# Host genetics and tuberculosis susceptibility

**P. Selvaraj**

Department of Immunology, Tuberculosis Research Centre, Indian Council of Medical Research, Mayor V. R. Ramanathan Road, Chennai 600 031, India

**Susceptibility to tuberculosis is multifactorial. The importance of host genetic factors on the susceptibility or resistance to tuberculosis has been emphasized by many workers. Host genetic factors such as human leucocyte antigens (HLA) and non-HLA genes that are associated with the susceptibility to tuberculosis will serve as genetic markers to predispose or predetermine the development of the disease. Such markers may be useful to understand the immune mechanism of susceptibility or resistance to tuberculosis. Association of various HLA and non-HLA genes with susceptibility to tuberculosis in various ethnic population has been established. HLA studies carried out in the Asian region, especially in India, revealed the association of HLA-DR2 and -DQ1 antigens with the susceptibility to pulmonary TB. Further, studies on DNA typing explored the association of DRB1 *1501 and *1502 (DR2 subtypes) in north Indian and DRB1 *1501, DRB1 *0601 (DQ1 subtype) and DPB1 *02 (DP2 subtype) in south Indian population. Various studies on non-classical major histocompatibility complex (MHC) genes and non-MHC/non-HLA gene polymorphisms such as transporter associated with antigen processing (TAP), tumour necrosis factor α and β (TNF α and β), mannose binding lectin (MBL), vitamin D receptor (VDR) (*Bsm*I, *Apa*I, *Taq*I and *Fok*I polymorphisms), Interleukin-1 receptor antagonist (IL-1RA) and natural resistance associated macrophage protein-1 (NRAMP-1) genes revealed the association of TAP2 gene variant along with HLA–DR2 and functional mutant homozygotes (FMHs) of MBL with the susceptibility to pulmonary TB. The polymorphic *Bsm*I, *Apa*I, *Taq*I and *Fok*I gene variants of VDR showed differential susceptibility and resistance with male and female subjects. These studies suggest that multicandidate genes are associated with the susceptibility to pulmonary tuberculosis in India.**

## Host genetics and susceptibility to disease

Host genetic factors explain, at least in part why some people resist infection more successfully than others. Rare gene disruptions cause fatal vulnerability to certain pathogens, but more subtle differences are common and arise from minor variations in many genes. To predict how much our genetic make up determines the different ways in which we respond to some infectious agents is a difficult task. This is especially difficult because of the many other contributory factors such as previous health status, acquired immunity and variability in the pathogen.

Analysis of the genetic basis of susceptibility to major infectious diseases is potentially a most complex area. Many immunogenetic loci influence susceptibility to several infectious pathogens. A genetic basis for interindividual variation in susceptibility to human infectious diseases has been indicated by twin, adoptee, pedigree and candidate gene studies[1].

## HLA and non-HLA and disease association hypotheses

Several hypotheses were put forward to explain the mechanisms of major histocompatibility complex MHC and non-MHC gene association with the diseases. HLA-A, -B, -C (class-I) and -DR, -DQ and -DP (class-II) antigens could act directly as disease susceptibility agents. For this, three possible mechanisms have been suggested: (a) There could be antigenic cross-reactivity or mimicry between infectious organisms and a given HLA antigen. This phenomenon is termed as 'molecular mimicry'. Serological cross-reaction between HLA-B27 antigen and the bacterial strains *Klebsiella* and *Shigella* has been identified[2,3]. This means that common antigenic determinants are shared by HLA-B27 and the bacteria; (b) HLA antigens could act as receptors for microorganism; and (c) HLA antigens could influence particular immune responses, acting as immune response (*Ir*) genes. It has been shown that *Ir* genes regulate the immune response to any antigen or pathogen[4].

## Immune response gene effects

Genetically controlled differences exist in the magnitude of immune responses. The genes, which are responsible for this variation, were called as immune response (*Ir*) genes initially, till it became clear that *Ir* genes were, in most cases, one and the same as *MHC* genes. Several HLA-linked examples of diseases are available and this provides a attractive mechanism to account for disease susceptibility. The three major mechanisms involved in *Ir* gene effects are:

### *Determinant selection*

The individual MHC molecule selects the determinant of an antigen that is displayed to T-cells restricted by that MHC molecule.

*For correspondence. (e-mail: psraj21@hotmail.com)

### *Holes in the T-cell repertoire*

Gaps or holes may be created in the exported repertoire of T-cells that manifest as a failure to recognize some intrinsic antigens.

### *T-cell mediated suppression*

Active regulation of potentially reactive cells takes place by a population of cells whose function is to suppress an immune response[5].

The MHC genes may be physically close to the chromosome region that carries a gene conferring susceptibility or resistance to a particular disease. This hypothesis may explain the lack of complete association and geographical variation in the association, due to linkage disequilibrium.

Though classical genetic studies in humans and experimental models have clearly documented the primary contribution of the MHC genes, these genes themselves appear to be insufficient in conferring susceptibility or resistance to disease and suggest the association of non-HLA genes.

## Host genetic factors and tuberculosis susceptibility/resistance

*Mycobacterium tuberculosis* is the causative pathogen for tuberculosis. Though environmental and socio-economic factors are primarily related, numerous studies have emphasized the importance of host resistance and hereditary susceptibility. It is estimated that one-third of the world's population is infected with *M. tuberculosis*. Among the infected only around 10% will ever develop clinical disease[6]. This raises the question 'What is different about those who succumb to tuberculosis?'. In 1926, accidental administration of live *M. tuberculosis* (in place of BCG) to babies in Lubeck, Germany left some babies unaffected whereas it led to severe disease and death in others[7]. This indicates that the majority of the population have effective innate resistance to tuberculosis.

Twin studies have supported a substantial role for host genetics in variable susceptibility to tuberculosis. These studies have compared the disease status among identical and non-identical twins, with the expectation that disease with genetically determined component. These twin studies have found higher concordance for tuberculosis among monozygotic twins compared to dizygotic twins[8,9].

The association of host genetic factors (HLA and non-HLA) with the susceptibility or resistance to tuberculosis has been studied using various methods such as case-control studies, candidate gene approach, family-based, genome-wide linkage studies.

Identifying HLA and non-HLA genes/gene products (antigens) which are associated with susceptibility or resistance to tuberculosis will serve to provide HLA genetic markers to predict the development or predispose tuberculosis. The protective association of HLA types will be useful for the development of new epitope-based vaccine. Studying the role of these markers in the immune mechanism underlying susceptibility or resistance to tuberculosis will be useful to understand the immunopathogenesis of the disease. Moreover, these studies may be useful for better management and control of the disease.

## HLA studies in tuberculosis

Racial differences in susceptibility to tuberculosis are well known. Several studies revealed the association of various HLA antigens with the disease susceptibility in different ethnic populations[10–16]. For this type of geographic variation, possible explanations have been put forward. It seems likely that evolutionary selection pressures have given rise to frequent polymorphisms in genes involved in resisting infectious pathogens and contributed to marked allele frequency differences at the same loci. When geographic variation in pathogen polymorphism is superimposed on host genetic heterogeneity, considerable variation may occur in detectable allelic association. Gene–environmental interactions are likely to introduce another layer of complexity. The genes involved in defense against infectious pathogens evolve more rapidly than others and excessive polymorphism in the human genome may result from selection pressures exerted by infectious diseases. Similarly, the causative organism *M. tuberculosis* also has genetic variation. During evolution, all these polymorphic forms might have evolved due to the host–parasite interaction[17].

