Rv3133c/*dosR* is a transcription factor that mediates the hypoxic response of *M. tuberculosis*

Heui-Dong Park[1,2], Kristi M. Guinn[1], Maria I. Harrell[1], Reiling Liao[1], Martin I. Voskuil[3], Martin Tompa[4], Gary K. Schoolnik[3] and David R. Sherman[1*]

[1]Departments of Pathobiology and [4]Computer Science and Engineering, University of Washington, Seattle, WA, 98195

[2]Current address: Department of Food Science and Technology, Kyungpook National University, Taegu 702-701, Korea.

[3]Division of Infectious Diseases and Geographic Medicine, Department of Medicine and Department of Microbiology and Immunology, Stanford Medical School, Stanford, CA 94305

For correspondence: dsherman@u.washington.edu; 206-221-5381; FAX 206-543-3873

Summary

Unlike many pathogens that are overtly harmful to their hosts, *Mycobacterium tuberculosis* can persist for years within humans in a clinically latent state. Latency is often linked to hypoxic conditions within the host. Among *M. tuberculosis* genes induced by hypoxia is a putative transcription factor, Rv3133c/DosR. We performed targeted disruption of this locus followed by transcriptome analysis of wild-type and mutant bacilli. Nearly all the genes powerfully regulated by hypoxia require Rv3133c/DosR for their induction. Computer analysis identified a consensus motif, a variant of which is located upstream of nearly all *M. tuberculosis* genes rapidly induced by hypoxia. Further, Rv3133c/DosR binds to the two copies of this motif upstream of the hypoxic response gene alpha-crystallin. Mutations within the binding sites abolish both Rv3133c/DosR binding as well as hypoxic induction of a downstream reporter gene. Also, mutation experiments with Rv3133c/DosR confirmed sequence-based predictions that the C-terminus is responsible for DNA binding and that the aspartate at position 54 is essential for function. Together these results demonstrate that Rv3133c/DosR is a transcription factor of the two-component response regulator class, and that it is the primary mediator of a hypoxic signal within *M. tuberculosis*.

Introduction

Tuberculosis is a global health emergency of staggering proportions. Worldwide, *Mycobacterium tuberculosis* (MTB) causes about 8 million new infections and two million deaths each year (Bloom and Small, 1998; Dye *et al.*, 1999). The remarkable success of MTB as a pathogen is closely associated with its ability to persist in humans for extended periods without causing disease. It is estimated that one-third of the world population, or about 1.9 billion people, harbors latent MTB infections (Dye *et al.*, 1999; Enarson and Murray, 1996), which can last for years or decades (Manabe and Bishai, 2000). This enormous reservoir of latent disease greatly complicates efforts at tuberculosis control.

Despite significant effort in recent years, progress has been slow in understanding the natural history of latent tuberculosis and reactivation (Parrish *et al.*, 1998). Important unresolved questions include the metabolic state of bacilli during latency, the role that metabolically dampened MTB may play in lengthening the time necessary for effective chemotherapy, and the nature of the bacterial genetic program and host responses that underlie long-term persistence. The failure of bacteria to increase in numbers during latency, the lack of clinical sequelae and the enhanced resistance of latent TB to chemotherapy argue that the bacilli may be metabolically dormant (Gangadharam, 1995; Gupta and Katoch, 1997; Hu *et al.*, 1998; Michele *et al.*, 1999; Mitchison, 1992; Wayne and Sramek, 1994). However, there is no direct evidence from the genome (Cole *et al.*, 1998) or the laboratory (McCune *et al.*, 1966; Parrish *et al.*, 1998; Robertson, 1933) that MTB is capable of a truly dormant, spore-like state. In addition, chemotherapy can reduce the rate of reactivation in persons with latent TB (Comstock and Woolpert, 1972; Comstock *et al.*, 1979), and immunotherapy can protect against reactivation in mice (Lowrie *et al.*, 1999). It is hard to see how these therapies would have any effect in the complete absence of mycobacterial metabolism.

Oxygen tension is one factor frequently associated with the establishment and maintenance of latent TB (Wayne and Sohaskey, 2001). *In vivo*, the number of bacilli in a lesion generally correlates well with the degree of oxygenation (Canetti, 1955), suggesting that O_2 supply may limit MTB growth during infections. Also, inhibition of MTB growth *in vivo* is associated with formation of hard, fibrous, hypoxic granulomas (Dannenberg, 1993; Yeager *et al.*, 1996). Replication of MTB requires oxygen, but bacilli show a remarkable ability to survive for years without oxygen *in vitro* (Canetti, 1955; Corper and Cohn, 1933). MTB maintained under anaerobic conditions *in vitro* lose their acid-fast character (Gillespie *et al.*, 1986) and some human studies (Parrish *et al.*, 1998) have associated latent TB with tubercle bacilli that were no longer acid-fast.

Based on these observations, Wayne has pioneered use of hypoxic culture conditions to generate non-replicating persistent bacilli *in vitro* as a model for latency (Wayne and Diaz, 1967; Wayne and Sramek, 1994; Wayne and Hayes, 1996). Variants of this model have been used to identify MTB genes potentially important for development or maintenance of the latent state (Hu *et al.*, 1999; Imboden and Schoolnik, 1998; Lim *et al.*, 1999; Yuan *et al.*, 1998). One such gene is *acr* (also known as *hspX*, Rv2031), which encodes alpha-crystallin. MTB Acr is a dominant antigen *in vivo*, recognized by most TB patient sera (Lee *et al.*, 1992; Verbon *et al.*, 1992). Acr is a member of the small heat shock protein family that forms high MW aggregates and has chaperone activity *in vitro* (Yuan *et al.*, 1996). Under hypoxic conditions, Acr expression is dramatically and rapidly increased (Florczyk *et al.*, 2001; Manabe *et al.*, 1999; Sherman *et al.*, 2001; Yuan *et al.*, 1996; Yuan *et al.*, 1998).

We have exploited the powerful regulation of Acr under reduced O_2 tension to provide insight into the nature of the genetic program by which MTB adapts to potentially hypoxic microenvironments within the host. Previously we described the subset of MTB genes (including *acr*) that respond rapidly to hypoxia and demonstrated that the MTB protein Rv3133c is required for hypoxic induction of *acr* (Sherman *et al.*, 2001). Recently it was demonstrated that long-term hypoxic survival of *M. bovis* BCG required Rv3133c, which was named DosR for dormancy survival regulator (Boon and Dick, 2002). The DosR sequence shows homology to transcription factors of the two-component response regulator class, allowing us to make and test several hypotheses about the behavior of this protein. Here, we define a sequence motif to which DosR binds. Transcriptome analysis of wild-type and DosR-mutant MTB strains indicates that this protein is needed to induce nearly all of the MTB genes that respond powerfully to a hypoxic signal. We show that DosR binds to motif sequences upstream of the *acr* coding region, and that binding is necessary but not sufficient for hypoxic gene induction. We conclude that Rv3133c/DosR is the primary transcription factor to mediate the genetic response to reduced oxygen tension in MTB.

