7400–7414 *Nucleic Acids Research, 2011, Vol. 39, No. 17* *Published online 7 June 2011*
doi:10.1093/nar/gkr375

Comprehensive insights into *Mycobacterium tuberculosis* DevR (DosR) regulon activation switch

Santosh Chauhan[1], Deepak Sharma[2], Alka Singh[1], Avadhesha Surolia[2] and Jaya Sivaswami Tyagi[1,*]

[1]Department of Biotechnology, All India Institute of Medical Sciences, Ansari Nagar, New Delhi 110029 and [2]Centre for Molecular Medicine, National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi 110067, India

Received February 20, 2011; Revised April 27, 2011; Accepted May 2, 2011

ABSTRACT

DevR regulon function is believed to be crucial for the survival of *Mycobacterium tuberculosis* during dormancy. In this study, we undertook a comprehensive analysis of the DevR regulon. All the regulon promoters were assigned to four classes based on the number of DevR binding sites (Dev boxes). A minimum of two boxes are essential for complete interaction and their tandem arrangement is an architectural hallmark at all promoters. Initial interaction of DevR with the conserved box is essential for its cooperative binding to adjacent sites bearing low to very poor sequence conservation and is the universal mechanism underlying DevR-mediated transcriptional induction. The functional importance of tandem arrangement was established by analyzing promoter variants harboring Dev boxes with altered spacing. Conserved sequence logos were generated from 47 binding sequences which included 24 newly discovered Dev boxes. In each half site of an 18-bp binding motif, G_5 and C_7 are essential for DevR binding. Finally, we show that DevR regulon induction occurs in a temporal manner and genes that are induced early are also usually powerfully induced. The information theory-based approach along with binding and temporal expression studies provide us with comprehensive insights into the complex pattern of DevR regulon activation.

INTRODUCTION

Mycobacterium tuberculosis (Mtb) is one of the most successful pathogens in history which accounted for an estimated 1.7 million deaths in 2009 (1) making tuberculosis a global health emergency. The success of Mtb is attributed in large measure to its ability to cause and sustain a persistent and latent infection, sometimes even for decades. Recent studies showed the importance of DevR regulon genes as potential markers of latent infection (2–4). A clear understanding of the expression pattern of DevR regulon genes will facilitate the identification of early and late dormancy antigens and potential candidates for subunit vaccines against latent tuberculosis.

Hypoxia, nitric oxide and nutrient starvation are some of the conditions which are believed to be associated with initiation and maintenance of Mtb dormancy (5–7). Carbon monoxide and ascorbic acid have also been implicated recently in dormancy adaptation (8–10). Hypoxia, nitric oxide, carbon monoxide and ascorbic acid signals induce a set of ~48 genes via the DevRS two-component system [also called DosRS (8–11)]. Ascorbic acid mimics multiple intracellular stresses and exerts wide-ranging effects on Mtb gene expression, including induction of the DevR regulon. The modulation of gene expression is accompanied by growth arrest and a 'dormant' drug-tolerant phenotype under *in vitro* and *ex vivo* conditions (10). Due to its importance in virulence and dormancy (12–18), DevRS is undoubtedly the best characterized two-component system of Mtb. *In silico* analysis of DevR regulon promoters revealed the presence of one or more copies of a consensus binding sequence located in the upstream promoter regions of target genes (11). We showed that activated (phosphorylated) DevR binds cooperatively to specific DNA sequences (Dev boxes) to activate target gene expression (19–21). Although we have a fair idea of the mechanism of DevR-mediated transcription regulation (19–21), our understanding is incomplete

*To whom correspondence should be addressed. Tel: 91 11 26588491; Fax: 91 11 26588663; Email: jstyagi@aiims.ac.in
Present addresses:
Santosh Chauhan, Department of Cancer Biology, MD Anderson Cancer Center, Houston, TX 77030, USA
Alka Singh, Centre of Agricultural Bioinformatics (CABin), Indian Agricultural Statistics Research Institute (IASRI), Library Avenue, PUSA, New Delhi 110012, India.

due to the absence of experimental evidence of all constituent promoters' structures.

The crystal structure of $DevR_C$–DNA complex has been solved (22). In this structure, a dimer of the DevR C-terminal domain ($DevR_C$) interacts with $G_4G_5G_6A_7C_8T_9$ motif present in each half of a palindromic consensus sequence. However, the relative importance of each of these nucleotides in DevR interaction is not known which is crucial for understanding DevR–DNA interaction in the natural genomic context wherein consensus sites are absent and moreover, target genes are expressed at various levels (11). We recently showed that the N-terminal domain of DevR plays a decisive role in cooperative binding of DevR to DNA and in supporting robust gene induction particularly during hypoxia (23). These observations emphasize the relevance of examining DevR regulon activation mechanisms in the context of full-length phosphorylated DevR–DNA interaction (rather than with $DevR_C$).

We recently showed there was a disparity between the number of predicted binding sites (11) and experimentally determined DevR binding sites in two regulon members, namely, *tgs1-Rv3131* and *fdxA* genes. Moreover, the newly discovered binding sites were functionally important (21,24). This raised a need to examine other DevR target promoters to gain a comprehensive understanding of regulon activation mechanisms. Here, we experimentally determined the promoter structure of all the known target genes of this regulon. Forty-seven experimentally determined sites were used to generate a sequence logo and the functional importance of each nucleotide at critical positions was assessed. We show that tandem positioning and helical phasing of DevR binding sites are essential for cooperative binding and synergistic transcriptional activation. Our Green Fluorescent Protein (GFP) reporter data suggests that DevR regulon genes are temporally and differentially regulated in a complex manner.

MATERIALS AND METHODS

Bacterial strains, plasmids and primers

Mycobacterium tuberculosis strains were cultured at 37°C in Dubos medium containing 0.05% Tween-80 plus 0.5% Albumin-0.75% Dextrose-0.085 % NaCl (ADS complex). *Escherichia coli* strains that were used and their culture conditions were as described earlier (25). All the plasmids used in this study are listed in Supplementary Table S1. All primers used in the construction of *gfp* transcriptional fusions and in the preparation of DNA fragments for DNase I footprinting are listed in Supplementary Table S2.

Construction of *gfp* transcriptional fusions

The putative promoter regions of all DevR regulon genes were amplified from *M. tuberculosis* H37Rv DNA using primers (listed in Supplementary Table S2) and cloned into the promoterless GFP reporter plasmid, pFPV27 (26) at the EcoRI site to place the promoters upstream of the GFP open reading frame. Dev box mutants and variants were generated by the mega primer method of mutagenesis as described (27). The sequences of all the cloned inserts and mutations were confirmed by DNA sequencing. The promoter plasmids were electroporated into *M. tuberculosis* H37Rv and GFP reporter assays were carried out as described below.

GFP reporter assay

GFP reporter assays were carried out as described previously (19). Briefly, Mtb strains carrying various promoter constructs were grown simultaneously under aerobic conditions and subcultured two to three times till all the strains were growing at apparently similar growth rates according to A_{595} measurements. At A_{595} = ~0.3–0.4, the cultures were diluted to an Optical Density at 595 nm = ~0.05 and standing cultures were established by dispensing 200-µl culture aliquots in quadruplicate into 96-well black, clear-bottom microtiter plates (Becton, Dickinson and Co., UK) and the plates were incubated at 37°C. For each time point separate plates were used. GFP fluorescence in Mtb cultures was measured in a spectroflourimeter using excitation and emission wavelengths of 483 and 515 nm, respectively and expressed as relative fluorescence units (RFU) per unit Optical Density at 595 nm (OD) after correcting for background fluorescence of bacteria harboring pFPV27 control vector which varied between ~30 and 45 RFU/OD. Fold induction is a ratio of RFU/OD of standing cultures versus that of aerobic shaking cultures at 48/72 h (without subtracting vector background).

Gel-shift assay and DNase I footprinting

DevR protein was purified as described previously (24). Phosphorylated DevR was prepared using acetyl phosphate for all interaction experiments because we showed previously that unphosphorylated or phosphorylation defective protein (DevR D54Vprotein) does not bind to DNA in a sequence-specific manner (19). $DevR_C$ protein (141–217 amino acids of DevR) was expressed and purified from pUS-$DevR_C$ as described (23) and was generously provided by Dr U.S. Gautam. DNase I footprinting and gel-shift assays were carried out as described previously (19,21). The fraction of bound DNA was estimated using Quantity One software (Bio-Rad).

Computational analysis

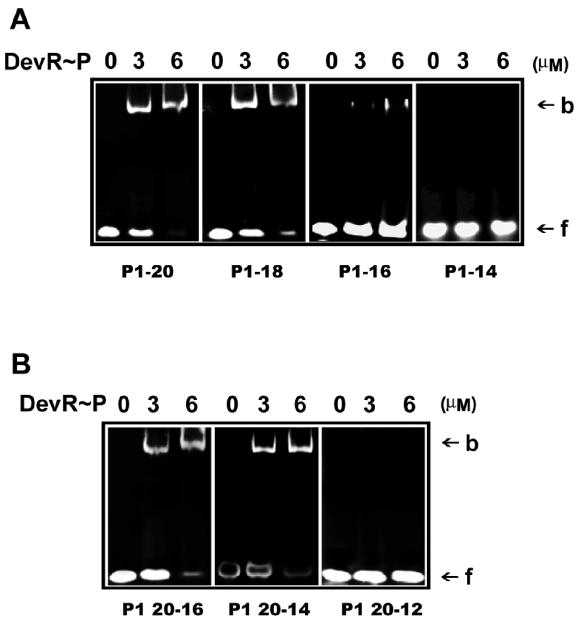
For the 25 DevR primary binding sites, the Delila programs *mkdb*, *dbbk*, *catal*, *delila* and *alist* were used to convert the sequences into Delila format and to align them (28,29). An information curve was made using *encode* and *rseq*, and the average shown as a sequence logo using *dalvec* and *makelogo*. Once the logo was made, the *ri* program was used with the same sequences to create an individual information weight matrix. Finally, the *makewalker* program was used to generate the walkers for the primary and secondary DevR binding sites. Three numbers are reported in the vertical box above or below the zero line opposite the base. The first is the position of the box on the sequence. The second is the sequence conservation of the entire binding site, given in bits. The third is the *Z*-score, which conveys the probability that a particular sequence is a member of the sites used to create the matrix.

