

# From *Corynebacterium glutamicum* to *Mycobacterium tuberculosis*—towards transfers of gene regulatory networks and integrated data analyses with MycoRegNet

Justina Krawczyk<sup>1,2</sup>, Thomas A. Kohl<sup>2</sup>, Alexander Goesmann<sup>1,2</sup>, Jörn Kalinowski<sup>2</sup> and Jan Baumbach<sup>3,\*</sup>

<sup>1</sup>Computational Genomics, <sup>2</sup>Center for Biotechnology, Bielefeld University, Bielefeld, Germany and <sup>3</sup>International Computer Science Institute, Berkeley, CA, USA

Received April 13, 2009; Revised May 12, 2009; Accepted May 13, 2009

## ABSTRACT

Year by year, approximately two million people die from tuberculosis, a disease caused by the bacterium *Mycobacterium tuberculosis*. There is a tremendous need for new anti-tuberculosis therapies (antitubercutic) and drugs to cope with the spread of tuberculosis. Despite many efforts to obtain a better understanding of *M. tuberculosis*' pathogenicity and its survival strategy in humans, many questions are still unresolved. Among other cellular processes in bacteria, pathogenicity is controlled by transcriptional regulation. Thus, various studies on *M. tuberculosis* concentrate on the analysis of transcriptional regulation in order to gain new insights on pathogenicity and other essential processes ensuring mycobacterial survival. We designed a bioinformatics pipeline for the reliable transfer of gene regulations between taxonomically closely related organisms that incorporates (i) a prediction of orthologous genes and (ii) the prediction of transcription factor binding sites. In total, 460 regulatory interactions were identified for *M. tuberculosis* using our comparative approach. Based on that, we designed a publicly available platform that aims to data integration, analysis, visualization and finally the reconstruction of mycobacterial transcriptional gene regulatory networks: MycoRegNet. It is a comprehensive database system and analysis platform that offers several methods for data exploration and the generation of novel hypotheses. MycoRegNet is publicly available at <http://mycoregnet.cebitec.uni-bielefeld.de>.

## INTRODUCTION

Year by year, approximately two million people die worldwide from tuberculosis (1) and one-third of the world's total population suffer from this communicable disease (<http://www.who.int>) caused by the bacterium *Mycobacterium tuberculosis*. Tuberculosis is the leading cause of death to people living with HIV and claims on average 200 000 lives every year, most of them in Africa. Persons infected with tuberculosis will not directly develop the characteristic full-blown clinical picture, but in most cases the latent form, which can progress to an active condition after years. About 10–15 people can be infected by a person with active tuberculosis a year, if she or he is left untreated (<http://www.who.int>). Although there is effective treatment to cure patients with tuberculosis, and new strategies have been developed to stop its further dissemination, its containment is still a serious problem (2). The number of multi-resistant strains not responding to standard drug treatments is increasing constantly worldwide (3,4). Consequently, there is a tremendous need for new anti-tuberculosis therapies (antitubercutic) and drugs to cope with the spread of tuberculosis. Despite many efforts to obtain a better understanding of the pathogenicity of *M. tuberculosis* and its survival strategy in humans, many questions are still unresolved. The molecular mechanisms responsible for resisting the human immune system and their activation are not perceived sufficiently so far; most notably, its ability to remain within the human host for years in a clinically latent state (5). Among other cellular processes in bacteria, pathogenicity is controlled by transcriptional regulation. Thus, various studies on *M. tuberculosis* concentrate in the analysis of transcriptional regulation in order to gain new insights on pathogenicity and other essential processes

\*To whom correspondence should be addressed. Tel: +1 510 666 2976; Fax: +1 510 666 2956; Email: [jbaumbac@icsi.berkeley.edu](mailto:jbaumbac@icsi.berkeley.edu)

ensuring mycobacterial survival. The identification and characterization of transcriptional regulation on a genome-wide level will enable a better understanding of drug metabolism in *M. tuberculosis* and facilitate the development of new antibiotics, which are urgently needed. At present, studies focus mainly on the analysis of single regulons, or distinct subunits of the complex transcriptional regulatory network of *M. tuberculosis* [see e.g. (3,5,6)]. Bioinformatics platforms for data storage and public access of transcriptional regulation exist for *M. tuberculosis*, similar to other organisms such as *Escherichia coli* [RegulonDB (7)] or *Corynebacterium glutamicum* [CoryneRegNet (8)]. MtbRegList (9) and MTBreg (<http://www.doe-mbi.ucla.edu/services/mtbreg>) offer information relevant to regulatory interactions in *M. tuberculosis* H37Rv (MT) accumulated from literature or attained from computational predictions. While MtbRegList contains predicted and characterized regulatory DNA motifs cross-referenced with transcription factors (TFs), MTBreg combines a collection of conditionally regulated proteins together with information about selected TFs. However, both systems are designed as data repositories and only provide nonsatisfying bioinformatics support necessary for transcriptional gene regulatory network visualization, analysis, and reconstruction. Recently, the TB database has become available. This integrated online platform for tuberculosis research combines the annotated genome and expression data with a suite of bioinformatic tools for data analysis (10). The scope of TB database is placed on investigating and providing expression data, while little support is given for the reconstruction of regulatory networks based on these findings. Hence, there is currently no online platform or database system available, which aims to an appropriate data handling and analysis of transcriptional regulation in *M. tuberculosis* on a genome-wide level.

Here, we introduce MycoRegNet, an online accessible, user-friendly platform dedicated to the biomedical researcher, who is interested in the regulation of gene expression in the human pathogen *M. tuberculosis*. MycoRegNet is online available at <http://mycoregnet.cybitec.uni-bielefeld.de>.

The first idea of our approach is based on the assumption that orthologous TFs tend to regulate the expression of orthologous target genes for taxonomically closely related species (11–13). *Corynebacterium glutamicum* and *M. tuberculosis* are taxonomically classified into the sub-order Corynebacterineae of the Actinobacteria phylum and are thus taxonomically closely related (14). Hence, the industrially important amino acid producer *C. glutamicum* has been successfully applied as model organism, e.g. for investigating cell envelope synthesis of *M. tuberculosis* (15–17). We therefore started with the well-examined regulatory network of *C. glutamicum* ATCC 13032 (CG) (18), which is stored in the corynebacterial reference database CoryneRegNet (8). Our comparative genomics approach aims for a reliable transfer of known regulatory interactions from CG to MT. Instead of relying exclusively on the detection of orthologous genes, we consider further evidence by means of an integrated TF binding site (TFBS) prediction. The resulting data were

subsequently stored in an online platform designed for the visualization and analysis of the deduced transcriptional regulatory network, which enables the execution of bioinformatics tools for further hypotheses generation: MycoRegNet. The remainder of this article is structured as follows: we first describe the workflow used for the transfer of *C. glutamicum* data to *M. tuberculosis* in detail. The design of MycoRegNet is briefly introduced afterwards. It aims to overcome typical data integration problems and to supply online visualization and hypotheses generation tools. In the last section, we illustrate and discuss these functionalities. We finally conclude that MycoRegNet is an appropriate reference database and platform for gene regulatory network analysis of *M. tuberculosis*.

## MATERIALS AND METHODS

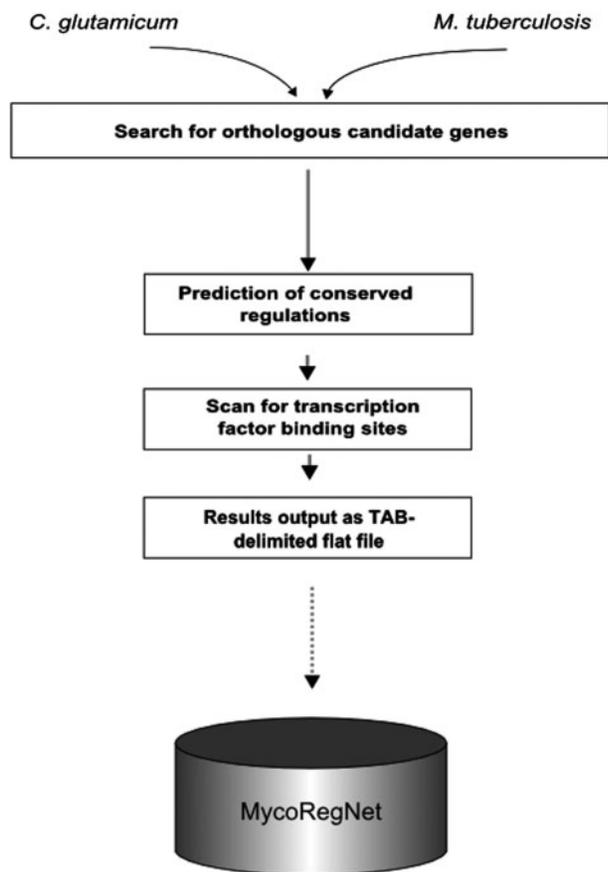
The network reconstruction pipeline mainly consists of the detection of (i) conserved genes between CG and MT and (ii) binding sites upstream the conserved genes in MT. Based on the corresponding results, a list of putative gene regulatory interactions in MT is generated and imported into the MycoRegNet database back-end (see Figure 1 for a graphical overview of the workflow).

### Detection of orthologous genes

Generally, the detection of orthologous genes is not straightforward, since analysis can be perturbed by factors like paralogs or sequence divergences in the genomes of interest. To reduce such effects, we searched for orthologous genes by performing bidirectional BLASTP (19) searches on the corresponding protein sequences. Therefore, we scanned the CG genome for sequence similarities with the MT genome and vice versa, performing BLASTP with an *E*-value cut-off of  $10^{-4}$  in both directions. As a result, we obtained amino acid sequence pairs, so called bidirectional best hits (BBHs), representing the reciprocal best alignments of respective protein sequences. Thus, identified BBHs were considered to be putative orthologous proteins in CG and MT, which in turn indicates the respective genes to be regulated in both bacteria by orthologous TFs.

### Transfer of regulatory interactions

Based on the previously identified BBHs, regulatory interactions characterized in CG were transferred to MT. We utilized the comprehensive data on transcriptional regulation in CG collected in the corynebacterial reference database CoryneRegNet (8), which contains 806 regulatory interactions of 72 TFs and 544 regulated target genes on CG (status: January 2009). For each regulatory interaction taken from CG, both the gene encoding the TF and the target gene were compared to the list of predicted orthologs in the MT genome. Only if both, the TF as well as its target gene, were identified as BBH, the regulatory interaction was transferred from CG to the orthologous counterparts in MT and was considered as a candidate transcriptional regulation in MT. Furthermore, we assume the regulatory role of the TF (activation or



**Figure 1.** Diagram of the prediction pipeline. The diagram shows the main steps performed during transfer of gene regulations from *C. glutamicum* to *M. tuberculosis*. Starting with an orthology detection, the next step was a prediction of conserved regulations. Based on that, a TFBSs prediction provided further evidence. Finally, the results can be exported as TAB-delimited files and imported into the MycoRegNet data repository.

repression) to be conserved as well, including known autoregulations.