Results

Microarray analysis defines the DosR regulon

We have previously demonstrated that the initial hypoxic response of MTB involves 47 induced genes and results in growth arrest (Sherman *et al.*, 2001). Recently we showed that exposure to low levels of nitric oxide (NO) also produces growth arrest and that the same set of MTB genes is induced by both NO and hypoxia (Voskuil, submitted). Coordinate expression following multiple stimuli suggests that these genes may comprise a regulon under control of a common transcription factor. Our previous work demonstrated that induction of the hypoxic response gene *acr* depends upon expression of the putative two-component response regulator DosR (Sherman *et al.*, 2001). To investigate its role in expression of other genes, we targeted DosR for deletion from the mycobacterial chromosome. Using a two-step method and *sacB*/sucrose counter-selection (Parish and Stoker, 2000; Pelicic *et al.*, 1996), we have generated a strain of H37Rv in which DosR has been replaced by a kanamycin-resistance determinant. This deletion (bp 3499282 to bp 3499933 of the H37Rv genome) extends from 18 bp upstream of the putative translation start site to 17 bp upstream of the putative stop codon. Successful gene replacement was confirmed both by site-specific PCR and by Southern blot (data not shown).

To test the role of DosR in the hypoxic response, we performed whole-genome expression profiling on wild-type H37Rv and mutant H37Rv:Δ*dosR*::*kan* (Δ*dosR*::*kan*) under both normoxic (normal oxygen) and hypoxic conditions *in vitro*. Whole-genome microarray technology provides a robust tool to assess expression of many genes simultaneously (DeRisi *et al.*, 1997). The MTB microarray has provided a powerful way to monitor genome-wide changes in MTB gene expression in response to the drug isoniazid (Wilson *et al.*, 1999), to study regulatory mutants (Kaushal *et al.*, 2002; Manganelli *et al.*, 2001; Manganelli *et al.*, 2002; Rodriguez *et al.*, 2002), and to assess the effects of environmental perturbations (Betts *et al.*, 2002; Fisher *et al.*, 2002; Sherman *et al.*, 2001). To determine if genes in addition to *acr* are under control of DosR, MTB

strains H37Rv and Δ*dosR*::*kan* were maintained for two hours at 0.2% O_2 in N_2 as described previously (Sherman *et al.*, 2001; Yuan *et al.*, 1998). RNA was isolated, labeled and applied to the array surface. The resulting hybridized array was scanned and fluorescent intensities identified regulated ORFs. Data for all ORFs induced or repressed at least 1.8 fold are available on the Journal's Web site, with highlights of the data described below. Consistent with previous results (Sherman *et al.*, 2001), incubation of MTB for 2 hr. under hypoxic conditions resulted in induction of numerous genes relative to expression under normoxic conditions (supplementary table 1). However, this response was dramatically muted in the *dosR* mutant (supplementary table 2). Genes whose induction depends on DosR are listed in Table 1. Notably, of the 27 genes most powerfully induced by hypoxia (≥5.7 fold), 26 require the presence of DosR. Genetic complementation demonstrated that the effects on hypoxic gene induction are due to mutation of DosR and not some unrelated MTB gene. Introduction of DosR to the *dosR* mutant restored hypoxic gene induction as measured by Western blot for the Acr protein (data not shown). Also, in the closely related mutant H37Rv:Δ3134::*kan* that does not express *dosR* (Sherman *et al.*, 2001) or respond genetically to reduced O_2, complementation with *dosR* restored hypoxic induction of all DosR-dependent genes (data not shown). We conclude that DosR is a crucial component of the MTB hypoxic response.

A consensus motif upstream of hypoxic response genes

Consistent with its role in gene expression, the DosR protein sequence predicts that it is a transcription factor of the two-component response regulator class (Dasgupta *et al.*, 2000). To begin testing this prediction, we searched by computer analysis with the motif discovery program YMF (Sinha and Tompa, 2002) for shared sequence motifs upstream of hypoxic response genes. Since transcription units in bacteria are often multi-genic, a single binding site could be responsible for activation of contiguous genes. Our analysis revealed a 20 bp palindromic consensus sequence 5' TTSGGGACTWWAGTCCCSAA 3'(S = C/G; W = A/T; Table 2), a variant of which is upstream of 42/50 (84%) MTB genes that are rapidly induced at least two-fold by hypoxia. Using a scoring matrix whose entries are the log likelihood ratios of each base at each position, we determined the best matches to this motif in the H37Rv genome (supplementary table 4). The top 13 scoring sites are all upstream of genes induced by hypoxia and regulated by DosR (Table 3). The genes most powerfully induced by hypoxia are most likely to have this motif upstream: 94% of 16 genes induced at least ten-fold were preceded by this motif (Table 1). In addition, powerfully induced genes sometimes had multiple copies in their upstream regions.

DosR binds to DNA upstream of the hypoxic response gene acr

To see if DosR could alter gene expression by binding to the hypoxia motif sites, we assessed the ability of this protein to bind specific DNA sequences by electrophoretic mobility shift assay. We expressed the DosR protein in *E. coli* from an IPTG-inducible promoter (Fig. 1A, lanes 1-3) and mixed protein extracts with radiolabeled DNA corresponding to the 146 bp upstream of the *acr* coding sequence (Fig. 1B). Previously we have shown that this region contains all the information necessary to direct the hypoxic induction of *acr* (Sherman *et al.*, 2001; Yuan *et al.*, 1998). A band of retarded mobility, indicating protein bound to the DNA, was evident when *acr* promoter DNA was mixed with *E. coli* extracts induced to express DosR (Fig. 1B, lane 2). In addition, DosR is predicted to be a two-component response regulator of the RO_{III} sub-class (Dasgupta *et al.*, 2000; Parkinson and Kofoid, 1992). Typical of such proteins, the DosR sequence divides into two domains: a C-terminal half with a helix-turn-helix DNA binding motif, and an N-terminal receiver half predicted to participate in a histidine-to-aspartate phosphorelay (Dasgupta *et al.*, 2000). To begin testing these predictions, we separately expressed the N-terminal and C-terminal portions of DosR in *E. coli* (Fig. 1A, lanes 4-7), and assessed the binding of these portions to the *acr* promoter by mobility shift assay (Fig. 1B). As expected, a shift was evident upon incubation with extracts induced to express the C-terminus (amino acids 144 – 217, Fig. 1B lane 4) but not the N-terminus (amino acids 1 – 134, Fig. 1B lane 3). All mobility shift assays reported here included a large molar excess of unlabeled poly dI/dC and calf thymus DNA, indicating that binding was specific for the *acr* promoter. Further, we conclude that the protein binding to the *acr* promoter is DosR and not some unrelated *E. coli* protein induced by IPTG since *E. coli* extracts containing the N-terminus of DosR failed to produce a shift.

To facilitate analysis of the interaction between DosR and the *acr* promoter, we mapped the *acr* transcription start site by primer extension. A synthetic oligonucleotide of 16 bases that terminates at bp 7 of the *acr* coding region was used in extension reactions with RNA from oxygenated (rolling) and hypoxic (static) log-phase cultures of MTB. A single start site was visualized 29 bp upstream of the initiator ATG (Fig. 2). Reactions with additional aliquots of MTB RNA and a different primer confirmed this result (data not shown), which is four bp downstream of the putative *acr* transcription start site reported by Coates and colleagues (Hu and Coates, 1999). This discrepancy may be due to differences in technique, as Hu et al. sized their extension products on a heterologous sequencing ladder. All base numberings below are relative to the *acr* transcription start site defined here.