RESULTS

The DevR regulon comprises of ~48 genes and it plays a key role in Mtb adaptation to a variety of environmental cues including hypoxia, exposure to nitric oxide or carbon monoxide or ascorbic acid (6,8–11). These genes are arranged singly (e.g. *Rv2623*, *Rv0571c* etc.) or in clusters and operons (e.g. *Rv3134c-devRS*, *Rv2031c-Rv2028c*) or are transcribed in a divergent manner (e.g. *Rv1738-narK2*). The regulon comprises some well-known genes including the *devRS* two-component system itself, *hspX*, etc. Although the role of DevR in regulating the activity of a few target promoters such as *devRS*, *hspX*, *narK2, tgs1,* has been characterized (19–21), little is known about the regulatory control of most of the DevR regulon that comprises many genes of unknown function. Although a wide range of induction responses were noted (8,11) and the extent of induction appeared to correlate with the number of DevR binding sites present in the target promoters (11), some aspects were puzzling. Many strongly induced genes such as *Rv0079*, *Rv2628*, *fdxA*, reportedly contain only one binding site, while on the other hand, low levels of induction was associated with the presence of several binding sites in some other instances, e.g. *Rv1997*, *Rv1733c*, and raises a question regarding the relationship between number of DevR binding sites on one hand and activation response (timing and magnitude) on the other hand. Other emergent questions relate to the role of DevR in inducing disparate expression of divergently arranged genes, the role of the newly discovered binding sites in cooperative recruitment of DevR to weak secondary sites, the relevance of tandem spacing of DevR binding sites and the relevance of differences in sequence content among primary and secondary binding sites. On the basis of previous predictions by statistical models (30) and occurrence of small intergenic regions, many of the DevR regulon genes could be arranged in operons. In the present study, a comprehensive analysis of the DevR regulon comprising of ~43 genes transcribed from 27 putative promoters (Supplementary Figure S1) was undertaken to answer the questions posed above.

The minimal binding site of DevR

The Dev box was suggested to be 18- or 20-bp long on the basis of *in silico* analysis (11,31). Toward defining its size by experimental means, gel-shift assays were performed with double-stranded oligonucleotides of varying lengths harboring the primary *narK2* binding motif, P1 [previously named as D1 in (20); Figure 1]. The 20- and 18-bp boxes bound with nearly equivalent efficiency to DevR. Further shortening of the box to 16- and 14-bp led to progressively weaker binding and abrogation of binding, respectively (Figure 1A and C). Therefore, we conclude that the two peripheral nucleotides of the P1 box are dispensable for binding and in agreement with the *in silico* prediction of an 18-bp long binding motif (31). The substitution of the peripheral nucleotides (Figure 1B) demonstrated that DevR interacts with 16- and 14-bp long motifs but not with the 12-bp box (Figure 1B and C), suggesting that DevR intimately contacts only the core 14-bp sequence of the binding site although a minimum of 18-bp is required for stable binding.



C

Oligonucleotides used in EMSA		Relative DevR binding activity(%)
Name	Sequence (5'-3')	
narK2 P1-20	TTAGGGCCGGAAGTCCCCAA	100
narK2 P1-18	TAGGGCCGGAAGTCCCCA	93
narK2 P1-16	AGGGCCGGAAGTCCCC	6
narK2 P1-14	GGGCCGGAAGTCCC	0
narK2 P1 20-16	AGAGGGCCGGAAGTCCCCCT	94
narK2 P1 20-14	AGTGGGCCGGAAGTCCCACT	94
narK2 P1 20-12	AGTCGGCCGGAAGTCCGACT	0

Figure 1. Size of DevR binding site. (**A** and **B**) Electrophoretic mobility shift assay (EMSA) with double-stranded oligonucleotide variants (deletion and substitution) of the *narK2* P1 Dev box (P1-20). 'f' and 'b' refer to free and bound DNA, respectively. (**C**) Percentage of DevR binding to oligonucleotide variants in relation to P1-20 box at 6 μM protein concentration. For simplicity, only the top strand is shown. P1 sequences are underlined in P1-20 substitution variants.

The presence of primary and secondary binding sites is a universal feature of DevR regulon promoters

Most of the DevR regulated genes were predicted earlier to have a single Dev box in their promoter regions (Table 1). One additional binding site each, which was located adjacent to the predicted DevR binding site, was discovered during characterization of the *tgs1-Rv3131* and *fdxA* promoters. Importantly, the additional site was vital for gene expression (21,24). This finding raised the possibility that additional functional cryptic binding sites could also be present in other DevR regulon promoters. Toward gaining insights into the promoter structure of regulon genes, we analyzed DevR binding properties at all the putative regulon promoters by DNase I footprinting and searched the protected DNA regions for 18-bp Dev box or Dev box-like sequences. The most striking characteristic of all the regulon promoters was the presence of two to four tandemly arranged upstream binding sites. There was not a single regulon promoter that featured only one DevR binding site (Table 1, Figure 2, Supplementary Figure S1–S5).

We have recently shown that the C-terminal DNA binding domain of DevR ($DevR_C$) is defective in cooperative interaction although it binds with primary binding sites (23). In the present study, this property was exploited to distinguish between primary and secondary binding sites in target promoters. By the criterion that they were bound to DevR, but not to $DevR_C$ and by computational analysis (see below), 22 of the 24 newly discovered sites were classified as secondary binding sites (Table 1, Figure 2, Supplementary Figure S1–S5).

Computational analysis of DevR binding sites

DNase I protected regions of all the regulon promoters were analyzed for the presence of DevR binding sequences using computational analysis tools (see 'Materials and methods' section). The analysis led to the discovery of 24 new binding sites out of which 22 were poorly conserved. Forty-seven experimentally verified DevR binding sites including 25 primary sites and 22 secondary sites from this study and previous studies (19–21,24), were used to generate sequence logos and total information content (R_i value) for individual Dev boxes (Figure 3, Supplementary Tables S3–S5). As the information content (R_i value) for each binding site could be correlated with the binding energy (32,33), we can generally assume the Dev box at each promoter with high R_i value to be primary (P) and with low R_i value to be secondary site (S). These assumptions are considered to be largely valid on the basis of experimental evidence from DNase I footprinting analysis (see above). C_7 is the most conserved nucleotide in the Dev box and it is present in all 47 sites. Other highly conserved nucleotides are G_3, G_4 and G_5 in both strands of the primary binding site (Figure 3). In contrast, the first and last positions in the sequence logo are poorly conserved; i.e. they are not information rich. This is consistent with the results of gel-shift assay (Figure 1), where shortening of the binding motif to 16-bp is tolerated provided the overall length is maintained at 20-bp. The results of gel-shift assay and computational analysis suggest that a motif length of 18 bp is optimal. In the secondary boxes, nucleotides G_4, C_7 and G_{12} are well conserved and have scores of 1 bit or more (Figure 3). Because DevR does not bind to secondary sites in the absence of binding to the primary site, interactions with these nucleotides appear to enable cooperative recruitment of DevR to secondary binding sites in spite of their low information content.

The 'sequence walkers' is a graphical method for displaying how binding proteins interact with individual bases of nucleotide sequences. We used *makewalker* program to generate the walkers for all the DevR primary and secondary binding sites (Figure 4, Supplementary Figure S6). This allowed us to determine the contribution of each base (positive or negative) to the average sequence conservation of the binding site, as represented by a sequence logo.

Identification of nucleotides critical for DevR binding

It is evident from the sequence logo that G_4, G_5 and C_7 nucleotides [previously numbered as G_5, G_6 and C_8, respectively in (11,19–22)] are highly conserved in both strands of the primary binding sites. Toward analyzing the importance of each of these nucleotides in DevR–DNA interaction, mutant variants of a well-conserved primary binding site (*narK2* primary Dev box P1, $R_i = 20$) were assessed for binding to DevR using gel-shift assays (Figure 5). In each variant box, the wild-type (WT) nucleotide was mutated to the least frequently occurring nucleotide at that position (Supplementary Table S3). A severe loss in binding was observed on substitution at G_5 (M-5, 90%) and C_7 position (M-7, 93%) which indicates that their strong conservation in the binding site is of high functional consequence while a less-severe binding defect was noted on substitution at G_3 and G_4 positions. A complete loss of DevR binding was observed when substitutions were introduced in both halves of the Dev box at any one of these four positions (M-3 + 3, M-4 + 4, M-5 + 5 and M-7 + 7). Similarly, binding was completely abolished when mutations were introduced in combination at nucleotide positions 3 + 7, 4 + 7, 5 + 7, 3 + 5 and 4 + 5 in one half of the Dev box. Binding was not significantly impaired when the Dev box was mutated at position 8 (M-8, M-8 + 8). Intriguingly, the substitution at nine positions (M-9 + 9) resulted in a complete loss of binding. Although the nucleotide at position 9 is not highly conserved and is usually occupied by A or T, it could help in bending of DNA during DevR–DNA interaction (discussed later). The importance of C_7 in binding is again underscored by the failure of DevR to bind with the D4 motif at the *narK2-Rv1738* promoter which bears a natural mutation at position C_7 (20) and was previously suggested as a binding site (11).