#### Further evidence through conserved TFBSs

In the last step of our regulatory network prediction pipeline, we add further evidence to the orthology-based approach introduced above by combining the preliminary results with the prediction of TFBSs. Therefore, all known binding sites of characterized TFs of CG with potential orthologs in MT were utilized to create appropriate motif profiles. TF binding motifs were modeled as so called position weight matrices (PWMs), the most widely used model for that purpose. However, we applied only PWMs of corynebacterial TFs deduced from more than 20 binding sites, i.e. of the TFs GlxR, RamB, AmtR, DtxR and LexA. To detect instances of the respective motifs in MT, we employed the TFBS matching tool PoSSuMsearch (20) and scanned 580-bp long, noncoding DNA sequences upstream all genes and operons, which have been detected as potential orthologs to target genes of the respective TF. The upstream sequences ranged from

+ 20 bp relative to the transcription start. In our initial approach, we performed a restrictive search by setting the  $P$ -value threshold to  $10^{-5}$ . Due to the low number of detected binding sites in the first PoSSuMsearch runs, we decided to decrease the  $P$ -value threshold since the set  $P$ -value might be chosen too restrictive for our TFBS predictions. To determine a new  $P$ -value, we considered  $P$ -values of binding site matches of the PWM for GlxR upstream 26 target genes as marking value, where the binding of the GlxR ortholog Rv3676 in MT was experimentally verified (21–24). For  $P$ -value definition, we chose the binding site that match upstream one of these genes with the worst  $P$ -value. Thus, we finally set the  $P$ -value to  $<10^{-2}$  and defined for each target gene/operon in MT the TFBS match with the lowest  $P$ -value as prediction for the respective binding site.

Taken together, the outcome of the above introduced workflow is a list of transcriptional regulations for MT where (i) the TF is conserved, and (ii) the target gene is conserved between CG and MT as well, and additionally (iii) a binding site is predicted, if the target gene/operon is controlled by one of the five TFs where a TFBS search was performed for. Hence, the resulting predictions present most likely regulatory interactions in MT due to the taxonomically close relation between CG and MT. This is the data we aim to integrate into the MycoRegNet platform together with validated knowledge we have from (5,25–34).

#### Data integration with the MycoRegNet platform

Based on our experiences with CoryneRegNet, we designed MycoRegNet in a very similar way: as an ontology-based data warehouse for mycobacterial TFs and regulatory networks. We set it up as a sister project of CoryneRegNet to store, analyze and visualize the regulatory interactions in *M. tuberculosis* that are derived from the above introduced prediction pipeline. MycoRegNet is composed of two main parts: (i) A web front-end running on an Apache HTTP web server that manages user-database interactions as well as the execution of further online bioinformatics computations. (ii) The back-end consists of data preprocessing tools and a MySQL database that stores all data corresponding to the deduced and ontologically restructured mycobacterial gene regulatory interactions. This process comprises the integration of transcriptional regulations, the complete genome sequence of MT along with the genome annotation as stored in the GenBank database (NCBI) (35), operon predictions available from the Virtual Institute of Microbial Stress and Survival (VIMSS) (36), precalculated PWMs and other preprocessed data necessary for subsequent online TFBS detections, and stimulons derived from literature (25–31). The import and conversion software is implemented in Java, while the web pages generated at front-end level are developed in PHP. An embedded Java applet realizes the visualization of gene regulatory networks from the included data. A SOAP-based web Service (37) client/server system implemented by means of NuSOAP enables a bidirectional interconnection with GenDB (38) and EMMA (39). The server is open access and provides

well-structured data access via the SOAP interface to any other bioinformatics client. GenDB is an open source system for the annotation of prokaryotic genomes, while EMMA is a web-based application for the storage and analysis of transcriptomics data from microarrays. By means of the clients for GenDB and EMMA, data integrated in MycoRegNet is supplemented with up-to-date information on the genome annotation of MT (GenDB) and gene expression data preanalyzed with EMMA. To give one example, the Web Service client for GenDB facilitates the mapping of all genes controlled by a certain regulator to KEGG pathways (40) in order to provide an overview on the general nature of a TF of interest. Furthermore, the automatic annotation pipeline of GenDB can be used to regularly update gene function assignments.

## RESULTS AND DISCUSSION

Here, we first summarize the database content. Subsequently, we present and discuss the benefits of MycoRegNet from the end-user perspective. We first describe the web interface with special attention to the TFBS prediction feature and the network visualization and analysis capability. We briefly describe the Web Service access afterwards and finally demonstrate the platforms' visualization functionality by means of an application example.

### The database content

By using the above described transfer pipeline for regulatory interactions, we identified 1012 of 3991 proteins from MT as putative orthologs to proteins from CG. Based on the respective set of genes coding for the orthologous proteins, we detected 226 of 806 regulatory interactions from CG as likely conserved in MT (Table 1). Our initial findings reveal 24 partial conserved regulons affecting processes of the carbohydrate metabolism, cellular program, macroelement and metal homeostasis, SOS and stress response, specific biosynthesis as well as processes governed by sigma factors. By setting the *P*-value threshold to 102, we could put further evidence to 129 target genes 40 by predicting binding sites upstream the respective target genes/operons regulated by the TF orthologs of GlxR, RamB, AmtR, DtxR and LexA in MT (Table 2). All in all, we obtained a set of regulatory interactions which is based on good evidence. The database content comprises 618 regulatory interactions for 515 target genes regulated by 26 TFs. Several gene expression experiments are also directly stored within MycoRegNet's database back-end (data not shown). We also integrated genome annotation data of *M. tuberculosis* CDC1551 for future investigations concerning transcriptional regulation in another ecotype of *M. tuberculosis*.

### The user interface

As for other online databases, MycoRegNet's web interface provides the three major capabilities: browsing the database content, searching by specifying filter criteria and basic visualization possibilities. Furthermore, the

**Table 1.** Putative gene regulations of CG in MT

TF	Target genes
Carbohydrate metabolism	
Rv0465c	Carbohydrate metabolism Rv0211 ( <i>pckA</i> ), Rv0247c (-), Rv0363c ( <i>fba</i> ), Rv0408 ( <i>pta</i> ) Rv0409 ( <i>ackA</i> ), Rv0465c (-), Rv0467 ( <i>icl</i> ), Rv0896 ( <i>gltA2</i> ) Rv0904c ( <i>accD3</i> ), Rv0951 ( <i>sucC</i> ), Rv0952 ( <i>sucD</i> ), Rv1475c ( <i>acn</i> ) Rv1837c ( <i>glcB</i> ), Rv1862 ( <i>adhA</i> ), Rv2193 ( <i>ctaE</i> ), Rv2241 ( <i>aceE</i> ) Rv2332 ( <i>mez</i> ), Rv2967c ( <i>pca</i> ), Rv3318 ( <i>sdhA</i> ) Cell division and septation Rv1009 ( <i>rpjB</i> ) Specific biosynthesis pathways Rv0884c ( <i>serC</i> ), Rv1010 ( <i>ksgA</i> ), Rv1011 ( <i>ispE</i> ), Rv1379 ( <i>pyrR</i> ) Rv1380 ( <i>pyrB</i> ), Rv1381 ( <i>pyrC</i> )
Rv0792c	Carbohydrate metabolism Rv0753c ( <i>mmsA</i> )
Rv1719	Carbohydrate metabolism Rv0554 ( <i>bpoC</i> ), Rv1074c ( <i>fadA3</i> ), Rv1719 (-), Rv2503c ( <i>scoB</i> ), Rv2504c ( <i>scoA</i> )
Rv3676	Carbohydrate metabolism Rv0211 ( <i>pckA</i> ), Rv0247c (-), Rv0400c ( <i>fadE7</i> ), Rv0465c (-) Rv0467 ( <i>icl</i> ), Rv0896 ( <i>gltA2</i> ), Rv0904c ( <i>accD3</i> ), Rv0951 ( <i>sucC</i> ) Rv0952 ( <i>sucD</i> ), Rv1098c ( <i>fumC</i> ), Rv1130 (-), Rv1161 ( <i>narG</i> ) Rv1162 ( <i>narH</i> ), Rv1163 ( <i>narJ</i> ), Rv1436 ( <i>gap</i> ), Rv1437 ( <i>pgk</i> ) Rv1438 ( <i>tpi</i> ), Rv1475c ( <i>acn</i> ), Rv1837c ( <i>glcB</i> ), Rv1854c ( <i>ndh</i> ) Rv1862 ( <i>adhA</i> ), Rv1872c ( <i>lldD2</i> ), Rv2029c ( <i>pfkB</i> ), Rv2193 ( <i>ctaE</i> ) Rv2194 ( <i>qcrC</i> ), Rv2195 ( <i>qcrA</i> ), Rv2196 ( <i>qcrB</i> ), Rv2200c ( <i>ctaC</i> ) Rv2524c ( <i>fas</i> ), Rv2967c ( <i>pca</i> ), Rv3010c ( <i>pfkA</i> ), Rv3043c ( <i>ctaD</i> ) Rv3279c ( <i>birA</i> ), Rv3280 ( <i>accD5</i> ), Rv3318 ( <i>sdhA</i> ), Rv3548c (-), Rv3676 (-) Cell division and septation Rv1009 ( <i>rpjB</i> ), Rv2145c ( <i>wag31</i> ), Rv2201 ( <i>asnB</i> ) Macroelement and metal homeostasis Rv0820 ( <i>phoT</i> ), Rv0928 ( <i>pstS3</i> ), Rv0929 ( <i>pstC2</i> ), Rv0930 ( <i>pstA1</i> ), Rv2220 ( <i>glnA1</i> ) Rv2832c ( <i>ugpC</i> ), Rv2833c ( <i>ugpB</i> ), Rv2834c ( <i>ugpE</i> ), Rv2835c ( <i>ugpA</i> ), Rv2918c ( <i>glnD</i> ) Rv2919c ( <i>glnB</i> ), Rv2920c ( <i>ami</i> ), Rv3859c ( <i>gltB</i> ) SOS and stress response Rv0867c ( <i>rpjA</i> ), Rv3048c ( <i>nrdF2</i> ), Rv3217c(-), Rv3219 ( <i>whiB1</i> ), Rv3681c ( <i>whiB4</i> ) Specific biosynthesis pathways Rv0884c ( <i>serC</i> ), Rv1010 ( <i>ksgA</i> ), Rv1011 ( <i>ispE</i> ), Rv1092c ( <i>coaA</i> ) Rv3001c ( <i>ilvC</i> ) Rv3002c ( <i>ilvN</i> ), Rv3003c ( <i>ilvB1</i> )
Cellular Program	
RelA	Sigma factor module Rv1221 ( <i>sigE</i> ), Rv2710 ( <i>sigB</i> ), Rv3221A (-), Rv3911 ( <i>sigM</i> ) SOS and stress response Rv2720 ( <i>lexA</i> )
Macroelement and metal homeostasis	
Rv0485	Macroelement and metal homeostasis Rv0132c ( <i>fgd2</i> )
PhoP	Macroelement and metal homeostasis Rv0545c ( <i>pitA</i> ), Rv0757 ( <i>phoP</i> ), Rv0758 ( <i>phoR</i> ), Rv0820 ( <i>phoT</i> ) Rv0928 ( <i>pstS3</i> ), Rv0929 ( <i>pstC2</i> ), Rv0930 ( <i>pstA1</i> ), Rv1095 ( <i>phoH2</i> )

(continued)