## Studies in non-Asian countries

A large number of HLA association studies have been carried out in non-Asian countries. One of the first reports of an association between HLA and tuberculosis showed an increased frequency of HLA-B8 in Canada[10]. Other studies showed an increased frequency of HLA-B5, -B15 and -DR5 in the North American blacks[11,12], HLA-A2 and -B5 in the Egyptian population[14] and -B27 in the Greek population[15]. A negative association has been reported for -DR6 in American blacks[13].

### *Studies in Asian populations*

Several studies of HLA association with pulmonary tuberculosis have been carried out in Chinese[16], Indonesian[18] and Russian patients[19]. A significantly increased frequency of HLA-DR2 was seen in the major studies which have revealed HLA-DR2 association with higher susceptibility

to tuberculosis. In a small study of tuberculosis in Vietnam, a susceptibility association with the rare HLA-DQB1 *0503 allele was reported[20]. Another study carried out in Thais revealed the association of HLA-DQB1 *0502 (ref. 21).

Of the numerous Indian studies on HLA association with pulmonary tuberculosis, an increased frequency of HLA-DR2 and –DQ1 was shown to be associated with the susceptibility to pulmonary tuberculosis[22–24]. Molecular study has revealed that the allele DRB1 *1501 of HLA-DR2 was higher compared with DRB1 *1502 in north Indian patients[25]. Studies carried out in south Indian patients revealed that, HLA-DRB1 *1501, (refs 26, 27) HLA-DQB1 *0601 (a subtype of HLA-DQ1) and -DPB1 *02 were found to be positively associated with susceptibility to pulmonary tuberculosis while a negative association (preventive fractions associated with resistance) has also been identified (DRB1 *11(5), DRB1 *10, DQB1 *0501 and DRB1 *08). Haplotype analysis also supports the DRB1 *1501 -DQB1 *0601 association with susceptibility to pulmonary tuberculosis[26] (Table 1). Though HLA-DR2, DQ1 and their subtypes are significantly associated with the susceptibility to tuberculosis, they may not be the sole genetic markers to predispose tuberculosis (relative risk is around 2.5). This suggested to look for the association of various non-HLA gene polymorphic variants. Association of multi-candidate genes (HLA and non-HLA) has been suggested for various infectious diseases[17].

## Non-HLA studies in tuberculosis

In north Indian pulmonary tuberculosis patients, compared with control subjects, the 'Transporter' associated with antigen processing gene 2 (TAP2) has been shown to be associated with the susceptibility to pulmonary tuberculosis along with HLA-DR2 (ref. 28). Definite association between tuberculosis and the haptoglobin 2–2 phenotype has been shown in Russian patients[29]. No such association is observed in Indonesians[30] and Indians[31].

Genome-wide linkage studies on sib-pairs of families affected with tuberculosis enable the identification of several candidate genes that are associated with the susceptibility to tuberculosis[32]. Some of the non-HLA candidate genes are discussed below.

**Table 1.** Association of important candidate gene variants of HLA and non-HLA genes with the susceptibility or resistance to pulmonary tuberculosis in Indian population

| Candidate genes | Effect | Reference |
|---|---|---|
| HLA | | |
| HLA-DR2 | Susceptibility | 22, 23, 24 |
| Sub-type | | |
| – DRB1 *1501, *1502 | Susceptibility | 25 |
| – DRB1 *1501 | Susceptibility | 26, 27 |
| HLA-DQ1 | Susceptibility | 24, 26 |
| – DQB1 *0601 | Susceptibility | 26 |
| HLA-DP | | |
| – DPB1 *02 | Susceptibility | 26 |
| Haplotype: | | |
| DRB1 and 1501-DQB1 *0601 | Susceptibility | 26 |
| DRB1 *11(5), DRB1 *10, DQB1 *0501 | Resistance | 26 |
| Non-classical HLA | | |
| Transporter Associated with Antigen Processing (TAP) gene TAP2 and DR2 | Susceptibility | 28 |
| Non-HLA | | |
| Functional Mutant Homozygotes of Mannose Binding Lectin (MBL) gene (codon 52, 54 and 57) | Susceptibility | 36 |
| – Heterozygotes of MBL codon 57 | Resistance to bacteriological relapse | 36 |
| Vitamin D Receptor (VDR) gene variants (*Bsm*I, *Apa*I, *Taq*I and *Fok*I) | Differential susceptibility and resistance in males and females | 45, 46 |
| NRMAP1 [(CA)n, 823 C/T, TGTG+/del and D543N G/A] | No association with susceptibility or resistance | 59 |
| Cytokine gene | | |
| TNFα – 238, – 308 | No association | 60 |
| TNFβ | No association | 60 |
| Haplotypes | | |
| HLA-B17-TNFα-238/A<br>HLA-B17-TNFα-308/2<br>HLA-B17-TNFβ-2 | Associated with bacteriological relapse | 60 |

### *Mannose-binding protein*

Mannose-binding protein (MBP), also known as mannose-binding lectin (MBL) is an acute phase protein secreted by the liver. It binds mannose and *N*-acetylglucosamine terminated glycoproteins and plays an important role in host defence against pathogens. Upon binding with certain carbohydrate moieties, such as terminal *N*-acetyl glucosamine or mannose, on various pathogens, MBP activates complement via specific protease and acts directly as an opsonin using the Clq receptor on macrophages. Mutations are found at the coding regions of the MBP genes, i.e. at codons 52, 54 and 57 that lead to low or near absent serum MBP levels in heterozygote and homozygotes respectively. Low serum level of MBP is associated with a common opsonic defect and is frequent in recurrent infections during infancy and possibly infections in adult life.

Several groups have studied MBL genotypes and tuberculosis, following a suggestion that MBL deficiency might have been maintained evolutionarily by a reduced capacity of mycobacteria to invade macrophages in the absence of MBL, leading to resistance to tuberculosis[33]. A study carried out in South Africa suggested that MBL-54 heterozygotes may be associated with protection against tuberculous meningitis[34] but a larger study in Gambia found no genotypic association[35]. Our study in south Indian population revealed an increased genotype frequency of MBP functional mutant homozygotes (including codons 52, 54 and 57) in pulmonary tuberculosis (10.9%) compared with control subjects (1.8%). Analysis of association of MBP genes and HLA-DR2 has showed that these genes are associated with susceptibility to pulmonary tuberculosis, independent of each other[36] (Table 1). Recently, a Mexican study of surfactant genes expressing collectins that are evolutionarily and functionally related to MBL genes has been suggested to influence tuberculosis susceptibility[37].

### *Vitamin D receptor*

It has long been suspected that vitamin D may be important in immunity to *M. tuberculosis*. Prior to the availability of antituberculous drugs, vitamin D was used in the treatment of patients with cutaneous tuberculosis and was reported to have dramatic effects[38]. The prevalence of both vitamin-D deficiency and tuberculosis is high among Asian immigrants in the UK, suggesting that vegetarian diet is a risk factor for tuberculosis[39]. 1,25-dihydroxyvitamin $D_3$ ($1,25(OH)_2D_3$) is an important immuno-modulatory hormone which activates monocytes and suppresses lymphocyte proliferation, immunoglobulin production and cytokine synthesis[40,41]. *In vitro*, 1,25 $(OH)_2D_3$ enhances the ability of human monocytes to restrict the growth of *M. tuberculosis*[40,42]. The effects of vitamin D are exerted by interaction through vitamin D receptor (VDR). Various diallelic polymorphisms have been identified in the vitamin D receptor gene and these polymorphic variants have been shown to be associated with the susceptibility or resistance to tuberculosis.