DevR regulon promoters are assigned to four architectural classes

The complete experimental mapping of DevR binding sites enabled us to group all the DevR regulon promoters

Table 1. Architecture of DevR regulon promoters

Promoter class	No. of experimentally determined binding sites	No. of predicted sites[a]	Newly discovered site(s) in this study[b]	Organization of Dev boxes	Distance from ATG/GTG[c]	Distance of proximal Dev box from TSP	Other genes in same operon (30)	Reference
(A) Class I (2 Dev boxes)								
Rv0569	2	1	S	P-S	52	ND	Rv570	This study
Rv0571c	2	1	P1	P1-P2	33	ND	–	This study
Rv0574c	2	1	S	S-P	93	ND	Rv0573c	This study
Rv1996	2	1	S	S-P	57	ND	–	This study
Rv1997/ctpF	2	2	S	P-S	71	ND	–	This study
Rv2005c	2	1	S	P-S	52	ND	Rv2004c, Rv2003c	This study
Rv2006/otsB	2	1	S	S-P	26	ND	–	This study
Rv2007c/fdxA	2	1	S	P-S	35	34	–	(24)
Rv2623	2	1	S	P-S	90	ND	–	This study
Rv2626c	2	2	S	S-P	71	ND	Rv2625c, Rv2624c	This study
Rv2628	2	1	S	P-S	+6	ND	Rv2629, Rv2630, Rv2631	This study
Rv3130c/tgs1	2	1	S	S-P	75	33	–	(21)
Rv3131	2	1	S	P-S	70	33	–	(21)
Rv3134c	2	2		P-S	73	33	Rv3133c, Rv3132c	(19,51)
(B) Class II (3 Dev boxes)								
Rv0079	3	1	S1, S2	P-S1-S2	26	ND	Rv0080, Rv0081, Rv0082, Rv0083	This study
Rv1733c	3	2	S	P1-S-P2	23	ND	–	This study
Rv2031c/hspX	3	3		P-S-P	63	34	Rv2030c, Rv2029c, Rv2028c	This study
Rv2032	3	3		P-S-P	58	ND	–	(19), This study
(C) Class III (4 Dev boxes)								
Rv1737c/narK2	4	3	S2	P2-S2-S1-P1	92	33	Rv1736c	(20), This study
Rv1738	4	3	S2	P1-S1-S2-P2	114	33	–	(20), This study
Rv2627c	4	1	S1, S2, S3	S1-S2-S3-P	34	ND	–	This study
(D) Class IV (with complex structure)								
Rv1813c	3+E	1	P1, S, E	P1-P2-S-E	221	ND	–	This study
Rv1734c	2+E	1	S, E	P-S-E	+25	ND	–	This study
Rv3127	1+E	1	E	P-E	17	ND	–	This study

'P' 'S' and 'E', primary, secondary and extended DevR binding site, respectively.
[a]According to (11,31).
[b]Location of sites is indicated in Supplementary Figure S1.
[c]Distance of proximal Dev box from start codon (ATG/GTG).
TSP, Transcriptional Start Point; ND, not determined.

into four different classes (Table 1 and Supplementary Figures S2–S5). The simplest of them, the Class I promoters, contain two neighboring Dev boxes in either the P-S or S-P arrangement (secondary site proximal to ATG/GTG or primary site proximal to ATG/GTG, respectively, Figure 2). *Rv0571c* was an exceptional promoter of this class and possesses two binding sites that were both designated as primary sites (P1-P2) on the basis of their interaction with $DevR_C$ (Supplementary Figure S2, Supplementary Tables S4 and S5). Promoters containing three DevR binding sites are categorized as Class II promoters and have P-S-S arrangement (*Rv0079*, Supplementary Figure S3, panel A) or P-S-P configuration (*Rv1733c* promoter and *hspX-Rv2032* promoters, Figure 2, Supplementary Figure S3, panel B).

Class III promoters are those with four tandem Dev boxes. Interestingly, the *Rv2627c* promoter which was originally reported to have a single binding site actually contains four Dev boxes that include a primary site and three secondary sites (Table 1 and Supplementary Figure S4, panel A). Even more interesting among the Class III promoters are the divergent *narK2-Rv1738* genes whose intergenic region is bound by DevR (20). Three binding sites, namely, P1, S2 and P2 (earlier named as D1, D2 and D3) were previously shown to be required for transcription of both the divergent genes (20) and continuous protection over a ~80-bp stretch of DNA suggested the presence of an additional site. Accordingly, a search was made and a fourth potential binding site, named as S1, was detected by careful visual inspection and by walking down (using sequence walkers) from P1 towards S2 (Figure 2, Supplementary Figure S4, panel B and Figure 4). The first half (9 bp) of the S1 box is quite degenerate (G_{12} T_{14} $C_{15}T_{16}$ in lower strand), where as the second half is highly conserved ($G_3G_4G_5C_7$, Supplementary Figure S4, panel B and Figure 4). Mutation in the second half of the palindrome at G_4 and C_7 nucleotides abolished DevR binding and established S1 as a genuine secondary DevR binding site (Supplementary Figure S4, panel B). Interestingly, upon mutating the S1 site (mut-S1), binding to the adjacent site, P1, was also abolished at lower protein concentration, but

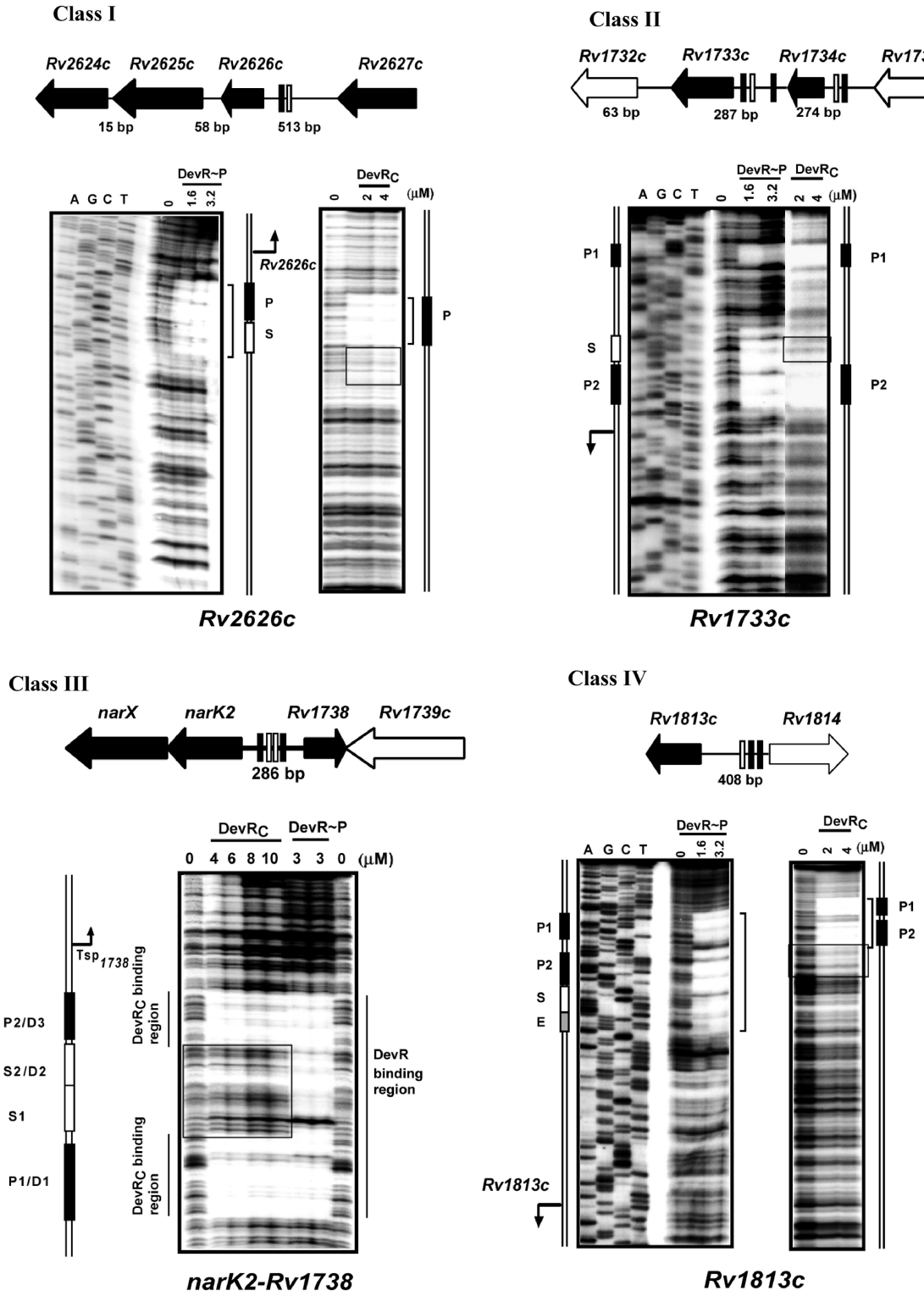


Figure 2. Representive DNase I footprints at promoter regions of *Rv2626*c (Class I), *Rv1733*c (Class II), *narK2-Rv1738* (Class III) and *Rv1813c* (Class IV). DNase I footprinting analyses of other members of each Class are shown in Supplementary Figure S2–S5. The genomic organization of each gene and its operon in Mtb is shown at the top where DevR regulon genes are indicated by black arrows. 'P' and 'S' refer to primary (black box) and secondary (white box) Dev box, respectively. Bent arrow indicates the direction and predicted position of translational start codon. Dideoxy sequencing reactions using the same primer and DNA template are also shown. The rectangle (black) marked inside the footprint indicates the box that was not bound by$DevR_C$.