Table 1. Continued

TF	Target genes
	<i>Rv2832c (ugpC)</i> , <i>Rv2833c (ugpB)</i> , <i>Rv2834c (ugpE)</i> , <i>Rv2835c (ugpA)</i>
Rv0827c	Macroelement and metal homeostasis <i>Rv0827c (-)</i>
Rv1994c	Macroelement and metal homeostasis <i>Rv1994c (-)</i>
IdeR	Carbohydrate metabolism <i>Rv0247c (-)</i> , <i>Rv3318 (sdhA)</i> Macroelement and metal homeostasis <i>Rv0827c (-)</i> , <i>Rv0844c (narL)</i> , <i>Rv1285 (cysD)</i> , <i>Rv1286 (cysN)</i> <i>Rv2391 (nirA)</i> , <i>Rv2392 (cysH)</i> , <i>Rv2393 (-)</i> , <i>Rv2895c (viuB)</i> <i>Rv3044 (fecB)</i> , <i>Rv3841 (bfrB)</i>
Rv3160c	Macroelement and metal homeostasis <i>Rv1848 (ureA)</i> , <i>Rv1849 (ureB)</i> , <i>Rv1850 (ureC)</i> , <i>Rv1852 (ureG)</i> <i>Rv2220 (glnA1)</i> , <i>Rv2918c (glnD)</i> , <i>Rv2919c (glnB)</i> , <i>Rv2920c (amt)</i> <i>Rv3664c (dppC)</i> , <i>Rv3665c (dppB)</i> , <i>Rv3666c (dppA)</i> , <i>Rv3859c (glbB)</i>
Rv3173c	Macroelement and metal homeostasis <i>Rv0132c (fgd2)</i> , <i>Rv0485 (-)</i> , <i>Rv1079 (metB)</i> , <i>Rv1133c (metE)</i> <i>Rv1175c (fadH)</i> , <i>Rv1285 (cysD)</i> , <i>Rv1286 (cysN)</i> , <i>Rv1294 (thrA)</i> <i>Rv1296 (thrB)</i> , <i>Rv1392 (metK)</i> , <i>Rv2124c (metH)</i> , <i>Rv2334 (cysKI)</i> <i>Rv2391 (nirA)</i> , <i>Rv2392 (cysH)</i> , <i>Rv2393 (-)</i> , <i>Rv3025c (iscS)</i> <i>Rv3028c (fixB)</i> , <i>Rv3029c (fixA)</i> , <i>Rv3173c (-)</i> , <i>Rv3340 (metC)</i> <i>Rv3341 (metA)</i>
Sigma factor module	
SigB	Carbohydrate metabolism <i>Rv0363c (fba)</i> , <i>Rv1023 (eno)</i> , <i>Rv1098c (fumC)</i> , <i>Rv1436 (gap)</i> <i>Rv1437 (pgk)</i> , <i>Rv1438 (tpi)</i> , <i>Rv3010c (pfkA)</i> SOS and stress response <i>Rv3132c (devS)</i> Specific biosynthesis pathways <i>Rv2210c (ilvE)</i>
SigM	SOS and stress response <i>Rv0384c (clpB)</i> , <i>Rv1464 (csd)</i> , <i>Rv1465</i> , <i>Rv1471 (trxBI)</i> <i>Rv3418c (groES)</i> , <i>Rv3913 (trxB2)</i> , <i>Rv3914 (trxC)</i>
SOS and stress response	
HspR	SOS and stress response <i>Rv0350 (dnaK)</i> , <i>Rv0351 (grpE)</i> , <i>Rv0352 (dnaJ1)</i> , <i>Rv0353 (hspR)</i> <i>Rv0384c (clpB)</i> , <i>Rv0440 (groEL)</i> , <i>Rv2745c</i>
HrcA	SOS and stress response <i>Rv0440 (groEL)</i> , <i>Rv3418c (groES)</i>
LexA	Cell division and septation <i>Rv2748c (ftsK)</i> SOS and stress response <i>Rv1235 (lpqY)</i> , <i>Rv1638 (uvrA)</i> , <i>Rv1696 (recN)</i> , <i>Rv2592c (ruvB)</i> <i>Rv2593c (ruvA)</i> , <i>Rv2594c (ruvC)</i> , <i>Rv2720 (lexA)</i> , <i>Rv2736c (recX)</i> <i>Rv2737c (recA)</i> , <i>Rv3370c (dnaE2)</i> , <i>Rv3395c</i> , <i>Rv3585 (radA)</i>
Rv2745c	SOS and stress response <i>Rv0782 (ptrBb)</i> , <i>Rv2460c (clpP2)</i> , <i>Rv2461c (clpP1)</i> , <i>Rv2725c (hflX)</i> <i>Rv3596c (clpC1)</i> , <i>Rv3715c (recR)</i> , <i>Rv3716c</i>
WhiB1	SOS and stress response <i>Rv3913 (trxB2)</i> , <i>Rv3914 (trxC)</i>
MtrA	SOS and stress response

(continued)

Table 1. Continued

TF	Target genes
CspA	<i>Rv0917 (betP)</i> , <i>Rv3476c (kgtP)</i> Carbohydrate metabolism <i>Rv1837c (glcB)</i>
Specific biosynthesis pathways	
PyrR	Specific biosynthesis pathways <i>Rv1379 (pyrR)</i> , <i>Rv1380 (pyrB)</i> , <i>Rv1381 (pyrC)</i> , <i>Rv2883c (pyrH)</i>
ArgR	Specific biosynthesis pathways <i>Rv1383 (carA)</i> , <i>Rv1384 (carB)</i> , <i>Rv1652 (argC)</i> , <i>Rv1653 (argJ)</i> <i>Rv1654 (argB)</i> , <i>Rv1655 (argD)</i> , <i>Rv1656 (argF)</i> , <i>Rv1657 (argR)</i> <i>Rv0488</i>

Putative gene regulations of CG in MT, predicted *in silico* by using the introduced MycoRegNet pipeline

front-end offers the execution of computational features. At the main page (Figure 2), one has the option to search or to browse the database content. The user may browse the data repository by clicking on an ecotype name of interest and is provided with an overview on the selected organism. Alternatively, using one of the provided options within the search form, the database can be searched for specific gene/protein identifiers, gene/protein names, regulator types or functional modules. The search results are presented in tabular form, listing all relevant information for subsequent investigation. Furthermore, the following built-in features can be accessed from the main page, directly: TFBSScan [for TFBS predictions; see below] and COMA (to check for contradictions within microarray gene expression studies, given the regulatory network stored in the database; refer to (8) for more details]. Detailed information on the results can be obtained via respective links at the result page. By selecting a particular gene, the corresponding gene details page is invoked. It presents a detailed overview of all available data attached to the gene of interest. Besides general information about the gene/protein (position in the genome, nucleotide sequence, etc.), it comprises a graphical representation of the genomic context, regulated target genes (if encoding a TF) including the TFBSs, etc., and stimulons that initiate a differential gene expression level. The integrated Web Service client for GenDB maintains the representation of up-to-date gene annotation data. General information (description, comments, an assigned function, etc.) is listed as well as the EC numbers for enzymes, and links to COG (41) and GO (42). Additionally, all target genes of a TF of interest, are linked to KEGG pathways and a list of regulated pathways is displayed.

**TFBS prediction.** With the integrated PoSSuMsearch software, MycoRegNet provides a statistically sound tool for the prediction of TFBSs based on PWMs, which have been precalculated during data import. To our knowledge, PoSSuMsearch is the only TFBSs profiling tool that offers exact *P*-value calculations and at the same time provides reasonable response times on

**Table 2.** Detected binding sites upstream transferred target genes of CG in MT

TF	Gene ID	Gene name	Operon	Binding motif
Rv0465c	Rv0211 <sup>a</sup>	<i>pckA</i>		ATAACTACGCAGG
	Rv0249c	–	Rv0249c-Rv0248c-Rv0247c <sup>a</sup>	AGTAGTTCGCGAT
	Rv0363c <sup>a</sup>	<i>fba</i>	–	CGTACTTCTCAAA
	Rv0407	<i>pta</i>	Rv0407-Rv0408 <sup>a</sup> -Rv0409 <sup>a</sup>	CGTGCTGTGCTCA
	Rv0465c	–	Rv0465c <sup>a</sup> -Rv0464c	CTAACTCTGCGAA
	Rv0467 <sup>a</sup>	<i>icl</i>	–	CAAAATTTGCAAA
	Rv0884c <sup>a</sup>	<i>serC</i>	–	ATGGCATGGCCGA
	Rv0896 <sup>a</sup>	<i>gltA2</i>	–	TGAGCAGATCACT
	Rv0904c <sup>a</sup>	<i>accD3</i>	–	ATTGCATGGCAAG
	Rv0951	<i>sucC</i>	Rv0951 <sup>a</sup> -Rv0952 <sup>a</sup>	AGTGCTAAGCCGT
	Rv1009	<i>rpfB</i>	Rv1009 <sup>a</sup> -Rv1010 <sup>a</sup> -Rv1011 <sup>a</sup>	TCTACTTACCAAA
	Rv1379	<i>pyrR</i>	Rv1379 <sup>a</sup> -Rv1380 <sup>a</sup> -Rv1381 <sup>a</sup> -Rv1382-Rv1383- Rv1384-Rv1385	AGTGCTACGCTGC
	Rv1475c	<i>acn</i>	Rv1475c <sup>a</sup> -Rv1474c	ACTGCTAGGCTGA
	Rv1837c <sup>a</sup>	<i>glcB</i>	–	TAGGCTGAGCAAT
	Rv1862 <sup>a</sup>	<i>adhA</i>	–	TGTGCTGGGCTAA
	Rv2193	<i>ctaE</i>	Rv2193 <sup>a</sup> -Rv2194-Rv2195-Rv2196	ACTACAAAGCGTC
	Rv2241	<i>aceE</i>	Rv2241 <sup>a</sup> -Rv2242	CAAACAGCGCAAG
	Rv2332 <sup>a</sup>	<i>mez</i>	–	TGCGCTCTGCGAA
	Rv2967c <sup>a</sup>	<i>pca</i>	–	CATGCAATGTCAA
	Rv3316	<i>sdhA</i>	Rv3316-Rv3317-Rv3318 <sup>a</sup> -Rv3319	GTTGCATTGCCCC
	IdeR	Rv0249c	–	Rv0249c-Rv0248c-Rv0247c <sup>a</sup>
Rv0827c <sup>a</sup>		–	–	CTATGGATCGCTGTACTAC
Rv0844c <sup>a</sup>		<i>narL</i>	–	CGACGAGCAGCTAAACTCA
Rv1285		<i>cysD</i>	Rv1285 <sup>a</sup> -Rv1286 <sup>a</sup>	GAGGGCGAGGCACACGTCA
Rv2391		<i>nirA</i>	Rv2391 <sup>a</sup> -Rv2392 <sup>a</sup> -Rv2393 <sup>a</sup>	TCAGGTGCGCGTCTCCAG
Rv2895c <sup>a</sup>		<i>viuB</i>	–	TAAGCGAAGCCGAACGCCA
Rv3044 <sup>a</sup>		<i>fecB</i>	–	GTAGACCAGGCTCCCTTG
Rv3316		<i>sdhA</i>	Rv3316-Rv3317-Rv3318 <sup>a</sup> -Rv3319	CTAAGAAAAGCCAGCCTAA
Rv3841 <sup>a</sup>		<i>bfrB</i>	–	CTAGGAAAAGCCTTTCCTGA
LexA		Rv1235	<i>lpqY</i>	Rv1235 <sup>a</sup> -Rv1236-Rv1237-Rv1238
	Rv1638 <sup>a</sup>	<i>uvrA</i>	–	TCGAATGTCAAGCTCGC
	Rv1696 <sup>a</sup>	<i>recN</i>	–	–
	Rv2594c	<i>ruvC</i>	Rv2594c <sup>a</sup> -Rv2593c <sup>a</sup> -Rv2592c <sup>a</sup>	TCGAACGATTGTTCCGG
	Rv2720 <sup>a</sup>	<i>lexA</i>	–	TCGAACACATGTTTGA
	Rv2737c	<i>recA</i>	Rv2737c <sup>a</sup> -Rv2736c <sup>a</sup>	TCGAACAGGTGTTCCGG
	Rv2748c <sup>a</sup>	<i>ftsK</i>	–	CCGACCAGGTGCTCGC
	Rv3370c <sup>a</sup>	<i>dnaE2</i>	–	TCGAACAATTGTTCCGA
	Rv3395c	–	Rv3395c <sup>a</sup> -Rv3394c	TCGAACATATTTCCGA
	Rv3160c	Rv1848	<i>ureA</i>	Rv1848 <sup>a</sup> -Rv1849-Rv1850 <sup>a</sup> -Rv1851-Rv1852 <sup>a</sup> - Rv1853
Rv2220 <sup>a</sup>		<i>glnA1</i>	–	CAACACGGGGTTGACTGACGGGCAATAT
Rv2920c		<i>amt</i>	Rv2920c <sup>a</sup> -Rv2919c <sup>a</sup> -Rv2918c <sup>a</sup>	AAGTTTTACGTTAATCCTGATGAAACAT
Rv3666c		<i>dppA</i>	Rv3666c <sup>a</sup> -Rv3665c <sup>a</sup> -Rv3664c <sup>a</sup> -Rv3663c- Rv3662c	GTGGTAGCTAACGGTCACCGGCGAGTGT
Rv3676	Rv3859c	<i>gltB</i>	Rv3859c <sup>a</sup> -Rv3858c	CGCTTGACGGACAGCCTATCGACAAGAC
	Rv0211 <sup>a</sup>	<i>pckA</i>		TGTGAGCAGGCTTATA
	Rv0249c	–	Rv0249c-Rv0248c-Rv0247c <sup>a</sup>	TGTGATCTGTAACACC
	Rv0400c <sup>a</sup>	<i>fadE7</i>	–	AGTGATGAGCACCCCG
	Rv0465c	–	Rv0465c <sup>a</sup> -Rv0464c	TTTGTGAGGCTCACG
	Rv0467 <sup>a</sup>	<i>ivl</i>	–	TGTACAACGCTCACA
	Rv0820 <sup>a</sup>	<i>phoT</i>	–	GGTGGTGATCCGCACC
	Rv0867c <sup>a</sup>	<i>rpfA</i>	–	TGTGACATTACCCACA
	Rv0884c <sup>a</sup>	<i>serC</i>	–	TGTGAGCTTGTTCACA
	Rv0896 <sup>a</sup>	<i>gltA2</i>	–	GGCGTTGAACATCAC
	Rv0904c <sup>a</sup>	<i>accD3</i>	–	CGTGAGTCGTATCACG
	Rv0928	<i>pstS3</i>	Rv0928 <sup>a</sup> -Rv0929 <sup>a</sup> -Rv0930 <sup>a</sup>	ACTGAATTGAACTCA
	Rv0951	<i>sucC</i>	Rv0951 <sup>a</sup> -Rv0952 <sup>a</sup>	TGTGAGTTGGATCACG
	Rv1009	<i>rpfB</i>	Rv1009 <sup>a</sup> -Rv1010 <sup>a</sup> -Rv1011 <sup>a</sup>	GGTGCCGCTATCACG
	Rv1092c <sup>a</sup>	<i>coaA</i>	–	TGCCACGTAGGTACCG
	Rv1099c	<i>fumC</i>	Rv1099c-Rv1098c <sup>a</sup> -Rv1097c	–
	Rv1130	–	Rv1130 <sup>a</sup> -Rv1131	TGTGGATAAGTCCAGG
	Rv1161	<i>narG</i>	Rv1161 <sup>a</sup> -Rv1162 <sup>a</sup> -Rv1163 <sup>a</sup> -Rv1164-Rv1165- Rv1166	TGCGTTGAACGGCACG
	Rv1436	<i>gap</i>	Rv1436 <sup>a</sup> -Rv1437 <sup>a</sup> -Rv1438 <sup>a</sup>	GGTTGTTAGCCAACA
Rv1475c	<i>acn</i>	Rv1475c <sup>a</sup> -Rv1474c	TGTAAGTCCGACATA	
Rv1837c <sup>a</sup>	<i>glcB</i>	–	AGGGATGCACTACACA	

