | MCan_19713 | HSE72590.1 |
| | MgIH | Mycobacterium gilum PYR-G0K | x | MGH_3206 | MGH_3206 | CP000856.1 |
| | Mme | Microbacterium streptomycitum MCI-155 | x | MME_2144 | MME_2143 | CP000810.1 |
| | MtbA6 | Mycobacterium tuberculosis H37Rv | x | MtbA6_2145 | MtbA6_2143 | CP000810.1 |
| | MtbA6 | Mycobacterium tuberculosis H37Rv | x | MtbA6_2145 | MtbA6_2143 | CP000810.1 |
| | MtbA6 | Mycobacterium tuberculosis H37Rv | x | MtbA6_2145 | MtbA6_2143 | CP000810.1 |
| | MtbA6 | Mycobacterium tuberculosis H37Rv | x | MtbA6_2145 | MtbA6_2143 | CP000810.1 |
| | Pfie | Propionibacterium freudenreichii subsp. shermanii CRM BIA1 | x | PFREU_0298 | PFREU_0297 | FN8607).1 |
| | Ropa | Rhodococcus opacus BA4 | x | ROP_02980 | ROP_02981 | AP011116.1 |
| | Xcel | Xylemonas cellulolytica DSM 1598 | x | Xcel_2929 | Xcel_2930 | CP000812.1 |
| | Tett | Thermotoga tetangica TMOD | x | Tett_0057 | Tett_0058 | CP000812.1 |
| | Tmar | Thermotoga maritima MS299 | x | Tmar_0330 | Tmar_0331 | AE000512.1 |
| | Trap | Thermotoga naphthophila RUK-10 | x | Trap_0472 | Trap_0473 | CP000819.1 |
| | Tnea | Thermotaeniella neapolitana DSM 4359 | x | Tnea_1254 | Tnea_1255 | CP000819.1 |
| | Tpet | Thermotaeniella petrophila RUK-1 | x | Tpet_1451 | Tpet_1452 | CP000819.1 |
| | Tbc1 | Thermotogales bacteria mesG1 Ag-4.2 ctg90 | x | Tbc1_1153 | Tbc1_1154 | CP000819.1 |
| | Tbc2 | Thermotogales bacteria mesG1 Ag-4.2 ctg90 | x | Tbc2_1153 | Tbc2_1154 | CP000819.1 |
| | Doac | Deinococcus radionicus R1 | x | Doac_0530 | Doac_0531 | AE000826.1 |
| | Oter | Cetalius terae DSM 11246 | x | Oter_0530 | Oter_0531 | CP001032.1 |
| | Synergist | Acoc | Acinobacter colombiensis DSM 12261 | x | Acoc_0944 | Acoc_0945 | CP001997.1 |
| 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 | 14,790    | 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             |
| Maf  | 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       |
| Mme  | 2212500-2232500    | 17,852 | 11,478    | Phage genes, transposase             |
| MubA | 2192500-2203500    | 15,566 | 13,415    | TA                                  |
| MubB | 2190000-2201000    | 18,18  | 15,778    | TA                                  |
| MubC | 2215000-2225000    | 19,425 | 15,433    | TA                                  |
| MubD | 2202500-2213500    | 15,956 | 13,857    | TA                                  |
| MubE | 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 | 9250000-957500     | 17,491 | 11,921    | Integrases                          |
| Pflb | 4565000-4585000    | 32,909 | 11,461    | PROPHAGE                            |
| Pfre | 3450000-371500     | 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 | 14525000-1470000   | 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
Fig. S3