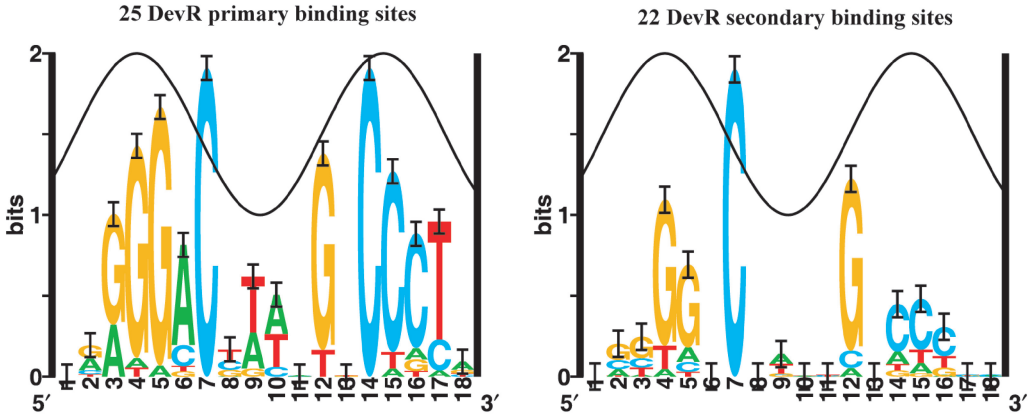


Figure 3. Sequence logos of DevR primary and secondary binding sites. The height of each letter is proportional to the frequency of base and the height of the letter stack is conservation in bits at that position. Error bars are shown at the top of the stacks. The sine wave represents the accessibility of a face of DNA (B-form, 10.6 bases of helical pitch) with the major groove centered at positions 4 and 14.6.

regained at higher concentration which shows that DevR binding to these sites is mutually cooperative (Supplementary Figure S4, panel B). Both of the secondary sites (S1 and S2) at *narK2-Rv1738* divergent promoter were not protected by $DevR_C$ (Figure 2). The binding defect at the mutated S1 site was associated with a functional defect too; introduction of the above-described mutations in pnarK2 GFP reporter plasmid (pnarK2 mut-S1) substantially reduced inducible *narK2* promoter activity by 78% (from 1208 ± 83 RFU/OD with wt pnarK2 to 265 ± 57 with pnarK2 mut-S1). In contrast, only a marginal 15% reduction in *Rv1738* divergent promoter activity was observed with mut-S1 (from 10 610 ± 88 RFU/OD with wt p1738 to 9040 ± 72 with p1738 mut-S1). These results conclusively establish the newly identified S1 box to be genuine DevR binding site and to be required for full hypoxic induction of the *narK2* promoter (which is proximal to it) and to some extent of the *Rv1738* promoter (which is distal to it). We conclude that the full induction of *narK2-Rv1738* promoters requires DevR interaction with all four boxes in the intergenic region.

Class IV regulon promoters have the most complex structure. They contain not only primary and secondary DevR binding sites but also display an extended DNase I protected region referred to as E (Figure 2, Supplementary Figure S1). The extended protected regions were always located adjacent to a secondary or primary site (Table 1, Figure 2, Supplementary Figures S1 and S5) and were from 13- to 34-bp long in *Rv1734c*, *Rv1813c* and *Rv3127* promoters. Binding to the E regions is likely due to high cooperative interaction that enables protein binding even to highly degenerate sites, as observed previously also in case of PhoP in *Streptomyces coelicolor* (34). The sequence walker method was employed to detect binding sites in the footprinted regions of these DNAs. Interestingly, the E regions in all three promoters showed two possible Dev box-like sequences and they possessed the highly conserved nucleotides like G_5 and C_7 (represented by walkers, gray rectangular boxes, Figure 4 and Supplementary Figure S6).

Tandem arrangement of Dev boxes is essential for DevR-mediated transcription

One of the most interesting findings of this study is that binding sites are arranged in an adjacent manner in the promoter regions of all DevR-regulated genes (with the exception of *Rv1733c*), suggesting that this arrangement holds the key for DevR-mediated gene activation. We assessed the importance of spacing by analyzing the *tgs1-Rv3131* promoters which belong to the Class I category. Here, the primary binding site P is placed 3-bp apart from the secondary binding site S and in the same helical phase (−42.5 and −63.5, Figure 6). We showed recently that cooperative DevR interaction with both sites is essential to activate divergent transcription (21). The functional importance of tandem and helical phase arrangement of binding sites was assessed by introducing 5, 10 or 15 bp of DNA sequences (corresponding to 0.5, 1 or 1.5 helical turns, Figure 6A) to alter the spacing between the P and S sites. When the P and S boxes were out of phase (0.5 and 1.5 turn in pTGS + 5 and pTGS + 15, respectively), DevR interacted with the P box but binding to the S site was abolished (Figure 6B) and was accompanied by a drastic reduction of *tgs1* promoter activity to ∼9% of WT promoter activity and complete abrogation of *Rv3131* promoter activity (Figure 6C). Restoration of the helical phase between P and S sites, i.e. insertion of 10-bp sequence, partially rescued cooperative DevR binding (Figure 6B) and also promoter activity in pTGS + 10 (18%) but not in p3131 + 10 (Figure 6C). There are two possible reasons for the abrogation of p3131 + 10 activity; first, a failure to recruit DevR cooperatively to the TSP-proximal box (S site) that is distanced by 10 bp in the mutant construct and second, it has an intrinsically weak promoter that supports minimal independent engagement of RNA polymerase. Taken together, these results show that proximity of two Dev boxes is essential for maximal binding and gene activation. Our results also suggest the importance of proper helical phasing in cooperative binding.

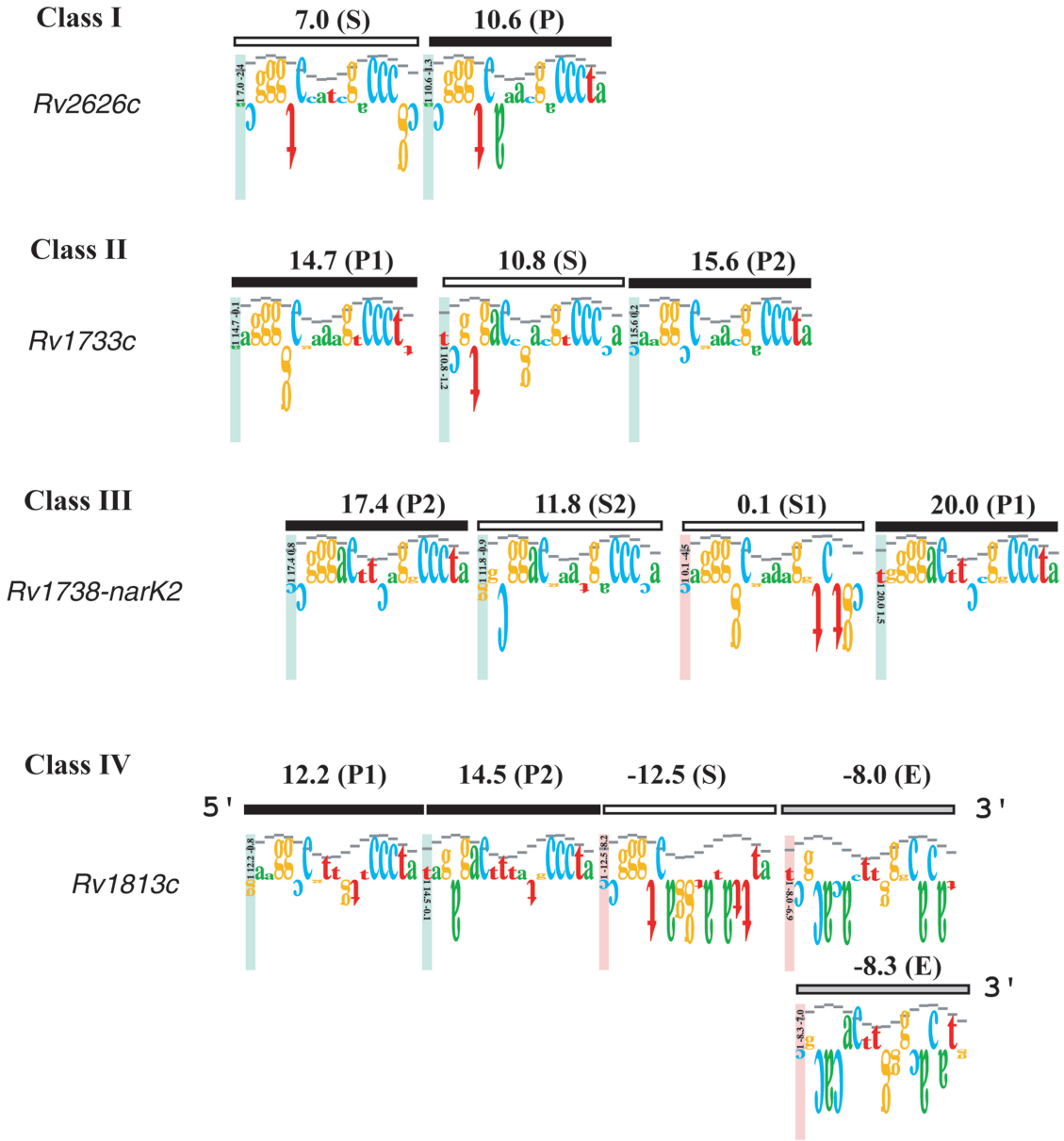
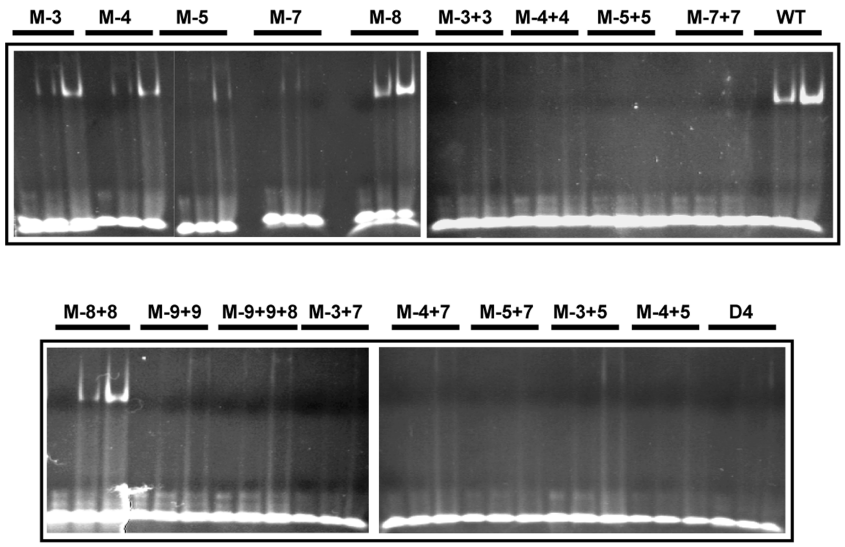


Figure 4. Computational analysis of Dev boxes belonging to representative Classes I–IV promoters using sequence walker. Primary and secondary boxes are in black and white, respectively. Extended sites (E) are in gray. The individual information (R_i) in bits is shown above each box. The height of the letters is the information content in bits (upper edge at +2 bits and lowest edge at −4 bits), and represents the contribution of each base to the conservation of the sequence.