(continued)

Table 2. Continued

TF	Gene ID	Gene name	Operon	Binding motif
	Rv1854c <sup>a</sup>	<i>ndh</i>	–	TGTGGCTGATGACACA
	Rv1862 <sup>a</sup>	<i>adhA</i>	–	CGTGGGGCGCCACACA
	Rv1872c <sup>a</sup>	<i>lldD2</i>	–	GATGCCGTAGCGCACT
	Rv2029c	<i>pfkB</i>	Rv2029c <sup>a</sup> -Rv2028c-Rv2027c-Rv2026c	GGTGACGAGTCGCGCA
	Rv2145c	<i>wag31</i>	–	CGTGACTGGCGTCCCA
	Rv2193	<i>ctaE</i>	Rv2193 <sup>a</sup> -Rv2194 <sup>a</sup> -Rv2195 <sup>a</sup> -Rv2196 <sup>a</sup>	GGTGGATAGGTTCCACC
	Rv2200c	<i>ctaC</i>	Rv2200c <sup>a</sup> -Rv2199c	TGTGATACAGGAGCGC
	Rv2201	<i>asnB</i>	–	GCTGTCTGAAGACCACG
	Rv2220 <sup>a</sup>	<i>glnA1</i>	–	TGTGACGAAAAGACG
	Rv2524c <sup>a</sup>	<i>fus</i>	–	CGTTACCCACGCACACG
	Rv2835c	<i>ugpA</i>	Rv2835c <sup>a</sup> -Rv2834c <sup>a</sup> -Rv2833c <sup>a</sup> -Rv2832c <sup>a</sup>	GGTGATGCCGGGACACG
	Rv2920c	<i>am</i>	Rv2920c <sup>a</sup> -Rv2919c <sup>a</sup> -Rv2918c <sup>a</sup>	AGTGGACCAATTCCCC
	Rv2967c <sup>a</sup>	<i>pca</i>	–	CGTGGTGGTGGTCACC
	Rv3003c <sup>a</sup>	<i>ilvB1</i>	Rv3003c <sup>a</sup> -Rv3002c <sup>a</sup> -Rv3001c <sup>a</sup>	TGTGGTGGCCACCCCA
	Rv3010c <sup>a</sup>	<i>pfkA</i>	–	GGTGATCGCGATGACC
	Rv3043c	<i>ctaD</i>	Rv3043c <sup>a</sup> -Rv3042c	AGTGGATCGCATCCCCG
	Rv3048c <sup>a</sup>	<i>nrdF2</i>	–	GGTGACTGGAACGCA
	Rv3217c <sup>a</sup>	–	–	TGTGGTGGCGGTTCGCA
	Rv3219 <sup>a</sup>	<i>whiB1</i>	–	AGTGAGATAGCCACG
	Rv3279c	<i>birA</i>	Rv3279c <sup>a</sup> -Rv3278c	TATCGGCTGCCGCACA
	Rv3280	<i>accD5</i>	Rv3280 <sup>a</sup> -Rv3281-Rv3282	CGGGACGTCGACCACA
	Rv3316	<i>sdhA</i>	Rv3316-Rv3317-Rv3318 <sup>a</sup> -Rv3319	CGAGACGTTTTCCACG
	Rv3549c	–	Rv3549c-Rv3548c <sup>a</sup>	GGTGATCGGCATTGCA
	Rv3676 <sup>a</sup>	–	–	TGTCACCTACGACAGA
	Rv3681c <sup>a</sup>	<i>whiB4</i>	–	TGAGATACAGGTAACA
	Rv3859c	<i>gltB</i>	Rv3859c <sup>a</sup> -Rv3858c	TGCTCCGGATTTACA

Detected binding sites of GlxR (ortholog in MT: Rv3676/Crp), RamB (ortholog in MT: Rv0465c), AmtR (ortholog in MT: Rv3160c), DtxR (ortholog in MT: IdeR/Rv3173c) and LexA (ortholog in MT: Rv2720/LexA) orthologs of CG in MT. Code:

<sup>a</sup>Transferred target gene of CG in MT.

genome-wide runs. There are three ways to access this feature through the MycoRegNet web site: (1) The TFBSscan button at the main page offers the possibility to upload user-defined sequences in FASTA format. (2) At any gene details page the user can predict TFBSs in the upstream sequence of the selected gene. (3) If the gene of interest encodes a TF, the PWM learned from the known TFBSs of the TF may be used to scan for further TFBSs in the upstream sequences of all other mycobacterial genes. The predicted results may further be visualized as graphs. The interface is easy to use: one just has to choose a background model (nucleotide distribution) and a *P*-value threshold. For further details regarding the prediction of prokaryotic TFBSs by utilizing PoSSuMsearch, the reader is referred to (20,43).

**Gene regulatory network visualization.** As mentioned earlier, MycoRegNet also provides a network visualization toolkit: GraphVis. It is a Java applet, which graphically reconstructs regulatory networks as graphs based on selected genes and a user-defined graph depth cutoff. It traverses all regulatory interactions from the starting point until the graph depth cutoff has been reached. Finally, a Java applet window appears showing the regulatory network as graph, where nodes represent genes and edges regulatory interactions. GraphVis allows the user to zoom into the graph, apply different layout styles, remove selected elements or retrieve detailed information on selected genes. Furthermore, it is possible to extend the graph dynamically with more genes/regulations from the

database and to display the operon grouping of presented genes. Visualized networks may also be graphically compared between two species or between a predicted and an evidenced network by utilizing special comparative graph layout algorithms. In addition, GraphVis features the projection of gene expression data onto the genes of a visualized network. The user can choose to apply gene expression data from the stimulon repository of MycoRegNet or from own tab-delimited text or MS-Excel files, which can be uploaded to GraphVis directly. It is also possible to use expression data extracted from EMMA by means of the integrated Web Service interconnection. According to the differential expression level of the genes, the concerned nodes are resized within the graph. Thus, the user can achieve a comprehensive overview of the transcriptional response of *M. tuberculosis* to a certain stimulus.

#### Well-structured data access by using Web Services

Although no real standard in bioinformatics, a growing number of platforms offer SOAP-based Web Service access to their data repositories [refer to some EBI resources (44) or to the BRENDA database (45), just to name two of them]. Many databases still provide flat files for exchange with other data processing systems. Thus, the developers of novel tools and platforms have to perform updates in regular time intervals and to adjust the downloaded data for their special purpose. On that account, gene regulatory data stored in MycoRegNet can also be accessed via the integrated Web Service

MycoRegNet Release 1.0

Search options

Organism

Search

in field

sort by

and

or

in field

ascending

descending

[Start search](#)

Statistics		Organisms	
Element	Number	Organism name	Number of genes
Genes	8183	<a href="#">Mycobacterium tuberculosis CDC1551</a>	4191
Proteins	8177	<a href="#">Mycobacterium tuberculosis H37Rv</a>	3992
Modules	7		
Stimulons	12		
Regulations	460		
Regulators	24		
Regulated genes	372		
Binding motifs	180		
Position weight matrices	6		
Protein clusters	2747		
Organisms	2		

Please also visit the sister project

[CoryneRegNet](#)

[Statistics](#) [TFBScan](#) [COMA](#)

[Documentation](#) [Literature](#)

[Web Service](#)

**Figure 2.** MycoRegNet main page. The main page includes a typical search mask, a statistical overview of the database content, an entry point to browse the integrated organisms, and links to more specific statistics, the system documentation and a tutorial on how to use the MycoRegNet Web Service.

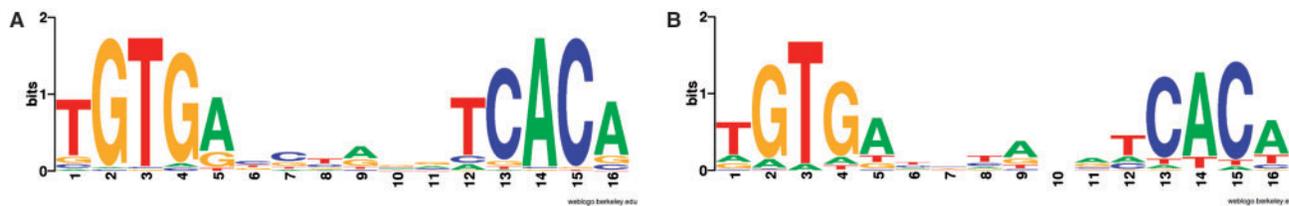
server. The data can be integrated directly into corresponding projects without further time-consuming data processing. Detailed information on how to use the MycoRegNet Web Services is available from the main page via the Web Service button.