To localize binding of DosR upstream of *acr*, we performed additional mobility shift assays with double stranded oligonucleotides corresponding to overlapping regions of the *acr* promoter. We detected protein binding in two distinct regions, a more distal 33 bp stretch from -111 to -79 and a more proximal 33 bp stretch from -66 to -34 (Fig 3A). Significantly, these regions each contain a copy of the motif defined above (Fig. 3B, 4A). To assess the specificity of DosR binding to these regions, we repeated the mobility shift assays after adding a 100-fold molar excess of various unlabeled competitor DNAs. In each case, binding was abolished by addition of the corresponding unlabeled DNA (Fig. 4B, lanes 3 and 9). In addition, the unlabeled DNA competed for binding to the other *acr* promoter region (Fig. 4B, lanes 5 and 11). To test specifically the role of the upstream motifs in DosR binding, we mutated four bases within each motif (-94 to -91 in the distal region and -53 to -50 in the proximal region; underlined in Fig. 4A). The mutated DNAs were no longer able to compete with the wild-type sequence for binding to DosR (Fig 4B, lanes 4, 6, 10, 12). Furthermore, when mutated versions of -111 to -79 and -66 to -34 were used in mobility shift assays shifted bands were abolished, indicating that the bases mutated are necessary for DosR binding (data not shown).

DosR binding is required for hypoxic gene induction

These results indicate that DosR binds to two distinct regions of the *acr* promoter, each of which contains an example of the hypoxia motif. To test the functional significance of this binding, we introduced the mutations described above into the *acr* upstream sequence and assessed the hypoxic responsiveness of the wild-type and mutant promoters by luciferase reporter gene assay (Sherman *et al.*, 2001; Yuan *et al.*, 1998)(Fig. 4C). None of the mutations had any discernable effect under normal oxygenation. However, mutating from -94 to -91, which eliminated DosR binding to the distal motif sequence, reduced hypoxic responsiveness roughly by half. Mutating from -53 to -50 to eliminate DosR binding to the proximal region reduced hypoxic responsiveness by nearly six-fold. Introducing both mutations simultaneously virtually abolished the induction, indicating that DosR binding to its motif is necessary for a proper hypoxic response in MTB.

Mutating DosR affects hypoxic gene induction

Based on alignment to known two-component response regulators, Asp-54 of DosR is predicted to be a site of phosphorylation that is essential for gene induction. To test the importance of this amino acid to DosR function, we mutated Asp-54 to Glu. This change should conserve the charge and structure of the protein but block the possibility of phosphorylation. Mutated DosR (Asp54Glu) could still bind to the *acr* promoter (Fig. 5A). However, in functional assays the mutated DosR was no longer able to mediate hypoxic induction of a luciferase reporter gene under control of the *acr* promoter (Fig. 5B). These results indicate that Asp-54 is essential to DosR activity, possibly because it is a site of phosphorylation.

Discussion

We have shown that DosR of MTB is necessary for induction of a robust genetic response to reduced oxygen tension. Further, we have defined a consensus sequence, a variant of which is upstream of nearly all of the hypoxic response genes. We have shown that the DosR protein

binds to this sequence upstream of the *acr* gene, and that binding is necessary but not sufficient for gene induction. Thus we conclude that DosR is a transcription factor that mediates induction of MTB genes in response to hypoxia. Further, consistent with the sequence-based prediction that DosR belongs to the two-component response regulator family of transcriptional activators we find that the C-terminal 73 amino acids are sufficient for DNA binding and that the aspartate at position 54 is essential for activity. Previously this protein was named DevR, because it is differentially expressed in virulent MTB strain H37Rv compared with avirulent H37Ra (Dasgupta *et al.*, 2000; Mayuri *et al.*, 2002; Saini *et al.*, 2002). Given the functional data presented elsewhere (Boon and Dick, 2002) and here, we believe that the name DosR is more appropriate.

Presence of an upstream motif sequence defined here strongly correlates with genes induced by hypoxia and regulated by DosR. Further, the number of motif sequences and their degree of fit to the consensus as defined by matrix score roughly predict the extent of hypoxic gene induction. Experiments are underway to determine whether the matrix score predicts strength of DNA binding. It will be interesting to see if a perfect match, which exists nowhere in the H37Rv genome, elicits stronger binding and/or more robust gene induction. Still, the link between motif sequence, regulation by hypoxia and by DosR is not exact. For example, a reasonable match to the motif is present upstream of Rv1976c and Rv1977 (score = 11.7), yet these genes are not induced by hypoxia. Conversely, a few genes (e.g. Rv3129 or Rv3841, see Table 1) appear to be regulated by DosR without an obvious upstream match to the consensus. These inconsistencies suggest that aspects of sequence or context may affect DosR-mediated gene expression in ways we have not yet discerned. Also, while nearly all the genes powerfully induced by hypoxia are controlled by DosR, there are numerous modestly induced genes for which DosR does not appear to play a role (supplementary table 1). We are now investigating whether another transcription factor or factors also affects the MTB hypoxic response.

We are interested in the signal transduction pathway by which a drop in oxygen tension results in altered MTB gene expression. For instance, it seems reasonable that some heme protein may play a role in monitoring available O_2, but the identity of this sensor is unknown. The identity of the kinase that activates DosR is also not yet clear. Directly adjacent to *dosR* on the MTB chromosome is a putative sensor kinase gene Rv3132c, but disruption of that gene had little effect as measured by accumulation of Acr protein (Sherman *et al.*, 2001) and little impact on survival during long-term hypoxic dormancy (Boon and Dick, 2002). Strong sequence homology between Rv3132c and Rv2027c has been noted previously (Dasgupta *et al.*, 2000; Sherman *et al.*, 2001). We are currently investigating whether either or both of these proteins participate in the phosphorelay that activates DosR.

Another outstanding question is the role of DosR-mediated gene expression and the hypoxic response during MTB infection of a mammalian host. *In vitro*, both hypoxia and low levels of NO cause MTB to enter a state of growth arrest (Wayne and Hayes, 1996, Voskuil et al, submitted). *In vivo*, MTB may be exposed to hypoxia and NO as granulomas form during a normal host response (Flynn and Chan, 2001; Wayne and Sohaskey, 2001). It is tempting to speculate that host-induced hypoxia/NO curtail bacterial replication, but the ensuing bacterial response results in long-term persistence rather than clearance. Indeed, we have shown recently that DosR response genes are powerfully induced during MTB infection of mice (Voskuil et al, submitted). However, despite its attractiveness the link between hypoxia, NO and mycobacterial latency *in vivo* remains circumstantial and will require further testing. Experiments to define the phenotype of the *dosR*-deletion mutant under hypoxic conditions *in vitro* and *in vivo* are underway.

Experimental Procedures

Mycobacteria and culture conditions

H37Rv (ATCC 27294), H37Rv:Δ*dosR::kan* and *M. bovis* BCG-Montreal (ATCC 35735) were grown to mid-log phase in 7H9 media with 0.05% Tween 80 and ADC supplement (Becton

Dickinson) and stored as 1 ml aliquots in 15% glycerol (final concentration) at -80ºC. For individual experiments, bacilli were grown in roller bottles in Middlebrook 7H9 medium (Becton Dickinson) with 0.05% Tween 80 and ADC supplement or on Middlebrook 7H10 plates at 37ºC as described (Sherman *et al.*, 1995). When needed, kanamycin was used at 30 μg/ml (50 μg/ml for *E. coli*) and hygromycin at 50 μg/ml (200 μg/ml for *E. coli*).

Targeted MTB Gene Disruption.