DevR regulon genes are temporally regulated under hypoxic conditions

The relation between promoter classes and their activation response was assessed next by a temporal analysis of 27 DevR regulon promoters controlling the expression of more than 43 genes of the regulon (arranged in operons) using GFP reporter assay over a period of 7 days. The standing hypoxia model was used wherein standing of aerobic cultures creates a gradual local hypoxic environment in settled bacteria. This model is widely used to understand the hypoxic response of Mtb (19,31,35–37). Considering the variations in binding site arrangements, it is not surprising that the DevR regulon promoters are not synchronously induced, but rather their activation occurs in a temporal fashion (Table 2 and Supplementary Figure S7). All the induced genes were placed in three groups based on their temporal response; two promoters of Classes II and III were induced earliest at 4 h (*hspX* and *Rv1738*, 'early'), 16 promoters from Classes I, II and III were induced at 6/8 h ('intermediate', including *Rv3134c*, which are divergent to the 'early' induced genes) and five 'late' promoters were induced at 12/24 h. Four promoters, namely, *Rv0572c*, *Rv1812c*, *Rv1734c* and *Rv3126c*, were not hypoxia inducible and our observations with regard to these four promoters are consistent with some previous reports (9,31,35,38). For most of the genes, hypoxic



Oligonucleotides used in EMSA		Relative DevR binding activity(%)
Name	Sequence (5'-3')	
	1 2 3 4 5 6 7 8 9 9 8 7 6 5 4 3 2 1	
WT	T A G G G C C G G A A G T C C C C A	100
M-3	— T —	38
M-4	— C —	25
M-5	— C —	10
M-7	— G —	7
M-8	— A —	95
M-3+3	— T — G —	0
M-4+4	— C — G —	0
M-5+5	— C — G —	0
M-7+7	— G — C —	0
M-8+8	— A — G —	85
M-9+9	— C G —	0
M-9+9+8	— C G G —	0
M-3+7	— T — G —	0
M-4+7	— C — G —	0
M-5+7	— C — G —	0
M-3+5	— T — C —	0
M-4+5	— C C —	0
D4	T G T G G A G G T T A C A C C C C A	0

Figure 5. Nucleotides critical for DevR binding. EMSA was carried out with DevR (2.3 and 4.6 μM final concentration) and 18-bp oligonucleotide carrying WT or mutant *narK2* P1 Dev box. The table shows the percent DevR binding to mutant oligonucleotides relative to DevR binding with wild type *narK2* P1 Dev box (100%).

induction peaked at 72–120 h in this assay set up (Table 2 and Supplementary Figure S7). Those genes that were induced early were also the most highly induced and this may reflect their importance for an effective dormancy response (Table 3).

Importantly, the induction of *hspX* and *Rv1738* precedes that of *devRS* (transcribed from *Rv3134c* promoter) and suggests that their induction merely requires activation of existing DevR molecules, and no new synthesis is necessary (see 'Discussion' section). The other genes of the regulon are induced along with or after induction of the *Rv3134c* promoter. This inducible promoter directs the synthesis of DevRS (from the *Rv3134c-devRS* operon) under hypoxia and it is autoregulated (19). Thus, the induction and sustained expression of the regulon can be explained by the increase in the intracellular levels of DevR due to positive autoregulation.

Strikingly, the temporal expression or the magnitude of induction of regulon genes does not appear to be dependent entirely on the number of DevR binding sites or their arrangement in the promoter regions (Table 3). For example, between *Rv0569* and *Rv0574c* which are both Class I promoters, the former is strongly induced early and strongly (at 6 h and 212-fold), while the latter is induced late and modestly (at 24 h and 3.5-fold, Table 2). Another example is that of the Class II *hspX-Rv2032* promoters which share three binding sites in their intergenic promoter region. While the former is induced early and strongly (at 4 h and ~300-fold), the latter is induced later at 8 h and to ~72-fold. To determine whether temporal regulation is related to the differential affinity of DevR for various target sites, gel-shift assays were performed with some temporally induced promoters (Supplementary Figure S8). We find that DevR interacts with promoters of all the groups (early, intermediate and late) at nearly similar protein concentrations (50–100 nM), suggesting that the temporal induction of regulon genes is not likely to be a sole attribute of the affinity of DevR for target promoters and likely involves other factors (see 'Discussion' section).

DISCUSSION

This study aimed at obtaining a comprehensive understanding of the DevR regulon transcriptional response. Although the role of DevR in mediating activation of a few target gene promoters is quite well characterized (19–21), regulatory control of most DevR-regulated genes remains vastly underexplored. Using the complementary approaches of DevR–DNA interaction studies, computational analysis and temporal measurements of gene expression, we defined the properties of the DevR regulon activation response in Mtb. By gel-shift analysis of Dev boxes of varying lengths, we defined a DevR binding site to be at least 18-bp long. By DNase I footprinting of all regulon promoters, we mapped hitherto unknown additional binding sites in most promoters. By computational analysis, we defined the logos for primary and secondary binding sites. By binding and expression studies, we showed that a minimum of two properly spaced binding sites in helical phase are required for optimum induction. Finally, we sought to determine the relationship between binding sites and the timing and magnitude of the activation response.

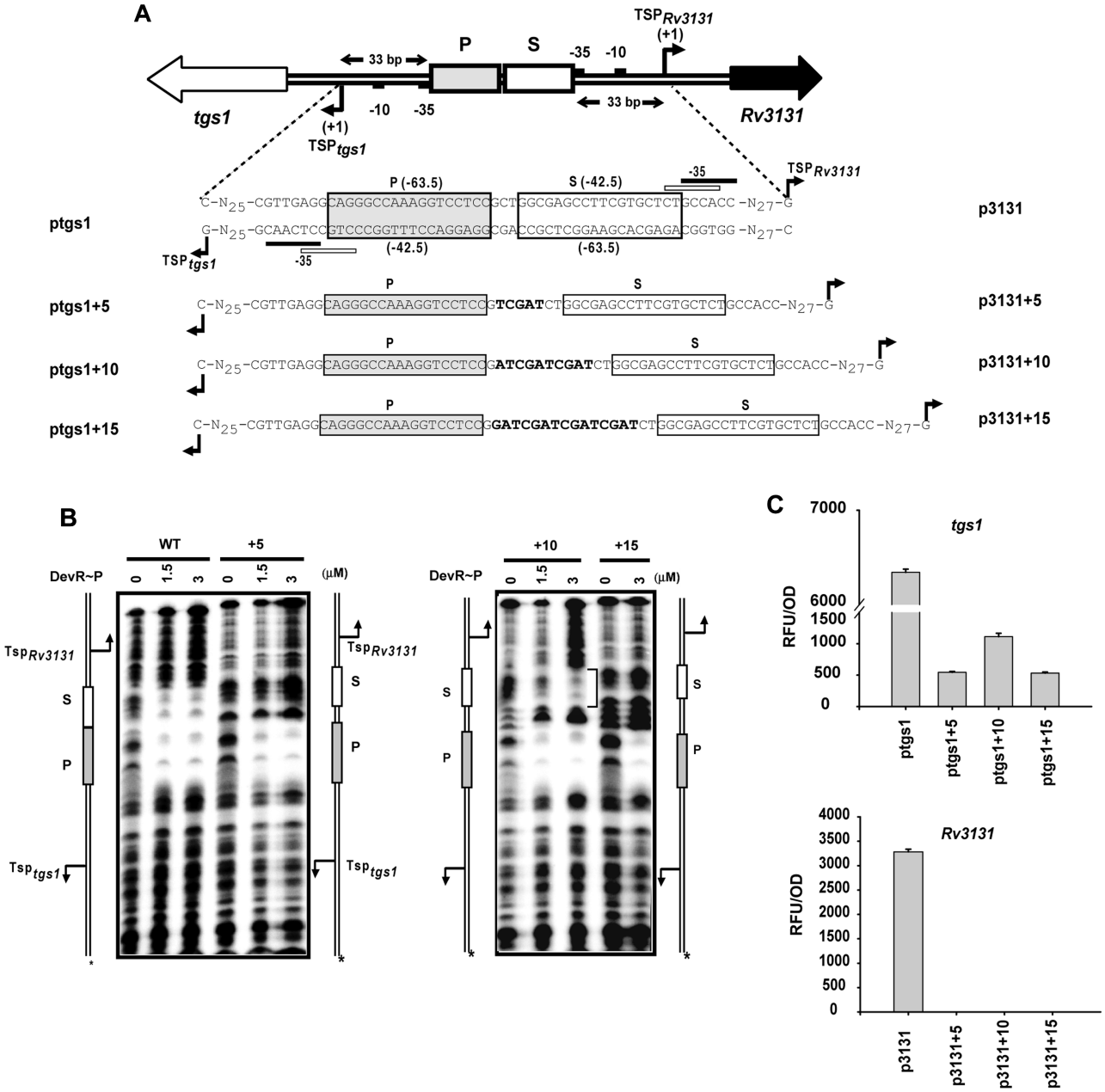


Figure 6. Adjacent location and helical phasing of Dev boxes are essential for maximal induction. (**A**) Schematic representation *tgs1-Rv3131* divergent promoter. (**B**) DNase I footprinting with DevR and WT, +5, +10 and +15 DNA. (**C**) GFP reporter assay of WT and altered spacer variant promoter constructs (+5, +10 and +15). Shown are the average values (± standard deviation) of GFP fluorescence (RFU/OD) from three experiments each performed in triplicate.