#### Application example—the regulatory network of the GlxR ortholog Rv3676 (Crp<sub>MT</sub>) in MT

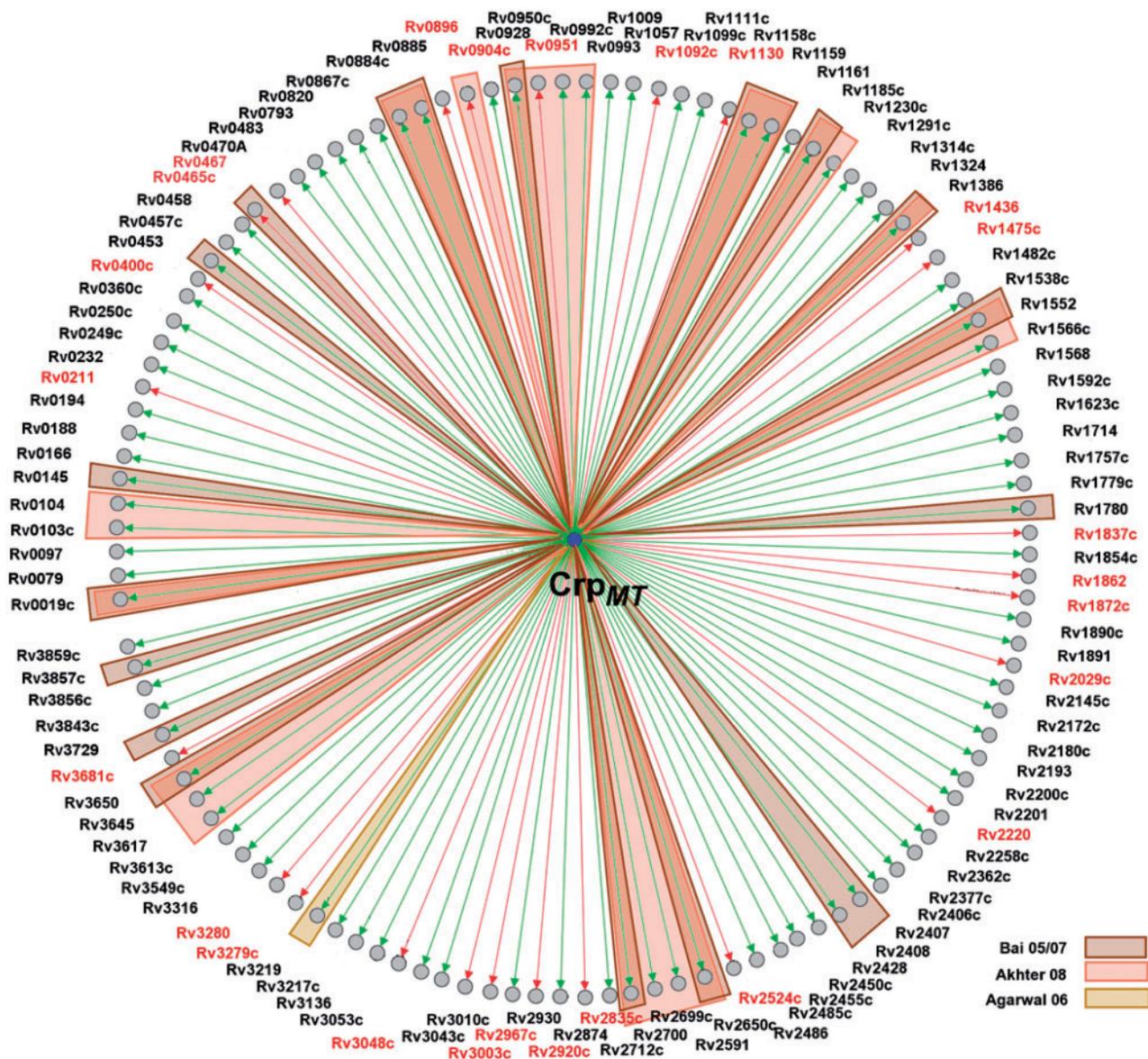
Both GlxR (Cg0350) of *C. glutamicum* and CRP<sub>MT</sub> (Rv3676) of *M. tuberculosis* belong to the Crp-Fnr family of TFs (46) and have been characterized as

cAMP sensing homologs of *E. coli* Crp (23,47,48). Crp-cAMP-dependent gene regulation is commonly involved in carbon catabolite repression and forms one of the possible connections between carbon metabolism and virulence (49,50).

In mycobacteria, cAMP signalling is the subject of intensive research, as it may be related to virulence of these strains (51,52). It is noteworthy that *M. tuberculosis* contains 16 putative adenylate cyclases, as well as 10 putative cyclic nucleotide binding proteins (53,54), hinting at a crucial and diverse role for cAMP signalling in mycobacteria.



**Figure 3.** Sequence logo of the predicted  $Crp_{MT}$  binding sites (A) in comparison to the sequence logo of GlxR (B). The sequence logo models the binding site motif of  $Crp_{MT}$ . It was deduced from the predicted binding sites in Table 3. The height of each letter within an individual stack represents the nucleotide's frequency relative to the particular motif position; thus, the degree of a nucleotide's conservation is indicated by the stack according to the respective position.



**Figure 4.** Reconstructed network of the GlxR ortholog  $Crp_{MT}$ . The network reconstruction of the  $Crp_{MT}$  regulon is based on the 121 transcription units presented in Table 3. It was generated by the integrated network reconstruction tool GraphVis of MycoRegNet. Transcription units relying on binding site predictions/experimental verifications that were reported previously in (22–24,60) and correspond with our findings are colored according to the appropriate publication. Arrows and gene IDs (node labels) coloured in red indicate a repressive regulation of  $Crp_{MT}$ , green arrows correspond to an activating regulation.

GlxR of *C. glutamicum* has been in the focus of interest in the last years (48,55–59), and available data indicates GlxR as global regulator with about 150 target genes in functional diverse network modules, such as carbohydrate metabolism, aerobic and anaerobic respiration, fatty acid metabolism, aromatic compound degradation, glutamate uptake and nitrogen assimilation, the cellular stress response and resuscitation.

Previous studies suggested a similar vital role for Crp<sub>MT</sub> in *M. tuberculosis*. Published data implicate Crp<sub>MT</sub> in virulence, hypoxia and nutrient starvation (21,23,24,60). Deletion of Crp<sub>MT</sub> altered the expression of 16 genes, and caused an impaired growth phenotype in bone marrow-derived macrophages as well as in tuberculosis mouse models (24). Several suggestions for a putative Crp<sub>MT</sub> regulon have been made, although these studies relied solely on the detection of putative binding sites (23,24,60).

As part of our pipeline, the known regulatory interactions of GlxR collected in CoryneRegNet have been used to reconstruct the regulon of the orthologous TF Crp<sub>MT</sub>. Due to the apparent vital role of these regulators in their respective organisms, and the available data on putative target genes and characterized binding sites, we chose them as application case for our analysis. Employing our pipeline, regulatory interactions with 64 target genes could be transferred from *C. glutamicum* GlxR to *M. tuberculosis* Crp<sub>MT</sub>. Furthermore, we considered

26 genes with experimental evidence of regulation by Crp<sub>MT</sub> as potential target genes (21–24).

Based on experimentally verified binding sites of Crp<sub>MT</sub> (21–23) together with binding sites predicted by the TFBS search of our pipeline, we complemented the suggested regulon with the prediction of Crp<sub>MT</sub> binding sites in the upstream regions of putative target genes.

In contrast to the TFBS searched within our pipeline, we created an adopted and optimized PWM for Crp<sub>MT</sub> from experimentally verified and predicted binding sites, and applied it for TFBS search. To detect the novel binding sites, we set the *P*-value threshold to 10<sup>-5</sup> and performed a restrictive search on sequences upstream genes/operons concerning the whole genome of MT. Again, we used PoSSuMsearch for binding site prediction scanning 580-bp long upstream sequence, ranging from +20 bp relative to the transcription start site. Using Weblogo (61), we generated a sequence logo from the detected binding sites of Crp<sub>MT</sub> and from the appropriate binding sites of GlxR that were used for PWM creation (see Methods section). The resulting sequence logos are shown in Figure 3.

In total, we identified 207 putative target genes of Crp<sub>MT</sub>, organized in 121 transcription units (see Table 3 and Figure 4). Of this set, 17 genes belong to the mycobacterial core regulon (62) and 41 were reported as essential for *M. tuberculosis* (63,64). Furthermore, at least 17

**Table 3.** Predicted Crp<sub>MT</sub> binding sites

Gene ID	Gene	Motif position <sup>d</sup>	Motif sequence	Operon
<b>Carbohydrate metabolism</b>				
Rv0211 <sup>a</sup>	<i>pckA</i>	-166	TGTGAGCAGGCTTATA	–
Rv0249c	<i>sdhCD</i>	-104	TGTGATCTGTAACACC	<i>Rv0249c-Rv0248c-Rv0247c<sup>a</sup></i>
Rv0249c	<i>sdhCD</i>	-410	GGTGTCCGAGGTCACA	<i>Rv0249c-Rv0248c-Rv0247c<sup>a</sup></i>
Rv0458	<i>adhA</i>	-41	TGTGAGCTGTATTACA	<i>Rv0458-Rv0459</i>
Rv0465c	–	-167	TTTGTGCGAGGCTCAG	<i>Rv0465c<sup>a</sup>-Rv0464c<sup>b</sup></i>
Rv0467 <sup>a,g</sup>	<i>icl</i>	-341	TGTTACAACGCTCACA	–
Rv0896 <sup>a</sup>	<i>gltA2</i>	-356	GGCGTTGAACATCACC	–
Rv0951	<i>sucC</i>	-173	TGTGAGTTGGATCAG	<i>Rv0951<sup>a,f</sup>-Rv0952<sup>a,f</sup></i>
Rv1099c	–	-515	GCTGATGAATCCCAG	<i>Rv1099c-Rv1098c<sup>a,f</sup>-Rv1097c</i>
Rv1130	<i>prpD2</i>	-152	TGTGGATAAGTCCAGG	<i>Rv1130<sup>a</sup>-Rv1131</i>
Rv1436	<i>gap</i>	-48	GGTTGTTTAGCCAACA	<i>Rv1436<sup>a,f</sup>-Rv1437<sup>a,f</sup>-Rv1438<sup>a,f</sup></i>
Rv1475c	<i>acn</i>	-462	TGTAACGTCGACATA	<i>Rv1475c<sup>a,f</sup>-Rv1474c</i>
Rv1552	<i>frdA</i>	-284	TGTGATCTAGGTCAG <sup>b</sup>	<i>Rv1552-Rv1553-Rv1554-Rv1555</i>
Rv1837 <sup>a</sup>	<i>glcB</i>	-381	AGGGATGCACTACACA	–
Rv1862 <sup>a</sup>	<i>adhA</i>	-227	CGTGGGGCGCCACACA	–
Rv1872 <sup>a</sup>	<i>lldD2</i>	-200	GATGCCGTAGCGCACT	–
Rv2029c <sup>a</sup>	<i>pfkB</i>	-410	GGTGACGAGTCGCGCA	<i>Rv2029c-Rv2028c-Rv2027c-Rv2026c<sup>f</sup></i>
Rv2967c <sup>a,f</sup>	<i>pca</i>	-389	CGTGGTGGTGGTCACC	–
Rv3010c <sup>a</sup>	<i>pfkA</i>	-532	CGTGATGGCGATGACC	–
Rv3316	<i>sdhC</i>	-386	GAGACGTTTTCCACG	<i>Rv3316-Rv3317-Rv3318<sup>a</sup>-Rv3319</i>
Rv3676 <sup>a</sup>	CRP	-538	TGTCACCTACGACAGA	–
<b>Fatty acid metabolism</b>				
Rv0097	–	-526	TGTCACGCCGGCCACG	<i>Rv0097-Rv0098<sup>c</sup>-Rv0099<sup>c</sup>-Rv0100<sup>c</sup>-Rv0101</i>
Rv0166	<i>fadD5</i>	-84	TGTGACCCAGACAACA	–
Rv0400c <sup>a,f</sup>	<i>fadE7</i>	-5	AGTGATGAGCACCCCG	–
Rv1185c	<i>fadD21</i>	-168	CGTGACGCCCTCAG	–
Rv1714	–	-405	GGTGACGGCGGCCACA	<i>Rv1714<sup>f</sup>-Rv1715<sup>f</sup>-Rv1716-Rv1717-Rv1718</i>
Rv2485c <sup>c</sup>	<i>lipQ</i>	-91	TGTGATCCTCGACACA	–
Rv2486	<i>echA14</i>	-287	TGTGTCGAGGATCACA	–
Rv2524c <sup>a,f</sup>	<i>fas</i>	-259	CGTTACCCACGACAG	–