In a study carried out in the Gambian (West Africa) pulmonary TB patients, the *tt* genotype of *Taq*I polymorphism of *VDR* gene was found less frequently in cases of pulmonary TB, suggesting that this genotype may be associated with resistance to pulmonary TB whereas *Apa*I polymorphism showed no association[43]. The variant *ff* genotype (homozygote) of *Fok*I polymorphism of *VDR* gene and 25-hydroxylcholecalciferol deficiency have been shown to be strongly associated with pulmonary tuberculosis in Gujarati Indians living in London[44]. Our preliminary studies in south Indian pulmonary TB patients on *Bsm*I, *Apa*I, *Taq*I and *Fok*I polymorphisms of *VDR* gene showed an increased frequency of the genotypes Bb (heterozygote) of *Bsm*I, TT (homozygote) of *Taq*I and FF (homozygote) of *Fok*I polymorphism, in males and *tt* genotype (homozygote) of *Taq*I polymorphism in female patients suggesting the association with the susceptibility to TB[45,46]. Whereas genotypes BB (homozygotes) of *Bsm*I and AA (homozygous) of *Apa*I polymorphism are associated with resistance to pulmonary tuberculosis in male subjects[45,46]. The variant genotypes of *Bsm*I, *Apa*I, *Taq*I and *Fok*I sites of *VDR* gene either alone or in combination with each other as haplotype may be associated with susceptibility or resistance to pulmonary tuberculosis in males or females (Table 1). This type of differential susceptibility with variant genotypes of *VDR* gene in male and female subjects may be due to the circulating level of vitamin $D_3$, dietary intake of vitamin $D_3$, level of vitamin D receptor expression and other host factors. Further, studies on the level of circulating vitamin $D_3$, vitamin D receptor expression and the variant genotypes of vitamin D receptor will explore the mechanism of tuberculosis susceptibility in males and females. It is well established that the prevalence of tuberculosis is more in males[47]. Recently, an X chromosome susceptibility gene has been suggested which may contribute to the excess of males with tuberculosis observed in many populations[48].

### *Natural resistance associated macrophage protein 1 (NRAMP1)*

NRAMP1 (recently renamed as SLC11A1-solute carrier family 11, member 1) was identified by several groups working on a mouse locus that confers susceptibility to intracellular infections, such as *Leishmania*, *Salmonella* and the BCG strain of *Mycobacterium bovis*[49]. NRAMP1, like the related NRAMP2 (SLC11A2), is probably a divalent cation transporter and is found in the membrane of the phagolysosomes[50]. In mouse models, NRAMP1 is important in resistance to several intracellular infections.

The human *NRAMP1* gene has several polymorphisms[51]. The effects of *NRAMP1* gene variants seem more modest, association has been found between tuberculosis susceptibility and NRAMP1 in populations as diverse as West Africans[52,53], Japanese and Koreans[54]. A study carried out in Taiwanese population revealed no association of *NRAMP1* gene variants with the susceptibility to tuberculosis[55]. Linkage between tuberculosis and the *NRAMP1* locus has been shown in a large Canadian pedigree[56], but linkage was not seen in Brazilian, West African or South African populations[57,58].

Our studies on *NRAMP1* gene polymorphism [$(CA)_n$, 823 C/T, TGTG+/del and D543N G/A] in south Indian pulmonary and spinal tuberculosis patients revealed no association with the susceptibility to pulmonary and spinal TB in Indian population. It was suggested that MHC and other non-MHC gene polymorphic variants may be associated[59] (Table 1).

### *Cytokine genes and receptors*

An analysis of the course of infection in gene-knock-out mice has provided examples of the potential relevance of polymorphism in cytokine and cytokine receptor genes to infectious disease susceptibility in humans.

*Tumour necrosis factor-α and β*: Increased production of inflammatory cytokines, such as tumour necrosis factor-α (TNF-α) has been found in tuberculosis and various other infectious diseases. TNF-α, is mainly produced by monocytes and macrophages and TNF-β by T-lymphcoytes. Variant genotypes of TNF-α are associated with increased production of TNF-α. Association studies have been carried out on polymorphisms in and near the tumour necrosis factor (TNF) gene located in class III region of MHC. Our studies on TNF-α (– 238 and – 308) and TNF-β gene polymorphisms in Indian pulmonary tuberculosis patients revealed no association either with susceptibility or resistance[60] (Table 1). A study carried out in Cambodian tuberculosis patients also revealed no association with TNF-α[61].

*Interleukin-1(IL-1)*: Interleukin-1 (α and β), another inflammatory cytokine, gene polymorphism has been studied in Gambians[62], Gujarati Indians[63] and Cambodians[61]. These studies revealed no association with the susceptibility to tuberculosis.

*IL-1 receptor antagonist (IL-1RA)*: Interleukin-1 receptor antagonist (IL-1RA) is another cytokine factor which competes for the IL-1 binding site. The association of *IL-1RA* gene variants in various diseases has been studied. Macrophages from carriers of IL-1RA alleles have been shown to produce more IL-1RA and less IL-1α than other genotypes. *IL-1RA* gene variants are not associated with the susceptibility to pulmonary tuberculosis. However, association of the haplotype IL-1 Ra $A2^-$/IL -1β (+ 3953) $A1^+$ with the susceptibility has been reported with tuberculous pleurisy[63]. Our study on *IL-1RA* gene polymorphism in Indian pulmonary tuberculosis patients revealed no association with any of the genotypes but spinal tuberculosis patients showed a trend towards an increased frequency of genotype 22 compared with the control subjects[45,64].

*Interleukin-10*: This is a macrophage-deactivating cytokine. *NRAMP1* gene has been suggested to influence tuberculosis susceptibility by regulation[65] of interleukin-10. In Cambodian patients, association of heterozygosity for the -1082 polymorphism of the IL-10 promoter with TB susceptibility has been reported[61].

*Interleukin-12 receptor (IL-12R)*: Interleukin-12, a cytokine associated with increased production of Th1 type of cytokines, binds to interleukin-12 receptor. A case control study carried out in Japanese tuberculosis patients revealed the association of homozygosity for R214-T365-R378 allele (genotype 2/2) with the susceptibility to tuberculosis. This genetic variation has been suggested to predispose individuals to tuberculosis infection by diminishing receptor responsiveness to IL-12 and to IL-23, leading to partial dysfunction of interferon-gamma-mediated immunity[66].

*Interferon-γ receptor (IFN-γR)*: Interferon-γ receptor (IFN- γR)/gene variants has been shown to be associated with the susceptibility to atypical mycobacterial infection with *M. fortuitum*, *M. cheloni* and *M. avium*[67]. A different mutation, IFN-γR1, was identified in a child with fatal disseminated BCG infection[68].

## Conclusions

Developments in modern genetics and genomics have contributed to our understanding of the pathogenic processes that underlie major infectious diseases by allowing a more systematic study of the genetic influences. The number of candidate susceptibility genes is expanding rapidly. Moreover, genome-wide linkage analysis is also beginning to provide insights into complex disease. Advances in single nucleotide polymorphism (SNP) typing, microarray technology and bioinformatics will be helpful in the study of infectious diseases.