Consensus sequences are thought to present a somewhat misleading view of binding sites as they frequently fail to identify genuine binding sites or predict sites where there are none (39). This is because they represent prominent bases at each position that are often absent or less prominent in the natural context. The sequence logo on the other hand allows plotting the conservation across all the positions in the set of aligned binding sites and provides a quantitative measure for affinity of the binding site. The total information ($R_{sequence}$) for a perfect Dev box is 22.29 and interestingly, such a sequence is not present in the entire Mtb genome. The G_4, G_5 and C_7 nucleotides are highly conserved (>1 bit) in both halves of the binding site (Figure 3) and their functional relevance was established by gel-shift analysis. Taking together the sequence logo and the results of previous studies (11,19–21), we can say that G_3, G_4, G_5 and C_7 bases in both strands of the box are the most critical nucleotides which are recognized by 'direct' read out mechanism (22). Furthermore, among these four bases, G_5 and C_7 were most important and mutating them individually reduced DevR–DNA interaction by ~90% (Figure 5).

Table 2. Temporal expression of DevR regulon promoters (up to 168 h)

Genes	Class[a]	Hours									
		0	4	6	8	12	24	48	72	120	168
Rv1738	III	0	234[b]	1109	1567	2098	5433	12456	14678	10945	9243
hspX/Rv2031c	II	0	334	886	1632	2343	4588	9987	11903	10982	10765
Rv0079	II	0	0	124	518	1418	4361	6953	8506	9280	7753
Rv0569	I	0	0	256	614	1505	3772	5780	8497	8266	8694
Rv1733c	II	0	0	223	397	648	1217	2053	3998	3146	2811
ctpF/Rv1997	I	0	0	344	566	1248	1463	2191	3215	3048	3435
Rv2005c	I	0	0	467	758	1772	2663	2596	6088	4300	4173
fdxA/Rv2007c	I	0	0	455	1259	1209	6085	7447	10534	8030	8149
Rv2627c	III	0	0	804	1248	1362	2290	3012	6924	5207	4534
Rv2628	I	337	345	734	1716	1652	2700	4914	8986	6730	5365
tgs1/Rv3130c	I	0	0	677	1452	1318	4695	5922	10699	7391	6256
Rv3131	I	0	0	223	534	878	1311	3360	7126	4256	3851
Rv3134c	I	0	0	445	1284	1956	3017	4273	8070	6837	5880
narK2/Rv1737c	III	0	0	0	514	734	1411	1200	1473	1154	1159
acg/Rv2032	II	0	0	0	855	1131	1508	2219	2899	3398	2973
Rv2623	I	0	0	0	1345	1332	2706	3567	6170	5511	4072
Rv2626c	I	0	0	0	398	772	1265	2934	4887	4993	4099
Rv3127	IV	0	0	0	320	408	728	996	2098	1996	1576
Rv1813c	IV	0	0	0	0	357	1311	1300	2509	2649	2593
Rv0571c	I	0	0	0	0	0	56	67	84	159	196
Rv0574c	I	0	0	0	0	0	74	83	144	105	188
Rv1996	I	0	0	0	0	0	103	493	1124	795	906
otsB/Rv2006	I	0	0	0	0	0	70	220	294	204	219
Rv0572c		136	ND	ND	ND	ND	ND	163	143	ND	ND
Rv3126c		0	0	0	0	0	0	0	0	0	0
Rv1734c	IV	0	0	0	0	0	0	0	0	0	0
Rv1812c		1021	ND	ND	ND	ND	ND	865	946	ND	ND

Data of a representative experiment out of three experiments (whose results vary <10%) is shown.
The heat map (green to red) indicates zero expression to maximum expression as measured by GFP fluorescence.
DevR regulon genes are classified into three groups based on the time when induction was first observed: 'early', 'intermediate' and 'late' genes (in blue shades).
[a]Classes I, IV, etc. refer to promoter architecture (see Table 1).
[b]Reporter GFP fluorescence in RFU/OD units (see 'Materials and Methods' section for details).
ND, not determined.

The simultaneous mutation of two critical nucleotides in either of the two halves of the palindrome abolished binding completely. Interestingly, on mutating the nucleotide at ninth position, DevR–DNA interaction was completely abolished although this position is not highly conserved but usually occupied by A or T nucleotide (Figure 3). This nucleotide is shown not to interact directly with $DevR_C$ (22), and it is possible that these nucleotides are recognized by 'indirect' read out mechanism as observed in case of *E. coli* cAMP receptor protein (CRP) where $T:A_6$ is recognized by 'indirect' read out mechanism and replacement of $T:A_6$ in CRP binding site with $C:G_6$ causes an 80-fold reduction in CRP affinity by increasing the free energy required to bend the DNA (40,41). Further studies in this regard would help to understand the underlying mechanism of the binding defect observed here.

This study reveals several novel and universal features of DevR-mediated activation. DevR binding to the primary sites is mediated primarily by the strength of DevR–DNA interaction. We show that secondary sites are ubiquitous at known DevR regulon promoters and contrary to previous predictions, none of studied target promoters feature a single Dev box. The protection of the secondary site appears to be cooperative since it is severely reduced by distancing it from the primary site (this study) or by destroying the neighboring primary site (20,21). Thus, binding to the secondary sites can be seen as being less dependent on DNA sequence and highly dependent on cooperative interactions with neighboring DevR molecules. Some of the promoters were observed to have three or even more binding sites suggesting the likelihood of highly cooperative interactions between as many DevR dimer molecules. Previously, we have shown

Table 3. Induction profile of DevR regulon promoters (at 72 h)

Gene name	Class	Probable gene function	Fold induction (72 h)[a]	No. of Dev boxes	Cloned promoter coordinates[b]
(A) Early genes (induced at 4 h)					
Rv1738	III	CHP	367	4	−203 to +74
hspX/acr	II	Alpha-crystallin chaperone protein	297	3	−132 to +48
(B) Intermediate genes (induced at 6/8 h)					
Rv0079	II	Probable cytoplasmic translation factor	212	3	−178 to +63
Rv0569	I	CHP	212	2	−198 to +20
Rv1733c	II	Probable transmembran-e protein	100	3	−200 to +120
Rv1997/ctpF	I	Metal cation-transporting ATPase	80	2	−195 to +9
Rv2005c	I	Probable ATP binding protein, USP	152	2	−195 to +27
Rv2007c/fdxA	I	probable ferredoxin	263	2	−200 to +30
Rv2627c	III	CHP	173	4	−339 to +42
Rv2628	I	CHP	224	2	−351 to +120
Rv3130c/tgs1	I	Triacylglycerol synthase	267	2	−143 to +45
Rv3131	I	Probable nitroreductase	178	2	−150 to +38
Rv3134c	I	USP	201	2	−348 to +90
Rv1737c/narK2	III	Nitrate/nitrite transporter	37	4	−220 to +57
Rv2032/acg	II	Probable nitroreductase	72	3	−215 to −1
Rv2623	I	ATP binding protein (USP)	154	2	−200 to +90
Rv2626c	I	CHP	122	2	−280 to +10
Rv3127	IV	Probable nitroreductase	52	1 + E	−300 to +112
(C) Late genes (induced at 12/24 h)					
Rv1813c	IV	CHP	62	3 + E	−351 to +18
Rv0571c	I	CHP	4	2	−178 to +90
Rv0574c	I	CHP	3.5	2	−183 to +10
Rv1996	I	USP	28	2	−140 to +19
Rv2006/otsB	I	Probable trehalose-6-phosphate phosphatase	7.5	2	−145 to +77
(D) Genes not induced					
Rv0572c		HP		0	−200 to +30
Rv1812c		Probable dehydrogenase		0	−300 to +30
Rv1734c-1	IV	CHP		2 + E	−272 to +14
Rv1734c-2				2 + E	−272 to +90
Rv3126c		HP		0	−136 to +276

[a]Fold induction at 72 h is calculated from data of a representative experiment out of three independent experiments and results of which deviate <10%.
[b]Refers to promoter coordinates with respect to their corresponding putative translational start sites, all constructs are transcriptional fusions with respect to GFP.
USP, universal stress protein; CHP, conserved hypothetical protein; HP, hypothetical protein.

at some DevR targets that cooperative interaction of DevR with two or more sites synergistically activates gene transcription (19–21). These and the present studies collectively suggest that cooperative binding and synergistic activation could be the universal approach used by DevR to bring about maximal induction of the regulon.