Targeted disruptions were performed as described previously (Sherman *et al.*, 2001). Genomic regions (~900 bp each) flanking *dosR* were amplified by PCR and cloned into the plasmid pKO so as to flank the kan^R determinant. Fidelity of the cloned flanking sequences was confirmed by sequence analysis. Constructs were electroporated into mycobacteria as described (Wards and Collins, 1996) and selected on 7H10 plates with hygromycin. In the first screening step, each colony was tested by PCR with two primer pairs, one specific for integration upstream of the gene of interest and the other specific for integration downstream. Colonies positive at either end by PCR were grown to an OD_{600} = ~1.0 and plated onto 7H10 plates containing 10% sucrose. Bacilli that grow on sucrose generally have either mutated copies of *sacB* or have lost the integrated plasmid. A portion of those in which the plasmid is lost will also lose the gene of interest. Colonies appearing on sucrose plates were picked into media and patched separately onto 7H10 plates with kanamycin and hygromycin. Sucrose-resistant, hygromycin-sensitive, kanamycin-resistant colonies (indicating loss of the integrated plasmid) were screened for loss of the gene of interest by PCR and confirmed by Southern blot.

Microarray analysis

RNA isolation was performed as described (Sherman *et al.*, 2001). Steps in *M. tuberculosis* DNA microarray gene expression analysis were performed as described (Ehrt *et al.*, 2002; Schoolnik *et al.*, 2001). A 70-mer oligonucleotide-based microarray (tuberculosis oligonucleotide set [Qiagen]) were used. Labeled cDNA was prepared as follows: 2 μg of total RNA and 4.4 μg of random oligonucleotide hexamers were incubated 2 m at 98°C, cooled on ice, combined with Stratascript RTase buffer, 0.5 mM dA, G, CTP, 0.02 mM dTTP, 1.5 nmol Cy3 or Cy5-dUTP (Amersham), and 1.8 μl Stratascript RTase (Stratagene) in a total volume of 25 μl and incubated 10 m at 25°C and 90 m at 42°C. cDNA was purified by microcon-10 (Amicon) filtration. 10 μl hybridization solution (labeled cDNA, 5 μg tRNA, 2X SSC, 25% formamide, 0.1% SDS) was hybridized at 52°C. Microarray-determined ratios were calculated from 3 biological replicates with 2 microarrays for each biological replicate. Microarrays were scanned using a GenePix 4000A (Axon Instruments). The intensities of the two dyes at each spot were quantified using ScanAlyze (Michael Eisen, http://rana.lbl.gov/EisenSoftware.htm). All gene specific spots on the microarray were used to normalize the intensities of Cy3 and Cy5 from each spot. Additional methods and analysis details are available at the following website: http://schoolniklab.stanford.edu/projects/tb.html.

Computer analysis

Of the genes whose induction by hypoxia requires DosR (Table 1), 27 have noncoding upstream regions of at least 25 bp. For each of these, 250 bp of upstream sequences was collected. The sequences for Rv1737c and Rv2032 were discarded, as they share approximately 250 bp upstream sequence with Rv1738 and *acr*, respectively, which are already in this set. The motif discovery program YMF (http://bio.cs.washington.edu/software.html) (Sinha and Tompa, 2002) was applied to the remaining 25 upstream sequences. The four to six motifs with greatest *z*-scores produced for each of the lengths 6, 7, 8, and 9 were aligned, resulting in 19 high scoring YMF motifs (supplementary table 3). All instances of these motifs in the 25 upstream sequences, together with the reverse complements of such instances, were collected and similarly aligned, resulting in 100 aligned motif instances from the upstream regions, each of length 20. These induced the 4×20 log likelihood ratio matrix M (first four columns of Table 2), where $M_{rj} = \log_2(p_{rj} / q_r)$ for $r \in \{A,T,G,C\}$ and $1 \le j \le 20$; $p_{rj} = (n_{rj} + q_r) / (n+1)$ is the likelihood of residue r in position j; n_{rj} is the number of occurrences of residue r in position j among the 100 aligned motif instances; n = 100 is the number of aligned motif instances; and q_r is the background frequency of residue r in the noncoding portion of MTB ($q_A = q_T = 0.19$, $q_G = q_C = 0.31$). The log likelihood ratio matrix M

was used to score every 20-mer in the noncoding portion of the MTB genome, where the score of $x[1]x[2]\ldots x[20]$ is given by $\Sigma_j M_{x[j],j}$. The highest scoring 20-mers are shown in Table 3. Additional high-scoring 20-mers are shown in supplementary table 4.

Expression of DosR in E. coli

To express DosR in *E. coli*, a DNA fragment of full-length, N-terminal or C-terminal portions of the DosR ORF was amplified by PCR and cloned into pET-22b(+)(Novagen). The resultant plasmid was introduced to *E. coli* BL21 (DE3) (Novagen). To induce expression of DosR, cultures grown to OD_{600} = 0.6 were treated with 1 mM IPTG at room temperature for 5 hours. To prepare bacterial extracts, cells were harvested and washed with 5X sonication buffer (20 mM Tris·Cl (pH 8.0), 100 mM KCl, 10 mM $MgCl_2$, 25% glycerol). After washing, cells were resuspended in the same buffer and disrupted by three 45 sec pulses in a bead-beater (Qbiogene Fast prep FP120). Cell extract was obtained by centrifugation at 14,000 x g at 4°C for 30 min.

Electrophoretic mobility shift assay

To assess protein binding, *E. coli* extracts (prepared as described above) were used as protein source. Oligonucleotides were used as DNA probes, designed so that when annealed, both ends had 5' "a" extensions that could be labeled with [^{32}P]-dTTP by a Klenow fill-in reaction as described (Winterling *et al.*, 1998). DNA probes were labeled in a 20 μl reaction mixture containing 2 μg of DNA. Gel mobility shift assay was carried out as described (Dhandayuthapani *et al.*, 1997) with some modifications. Binding of protein to DNA fragments was carried out by incubation at room temperature for 10 min in a 20 μl reaction mixture. The mixture was composed of labeled probe (10,000 cpm), 5-10 μg of protein, 25 mM Tris-HCl (pH 8.0), 0.5 mM EDTA, 20 mM KCl, 6 mM $MgCl_2$, 1 mM DTT, 50 ng μl^{-1} poly (dI/dC), 50 ng μl^{-1} calf thymus DNA and 5% glycerol. After incubation, the reaction was resolved in a 5% non-denaturing TBE polyacrylamide gel (BioRad).

Primer extension

Primer extension was performed as described (Sambrook *et al.*, 1989), using a synthetic oligonucleotide, sequence 5'-gaacgggaagggtggt-3', and RNA (5 – 10 μg/rxn)from oxygenated (rolling) and hypoxic (static) log-phase cultures of MTB, isolated as described (Sherman *et al.*, 2001).

Luciferase reporter gene assay

Assay was performed as described previously (Sherman *et al.*, 2001; Yuan *et al.*, 1998). Briefly, *acr* promoter (with described mutations) was amplified by PCR and cloned into the integrating mycobacterial shuttle plasmid pMH66 upstream of the firefly luciferase gene. The resultant construct was electroporated into BCG, where it integrated into the phage L5 attachment site. Luciferase activity was assayed in triplicate for 15 sec by TD-20/20 luminometer (Turner Designs).

Other procedures

Mutated versions of *dosR* and the *acr* promoter sequence were generated by mutagenic gene fusion PCR (Yon and Fried, 1989). Mycobacterial transformations were carried out by electroporation as described previously (Sherman *et al.*, 2001; Wards and Collins, 1996). *E. coli* strain DH5α (Invitrogen) was used for routine DNA manipulations.