(continued)

Table 3. Continued

Gene ID	Gene	Motif position <sup>d</sup>	Motif sequence	Operon
Rv2930 <sup>c</sup>	<i>fadD26</i>	-498	<b>TGTTAATCTCGTCACA</b>	<i>Rv2930-Rv2931-Rv2932<sup>f</sup>-Rv2933-Rv2934-Rv2935-Rv2936<sup>g</sup>-Rv2937<sup>g</sup>-Rv2938-Rv2939</i>
Rv3279c	<i>birA</i>	-38	<b>TATCGGCTGCCGCACA</b>	<i>Rv3279c<sup>a</sup>-Rv3278c<sup>e</sup></i>
Rv3280	<i>accD5</i>	-331	<b>CGGGACGTCGACCACA</b>	<i>Rv3280<sup>a</sup>-Rv3281<sup>e,f</sup>-Rv3282</i>
Rv3549c	-	-67	<b>GGTGATCGGCATTGCA</b>	<i>Rv3549c-Rv3548c<sup>a</sup></i>
Nitrogen assimilation				
Rv1538c	<i>ansA</i>	-187	<b>TGTGAGCACCACCACA</b>	-
Rv2220 <sup>a,f,g</sup>	<i>glnA1</i>	-1	<b>TGTGACGGAAAAGACG</b>	-
Rv2920c	<i>amt</i>	-2	<b>AGTGGACCAATTCCCC</b>	<i>Rv2920c<sup>a</sup>-Rv2919c<sup>a,g</sup>-Rv2918c<sup>a</sup></i>
Rv3859c	<i>gltB</i>	-398	<b>TGCTCCGGATTTTACA</b>	<i>Rv3859c<sup>a,f</sup>-Rv3858c<sup>f</sup></i>
PGRS				
Rv0453	PPE11	-269	<b>GGTGACCAAACCTCAGG</b>	-
Rv1386	PE15	-133	<b>TGTGACCAAACCTCACC<sup>b</sup></b>	<i>Rv1386c-Rv1387c<sup>c</sup></i>
Rv2408	PE24	-213	<b>GGTGATCGGGCGTCAGG</b>	-
Rv2591	P_PGRS44	-38	<b>CGTGACATGTGTCACA</b>	-
Rv3136 <sup>c</sup>	PPE51	-16	<b>AAGGAGCTGAGACACA</b>	-
Rv3650	PE33	-83	<b>TGTGATGCACCTTGACA</b>	-
Respiration				
Rv1161	<i>narG</i>	-512	<b>TGCGTTGAACGGCAGG</b>	<i>Rv1161<sup>a</sup>-Rv1162<sup>a</sup>-Rv1163<sup>a</sup>-Rv1164-Rv1165-Rv1166<sup>f</sup></i>
Rv1623c <sup>e</sup>	<i>cydA</i>	-181	<b>CGTGGTGATCGGCACA</b>	-
Rv1854c <sup>a</sup>	<i>ndh</i>	-109	<b>TGTGGCTGATGACACA</b>	-
Rv2193	<i>ctaE</i>	-517	<b>GGTGGATAGGTTTACC</b>	<i>Rv2193<sup>a,f</sup>-Rv2194<sup>a,f</sup>-Rv2195<sup>a,f</sup>-Rv2196<sup>a,f</sup></i>
Rv2200c	<i>ctaC</i>	-23	<b>TGTGATACAGGAGGCG</b>	<i>Rv2200c<sup>a,f</sup>-Rv2199c</i>
Rv3043c	<i>ctaD</i>	-227	<b>AGTGGATCGCATCCCG</b>	<i>Rv3043c<sup>a,f</sup>-Rv3042c<sup>f</sup></i>
Other cellular processes				
Rv0019c <sup>e</sup>	<i>fhaB</i>	-69	<b>CGTGACTTTGCTGACG<sup>b</sup></b>	-
Rv0079	-	-110	<b>GGTGACACAGCCCACA</b>	<i>Rv0079-Rv0080</i>
Rv0103c	<i>ctpB</i>	-159	<b>TGTGACGGGGCGTCACA</b>	-
Rv0104	-	-1	<b>TGTGACGCCCCGTCACA</b>	-
Rv0145	-	-59	<b>AGTGATGTGCCACACA<sup>b</sup></b>	<i>Rv0145-Rv0146</i>
Rv0188c	-	-356	<b>AGAGAACAACGTCGCA</b>	-
Rv0194	-	-517	<b>TGTCATCTAGATCAGG</b>	-
Rv0232	-	-53	<b>CGTGATGCAGCGCACA</b>	<i>Rv0232-Rv0233</i>
Rv0250c <sup>e</sup>	-	-37	<b>TGTGATCTGTAACACC</b>	-
Rv0360c	-	-2	<b>CGTGACCAAGCGCACA</b>	-
Rv0457c	-	-43	<b>TGTAATACAGCTCACA</b>	-
Rv0470A	-	-212	<b>TGTGGTGGAATCACA</b>	-
Rv0483	<i>lprQ</i>	-116	<b>TGTGTTTGGTATCACA</b>	-
Rv0793	-	-375	<b>TGTGATGGTGCGCAGG</b>	-
Rv0820 <sup>a,g</sup>	<i>phoT</i>	-538	<b>GGTGGTGATCCGCACC</b>	-
Rv0867c <sup>a</sup>	<i>rpfA</i>	-443	<b>TGTGACATTACCCACA<sup>b</sup></b>	-
Rv0884c <sup>a,f</sup>	<i>serC</i>	-91	<b>TGTGAGCTTGTTTACA<sup>b</sup></b>	-
Rv0885	-	-133	<b>TGTGAACAAGCTCACA</b>	<i>Rv0885-Rv0886</i>
Rv0904c <sup>a</sup>	<i>accD3</i>	-2	<b>CGTGAGTCGTATCAGG</b>	-
Rv0928	<i>pstS3</i>	-6	<b>ACTGAATTGAAACTCA</b>	<i>Rv0928<sup>a,g</sup>-Rv0929<sup>a,g</sup>-Rv0930<sup>a,g</sup></i>
Other cellular processes				
Rv0993	<i>galU</i>	-8	<b>TGTGAACGATGTCAGG</b>	<i>Rv0993<sup>f</sup>-Rv0994<sup>g</sup>-Rv0995<sup>g</sup></i>
Rv0950c	-	-153	<b>CGTGATCCAACCTCACA<sup>b</sup></b>	-
Rv0992c	-	-109	<b>CGTGACATCGTTTACA</b>	<i>Rv0992-Rv0991-Rv0990</i>
Rv1009	<i>rpfB</i>	-271	<b>GGTGGCGCTCATCACC</b>	<i>Rv1009<sup>a</sup>-Rv1010<sup>a,g</sup>-Rv1011<sup>a,f</sup></i>
Rv1057	-	-248	<b>CGTGACCTAGGTAACA</b>	-
Rv1092c <sup>a,f</sup>	<i>coaA</i>	-242	<b>TGCCACGTAGGTCAGG</b>	-
Rv1111c <sup>e</sup>	-	-411	<b>GGTGACATGAGTCAGG</b>	-
Rv1158c	-	-69	<b>TGTCACTTGAGTCACA<sup>b</sup></b>	<i>Rv1158c<sup>e</sup>-Rv1157c<sup>e</sup></i>
Rv1159	<i>pimE</i>	-77	<b>TGTGACTCAAGTGACA</b>	-
Rv1230c	-	-79	<b>GGTGATCTAGTTTACA<sup>b</sup></b>	-
Rv1291c	-	-323	<b>TGTGATCGGGCCACC</b>	-
Rv1314c	-	-294	<b>GGTGATCCGGGCCAGG</b>	-
Rv1324c	-	-104	<b>TGTGATCTTGGTCATA</b>	-
Rv1482c	-	-23	<b>TGTGACTCAGCACACT</b>	-
Rv1566c	-	-235	<b>CGTGACTGAAATCACA</b>	-
Rv1568	<i>bioA</i>	-553	<b>TGTGATTTTCAAGTCAGG</b>	<i>Rv1568-Rv1569<sup>g</sup>-Rv1570-Rv1571</i>
Rv1592c <sup>e</sup>	-	-215	<b>TGTGATAGGCGCCAGG</b>	-
Rv1757c	-	-351	<b>TGTGACGGCGGCCAGG</b>	-
Rv1779c	-	-89	<b>TGTGAACAACACCACA</b>	-

(continued)