All DevR regulon promoters were placed into four classes based on the configuration of binding sites. The major category, Class I, includes 14 promoters (controlling ~25 genes); each having two Dev boxes in its promoter region in tandem arrangement. This arrangement of the binding sites is functionally important because insertion of 5 or 15 bp between the boxes resulted in absence of S site binding at the *tgs1-Rv3131* Class I promoter. Cooperative interaction is apparently facilitated by spacing of the binding sites at a distance that permits interaction with DevR dimer molecules on the same face of the DNA double helix. Supportive evidence comes from the partial restoration of DevR binding to the S site in the *tgs1-Rv3131* promoter variant carrying a 10-bp in-phase insertion between P and S sites. The layout of the *Rv2005-otsB* promoter resembles the *tgs1-Rv3131* promoter and they could be regulated in a similar way. Thus, the salient features of Class I regulon promoters can be summarized; two binding sites are essential, they are tandemly placed to maintain them on the same face of DNA, a primary binding site with the conserved sequence is essential and considerable degeneracy is tolerated in the secondary site sequence provided that G_4, C_7 and G_{12} are conserved in it.

In Classes II, III and IV promoters that contain three or even four binding sites sometimes in a complex arrangement, a primary site binds to DevR and facilitates the cooperative binding of additional DevR molecules to the remaining sites even if they are poorly conserved and have low information content. As the information content is correlated with binding energy, it is likely that binding of DevR to tandem sites of varying affinity is mutually stabilized through protein–protein interactions between bound molecules as seen in the case of *Rv3134c, narK2* and *tgs1* promoters (19–21). These cooperative interactions are expected to decrease the total energy for DevR binding and may be a regulatory strategy for efficient gene induction. Other examples of Class II promoters, namely *hspX* and *Rv2032*, are likely to be

regulated like *narK2-Rv1738* promoters (20), which was shown to contain an additional fourth site, S1 (Figure 2C, Supplementary Figure S4). Along similar lines, it is possible that an additional degenerate binding site may be located in the 17-bp region between P1 and S sites in the *hspX-Rv2032* intergenic region (Supplementary Figures S1 and S3, panel B). *Rv1733c* promoter is exceptional as it contains an isolated Dev box (P1) placed 63-bp upstream of the two tandem Dev boxes (S-P2). It would be interesting to investigate the role of the isolated Dev box in transcription, especially since it is placed ~6 helical turns apart and suggests the possibility of DNA looping mediated by DevR interaction with the P1 and downstream binding sites. Class IV promoters include *Rv1813c*, *Rv1734c* and *Rv3127* and have a complex structure that includes at least one binding site (E) with very low to poor information content in an extended footprinted region, E (Table 1). Interestingly, and consistent with previous observations (31), *Rv1734c* was not induced under hypoxic conditions (Table 2). Possible reasons may be the absence of an active promoter or blocking of transcription due to DevR bound at the E site. Other genes, namely, *Rv0572*, *Rv1812c* and *Rv3126c* are also not induced under hypoxia (Tables 2 and 3, Supplementary Figure S7) and this property is consistent with the absence of Dev box-like sequences in their upstream regions.

Temporal analysis of promoter activity has provided valuable insights into DevR regulon activity and also raised intriguing questions. The genes to be reproducibly induced the earliest (at 4 h in this model) were *hspX* and *Rv1738*. Notably, their induction preceded that of *devRS* (transcribed from the *Rv3134c* promoter). The rapid activation of these genes implies a very early role for them during initiation of the dormancy response. The pattern of gene induction observed in the present study highlights important aspects of the DevR regulon response. First, mild hypoxia is adequate to trigger the DevRS-DosT signaling pathway and to activate DevR (through phosphorylation). Second, the induction of *hspX* and *Rv1738* is mediated by the activation of existing DevR molecules (from basal-level expression) and not those produced post-induction The presence of a high density of binding sites including two primary sites most likely enables cooperative binding of DevR at basal protein concentrations and very early recruitment of RNA polymerase to bring about rapid gene induction. Third, the subsequent activation of other promoters including *Rv3134c-devRS* suggests that a higher level of DevR~P is required for this temporal program. This requirement can be met by positive autoregulation which results in an increase of intracellular DevR protein concentration under inducing conditions (13,19,36). Temporal expression suggests that target genes may require a threshold level of DevR~P for activation and therefore their expression profile is switch-like; transcription is absent below this threshold and induced above it. The occurrence of a threshold response has been suggested by a recent study where we showed that DevR regulon activation and hypoxia adaptation is compromised by diversion of the activating phosphosignal from DevR (13). Thus, although 'adequate' intracellular DevR~P levels are crucial to switching the regulon genes 'on', the lack of a direct relationship between the nature and number of binding sites on one hand and timing and magnitude of the activation response on the other is striking.

We show that DevR regulon genes are temporally and differentially regulated but temporal regulation does not appear to correlate completely and exactly with the affinity of DevR for target genes. Thus, promoters appear to bind DevR with equivalent affinity (Supplementary Figure S8), yet show differences in temporal expression (Table 2 and Supplementary Figure S7). The absence of a clear relationship between binding sites and the induction response suggest us to propose that the temporal nature of the induction response is not a sole attribute of the affinity of DevR for various target promoters and may involve other factors such as variation in intrinsic promoter strengths, participation of other *cis* elements and *trans*-acting factors, or proteins other than DevS/DosT that may influence the phosphorylation (activation) state of DevR. For example, the involvement of some negative *cis*/*trans* regulators was suggested for modulating *narK2* expression (20,42). Furthermore, recent studies have suggested that DevR regulon expression could be fine tuned through MprA, PhoP and PknH (42–45). However, the exact molecular mechanisms involved are yet unknown. Such a complex interplay of multiple regulators is well understood in the case of CRP in *E. coli* wherein co-regulators like MelR, AraC, CytR and GalR modulate the expression of CRP-controlled promoters (46–49). Complex interactions between multiple regulators at individual promoters could explain the lack of correlation between CRP-dependent promoter activity and the quality of the CRP binding site (50). It would be exciting and relevant to decipher the apparently complex regulation of the DevR regulon.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

J.S.T. is thankful to the Department of Biotechnology, Government of India for funding and for the Tata Innovation Fellowship. A.S. is thankful to the Department of Biotechnology, Government of India for its Centre for Molecular Medicine. S.C. is thankful to the Council of Scientific and Industrial Research, Government of India for research fellowship during this work, D.S. is grateful to Indian National Science Academy (INSA) for the research fellowship. Funding for open access charge: National Institute of Immunology.

Conflict of interest statement. None declared.

REFERENCES

1. WHO. *Tuberculosis, Fact sheet No. 104*, November 2010. Available at http://www.who.int/mediacentre/factsheets/fs104/en/index.html.