Acknowledgements

We thank Mark Hickey for mapping the *acr* transcription start site by primer extension, and Carleen Collins, Wim Hol, Goragot Wisedchaisri, David Roberts, Mark Hickey and members of the Sherman lab for helpful discussions. We gratefully acknowledge funding from NIH (grants

AI47744 to DRS, AI44826 to GKS, and HG02602-01 to MT), NSF (grant DBI-9974498 to MT) and the Sequella Global Tuberculosis Foundation, of which DRS is a Core Scientist.

References

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.

Bloom, B.R., and Small, P.M. (1998) The evolving relation between humans and *Mycobacterium tuberculosis*. *New Engl J Med.* **338**: 677-678.

Boon, C., and Dick, T. (2002) *Mycobacterium bovis* response regulator essential for hypoxic dormancy. *J Bacteriol* **184**: 6760-6767.

Canetti, G. (1955) Growth of the tubercle bacilus in the tuberculosis lesion. In *The tubercle bacillus in the pulmonary lesion of man*. New York: Springer Publishing Co., pp. 111-126.

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., and Barrell, B.G. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**: 537-544.

Comstock, G.W., and Woolpert, S.F. (1972) Preventive treatment of untreated, non-active tuberculosis in an Eskimo population. *Arch Env Health* **25**: 333-337.

Comstock, G.W., Baum, C., and Snider, D.E. (1979) Isoniazid prophylaxis among alaskan eskimos: a final report of the Bethel isoniazid studies. *Am Rev Resp Dis* **119**: 827-830.

Corper, H.J., and Cohn, M.L. (1933) The viability and virulence of old cultures of tubercle bacilli. *Ann Rev Tuberc* **28**: 856-874.

Dannenberg, A.M., Jr. (1993) Immunopathogenesis of pulmonary tuberculosis. *Hosp Pract* **28**: 51-58.

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.

DeRisi, J.L., Iyer, V.R., and Brown, P.O. (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**: 680-686.

Dhandayuthapani, S., Mudd, M., and Deretic, V. (1997) Interactions of OxyR with the promoter region of the *oxyR* and *ahpC* genes from *Mycobacterium leprae* and *Mycobacterium tuberculosis*. *J Bacteriol* **179**: 2401-2409.

Dye, C., Scheele, S., Dolin, P., Pathania, V., and Raviglione, M.C. (1999) Consensus statement. Global burden of tuberculosis: estimated incidence, prevalence, and mortality by country. WHO Global Surveillance and Monitoring Project. *Jama* **282**: 677-686.

Ehrt, S., Voskuil, M.I., Schoolnik, G.K., and Schnappinger, D. (2002) Genome-wide expression profiling of intracellular bacteria: the interaction of *Mycobacterium tuberculosis* and macrophages. In *Methods in Microbiology*. Vol. 32. Kabelitz, K. (ed): Academic Press, pp. 169-180.

Enarson, D.A., and Murray, J.F. (1996) Global epidemiology of tuberculosis. In *Tuberculosis*. Rom, W.N. and Garay, S. (eds). Boston: Little, Brown and Co., pp. 55-75.

Fisher, M.A., Plikaytis, B.B., and Shinnick, T.M. (2002) Microarray analysis of the Mycobacterium tuberculosis transcriptional response to the acidic conditions found in phagosomes. *J Bacteriol* **184**: 4025-4032.

Florczyk, M.A., McCue, L.A., Stack, R.F., Hauer, C.R., and McDonough, K.A. (2001) Identification and characterization of mycobacterial proteins differentially expressed under standing and shaking culture conditions, including Rv2623 from a novel class of putative ATP-binding proteins. *Infect Immun* **69**: 5777-5785.

Flynn, J.L., and Chan, J. (2001) Immunology of tuberculosis. *Annu Rev Immunol* **19**: 93-129.

Gangadharam, P.R. (1995) Mycobacterial dormancy. *Tuber Lung Dis* **76**: 477-479.

Gillespie, J., Barton, L.L., and Rypka, E.W. (1986) Phenotypic changes in mycobacteria grown in oxygen-limited conditions. *J Med Microbiol* **21**: 251-255.

Gupta, U.D., and Katoch, V.M. (1997) Understanding the phenomenon of persistence in mycobacterial infections. *Indian J Lepr* **69**: 385-393.

Hu, Y., and Coates, A.R. (1999) Transcription of the stationary-phase-associated *hspX* gene of *Mycobacterium tuberculosis* is inversely related to synthesis of the 16- kilodalton protein. *J Bacteriol* **181**: 1380-1387.

Hu, Y.M., Butcher, P.D., Sole, K., Mitchison, D.A., and Coates, A.R. (1998) Protein synthesis is shutdown in dormant *Mycobacterium tuberculosis* and is reversed by oxygen or heat shock. *FEMS Microbiol Lett* **158**: 139-145.

Hu, Y.M., Butcher, P.D., Mangan, J.A., Rajandream, M.A., and Coates, A.R.M. (1999) Regulation of *hmp* gene transcription in *Mycobacterium tuberculosis*: Effects of oxygen limitation and nitrosative and oxidative stress. *J Bacteriol* **181**: 3486-3493.

Imboden, P., and Schoolnik, G.K. (1998) Construction and characterization of a partial *Mycobacterium tuberculosis* cDNA library of genes expressed at reduced oxygen tension. *Gene* **213**: 107-117.

Kaushal, D., Schroeder, B.G., Tyagi, S., Yoshimatsu, T., Scott, C., Ko, C., Carpenter, L., Mehrotra, J., Manabe, Y.C., Fleischmann, R.D., and Bishai, W.R. (2002) Reduced immunopathology and mortality despite tissue persistence in a Mycobacterium tuberculosis mutant lacking alternative sigma factor, SigH. *Proc Natl Acad Sci U S A* **99**: 8330-8335.

Lee, B.Y., Hefta, S.A., and Brennan, P.J. (1992) Characterization of the major membrane protein of virulent *Mycobacterium tuberculosis*. *Infect Immun* **60**: 2066-2074.

Lim, A., Eleuterio, M., Hutter, B., Murugasu-Oei, B., and Dick, T. (1999) Oxygen depletion-induced dormancy in *Mycobacterium bovis* BCG. *J Bacteriol* **181**: 2252-2256.

Lowrie, D.B., Tascon, R.E., Bonato, V.L., Lima, V.M., Faccioli, L.H., Stavropoulos, E., Colston, M.J., Hewinson, R.G., Moelling, K., and Silva, C.L. (1999) Therapy of tuberculosis in mice by DNA vaccination. *Nature* **400**: 269-271.

Manabe, Y.C., Chen, J.M., Ko, C.G., Chen, P., and Bishai, W.R. (1999) Conditional sigma factor expression, using the inducible acetamidase promoter, reveals that the Mycobacterium tuberculosis sigF gene modulates expression of the 16-kilodalton alpha-crystallin homologue. *J Bacteriol* **181**: 7629-7633.

Manabe, Y.C., and Bishai, W.R. (2000) Latent mycobacterium tuberculosis-persistence, patience, and winning by waiting. *Nat Med* **6**: 1327-1329.

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.

Mayuri, Bagchi, G., Das, T.K., and Tyagi, J.S. (2002) Molecular analysis of the dormancy response in Mycobacterium smegmatis: expression analysis of genes encoding the DevR-DevS two-component system, Rv3134c and chaperone alpha-crystallin homologues. *FEMS Microbiol Lett* **211**: 231-237.

McCune, R.M., Feldmann, F.M., and McDermott, W. (1966) Microbial persistence. II. Characteristics of the sterile state of tubercle bacilli. *J Exp Med* **123**: 469-486.