The development of tuberculosis or other mycobacterial diseases is the result of a complex interaction between the host and pathogen influenced by environmental factors. Numerous host genes are likely to be involved in this process. Using a variety of study methods, substantial progress has already been made in advancing our understanding of genetic susceptibility to tuberculosis. However, only a small part of the total familiar clustering observed in tuberculosis can be explained by the host

genes identified to date. There is much work still to be done as there are likely to be many more tuberculosis-susceptibility genes to be identified[69].


1. Hill, A., The genomics and genetics of human infectious disease susceptibility. *Annu. Rev. Genomics Hum. Genet.*, 2001, **2**, 373–400.
2. van Bohemen, C. G., Grumet, F. C. and Zanen, H. C., Identification of HLA-B27M1 and -M2 cross-reactive antigens in Klebsiella, Shigella and Yersinia. *Immunology*, 1984, **52**, 607–610.
3. Schwimmbeck, P. L., Yu, D. T. and Oldstone, M. B., Autoantibodies to HLA B27 in the sera of HLA B27 patients with ankylosing spondylitis and Reiter's syndrome. Molecular mimicry with Klebsiella pneumoniae as potential mechanism of autoimmune disease. *J. Exp. Med.*, 1987, **166**, 173–181.
4. McDevitt, H., Ir genes: towards the year 2000. *Res. Immunol.*, 1991, **142**, 509–513.
5. Gershon, R. K., T cell control of antibody production. *Contemp. Top. Immunobiol.*, 1974, **3**, 1–40.
6. Murray, C. J., Styblo, K. and Rouillon, A., Tuberculosis in developing countries: burden, intervention and cost. *Bull. Int. Union. Tuberc. Lung. Dis.*, 1990, **65**, 6–24.
7. Dubos, R. and Dubos, J., *The White plague: Tuberculosis, Man and Society*, Little, Brown and Co., Boston, 1952.
8. Kallmann, F. J. and Reisner, D., Twin studies on the significance of genetic factors in tuberculosis. *Am. Rev. Tuberc.*, 1942, **47**, 549–574.
9. Comstock, G. W., Tuberculosis in twins: a re-analysis of the Prophit survey. *Am. Rev. Respir. Dis.*, 1978, **117**, 621–624.
10. Selby, R., Barnard, J. M., Buehler, S. K., Crumley, J., Larsen, B. and Marshall, W. H., Tuberculosis associated with HLA-B8, BfS in a Newfoundland community study. *Tissue Antigens*, 1978, **11**, 403–408.
11. Al-Arif, L. I., Goldstein, R. A., Affronti, L. F. and Janicki, B. W., HLA-Bw15 and tuberculosis in a North American black population. *Am. Rev. Respir. Dis.*, 1979, **120**, 1275–1278.
12. Cox, R. A., Arnold, D. R., Cook, D. and Lundberg, D. I., HLA phenotypes in Mexican Americans with tuberculosis. *Am. Rev. Respir. Dis.*, 1982, **126**, 653–655.
13. Hwang, C. H., Khan, S., Ende, N., Mangura, B. T., Reichman, L. B. and Chou, J., The HLA-A, -B and -DR phenotypes and tuberculosis. *Am. Rev. Respir. Dis.*, 1985, **132**, 382–385.
14. Hafez, M., el-Salab, S., el-Shennawy, F. and Bassiony, M. R., HLA-antigens and tuberculosis in the Egyptian population. *Tubercle*, 1985, **66**, 35–40.
15. Zervas, J., Constantopoulos, C., Toubis, M., Anagnostopoulos, D. and Cotsovoulou, V., HLA-A and B antigens and pulmonary tuberculosis in Greeks. *Br. J. Dis. Chest*, 1998, **81**, 147–149.
16. Hawkins, B. R., Higgins, D. A., Chan, S. L., Lowrie, D. B., Mitchison, D. A. and Girling, D. J., HLA typing in the Hong Kong Chest Service/British Medical Research Council study of factors associated with the breakdown to active tuberculosis of inactive pulmonary lesions. *Am. Rev. Respir. Dis.*, 1988, **138**, 1616–1621.
17. Hill, A. V., The immunogenetics of human infectious diseases. *Annu. Rev. Immunol.*, 1998, **16**, 593–617.
18. Bothamley, G. H., Beck, J. S., Schreuder, G. M., D'Amaro, J. and de Vries, R. R., Association of tuberculosis and M. tuberculosis-specific antibody levels with HLA. *Infect. Dis.* (eds Kardjito, T. and Ivanyi, J. J.), 1989, **159**, 549–555.
19. Khomenko, A. G., Litvinov, V. I., Chukanova, V. P. and Pospelov, L. E., Tuberculosis in patients with various HLA phenotypes. *Tubercle*, 1990, **71**, 187–192.
20. Goldfeld, A. E., *et al.*, Association of an HLA-DQ allele with clinical tuberculosis. *JAMA*, 1998, **279**, 226–228.
21. Vejbaesya, S., Chierakul, N., Luangtrakool, K., Srinak, D. and Stephens, H. A., Associations of HLA class II alleles with pulmonary tuberculosis in Thais. *Eur. J. Immunogenet.*, 2002, **29**, 431–434.
22. Singh, S. P, Mehra, N. K., Dingley, H. B., Pande, J. N. and Vaidya, M. C., Human leukocyte antigen (HLA)-linked control of susceptibility to pulmonary tuberculosis and association with HLA-DR types. *J. Infect. Dis.*, 1983, **148**, 676–681.
23. Brahmajothi, V. *et. al.*, Association of pulmonary tuberculosis and HLA in south India. *Tubercle*, 1991, **72**, 123–132.
24. Selvaraj, P. *et. al.*, HLA antigen profile in pulmonary tuberculosis patients and their spouses. *Indian J. Med. Res.*, 1998, **107**, 155–158.
25. Rajalingam, R., Mehra, N. K., Jain, R. C., Myneedu, V. P. and Pande, J. N. J., Polymerase chain reaction-based sequence-specific oligonucleotide hybridization analysis of HLA class II antigens in pulmonary tuberculosis: relevance to chemotherapy and disease severity. *Infect Dis.*, 1996, **173**, 669–676.
26. Ravikumar, M. *et. al.*, Associations of HLA-DRB1, DQB1 and DPB1 alleles with pulmonary tuberculosis in south India. *Tuber. Lung Dis.*, 1999, **79**, 309–317.
27. Uma, S., Selvaraj, P., Kurian, S. M., Reetha, A. M. and Narayanan, P. R., Indian HLA-DR2 subtypes and immune responses in pulmonary tuberculosis. *Indian J. Med. Res.*, 2001, **113**, 117–124.
28. Rajalingam, R., Singal, D. P. and Mehra, N. K, Transporter associated with antigen-processing (TAP) genes and susceptibility to tuberculoid leprosy and pulmonary tuberculosis. *Tissue Antigens*, 1997, **49**, 168–172.
29. Kharakter, Zh, Z., Mazhak, K. D. and Pavlenko, A. V., *Probl. Tuberk.*, (The role of genetically determined haptoglobin phenotypes in patients with destructive pulmonary tuberculosis) 1990, **7**, 50–52.
30. Grange, J. M., Kardjito, T., Beck, J. S., Ebeid, O., Kohler, W. and Prokop, O., Haptoglobin: an immunoregulatory role in tuberculosis? *Tubercle*, 1985, **66**, 41–47.
31. Papiha, S. S. *et al.*, Association of HLA and other genetic markers in South Indian patients with pulmonary tuberculosis. *Tubercle*, 1987, **68**, 159–167.
32. Bellamy, R. *et al.*, Genetic susceptibility to tuberculosis in Africans: a genome-wide scan. *Proc. Natl. Acad. Sci. USA*, 2000, **97**, 8005–8009.
33. Garred, P., Harboe, M., Oettinger, T., Koch, C. and Svejgaard, A., Dual role of mannan-binding protein in infections: another case of heterosis? *Eur. J. Immunogenet.*, 1994, **21**, 125–131.
34. Hoal-Van Helden, E. G. *et al.*, Mannose-binding protein B allele confers protection against tuberculous meningitis. *Pediatr. Res.*, 1999, **45**, 459–464.
35. Bellamy, R. *et al.*, Mannose binding protein deficiency is not associated with malaria, hepatitis B carriage nor tuberculosis in Africans. *Q. J. Med.*, 1998, **91**, 3–8.
36. Selvaraj, P., Narayanan, P. R. and Reetha, A. M., Association of functional mutant homozygotes of the mannose binding protein gene with susceptibility to pulmonary tuberculosis in India. *Tuber. Lung Dis.*, 1999, **79**, 221–227.
37. Floros, J. *et al.*, Surfactant protein genetic marker alleles identify a subgroup of tuberculosis in a Mexican population. *J. Infect Dis.*, 2000, **182**, 1473–1478.
38. Dowling, G. B. and Prosser-Thomas, E. W., Treatment of lupus vulgaris with calciferol. *Lancet*, 1946, 919–922.
39. Strachan, D. P., Powell, K. J., Thaker, A., Millard, F. J. and Maxwell, J. D., Vegetarian diet as a risk factor for tuberculosis in immigrant south London Asians. *Thorax*, 1995, **50**, 175–180.
40. Rook, G. A. *et al.*, Vitamin D3, gamma interferon and control of proliferation of Mycobacterium tuberculosis by human monocytes. *Immunology*, 1986, **57**, 159–163.
41. Lemire, J. M., Adams, J. S., Sakai, R. and Jordan, S. C., 1-alpha, 25-dihydroxyvitamin D3 suppresses proliferation and immunoglo-