2. Black,G.F., Thiel,B.A., Ota,M.O., Parida,S.K., Adegbola,R., Boom,W.H., Dockrell,H.M., Franken,K.L., Friggen,A.H., Hill,P.C. *et al.* (2009) Immunogenicity of novel DosR regulon-encoded candidate antigens of Mycobacterium tuberculosis in three high-burden populations in Africa. *Clin. Vaccine Immunol.*, **16**, 1203–1212.
3. Leyten,E.M., Lin,M.Y., Franken,K.L., Friggen,A.H., Prins,C., van Meijgaarden,K.E., Voskuil,M.I., Weldingh,K., Andersen,P., Schoolnik,G.K. *et al.* (2006) Human T-cell responses to 25 novel antigens encoded by genes of the dormancy regulon of Mycobacterium tuberculosis. *Microbes Infect.*, **8**, 2052–2060.
4. Roupie,V., Romano,M., Zhang,L., Korf,H., Lin,M.Y., Franken,K.L., Ottenhoff,T.H., Klein,M.R. and Huygen,K. (2007) Immunogenicity of eight dormancy regulon-encoded proteins of Mycobacterium tuberculosis in DNA-vaccinated and tuberculosis-infected mice. *Infect. Immun.*, **75**, 941–949.
5. 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.
6. Voskuil,M.I., Schnappinger,D., Visconti,K.C., Harrell,M.I., Dolganov,G.M., Sherman,D.R. and Schoolnik,G.K. (2003) Inhibition of respiration by nitric oxide induces a Mycobacterium tuberculosis dormancy program. *J. Exp. Med.*, **198**, 705–713.
7. Wayne,L.G. and Sohaskey,C.D. (2001) Nonreplicating persistence of mycobacterium tuberculosis. *Annu. Rev. Microbiol.*, **55**, 139–163.
8. Kumar,A., Deshane,J.S., Crossman,D.K., Bolisetty,S., Yan,B.S., Kramnik,I., Agarwal,A. and Steyn,A.J. (2008) Heme oxygenase-1-derived carbon monoxide induces the Mycobacterium tuberculosis dormancy regulon. *J. Biol. Chem.*, **283**, 18032–18039.
9. Shiloh,M.U., Manzanillo,P. and Cox,J.S. (2008) Mycobacterium tuberculosis senses host-derived carbon monoxide during macrophage infection. *Cell Host. Microbe*, **3**, 323–330.
10. Taneja,N.K., Dhingra,S., Mittal,A., Naresh,M. and Tyagi,J.S. (2010) Mycobacterium tuberculosis transcriptional adaptation, growth arrest and dormancy phenotype development is triggered by vitamin C. *PLoS One*, **5**, e10860.
11. 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.
12. Boon,C. and Dick,T. (2002) Mycobacterium bovis BCG response regulator essential for hypoxic dormancy. *J. Bacteriol.*, **184**, 6760–6767.
13. Majumdar,S.D., Sharma,D., Vashist,A., Kaur,K., Taneja,N.K., Chauhan,S., Challu,V.K., Ramanathan,V.D., Balasangameshwara,V., Kumar,P. *et al.* (2010) Co-expression of DevR and DevR(N)-Aph proteins is associated with hypoxic adaptation defect and virulence attenuation of Mycobacterium tuberculosis. *PLoS One*, **5**, e9448.
14. Malhotra,V., Sharma,D., Ramanathan,V.D., Shakila,H., Saini,D.K., Chakravorty,S., Das,T.K., Li,Q., Silver,R.F., Narayanan,P.R. *et al.* (2004) Disruption of response regulator gene, devR, leads to attenuation in virulence of Mycobacterium tuberculosis. *FEMS Microbiol. Lett.*, **231**, 237–245.
15. Parish,T., Smith,D.A., Kendall,S., Casali,N., Bancroft,G.J. and Stoker,N.G. (2003) Deletion of two-component regulatory systems increases the virulence of Mycobacterium tuberculosis. *Infect. Immun.*, **71**, 1134–1140.
16. Reed,M.B., Gagneux,S., Deriemer,K., Small,P.M. and Barry,C.E. 3rd (2007) The W-Beijing lineage of Mycobacterium tuberculosis overproduces triglycerides and has the DosR dormancy regulon constitutively upregulated. *J. Bacteriol.*, **189**, 2583–2589.
17. Leistikow,R.L., Morton,R.A., Bartek,I.L., Frimpong,I., Wagner,K. and Voskuil,M.I. (2010) The Mycobacterium tuberculosis DosR regulon assists in metabolic homeostasis and enables rapid recovery from nonrespiring dormancy. *J. Bacteriol.*, **192**, 1662–1670.
18. Converse,P.J., Karakousis,P.C., Klinkenberg,L.G., Kesavan,A.K., Ly,L.H., Allen,S.S., Grosset,J.H., Jain,S.K., Lamichhane,G., Manabe,Y.C. *et al.* (2009) Role of the dosR-dosS two-component regulatory system in Mycobacterium tuberculosis virulence in three animal models. *Infect. Immun.*, **77**, 1230–1237.
19. Chauhan,S. and Tyagi,J.S. (2008) Cooperative binding of phosphorylated DevR to upstream sites is necessary and sufficient for activation of the Rv3134c-devRS operon in Mycobacterium tuberculosis: implication in the induction of DevR target genes. *J. Bacteriol.*, **190**, 4301–4312.
20. Chauhan,S. and Tyagi,J.S. (2008) Interaction of DevR with multiple binding sites synergistically activates divergent transcription of narK2-Rv1738 genes in Mycobacterium tuberculosis. *J. Bacteriol.*, **190**, 5394–5403.
21. Chauhan,S. and Tyagi,J.S. (2009) Powerful induction of divergent tgs1-Rv3131 genes in Mycobacterium tuberculosis is mediated by DevR interaction with a high-affinity site and an adjacent cryptic low-affinity site. *J. Bacteriol.*, **191**, 6075–6081.
22. Wisedchaisri,G., Wu,M., Rice,A.E., Roberts,D.M., Sherman,D.R. and Hol,W.G. (2005) Structures of Mycobacterium tuberculosis DosR and DosR-DNA complex involved in gene activation during adaptation to hypoxic latency. *J. Mol. Biol.*, **354**, 630–641.
23. Gautam,U.S., Chauhan,S. and Tyagi,J.S. (2011) Determinants outside the DevR C-terminal domain are essential for cooperativity and robust activation of dormancy genes in Mycobacterium tuberculosis. *PLoS One*, **6**, e16500.
24. Gupta,R.K., Chauhan,S. and Tyagi,J.S. (2011) K182G substitution in DevR or C8G mutation in the Dev box impairs protein-DNA interaction and abrogates DevR-mediated gene induction in Mycobacterium tuberculosis. *FEBS J.*, April 22 (doi:10.1111/j.1742-4658.2011.08130.x; Epub ahead of print).
25. Bagchi,G., Chauhan,S., Sharma,D. and Tyagi,J.S. (2005) Transcription and autoregulation of the Rv3134c-devR-devS operon of Mycobacterium tuberculosis. *Microbiology*, **151**, 4045–4053.
26. Valdivia,R.H., Hromockyj,A.E., Monack,D., Ramakrishnan,L. and Falkow,S. (1996) Applications for green fluorescent protein (GFP) in the study of host-pathogen interactions. *Gene*, **173**, 47–52.
27. Sambrook,J. and Russell,D.W. (2001) *Molecular Cloning: A Laboratory Manual*, 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
28. Schneider,T.D. and Stephens,R.M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.*, **18**, 6097–6100.
29. Schneider,T.D. (1997) Sequence walkers: a graphical method to display how binding proteins interact with DNA or RNA sequences. *Nucleic Acids Res.*, **25**, 4408–4415.
30. Roback,P., Beard,J., Baumann,D., Gille,C., Henry,K., Krohn,S., Wiste,H., Voskuil,M.I., Rainville,C. and Rutherford,R. (2007) A predicted operon map for Mycobacterium tuberculosis. *Nucleic Acids Res.*, **35**, 5085–5095.
31. 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.
32. Schneider,T.D. (1997) Information content of individual genetic sequences. *J. Theor. Biol.*, **189**, 427–441.
33. Schneider,T.D. (1999) Measuring molecular information. *J. Theor. Biol.*, **201**, 87–92.
34. Sola-Landa,A., Rodriguez-Garcia,A., Apel,A.K. and Martin,J.F. (2008) Target genes and structure of the direct repeats in the DNA-binding sequences of the response regulator PhoP in Streptomyces coelicolor. *Nucleic Acids Res.*, **36**, 1358–1368.
35. Kendall,S.L., Movahedzadeh,F., Rison,S.C., 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*, **84**, 247–255.
36. Saini,D.K., Malhotra,V., Dey,D., Pant,N., Das,T.K. and Tyagi,J.S. (2004) DevR-DevS is a bona fide two-component system of Mycobacterium tuberculosis that is hypoxia-responsive in the absence of the DNA-binding domain of DevR. *Microbiology*, **150**, 865–875.
37. Sherman,D.R., Voskuil,M., Schnappinger,D., Liao,R., Harrell,M.I. and Schoolnik,G.K. (2001) Regulation of the Mycobacterium tuberculosis hypoxic response gene

encoding alpha-crystallin. *Proc. Natl Acad. Sci. USA*, **98**, 7534–7539.
38. Zhang,Y., Hatch,K.A., Wernisch,L. and Bacon,J. (2008) A Bayesian change point model for differential gene expression patterns of the DosR regulon of Mycobacterium tuberculosis. *BMC Genomics*, **9**, 87.
39. Schneider,T.D. (2002) Consensus sequence Zen. *Appl. Bioinformatics*, **1**, 111–119.
40. Chen,S., Gunasekera,A., Zhang,X., Kunkel,T.A., Ebright,R.H. and Berman,H.M. (2001) Indirect readout of DNA sequence at the primary-kink site in the CAP-DNA complex: alteration of DNA binding specificity through alteration of DNA kinking. *J. Mol. Biol.*, **314**, 75–82.
41. Chen,S., Vojtechovsky,J., Parkinson,G.N., Ebright,R.H. and Berman,H.M. (2001) Indirect readout of DNA sequence at the primary-kink site in the CAP-DNA complex: DNA binding specificity based on energetics of DNA kinking. *J. Mol. Biol.*, **314**, 63–74.
42. Vasudeva-Rao,H.M. and McDonough,K.A. (2008) Expression of the Mycobacterium tuberculosis acr-coregulated genes from the DevR (DosR) regulon is controlled by multiple levels of regulation. *Infect. Immun.*, **76**, 2478–2489.
43. Chao,J.D., Papavinasasundaram,K.G., Zheng,X., Chavez-Steenbock,A., Wang,X., Lee,G.Q. and Av-Gay,Y. (2010) Convergence of Ser/Thr and two-component signaling to coordinate expression of the dormancy regulon in Mycobacterium tuberculosis. *J. Biol. Chem.*, **285**, 29239–29246.
44. Gonzalo-Asensio,J., Mostowy,S., Harders-Westerveen,J., Huygen,K., Hernandez-Pando,R., Thole,J., Behr,M., Gicquel,B. and Martin,C. (2008) PhoP: a missing piece in the intricate puzzle of Mycobacterium tuberculosis virulence. *PLoS One*, **3**, e3496.
45. Pang,X., Vu,P., Byrd,T.F., Ghanny,S., Soteropoulos,P., Mukamolova,G.V., Wu,S., Samten,B. and Howard,S.T. (2007) Evidence for complex interactions of stress-associated regulons in an mprAB deletion mutant of Mycobacterium tuberculosis. *Microbiology*, **153**, 1229–1242.
46. Belyaeva,T.A., Wade,J.T., Webster,C.L., Howard,V.J., Thomas,M.S., Hyde,E.I. and Busby,S.J. (2000) Transcription activation at the Escherichia coli melAB promoter: the role of MelR and the cyclic AMP receptor protein. *Mol. Microbiol.*, **36**, 211–222.
47. Irani,M.H., Orosz,L. and Adhya,S. (1983) A control element within a structural gene: the gal operon of Escherichia coli. *Cell*, **32**, 783–788.
48. Lobell,R.B. and Schleif,R.F. (1991) AraC-DNA looping: orientation and distance-dependent loop breaking by the cyclic AMP receptor protein. *J. Mol. Biol.*, **218**, 45–54.
49. Valentin-Hansen,P., Sogaard-Andersen,L. and Pedersen,H. (1996) A flexible partnership: the CytR anti-activator and the cAMP-CRP activator protein, comrades in transcription control. *Mol. Microbiol.*, **20**, 461–466.
50. Zheng,D., Constantinidou,C., Hobman,J.L. and Minchin,S.D. (2004) Identification of the CRP regulon using in vitro and in vivo transcriptional profiling. *Nucleic Acids Res.*, **32**, 5874–5893.
51. Dasgupta,N., Kapur,V., Singh,K.K., Das,T.K., Sachdeva,S., Jyothisri,K. and Tyagi,J.S. (2000) Characterization of a two-component system, devR-devS, of Mycobacterium tuberculosis. *Tuber Lung Dis.*, **80**, 141–159.