Michele, T.M., Ko, C., and Bishai, W.R. (1999) Exposure to antibiotics induces expression of the *Mycobacterium tuberculosis sigF* gene: implications for chemotherapy against mycobacterial persistors. *Antimicrob Agents Chemother* **43**: 218-225.

Mitchison, D.A. (1992) The Garrod Lecture: understanding the chemotherapy of tuberculosis -- current problems. *J Antimicrob Chemother* **29**: 477-493.

Parish, T., and Stoker, N.G. (2000) Use of a flexible cassette method to generate a double unmarked Mycobacterium tuberculosis tlyA plcABC mutant by gene replacement. *Microbiology* **146**: 1969-1975.

Parkinson, J.S., and Kofoid, E.C. (1992) Communication modules in bacterial signaling proteins. *Ann Rev Genet* **26**: 71-112.

Parrish, N.M., Dick, J.D., and Bishai, W.R. (1998) Mechanisms of latency in *Mycobacterium tuberculosis*. *Trends Microbiol* **6**: 107-112.

Pelicic, V., Reyrat, J.M., and Gicquel, B. (1996) Generation of unmarked directed mutations in mycobacteria, using sucrose counter-selectable suicide vectors. *Mol Microbiol* **20**: 919-925.

Robertson, H.E. (1933) The persistence of tuberculosis infections. *Am J Pathol* **9**: 711-718.

Rodriguez, G.M., Voskuil, M.I., Gold, B., Schoolnik, G.K., and Smith, I. (2002) ideR, An essential gene in mycobacterium tuberculosis: role of IdeR in iron-dependent gene expression, iron metabolism, and oxidative stress response. *Infect Immun* **70**: 3371-3381.

Saini, D.K., Pant, N., Das, T.K., and Tyagi, J.S. (2002) Cloning, Overexpression, Purification, and Matrix-Assisted Refolding of DevS (Rv 3132c) Histidine Protein Kinase of Mycobacterium tuberculosis. *Protein Expr Purif* **25**: 203-208.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular cloning: A laboratory manual*: Cold Spring Harbor Laboratory Press.

Schoolnik, G.K., Voskuil, M.I., Schnappinger, D., Yildiz, F.H., Meibom, K., Dolganov, N.A., Wilson, M.A., and Chong, K.H. (2001) Whole genome DNA microarray expression analysis of biofilm development by *Vibrio cholerae* O1 E1 Tor. In *Methods Enzymol* Vol. 336. Doyle, R. (ed): Academic Press, pp. 3-18.

Sherman, D.R., Sabo, P.J., Hickey, M.J., Arain, T.M., Mahairas, G.G., Yuan, Y., Barry III, C.E., and Stover, C.K. (1995) Disparate responses to oxidative stress in saprophytic and pathogenic mycobacteria. *Proc Nat Acad Sci USA* **92**: 6625-6629.

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 U S A* **98**: 7534-7539.

Sinha, S., and Tompa, M. (2002) Discovery of novel transcription factor binding sites by statistical overrepresentation. *Nucleic Acids Res* **30**: 5549-60.

Verbon, A., Hartskeerl, R.A., Schuitema, A., Kolk, A.H., Young, D.B., and Lathigra, R. (1992) The 14,000-molecular-weight antigen of Mycobacterium tuberculosis is related to the alpha-crystallin family of low-molecular-weight heat shock proteins. *J Bacteriol* **174**: 1352-1359.

Wards, B.J., and Collins, D.M. (1996) Electroporation at elevated temperatures substantially improves transformation efficiency of slow-growing mycobacteria. *FEMS Microbiol Lett* **145**: 101-105.

Wayne, L.G., and Diaz, G.A. (1967) Autolysis and secondary growth of *Mycobacterium tuberculosis* in submerged culture. *J Bacteriol* **93**: 1374-1381.

Wayne, L.G., and Sramek, H.A. (1994) Metronidazole is bactericidal to dormant cells of *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother* **38**: 2054-2058.

Wayne, L.G., and Hayes, L.G. (1996) An *in vitro* model for sequential study of shiftdown of *Mycobacterium tuberculosis* through two stages of nonreplicating persistence. *Infect Immun* **64**: 2062-2069.

Wayne, L.G., and Sohaskey, C.D. (2001) Nonreplicating persistence of *Mycobacterium tuberculosis*. *Ann Rev Microbiol* **55**: 139-163.

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.

Winterling, K.W., Chafin, D., Hayes, J.J., Sun, J., Levine, A.S., Yasbin, R.E., and Woodgate, R. (1998) The Bacillus subtilis DinR binding site: redefinition of the consensus sequence. *J Bacteriol* **180**: 2201-2211.

Yeager, H.J., Azumi, N., and Underhill, C.B. (1996) Fibrosis: the formation of the granuloma matrix. In *Tuberculosis*. Rom, W.N. and Garay, S.M. (eds). Boston: Little, Brown and Co., pp. 363-372.

Yon, J., and Fried, M. (1989) Precise gene fusion by PCR. *Nucleic Acids Res* **17**: 4895.

Yuan, Y., Crane, D.D., and Barry III, C.E. (1996) Stationary phase-associated protein expression in *Mycobacterium tuberculosis*: function of the mycobacterial alpha-crystallin homolog. *J Bacteriol* **178**: 4484-4492.

Yuan, Y., Crane, D.D., Simpson, R.M., Zhu, Y., Hickey, M.J., Sherman, D.R., and Barry III, C.E. (1998) The 16-kDa α-crystallin (Acr) protein of *Mycobacterium tuberculosis* is required for growth in macrophages. *Proc Nat Acad Sci USA* **95**: 9578-9583.

Table 1. MTB genes whose induction by hypoxia requires DosR.