bulin production by normal human peripheral blood mononuclear cells. *J. Clin. Invest.*, 1984, **74**, 657–661.
42. Rook, G. A., Taverne, J., Leveton, C. and Steele, J., The role of gamma-interferon, vitamin D3 metabolites and tumour necrosis factor in the pathogenesis of tuberculosis. *Immunology*, 1987, **62**, 229–234.
43. Bellamy, R. *et al.*, Tuberculosis and chronic hepatitis B virus infection in Africans and variation in the vitamin D receptor gene. *J. Infect. Dis.*, 1999, **179**, 721–724.
44. Wilkinson, R. J. *et al.*, Influence of vitamin D deficiency and vitamin D receptor polymorphisms on tuberculosis among Gujarati Asians in west London: a case-control study. *Lancet*, 2000, **355**, 618–621.
45. Selvaraj, P., Narayanan, P. R. and Reetha, A. M., Association of vitamin D receptor genotypes with the susceptibility to pulmonary tuberculosis in female patients and resistance in female contacts. *Indian J. Med. Res.*, 2000, **111**, 172–179.
46. Selvaraj, P., Chandra, G., Kurian, S. M., Reetha, A. M. and Narayanan, P. R., Association of vitamin D receptor gene variants of *Bsm*I, *Apa*I and *Fok*I polymorphisms with susceptibility or resistance to pulmonary tuberculosis. *Curr. Sci.*, 2003, **84**, 1564–1568.
47. Fifteen year follow up of trial of BCG vaccines in south India for tuberculosis prevention. Tuberculosis Research Centre (ICMR), Chennai. *Indian J. Med. Res.*, 1999, **110**, 56–69.
48. Bellamy, R., Identifying genetic susceptibility factors for tuberculosis in Africans: a combined approach using a candidate gene study and a genome-wide screen. *Clin. Sci.* (*London*), 2000, **98**, 245–250.
49. Vidal, S. M., Malo, D., Vogan, K., Skamene, E. and Gros, P., Natural resistance to infection with intracellular parasites: isolation of a candidate for Bcg. *Cell*, 1993, **73**, 469–485.
50. Gruenheid, S. and Gros, P., Genetic susceptibility to intracellular infections: Nramp1, macrophage function and divalent cations transport. *Curr. Opin. Microbiol.*, 2000, **3**, 43–48.
51. Liu, J. *et al.*, Identification of polymorphisms and sequence variants in the human homologue of the mouse natural resistance-associated macrophage protein gene. *Am. J. Hum. Genet.*, 1995, **56**, 845–853.
52. Bellamy, R. *et al.*, Variations in the *NRAMP1* gene and susceptibility to tuberculosis in West Africans. *N. Engl. J. Med.*, 1998, **338**, 640–644.
53. Cervino, A. C., Lakiss, S., Sow, O. and Hill, A. V., Allelic association between the *NRAMP1* gene and susceptibility to tuberculosis in Guinea-Conakry. *Ann. Hum. Genet.*, 2000, **64**, 507–512.
54. Gao, P. S. *et al.*, Genetic variants of NRAMP1 and active tuberculosis in Japanese populations. International Tuberculosis Genetics Team. *Clin. Genet.*, 2000, **58**, 74–76.
55. Liaw, Y. S. *et al.*, Variations in the *NRAMP1* gene and susceptibility of tuberculosis in Taiwanese. *Int. J. Tuberc. Lung Dis.*, 2002, **6**, 454–460.
56. Greenwood, C. M. *et al.*, Linkage of tuberculosis to chromosome 2q35 loci, including NRAMP1, in a large aboriginal Canadian family. *Am. J. Hum. Genet.*, 2000, **67**, 405–416.
57. Shaw, M. A. *et al.*, Evidence that genetic susceptibility to Mycobacterium tuberculosis in a Brazilian population is under oligogenic control: linkage study of the candidate genes NRAMP1 and TNFA. *Tuber. Lung Dis.*, 1997, **78**, 35–45.
58. Bellamy R *et al.*, Genetic susceptibility to tuberculosis in Africans: a genome-wide scan. *Proc. Natl. Acad. Sci. USA,* 2000, **97**, 8005–8009.
59. Selvaraj, P., Chandra, G., Kurian, S. M., Reetha, A. M., Charles, N. and Narayanan, P. R., *NRAMP1* gene polymorphism in pulomonary and spinal tuberculosis. *Curr. Sci.*, 2002, **82**, 451–454.
60. Selvaraj, P., Sriram, U., Mathan Kurian, S., Reetha, A. M. and Narayanan, P. R., Tumour necrosis factor alpha (– 238 and – 308) and beta gene polymorphisms in pulmonary tuberculosis: haplotype analysis with HLA-A, B and DR genes. *Tuberculosis (Edinb)*, 2001, **81**, 335–341.
61. Delgado, J. C., Baena, A., Thim, S. and Goldfeld, A. E., Ethnic-specific genetic associations with pulmonary tuberculosis. *J. Infect. Dis.,* 2002, **186**, 1463–1468.
62. Bellamy, R., Ruwende, C., Corrah, T., McAdam, K. P., Whittle, H. C. and Hill, A. V., Assessment of the Interleukin 1 gene cluster and other candidate gene polymorphisms in host susceptibility to tuberculosis. *Tuber. Lung Dis.*, 1998, **79**, 83–89.
63. Wilkinson, R. J. *et al.*, Influence of polymorphism in the genes for the interleukin (IL)-1 receptor antagonist and IL-1 beta on tuberculosis. *J. Exp. Med.*, 1999, **189**, 1863–1874.
64. Selvaraj, P., Kurian, S. M., Reetha, A. M., Charles, N. and Narayanan, P. R., Vitamin D receptor and Interleukin-I receptor antagonist gene polymorphism in spinal tuberculosis. *Curr. Sci.*, 2000, **79**, 986–989.
65. Awomoyi, A. A., Marchant, A., Howson, J. M., McAdam, K. P., Blackwell, J. M. and Newport, M. J., Interleukin-10, polymorphism in SLC11A1 (formerly NRAMP1) and susceptibility to tuberculosis. *J. Infect. Dis.*, 2002, **186**, 1808–1814.
66. Akahoshi, M. *et al.*, Influence of interleukin-12 receptor beta1 polymorphisms on tuberculosis. *Hum. Genet.*, 2003, **112**, 237–243.
67. Newport, M. J., Huxley, C. M., Huston, S., Hawrylowicz, C. M., Oostra, B. A. and Williamson, R., A mutation in the interferon-gamma-receptor gene and susceptibility to mycobacterial infection. *N. Engl. J. Med.*, 1996, **335**, 1941–1949.
68. Jouanguy, E. *et al.*, Interferon-gamma-receptor deficiency in an infant with fatal bacille Calmette-Guerin infection. *N. Engl. J. Med.*, 1996, **335**, 1956–1961.
69. Bellamy, R., Susceptibility to mycobacterial infections: the importance of host genetics. *Genes Immun.*, 2003, **4**, 4–11.