Table 3. Continued

Gene ID	Gene	Motif position <sup>d</sup>	Motif sequence	Operon
Rv1780	–	–147	TGTGGTGTGTTTCACA	–
Rv1890c	–	–7	TGTGTCGTGGCCACA	–
Rv1891 <sup>e,g</sup>	–	–63	TGTGGGCCACGACACA	Rv1891–Rv1892–Rv1893 <sup>e</sup>
Rv2145c <sup>a,f</sup>	wag31	–463	CGTGACTGGCGTCCCA	–
Rv2172c <sup>e</sup>	–	–2	TGTGACCCCTCAACACG	–
Rv2180c	–	–304	TGTGTGGAACAACACA	–
Rv2201 <sup>a,f</sup>	asnB	–336	GCTGTCTGAAGACCACG	–
Rv2258c	–	–459	GGTGACGTGACCACG	–
Rv2362c	recO	–224	TGTGGGCTGGCTCACA	Rv2362c–Rv2361c <sup>f</sup> –Rv2360c
Rv2377c	mbiH	–268	TGTGGTTCACTCACT	–
Rv2406c	–	–34	TGTGAACCAGCTCACC	–
Rv2407	–	–242	GGTGAGCTGGTTCACA	–
Rv2428 <sup>c</sup>	ahpC	–93	GGTGTGATATATCACC	–
Rv2450c	rpfE	–509	TGTGGCGCAGGTCACC	–
Rv2450c	rpfE	–422	CGTGATTCGGCTCAG	–
Rv2455c	–	–237	AGTGACCAATACCACA	Rv2455c–Rv2454c–Rv2453c
Rv2650c	–	–305	CGTGAGGAGCCTCACC	–
Rv2699c	–	–116	TGTGATGTAAATCACA	–
Rv2700 <sup>e,f</sup>	–	–138	TGTGATTTACATCACA	–
Rv2712c	–	–296	GGTGAGGTAGAGCACA	–
Rv2835c	ugpA	–513	GGTGATGCCGGGCACG	Rv2835c–Rv2834c <sup>a</sup> –Rv2833c <sup>a,f</sup> –Rv2832c <sup>a,f</sup>
Other cellular processes				
Rv2874	dipZ	–351	TGTGGCGGAGTTCACA	–
Rv3003c	ilvB1	–335	TGTGGTGGCCACCCCA	Rv3003c <sup>a,f</sup> –Rv3002c <sup>a,f</sup> –Rv3001c <sup>a,f</sup>
Rv3048c <sup>a,f</sup>	nrdF2	–2	GGTGACTGAAAACGCA	–
Rv3053c	nrdH	–347	GGTGATCTGCGACACG	Rv3053c–Rv3052c–Rv3051c <sup>f</sup>
Rv3217c <sup>a,c</sup>	–	–278	TGTGGTGGCGGTTCGCA	–
Rv3219 <sup>a,c</sup>	whiB1	–176	AGTGAGATAGCCCACG <sup>b</sup>	–
Rv3613c <sup>c</sup>	–	–458	CGTGACGAATCCCCCA	–
Rv3617	ephA	–315	TGTGACCGGTGTCACT	Rv3617–Rv3618
Rv3645	–	–179	TGTGAGCCGAATCACC	–
Rv3681c <sup>a</sup>	whiB4	–106	TGAGATACAGGTAACA	–
Rv3729	–	–190	TGTGACCACGGCCACG	–
Rv3843c	–	–505	GGTGAGGTAAGTCACA	Rv3843c <sup>e</sup> –Rv3842c <sup>e</sup>
Rv3856c	–	–547	TGTGGGCTTCGTCACA	–
Rv3857c	–	–341	TGTGGGCTTCGTCACA <sup>b</sup>	–
Consensus			TGTGANNNNNNTCACA	

Crp<sub>MT</sub> binding sites detected by the TFBS search of the introduced pipeline and by the additional TFBS search with adopted and optimized PWMs. Bold letters indicate conserved pentamers of the motif. Codes:

<sup>a</sup>Transferred target gene from CG.

<sup>b</sup>Experimentally verified binding site by EMSA/ChIP/RT-PCR (21–23).

<sup>c</sup>Gene showed altered expression in microarray studies of  $\Delta$ Rv3676 versus WT (24).

<sup>d</sup>Motif position relative to the translation start site.

<sup>e</sup>Core gene.

<sup>f</sup>Essential gene.

<sup>g</sup>Gene involved in virulence processes

genes of the suggested regulon are connected to antibiotic resistance and virulence of *M. tuberculosis* (65–69). Based on annotation information for *M. tuberculosis* (69), knowledge about orthologous *C. glutamicum* genes and operon structures, we attributed individual target genes to distinct functional modules.

Similar to present knowledge on GlxR, results implicate Crp<sub>MT</sub> in the regulation of several functional modules such as carbohydrate metabolism (40 target genes), fatty acid metabolism (33 target genes), respiration (16 target genes) and nitrogen assimilation (7 target genes). Therefore, the position of the GlxR homolog Crp<sub>MT</sub> as global regulator in the transcriptional regulatory network seems to be conserved in *M. tuberculosis*. It is interesting to note that the suggested regulon comprises genes involved in essential functional modules, e.g. the citrate cycle, as well as genes involved in the synthesis of the

cellular envelope which plays an important role in the virulence of *M. tuberculosis*. Together with the supposed regulation of further virulence-associated genes this might explain why a functional Crp<sub>MT</sub> is required for virulence in model systems (24).

## CONCLUSIONS

With MycoRegNet, we have set up a system that allows researchers of the tuberculosis community to perform comprehensive analysis and visualizations of the gene regulatory network of MT. With its TFBS prediction it further provides easy access to a method that helps to generate new hypotheses *in silico*. As the sister project to CoryneRegNet, the MycoRegNet database content was generated through our comparative genomics pipeline, which provided us with reliable transfers of gene

regulatory interactions from the reference organism *C. glutamicum* to *M. tuberculosis*. With MycoRegNet, the corresponding data are publicly available and can be accessed easily through the web interface, or in a well-structured manner by using the MycoRegNet Web Service to maintain the reconstruction, visualization, and validation of mycobacterial regulatory networks at different hierarchical levels. Taken together, MycoRegNet is a reference resource for the tuberculosis community to gain a better understanding of the complex coherences of transcriptional gene control. It has the potential to assist researchers at the development of new vaccines and drugs to treat and prevent tuberculosis. Although MycoRegNet has been initially designed for MT, it may also serve for other mycobacterial strains in future, such as the already integrated *M. tuberculosis* CDC1551.

## FUNDING

GenoMik-Plus initiative of the German Federal Ministry of Education and Research (grant 0313805A); ERA-NET PathoGenoMics SPATELIS project (to J.K.); German Academic Exchange Service (to J.B. for his work at ICSI). Funding for open access charge: The work was financed in part by the GenoMik-Plus initiative of the German Federal Ministry of Education and Research, grant 0313805A.

*Conflict of interest statement.* None declared.

## REFERENCES

- Frieden, T.R., Sterling, T.R., Munsiff, S.S., Watt, C.J. and Dye, C. (2003) Tuberculosis. *Lancet*, **362**, 887–899.
- Raviglione, M.C. and Smith, I.M. (2007) XDR tuberculosis—implications for global public health. *N. Engl. J. Med.*, **356**, 656–659.
- Wilson, M., DeRisi, J., Kristensen, H.H., Imboden, P., Rane, S., Brown, P.O. and Schoolnik, G.K. (1999) Exploring drug-induced alterations in gene expression in *Mycobacterium tuberculosis* by microarray hybridization. *Proc. Natl Acad. Sci. USA*, **96**, 12833–12838.
- Singh, J.A., Upshur, R. and Padayatchi, N. (2007) XDR-TB in South Africa: no time for denial or complacency. *PLoS Med.*, **4**, e50.
- Park, H.-D., Guinn, K.M., Harrell, M.I., Liao, R., Voskuil, M.I., Tompa, M., Schoolnik, G.K. and Sherman, D.R. (2003) Rv3133c/dosR is a transcription factor that mediates the hypoxic response of *Mycobacterium tuberculosis*. *Mol. Microbiol.*, **48**, 833–843.
- Stewart, G.R., Snewin, V.A., Walzl, G., Hussell, T., Tormay, P., O'Gaora, P., Goyal, M., Betts, J., Brown, I.N. and Young, D.B. (2001) Overexpression of heat-shock proteins reduces survival of *Mycobacterium tuberculosis* in the chronic phase of infection. *Nat. Med.*, **7**, 732–737.
- Gama-Castro, S., Jimnez-Jacinto, V., Peralta-Gil, M., Santos-Zavaleta, A., Pealoza-Spinola, M.I., Contreras-Moreira, B., Segura-Salazar, J., Muiz-Rascado, L., Martnez-Flores, I., Salgado, H. et al. (2008) RegulonDB (version 6.0): gene regulation model of *Escherichia coli* K-12 beyond transcription, active (experimental) annotated promoters and Textpresso navigation. *Nucleic Acids Res.*, **36**, D120–D124.
- Baumbach, J. (2007) CoryneRegNet 4.0 – a reference database for Corynebacterial gene regulatory networks. *BMC Bioinformatics*, **8**, 429.
- Jacques, P.-E., Gervais, A.L., Cantin, M., Lucier, J.-F., Dallaire, G., Drouin, G., Gaudreau, L., Goulet, J. and Brzezinski, R. (2005) MtbRegList, a database dedicated to the analysis of transcriptional regulation in *Mycobacterium tuberculosis*. *Bioinformatics*, **21**, 2563–2565.
- Reddy, T.B.K., Riley, R., Wymore, F., Montgomery, P., DeCaprio, D., Engels, R., Gellesch, M., Hubble, J., Jen, D., Jin, H. et al. (2009) Tb database: an integrated platform for tuberculosis research. *Nucleic Acids Res.*, **37**, D499–D508.
- Babu, M.M. and Teichmann, S.A. (2003) Evolution of transcription factors and the gene regulatory network in *Escherichia coli*. *Nucleic Acids Res.*, **31**, 1234–1244.
- Babu, M.M., Luscombe, N.M., Aravind, L., Gerstein, M. and Teichmann, S.A. (2004) Structure and evolution of transcriptional regulatory networks. *Curr. Opin. Struct. Biol.*, **14**, 283–291.
- Babu, M.M., Teichmann, S.A. and Aravind, L. (2006) Evolutionary dynamics of prokaryotic transcriptional regulatory networks. *J. Mol. Biol.*, **358**, 614–633.
- Stackebrandt, E., Rainey, F.A. and Ward-Rainey, N.L. (1997) Proposal for a new hierarchic classification system, Actionbacteria classis nov. *Int. J. Sys. Bacteriol.*, **47**, 479–491.
- Eggeling, L., Besra, G.S. and Alderwick, L. (2008) Structure and synthesis of the cell wall. In Burkovski, A. (ed.), *Corynebacteria Genomics and Molecular Biology*, Caister Academic Press, pp. 267–294.
- Seidel, M., Alderwick, L.J., Sahn, H., Besra, G.S. and Eggeling, L. (2007) Topology and mutational analysis of the single Emb arabinofuranosyltransferase of *Corynebacterium glutamicum* as a model of Emb proteins of *Mycobacterium tuberculosis*. *Glycobiology*, **17**, 210–219.
- Portevin, D., Sousa-D'Auria, C.D., Houssin, C., Grimaldi, C., Chami, M., Daff, M. and Guilhot, C. (2004) A polyketide synthase catalyzes the last condensation step of mycolic acid biosynthesis in mycobacteria and related organisms. *Proc. Natl Acad. Sci. USA*, **101**, 314–319.
- Brinkrolf, K., Brune, I. and Tauch, A. (2007) The transcriptional regulatory network of the amino acid producer *Corynebacterium glutamicum*. *J. Biotechnol.*, **129**, 191–211.
- Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
- Beckstette, M., Homann, R., Giegerich, R. and Kurtz, S. (2006) Fast index based algorithms and software for matching position specific scoring matrices. *BMC Bioinformatics*, **7**, 389.
- Agarwal, N., Raghunand, T.R. and Bishai, W.R. (2006) Regulation of the expression of *whiB1* in *Mycobacterium tuberculosis*: role of cAMP receptor protein. *Microbiology*, **152** (Pt 9), 2749–2756.
- Bai, G., Gazdik, M.A., Schaak, D.D. and McDonough, K.A. (2007) The *Mycobacterium bovis* BCG cyclic AMP receptor-like protein is a functional DNA binding protein in vitro and in vivo, but its activity differs from that of its *M. tuberculosis* ortholog, Rv3676. *Infect. Immun.*, **75**, 5509–5517.
- Bai, G., McCue, L.A. and McDonough, K.A. (2005) Characterization of *Mycobacterium tuberculosis* Rv3676 (CRPmt), a cyclic amp receptor protein-like DNA binding protein. *J. Bacteriol.*, **187**, 7795–7804.
- Rickman, L., Scott, C., Hunt, D.M., Hutchinson, T., Menendez, M.C., Whalan, R., Hinds, J., Colston, M.J., Green, J. and Buxton, R.S. (2005) A member of the cAMP receptor protein family of transcription regulators in *Mycobacterium tuberculosis* is required for virulence in mice and controls transcription of the *rfpA* gene coding for a resuscitation promoting factor. *Mol. Microbiol.*, **56**, 1274–1286.
- Sun, R., Converse, P.J., Ko, C., Tyagi, S., Morrison, N.E. and Bishai, W.R. (2004) *Mycobacterium tuberculosis* ECF sigma factor sigC is required for lethality in mice and for the conditional expression of a defined gene set. *Mol. Microbiol.*, **52**, 25–38.
- Manganelli, R., Voskuil, M.I., Schoolnik, G.K. and Smith, I. (2001) The *Mycobacterium tuberculosis* ECF sigma factor sigmaE: role in global gene expression and survival in macrophages. *Mol. Microbiol.*, **41**, 423–437.
- Manganelli, R., Voskuil, M.I., Schoolnik, G.K., Dubnau, E., Gomez, M. and Smith, I. (2002) Role of the extracytoplasmic-function sigma factor sigma(H) in *Mycobacterium tuberculosis* global gene expression. *Mol. Microbiol.*, **45**, 365–374.