	ORF	Gene	Rv	*ΔdosR*	Gene product
*	Rv0079		22.2 ± 6.9	0.9 ± 0.1	HP
	Rv0080		7.8 ± 1.7	0.9 ± 0.1	HP
	Rv0081		4.9 ± 1.2	0.5 ± 0.1	transc. regulator
	Rv0082		3.1 ± 0.4	0.6 ± 0.1	prob. oxidored. sub.
	Rv0083		2.1 ± 0.4	1.1 ± 0.2	prob. oxidored. sub.
	Rv0569		9.0 ± 4.3	0.9 ± 0.1	CHP
	Rv0570	*NrdZ*	5.5 ± 3.0	1.1 ± 0.1	ribonuc. red. cl. II
*	Rv0571c		1.7 ± 0.5	1± 0.2	CHP
	Rv0572c		3.0 ± 0.8	1 ± 0.1	HP
*	Rv0574c		2.0 ± 0.5	1.1 ± 0.1	CHP
	MT0639		2.0 ± 0.4	1 ± 0.2	HP
***	Rv1733c		4.1 ± 1.3	Nd	poss. mem. prot.
*	Rv1734c		1.5 ± 0.1	1 ± 0.2	HP
	Rv1736c	*NarX*	3.7 ± 0.7	1 ± 0.2	fused nitrate red.
***	Rv1737c	*narK2*	8.5 ± 2.0	1.1 ± 0.2	nitrite extrusion prot.
***	Rv1738		22.8 ± 9.7	1.3 ± 0.1	CHP
	Rv1812c		3.6 ± 0.6	1 ± 0.1	HP
*	Rv1813c		11.4 ± 3.0	0.8 ± 0.2	CHP
*	Rv1996		7.9 ± 4.6	0.7	CHP
	Rv1997	*CtpF*	4.3 ± 2.2	0.8 ± 0.1	cation trans. ATPase
	Rv2003c		2.3 ± 0.6	0.9 ± 0.1	CHP
	Rv2004c		4.2 ± 1.4	1.2 ± 0.2	HP
*	Rv2005c		7.3 ± 3.7	1.3 ± 0.5	CHP
*	Rv2006	*OtsB*	2.2 ± 0.9	0.8 ± 0.1	trehalose phos.
*	Rv2007c	*FdxA*	25.9 ± 3.3	0.8 ± 0	ferredoxin
	Rv2028c		6.0 ± 1.7	0.9 ± 0.2	CHP
	Rv2029c	*PfkB*	13.3 ± 5.7	0.9 ± 0.1	phosphofructokin. II
	Rv2030c		27.3 ± 6.3	Nd	CHP
**	Rv2031c	*Acr*	27.9 ± 7.6	Nd	α-crystallin
**	Rv2032		15.1 ± 5.0	Nd	CHP
	Rv2623		18.8 ± 4.1	Nd	CHP
	Rv2624c		3.9 ± 1.3	0.6	CHP
	Rv2625c		3.0 ± 1.1	1.3 ± 0.1	CHP
**	Rv2626c		24.5 ± 4.6	1.2 ± 0.1	CHP
**	Rv2627c		12.4 ± 4.9	0.8	CHP
**	Rv2628		13.6 ± 10.8	0.8 ± 0	HP
	Rv2629		7.6 ± 7.4	1.4 ± 0.1	HP
	Rv2630		6.5 ± 4.6	2 ± 0.5	HP
	Rv2631		3.4 ± 2.1	1.4 ± 0.3	CHP
	Rv2830c		2.6 ± 0.7	1.2 ± 0.1	HP
	Rv3126c		1.7 ± 0.7	0.8 ± 0	HP
**	Rv3127		17.4 ± 2.4	0.8 ± 0.1	CHP
	Rv3128c		1.5 ± 0.5	0.8 ± 0.1	CHP
	Rv3129		2.7 ± 1.3	0.6 ± 0.1	CHP
*	Rv3130c		25.5 ± 9.4	Nd	CHP
*	Rv3131		34.1 ± 6.4	Nd	CHP
	Rv3132c		5.7 ± 1.1	0.8 ± 0.1	sensor hist. kinase
	Rv3133c	*DosR*	9.1 ± 3.3	1.1 ± 0.2	2-comp. resp. reg.
**	Rv3134c		22.2 ± 17.9	1.2 ± 0.2	CHP
	Rv3841	*BfrB*	5.2 ± 1.9	2.0 ± 1.3	bacterioferritin

The presence of one, two or three motif sequences (matrix score >9.5) upstream of a gene is indicated by *, **, or *** respectively. Values are fold induction +/- std dev. Gene descriptions are by the Pasteur Institute: http://genolist.pasteur.fr/TubercuList or by TIGR: http://www.tigr.org/tigr-scripts/CMR2/GenomePage3.spl?database=gmt. Nd = not detected (signal too low). (C)HP = (conserved) hypothetical protein.

Table 2. Scoring matrix from which the hypoxia consensus motif is derived.

A	T	G	C	relative entropy	consensus	
-0.247	0.391	-0.046	-0.0944	0.0172	5’	T
0.0719	0.789	-0.854	-0.046	0.116		T
-0.0014	-0.915	0.133	0.254	0.0474		C/G
-0.542	-2.2	1.25	-1.75	0.608		G
-1.23	-2.2	1.41	-2.56	.859		G
-1.63	-4.01	1.48	-2.31	1.01		G
1.23	-0.247	-1.13	-0.306	0.281		A
-0.915	-1.42	-0.854	1.08	0.379		C
-0.542	0.699	-0.144	-0.144	0.0654		T
0.789	0.914	-0.773	-1.23	0.29		T/A
0.914	0.789	-1.23	-0.773	0.29		A/T
0.699	-0.542	-0.144	-0.144	0.0654		A
-1.42	-0.915	1.08	-0.854	0.379		G
-0.247	1.23	-0.306	-1.13	0.281		T
-4.01	-1.63	-2.31	1.48	1.01		C
-2.2	-1.23	-2.56	1.41	0.859		C
-2.2	-0.542	-1.75	1.25	0.608		C
-0.915	-0.0014	0.254	0.133	0.0474		G/C
0.789	0.0719	-0.046	-0.854	0.116		A
0.391	-0.247	-0.0944	-0.046	0.0172	3’	A

Values are the log likelihood ratios of each base at each position. Relative entropy describes the relative contribution of that position within the motif to the consensus.

Table 3. Best matches to the hypoxia consensus in H37Rv genome.

Sequence	Score	Gene	Location
ggCGGGACgTAAGTCCCtAA	15.8	Rv2627c	-53
TTaGGGcCggAAGTCCCCAA	14.6	Rv1738	-196
cTGGGGACcgAAGTCCCCgg	14.4	Rv1734c	-49
TcGGGGACTTctGTCCCtAg	14.4	*acr*	-53
gcCGGGACTTcAGgCCCtAt	13.6	Rv1738	-134
acaGGGtCaATgGTCCCCAA	13.3	*acr*	-110
gaaGGGgCgAAAGTCCCttA	13.3	Rv1733c	-146
TTGaGGACcTTcGgCCCCAc	13.2	Rv0574c	-112
gTGGGGACcAAcGcCCCtgg	12.7	Rv3134c	-92
catGGGACTTTcGgCCCtgt	12.4	Rv0079	-89
aTaaGGACTAAcGgCCCtcA	11.9	Rv3134c	-113
caCGGGtCaAAcGaCCCtAg	11.9	Rv2626c	-90
TTaGaGACTTTAtgCCCtAc	11.8	Rv1813c	-272

Scores are the sum of the log likelihood ratios for each position in the motif. Location is relative to the translation start site as determined at: http://genolist.pasteur.fr/TubercuList, except for *acr*, where location is relative to the transcription start site defined here.

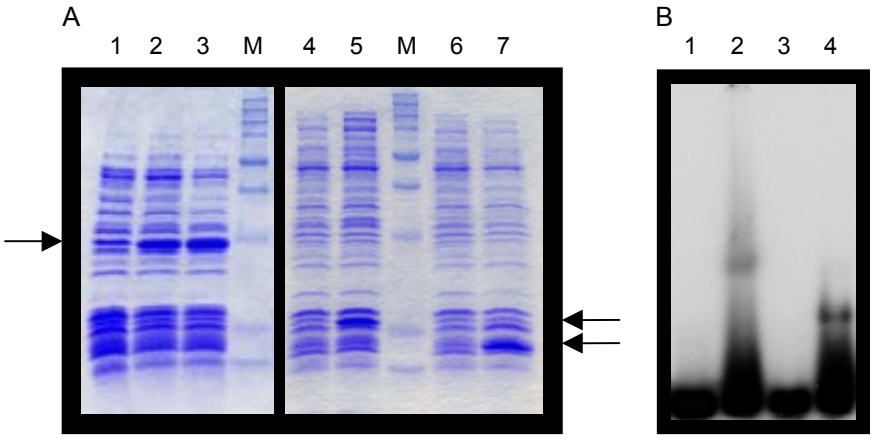

Figure 1. DosR binds to DNA upstream of the *acr* coding sequence. A. SDS-PAGE analysis of DosR expression in *E. coli*. The *dosR* coding sequence was cloned into a pET expression vector, and expression was induced by IPTG. Lanes (1-3) full length: 1) uninduced, 2) induced 2 hr, 3) induced 5 hr, lanes (4-5) aa 1-134: 4) uninduced, 5) induced 5 hr, lanes (6-7) aa 144-217: 6) uninduced, 7) induced 5 hr. Arrows indicate the position at which DosR is expected to migrate. B. Electrophoretic mobility shift assay (EMSA) *of E. coli* extracts with full-length, N-terminal and C-terminal portions of DosR and radiolabeled *acr* promoter DNA. Lanes: 1) no extract, 2) aa 1-217 (full length), 3) aa 1-134, 4) aa 144-217.