ACKNOWLEDGEMENTS. I thank Dr P. R. Narayanan, Director, Tuberculosis Research Centre, Chennai for his encouragement.

# Cellular signaling pathways and transcriptional regulation in *Mycobacterium tuberculosis*: Stress control and virulence

**Patricia A. Fontan, Shaun Walters and Issar Smith***

TB Center, The Public Health Research Institute at the International Center for Public Health, 225 Warren Street, Newark, NJ 07103-3535, USA

***Mycobacterium tuberculosis* is a successful pathogen that overcomes the numerous challenges presented by the immune system of the host. This bacterium usually establishes a chronic infection in the lung where it may silently persist inside a granuloma until a failure in host defenses can lead to the reactivation of the disease. An increasing number of molecular genetic tools available in the past few years have made possible the study of signal transduction and transcriptional regulatory mechanisms that contribute to the ability of *M. tuberculosis* to be so well adapted.**

**This review intends to summarize the current knowledge about the major molecular components involved in signaling pathways and in transcriptional regulation in *M. tuberculosis* with specific emphasis on their relation to bacterial pathogenesis and the ability to cope with stresses inside the host.**

*MYCOBACTERIUM TUBERCULOSIS* is the causative agent of tuberculosis (TB). This disease together with HIV and malaria is one of the main causes of mortality due to an infectious disease[1]. Two billon people are infected in the world, 10% of them will develop TB at some point in their lives and 2 million die each year from the disease[2]. Although TB has been the focus of medical research for more than 100 years, *M. bovis* BCG, the only vaccine developed and used for the last 70 years is not completely satisfactory due to great variation in its efficacy[3]. Currently used front-line antibiotics can be effective but these are not available in all places in the world, and there is also the severe problem of newly emerging drug-resistant strains due to the use of inferior drugs or noncompliance[4]. Infection of a mammalian host by *M. tuberculosis* occurs primarily by the aerosol route. The lung is usually the principal organ affected, and the bacteria initially reside in alveolar macrophages[5]. In those hosts that cannot control the bacterial infection, *M. tuberculosis* can overcome the hostile conditions and is usually able to replicate in the macrophages. To survive in this cell type, *M. tuberculosis* has developed strategies to arrest phagosome maturation at an early stage, maintaining a relative non-acidic pH and avoiding fusion with lysosomes[6]. An immuno-competent individual infected with *M. tuberculosis* is usually able to develop a strong immune response whereby numerous peripheral blood monocytes and T lymphocytes migrate to the lung and contain the focus of infection by forming a granuloma. In these stages of containment the bacteria may remain for several months or years usually until the person becomes immuno-compromised for some reason (e.g. immunosuppressive drugs, AIDS, aging, etc.). This state of dormancy and persistence is one of most intriguing phenomena of *M. tuberculosis* lung infection.

To be so well adapted to the numerous environmental conditions that the host offers, *M. tuberculosis* must be able to mount an array of functions, like inducing metabolic pathways to utilize the carbon source available inside the macrophage, scavenging oxygen radicals to avoid cellular damage, and acquiring iron, to name a few. Thus, to establish a successful infection, bacteria have to constantly sense the medium, and efficiently signal the changes that enable a quick adjustment to new conditions. This series of events occurs coordinately by means of numerous mechanisms of signal transduction and transcriptional regulation. Signal transduction in bacteria mainly involves the so-called two component systems, however mycobacteria also have several serine/threonine kinases and tyrosine phosphatases that were originally thought to participate only in eukaryotic signal transduction. After the signals indicating environmental changes are sensed and transduced, bacteria respond by synthesizing new proteins and down regulating others. This balanced regulation of gene expression in bacteria occurs primarily at the level of transcription, largely through the activity of DNA-binding proteins called repressors or activators. As discussed below, the coordinate operation of the molecular machinery involved in bacterial signaling and transcription is essential for *M. tuberculosis* to successfully adapt inside the host.

## Two-component systems

The two-component systems form a large family of proteins involved in signal transduction that allow bacteria to detect and respond to many different kinds of stimuli. This signaling mechanism is widespread throughout the prokaryote world and is also found in some eukaryotes. The basic two-component system consists of two proteins, a sensor and a receiver, that are involved in a

*For correspondence. (e-mail: smitty@phri.org)

phosphotransfer reaction. After interaction with the appropriate stimulating ligand, the sensor protein, called a histidine kinase, can bind and hydrolyse ATP, catalysing the autophosphorylation of a conserved histidine residue founded in all histidine kinases and producing a high-energy phosphoryl group[7–9]. The phosphate is then transferred to the associated receiver protein, known as response regulator, at a conserved aspartic acid residue generating a high-energy acyl phosphate[10,11]. Once the phosphotransfer occurs, the response regulator is activated, allowing it then to carry out its specific function[12–14]. In most cases, the response is the modulation of transcription though the phosphorylated response regulator interacting with DNA at specific binding sites located in target gene promoter regions. The total effect is changes in global gene expression that aid the organism in responding to the initial signal sensed by the histidine kinase. Most histidine kinases and response regulator are highly conserved, however, there is little sequence similarity in the sensing domain of different histidine kinases. This variety of different extracellular, intracellular and/or transmembrane sensor domains accounts for the different types of molecules that can initiate the signal.

Two-component systems are involved in the signal transfer from initial stimulus to cellular response for a multitude of different processes. Some examples include: pathogenesis and virulence (*Salmonella typhimurium* PhoP–PhoQ)[15,16], chemotaxis (*Escherichia coli* CheA–CheY/CheB)[17], pilus production (*Pseudomonas aeruginosa* PilS–PilR)[18], adhesion (*P. aeruginosa* FleS–FleR)[19], osmoregulation (*E. coli* EnvZ–OmpR)[20,21], phosphate sensing (*E. coli* PhoR–PhoB)[22–25] and sporulation (*Bacillus subtilis* KinA–KinB–Spo0F–Spo0B–Spo0A)[26–29].