28. Davis, E.O., Dullaghan, E.M. and Rand, L. (2002) Definition of the mycobacterial SOS box and use to identify LexA-regulated genes in *Mycobacterium tuberculosis*. *J. Bacteriol.*, **184**, 3287–3295.
29. Stewart, G.R., Wernisch, L., Stabler, R., Mangan, J.A., Hinds, J., Laing, K.G., Young, D.B. and Butcher, P.D. (2002) Dissection of the heat-shock response in *Mycobacterium tuberculosis* using mutants and microarrays. *Microbiology*, **148** (Pt 10), 3129–3138.
30. Parish, T., Smith, D.A., Roberts, G., Betts, J. and Stoker, N.G. (2003) The SenX3-RegX3 two-component regulatory system of *Mycobacterium tuberculosis* is required for virulence. *Microbiology*, **149** (Pt 6), 1423–1435.
31. Betts, J.C., Lukey, P.T., Robb, L.C., McAdam, R.A. and Duncan, K. (2002) Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling. *Mol. Microbiol.*, **43**, 717–731.
32. Kendall, S.L., Movahedzadeh, F., Rison, S.C.G., Wernisch, L., Parish, T., Duncan, K., Betts, J.C. and Stoker, N.G. (2004) The *Mycobacterium tuberculosis* dosRS two-component system is induced by multiple stresses. *Tuberculosis (Edinb)*, **84**, 247–255.
33. Kendall, S.L., Withers, M., Soffair, C.N., Moreland, N.J., Gurcha, S., Sidders, B., Frita, R., Bokum, A.T., Besra, G.S., Lott, J.S. *et al.* (2007) A highly conserved transcriptional repressor controls a large regulon involved in lipid degradation in *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*. *Mol. Microbiol.*, **65**, 684–699.
34. Florczyk, M.A., McCue, L.A., Purkayastha, A., Currenti, E., Wolin, M.J. and McDonough, K.A. (2003) A family of acr-coregulated *Mycobacterium tuberculosis* genes shares a common DNA motif and requires Rv3133c (dosR or devR) for expression. *Infect. Immun.*, **71**, 5332–5343.
35. Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. and Sayers, E.W. (2009) GenBank. *Nucleic Acids Res.*, **37**, D26–D31.
36. Alm, E.J., Huang, K.H., Price, M.N., Koche, R.P., Keller, K., Dubchak, I.L. and Arkin, A.P. (2005) The MicrobesOnline web site for comparative genomics. *Genome Res.*, **15**, 1015–1022.
37. Curbera, F., Duftler, M., Khalaf, R., Nagy, W., Mukhi, N. and Weerawarana, S. (2002) Unraveling the web services web: an introduction to SOAP, WSDL, and UDDI. *IEEE Internet Comput.*, **6**, 86–93.
38. Meyer, F., Goesmann, A., McHardy, A.C., Bartels, D., Bekel, T., Clausen, J., Kalinowski, J., Linke, B., Rupp, O., Giegerich, R. *et al.* (2003) GenDB—an open source genome annotation system for prokaryote genomes. *Nucleic Acids Res.*, **31**, 2187–2195.
39. Dondrup, M., Albaum, S., Griebel, T., Henckel, K., Jünemann, S., Kahlke, T., Kleindt, C., Küster, H., Linke, B., Mertens, D. *et al.* (2009) EMMA 2 – a MAGE-compliant system for the collaborative analysis and integration of microarray data. *BMC Bioinformatics*, **10**, 50.
40. Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S., Katayama, T., Araki, M. and Hirakawa, M. (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.*, **34**, D354–D357.
41. Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N. *et al.* (2003) The COG database: an updated version includes eukaryotes. *BMC Bioinformatics*, **4**, 41.
42. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.
43. Baumbach, J., Brinkrolf, K., Wittkop, T., Tauch, A. and Rahmann, S. (2006) CoryneRegNet 2: an integrative bioinformatics approach for reconstruction and comparison of transcriptional regulatory networks in prokaryotes. *J. Integr. Bioinform.*, **3**, 24.
44. Pillai, S., Silventoinen, V., Kallio, K., Senger, M., Sobhany, S., Tate, J., Velankar, S., Golovin, A., Henrick, K. and Rice, P. (2005) Soap-based services provided by the European Bioinformatics Institute. *Nucleic Acids Res.*, **33**, W25–W28.
45. Barthelme, J., Ebeling, C., Chang, A., Schomburg, I. and Schomburg, D. (2007) BRENDA, AMENDA and FRENDA: the enzyme information system in 2007. *Nucleic Acids Res.*, **35**, D511–D514.
46. Körner, H., Sofia, H.J. and Zumft, W.G. (2003) Phylogeny of the bacterial superfamily of Crp-Fnr transcription regulators: exploiting the metabolic spectrum by controlling alternative gene programs. *FEMS Microbiol. Rev.*, **27**, 559–592.
47. Akhter, Y., Tundup, S. and Hasnain, S.E. (2007) Novel biochemical properties of a CRP/FNR family transcription factor from *Mycobacterium tuberculosis*. *Int. J. Med. Microbiol.*, **297**, 451–457.
48. Kim, H.-J., Kim, T.-H., Kim, Y. and Lee, H.-S. (2004) Identification and characterization of GlxR, a gene involved in regulation of glyoxylate bypass in *Corynebacterium glutamicum*. *J. Bacteriol.*, **186**, 3453–3460.
49. Deutscher, J. (2008) The mechanisms of carbon catabolite repression in bacteria. *Curr. Opin. Microbiol.*, **11**, 87–93.
50. Görke, B. and Stülke, J. (2008) Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat. Rev. Microbiol.*, **6**, 613–624.
51. Shenoy, A.R., Capuder, M., Draskovic, P., Lamba, D., Visweswariah, S.S. and Podobnik, M. (2007) Structural and biochemical analysis of the Rv0805 cyclic nucleotide phosphodiesterase from *Mycobacterium tuberculosis*. *J. Mol. Biol.*, **365**, 211–225.
52. Shenoy, A.R. and Visweswariah, S.S. (2006) New messages from old messengers: cAMP and mycobacteria. *Trends Microbiol.*, **14**, 543–550.
53. Shenoy, A.R. and Visweswariah, S.S. (2006) Mycobacterial adenyl cyclases: biochemical diversity and structural plasticity. *FEBS Lett.*, **580**, 3344–3352.
54. McCue, L.A., McDonough, K.A. and Lawrence, C.E. (2000) Functional classification of cNMP-binding proteins and nucleotide cyclases with implications for novel regulatory pathways in *Mycobacterium tuberculosis*. *Genome Res.*, **10**, 204–219.
55. Kohl, T.A., Baumbach, J., Jungwirth, B., Phler, A. and Tauch, A. (2008) The GlxR regulon of the amino acid producer *Corynebacterium glutamicum*: *in silico* and *in vitro* detection of DNA binding sites of a global transcription regulator. *J. Biotechnol.*, **135**, 340–350.
56. Jungwirth, B., Emer, D., Brune, I., Hansmeier, N., Pühler, A., Eikmanns, B.J. and Tauch, A. (2008) Triple transcriptional control of the resuscitation promoting factor 2 (rpf2) gene of *Corynebacterium glutamicum* by the regulators of acetate metabolism RamA and RamB and the cAMP-dependent regulator GlxR. *FEMS Microbiol. Lett.*, **281**, 190–197.
57. Han, S.O., Inui, M. and Yukawa, H. (2008) Effect of carbon source availability and growth phase on expression of *Corynebacterium glutamicum* genes involved in the tricarboxylic acid cycle and glyoxylate bypass. *Microbiology*, **154** (Pt 10), 3073–3083.
58. Han, S.O., Inui, M. and Yukawa, H. (2007) Expression of *Corynebacterium glutamicum* glycolytic genes varies with carbon source and growth phase. *Microbiology*, **153** (Pt 7), 2190–2202.
59. Letek, M., Valbuena, N., Ramos, A., Ordez, E., Gil, J.A. and Mateos, L.M. (2006) Characterization and use of catabolite-repressed promoters from gluconate genes in *Corynebacterium glutamicum*. *J. Bacteriol.*, **188**, 409–423.
60. Akhter, Y., Yellaboina, S., Farhana, A., Ranjan, A., Ahmed, N. and Hasnain, S.E. (2008) Genome scale portrait of cAMP-receptor protein (CRP) regulons in mycobacteria points to their role in pathogenesis. *Gene*, **407**, 148–158.
61. Crooks, G.E., Hon, G., Chandonia, J.-M. and Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.*, **14**, 1188–1190.
62. Marmiesse, M., Brodin, P., Buchrieser, C., Gutierrez, C., Simoes, N., Vincent, V., Glaser, P., Cole, S.T. and Brosch, R. (2004) Macro-array and bioinformatic analyses reveal mycobacterial ‘core’ genes, variation in the ESAT-6 gene family and new phylogenetic markers for the *Mycobacterium tuberculosis* complex. *Microbiology*, **150** (Pt 2), 483–496.
63. Sassetti, C.M., Boyd, D.H. and Rubin, E.J. (2003) Genes required for mycobacterial growth defined by high density mutagenesis. *Mol. Microbiol.*, **48**, 77–84.
64. Lamichhane, G., Zignol, M., Blades, N.J., Geiman, D.E., Dougherty, A., Grosset, J., Broman, K.W. and Bishai, W.R. (2003) A postgenomic method for predicting essential genes at subsaturation levels of mutagenesis: application to *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA*, **100**, 7213–7218.
65. Rengarajan, J., Bloom, B.R. and Rubin, E.J. (2005) Genome-wide requirements for *Mycobacterium tuberculosis* adaptation and

- survival in macrophages. *Proc. Natl Acad. Sci. USA*, **102**, 8327–8332.
66. Tullius, M.V., Harth, G. and Horwitz, M.A. (2003) Glutamine synthetase GlnA1 is essential for growth of *Mycobacterium tuberculosis* in human THP-1 macrophages and guinea pigs. *Infect. Immun.*, **71**, 3927–3936.
67. McAdam, R.A., Quan, S., Smith, D.A., Bardarov, S., Betts, J.C., Cook, F.C., Hooker, E.U., Lewis, A.P., Woollard, P., Everett, M.J. *et al.* (2002) Characterization of a *Mycobacterium tuberculosis* H37Rv transposon library reveals insertions in 351 ORFs and mutants with altered virulence. *Microbiology*, **148** (Pt 10), 2975–2986.
68. McKinney, J.D., Höner zu Bentrup, K., Muñoz-Eliás, E.J., Miczak, A., Chen, B., Chan, W.T., Swenson, D., Sacchettini, J.C., Jacobs, W.R. and Russell, D.G. (2000) Persistence of *Mycobacterium tuberculosis* in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. *Nature*, **406**, 735–738.
69. Cole, S.T. (1999) Learning from the genome sequence of *Mycobacterium tuberculosis* H37Rv. *FEBS Lett.*, **452**, 7–10.