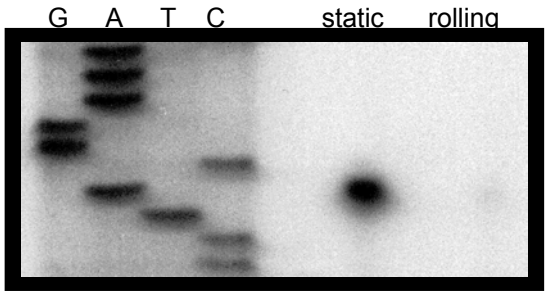

Figure 2. Localization of the *acr* transcription start site. Primer extension was performed with RNA from oxygenated (rolling) and hypoxic (static) log-phase cultures of MTB. Products were sized on a sequencing ladder of the *acr* promoter region.

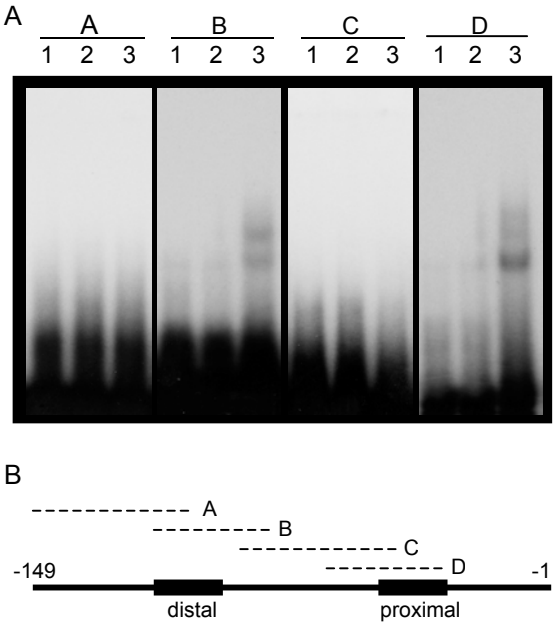

Figure 3. Localization of DosR binding within the *acr* promoter. A. EMSA of *E. coli* extracts and radiolabeled overlapping regions of *acr* promoter DNA. Labeled probe A (-149 to -96); B (-111 to -79); C (-89 to -49); D (-66 to -34). For each fragment, DNA was mixed with 1) no extract, 2) uninduced extract, and 3) induced extract. B. Schematic diagram of *acr* promoter (solid black line) with regions analyzed (dashed lines) relative to hypoxia motif sites (black boxes).

A distal: -110 to -91 acagggtcaatggtcccccaa
proximal: -53 to -34 tcggggacttctgtccctag

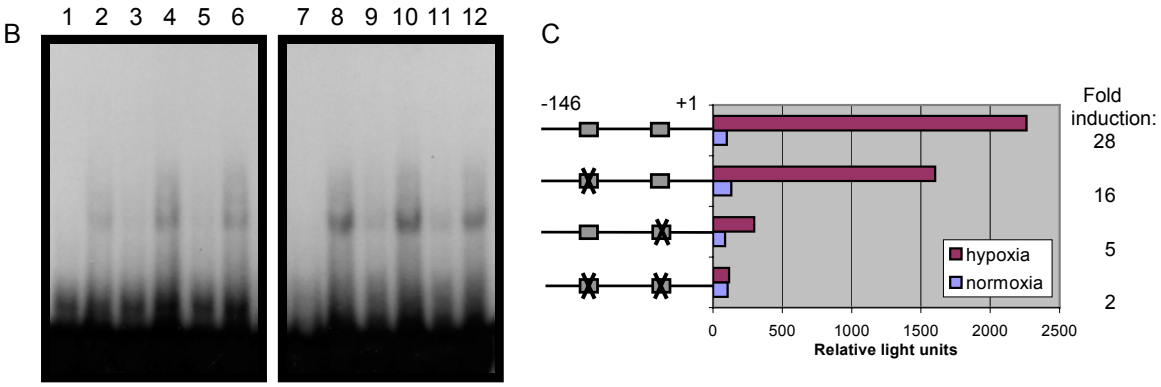

Figure 4. Effects of mutations in the hypoxia motifs upstream of the *acr* coding sequence. A. 20-mer hypoxia motifs in the native *acr* promoter. Positions are relative to transcription start site. Underlined bases were mutated as follows: "ccaa" ⟶ "ggtt," and "tcgg" ⟶ "agcc." B. EMSA with radiolabeled probes and specific unlabeled competitor DNAs. Lanes 1-6, labeled probe (-111 to -79); lanes 7-12, labeled probe (-66 to -34). Lanes 1 and 7: no extract control. Lanes 2 and 8: no competitor control. Lanes 3 and 11: unlabeled competitor (-111 to -79). Lanes 4 and 12: unlabeled competitor (mutant -111 to -79). Lanes 5 and 9: unlabeled competitor (-66 to -34). Lanes 6 and 10: unlabeled competitor (mutant -66 to -34). C. Hypoxic induction of *acr* promoter with mutations in the hypoxia motifs. Activity of *acr* promoter was measured by luciferase reporter gene assay. Indicated mutations in distal and proximal motifs were made as described above. Shown are representative data from one of two experiments, each of which was performed in duplicate.

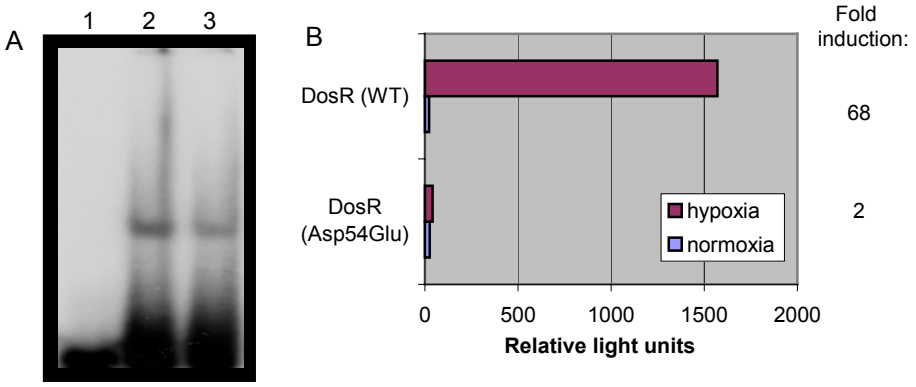

Figure 5. DosR Asp54 is required for gene induction but not for binding. A. EMSA of mutant DosR (Asp54Glu) with labeled *acr* promoter DNA. Lanes 1) no extract, 2) induced extract of DosR (Asp54Glu), 3) induced extract of DosR (WT). B. Hypoxic induction of *acr* by mutant DosR (Asp54Glu). Acr expression measured by luciferase reporter gene in BCG and in BCG with DosR (Asp54Glu) replacing DosR (WT). Shown are representative data from one of two experiments, each of which was performed in duplicate.