The published genome of *M. tuberculosis* H37Rv reveals the presence of eleven paired two-component systems, five unpaired response regulators, and two unpaired histidine kinases[30]. With the idea of identifying systems involved in virulence, a preliminary step in the characterization of the *M. tuberculosis* two-component systems involves mutagenesis of the individual systems and testing mutants for survival in macrophages and mice. Presently, many laboratories are mutating the *M. tuberculosis* two-component systems and performing such experiments. Of the eleven-paired systems in the *M. tuberculosis* genome, nine have been transcriptionally disrupted in either the response regulator or the histidine kinase and the resulting mutants' growth phenotype have been tested *in vivo*. In many cases these mutants have not been complemented, then the presence of suppressor mutations arising during the original mutagenesis, have not been ruled out.

## *M. tuberculosis* response regulator transcriptional regulation in macrophages

In an attempt to identify which two-component systems are induced during a macrophage infection, and hence possibly required for virulence, Deretic *et al.* characterized the gene expression of *M. tuberculosis* response regulators in 7H9 medium (*in vitro*) and in macrophages (*ex vivo*). Using green fluorescent protein (GFP) expression vectors fused with the promoters of select *M. tuberculosis* response regulators[31], constructs placed in *M. bovis* BCG and *M. tuberculosis* were screened for GFP expression during infection of macrophages. They found the response regulators *phoP* and *Rv0818* to be constitutively expressed *in vitro* and *ex vivo* in *M. bovis* and *M. tuberculosis*. The response regulator, *mtrA*, was induced in both *M. bovis* and *M. tuberculosis* in macrophages, while the response regulators *Rv0981* and *Rv3143* were induced in macrophages in *M. bovis* but not in *M. tuberculosis*. The following response regulators showed no induction in macrophages: *narL*, *Rv0903c*, *Rv1033c*, *Rv3765c*, *Rv1626*, and *Rv2884* (ref. 31). Since *mtrA* was the only response regulator induced during macrophage infection but not in broth culture, these workers attempted to delete the *mtrA–mtrB* system, but were unsuccessful, suggesting that this system is essential for growth[32]. While this work was an initial step in understanding the response regulators role in *M. tuberculosis* pathogenesis, a lack of gene induction in macrophages does not rule out their involvement in virulence since there are many aspects to *M. tuberculosis* infection and disease progression. Also, constitutively active response regulators may still be important for *M. tuberculosis* survival in macrophages, as seems to be the case with the PhoP–PhoR two-component system.

## PhoP–PhoR (Rv0757–Rv0758)

The role of two-component systems in the ability of a pathogen to successfully mount an infection has been well documented with the most extensively studied system involved in bacterial pathogenesis being the *Salmonella enterica serovar typhimurium* (*S. typhimurium*) PhoP–PhoQ system[15,16,33]. Knockout strains of either the response regulator PhoP or the histidine kinase PhoQ cause *S. typhimurium* to become avirulent[15,16]. The function of the PhoP–PhoQ two-component system is to detect the concentration of $Mg^{2+}$ and to a lesser extent calcium ($Ca^{2+}$) that is available to the bacterial cell[34]. In a low $Mg^{2+}$ environment, the PhoP–PhoQ system upregulates transcription of high affinity $Mg^{2+}$ transport systems allowing the bacteria to overcome $Mg^{2+}$ starvation.

The *M. tuberculosis* PhoP shows sequence similarity to the *S. typhimurium* PhoP response regulator[35]. In 2001, Perez *et al.* created an *M. tuberculosis* strain 103 mutant deleted in the PhoP–PhoR two-component system and showed that this mutant was unable to grow in murine bone marrow macrophages or in mice[36]. Complementation was performed using the response regulator and the wild type phenotype was restored when infected into macrophages, leading to the hypothesis that only *phoP* is

important for the virulence phenotype. This mutant is smaller in size and different than wild type in its cording properties when visually observed using the auramine (AFB) stain. In addition, mutant cells have an altered, rounded shape, and they show differences in levels of lipoarabinomannan derivatives, compared to the wild type[37].

Another mutant in the *PhoP–PhoR* two-component system in *M. tuberculosis* H37Rv is also attenuated in macrophages and mice (S. Walters *et al.*, unpublished results). *In vitro* growth experiments using this mutant reveal an inability to grow in low $Mg^{2+}$ containing medium, indicating that this system may be sensing $Mg^{2+}$ and is the functional homologue of the *S. typhimurium* PhoP–PhoQ system in *M. tuberculosis*. Genes requiring PhoP for their expression during growth in broth have been identified by DNA array transcription analyses, and among these 85 are genes encoding proteins involved in lipid metabolism, cell wall synthesis, membrane transport, and oxidative stress response (S. Walters *et al.*, unpublished results). However, the binding of PhoP to these genes has not been demonstrated yet, and it is not known which, if any of them are direct targets for this response regulator. A similar number of genes show higher expression in the *phoP* mutant, and the observation that many of these are also induced in hypoxic conditions and other stresses suggests that the absence of PhoP is, in itself, a severe stress condition. Current experiments are attempting to find which genes controlled by PhoP are necessary for virulence.

## MprA–MprB (Rv981–Rv982)

This two-component system was inactivated by Zahrt *et al.* and the mutant strain was tested for its ability to survive in activated and non-activated J774 macrophages and murine bone marrow-derived macrophages (BMM)[31]. The *mprAB* mutant grew better in non-activated macrophages than wild type, but displayed a similar survival phenotype to wild type in activated macrophages. In mice, the growth of the mutant is attenuated during the acute phase of infection in the spleen but not in the liver or in the lung. However, specifically in the lung and the spleen but not in the liver, the mutant failed to maintain viability and persist in the later stages of infection. These results collectively supported the role of the MprA–MprB two-component system in the establishment and maintenance of persistent infections in *M. tuberculosis*. The different phenotypes displayed by this mutant suggest that the histidine kinase MprB is sensing a tissue-specific ligand present during the many stages of the infection process or is able to respond to multiple signals. Future work on this interesting system will hopefully identify the stimulatory signal of MprB and the genes contained in the MprA regulon, which will provide many insights into the later stages of the *M. tuberculosis* pathogenic process.

## PrrA–PrrB (Rv903c–Rv902c)

By screening a transposon mutant library of *M. tuberculosis* Erdman for the presence of two-component system mutants, Ewann *et al.* isolated a transposon insertion in the promoter region of the response regulator *prrA*[38]. Since the separation between *prrA* and *prrB* is 14 bp, disruption of the *prrA* promoter causes a polar mutation on *prrB*, essentially mutating the entire system.

The *prrA* mutant was used to infect murine bone marrow macrophages and BALB/c mice and displayed a defect in early (day 1–6) intracellular growth in macrophages. This phenotype did not affect the final course of the infection, since a full growth capacity was restored after one week, and at nine days post inoculation the mutant showed a bacterial burden inside the macrophages that was similar to that observed for the wildtype strain. Complementation of the macrophage phenotype was performed by the re-addition of the *prrA–prrB* system into the *prrA* mutant.

When infected into mice, the *prrA* mutant was not attenuated. Promoter analysis of *prrA* showed an induction in *M. bovis* infected macrophages for the first four days of the infection and a gradual decrease in promoter activity. These results are consistent with the findings of Graham and Clark-Curtiss who isolated *prrA* cDNA from intracellular *M. tuberculosis* but not from *M. tuberculosis* broth cultures, but inconsistent with the findings of Deretic *et al.* that did not observe induction of this system in macrophages[39,31].

## RegX–SenX3 (Rv0491–Rv0490)

These genes were found to be polycistronic with the promoter region preceding[40] *senX3*. Biochemical experiments have shown that SenX3 can catalyse autophosphorylation, and that the phosphotransfer reaction occurs with[41] RegX3. Both unphosphorylated and phosphorylated RegX3 can bind to the DNA region directly upstream of *senX3* with over-expression of *senX3–regX3* in *M. smegmatis* increasing *senX3* expression. The potential RegX3 binding region in the *senX3* promoter is a palindromic sequence separated by one nucleotide consisting of GCTGTTTG located 101 bp from the translational start codon. These results suggest that RegX3 activates the expression of the *senX3-regX3* operon, and through this auto-regulation, signal amplification can occur. It is not currently known which environmental signal SenX3 is sensing or the genes that RegX3 controls, however, the role of this system in virulence has been tested. A transposon insertion in the *regX3* gene in *M. tuberculosis* Erdman has been created and the mutants ability to survive in bone marrow macrophages and mice proved to be similar to wild type[38]. Additionally, analysis of the *M. tuberculosis senX3* promoter expression during an infection of *M. bovis* BCG in macrophages showed no increase in expression during 14 days of infection.

## TrcR–TrcS (Rv1033c–Rv1032c)

The histidine kinase TrcS can directly phosphorylate the response regulator TrcR, and similar to the *regX3–senX3* system, autoregulates its own gene expression though binding of the phosphorylated TrcR to the *trcR* promoter region[42,43]. Mutational analyses of this interaction and DNAse footprinting have identified an A-T rich sequence that is essential for TrcR binding and *trcR* regulation. The stimulatory signal for activation of the TcrR–TrcS system is currently unknown, but one study comparing the global transcription profile of an *M. tuberculosis trcS* mutant with *M. tuberculosis* H37Rv growing exponentially in 7H9 medium, found 36 genes expressed at a higher level in the wildtype, while 14 genes were overexpressed in the mutant[44].

Transcriptional disruption of the *tcrS* gene in *M. tuberculosis* H37Rv by homologous recombination and by transposon insertion in *M. tuberculosis* Erdman has been performed separately by two laboratories[38,45]. These *tcrS* mutants leave the upstream *trcR* gene intact, and presumably still functional, since in many examples, unphosphorylated response regulators are able to bind DNA and cause gene expression, albeit at lesser levels than when phosphorylated[46–52]. Therefore, these *trcS* mutants are not ideal to fully study the function of the *trcR–trcS* system in *M. tuberculosis* without the accompanying *trcR* mutant. These mutants may still retain many wildtype functions even though the systems ability to react to the initial stimulus is detached from the response in TrcS absence. Both mutants were tested for survival *in vivo*, and neither strain was attenuated in macrophages or in immunocompetent mice. Interesting, when the *tcrS* H37Rv mutant was infected into SCID mice, a hypervirulence phenotype occurred, with the mutant causing a significant increase in the time to death of infected mice[45]. The most direct and logical explanation for this phenomenon is that this two-component system's ordinarily function to repress genes necessary for bacterial pathogenicity, and this new class of mutant presents exciting possibilities concerning *M. tuberculosis* pathogenesis.

## DosR–DosS (Rv3133c–Rv3132c)

DosR, also known as DevR, has been studied in *M. smegmatis*, *M. bovis* BCG and *M. tuberculosis* (reviewed by Park *et al.* and references, therein)[53]. It is required for the expression of a regulon containing approximately 50 genes that is induced during hypoxia or anoxia during hypoxic induced dormancy/persistence[53] and during NO stress (M. Voskuil, pers. commun.). Among the DosR requiring genes induced under these conditions are: *hxpX*, encoding the α-crystalline chaperone-like protein; *narX* and *narK2*, annotated as encoding respectively, a fused nitrate reductase and a nitrite extrusion protein; and *fdxA*, encoding ferredoxin, a protein involved in alternative respiratory pathways. DosR binds to a consensus sequence directly upstream of the promoter region of some of the genes, including *hspX* that require DosR for their induction under the conditions mentioned above[53]. There are two binding sites for DosR in the *hspX* promoter region, one at nucleotide positions – 110 to – 91 and the second, – 53 to – 34, relative to the transcription initiation nucleotide. Mutating the upstream site lowered *hxpX* expression after anoxic stress approximately 50% while mutating the downstream sequence essentially abolished all the induced *hspX* transcriptional activity. This data strongly suggests that DosR is an activator of *hspX*, presumably by interacting with the two DosR-binding sites, which in turn facilitates the binding of RNA polymerase to the *hspX* promoter. A *M. tuberculosis dosR* mutant showed an interesting virulence phenotype in that it initially grows better than the *M. tuberculosis* H37Rv parent strain in activated murine macrophages and immunocompetent mice. The mutant also kills SCID mice more rapidly than the wildtype[45], suggesting that the DosR–DosS system also functions to repress genes needed for pathogenesis as it is postulated for the TcrR–TcrS system.

## KdpE–KdpD (Rv1028c–Rv1027c)

Genes annotated as being part of the *M. tuberculosis* $K^+$ uptake machinery are induced by low $K^+$ levels[54]. The two-component system KdpE (reponse regulator), KdpD (histidine kinase) is believed to control $K^+$ acquisition components in *M. tuberculosis* as related proteins do in other bacteria[55,56]. An analysis of the interactions between the sensing domain of KdpD with other unknown proteins was performed using yeast two-hybrid and three-hybrid systems and verified using surface-enhanced laser desorption/ionization time-of-flight mass spectrometry[54]. These elegant studies shown that the sensing domain of KdpD interacts with two membrane lipoproteins, LprF and LprJ, and that a ternary complex can be formed between the histidine kinase domain of KdpD, and LprF, LprJ and the KdpD sensing domain. This interaction identifies the LprF and LprJ lipoproteins as the first accessory, adapter or ligand-binding proteins associated with a *M. tuberculosis* two-component system and leads to the speculation that these proteins may function to modulate the kinase or phosphatase activity of KdpD. Additionally, using yeast two hybrid assays these authors also found a direct interaction between KdpD and KdpE.

In the genome of *M. tuberculosis*, the *kdpD–kdpE* genes are adjacent to but divergent from the *kdpFABC*, a gene cluster believed to be involved in $K^+$ acquisition. In *E. coli*, the *kdpFABC* operon codes for a four-subunit, $K^+$-transporting P-type ATPase consisting of the KdpA, KdpB, KdpC and KdpF proteins, and this operon is induced during osmotic stress[56,57]. Overexpression of LprF and LprJ in *M. tuberculosis* increased expression of the low $K^+$ induced *kdpFABC* and mutating them modu-

lated expression of this regulon. While direct interactions of KdpE with the promoter region of these genes have not yet been demonstrated, these experiments indicate that KdpE is transcriptionally modulating their expression and may also auto-regulated the gene expression of *kdpD–kdpE*.

The KdpD–KdpE two-component system role in *M. tuberculosis* pathogenesis has been investigated, and a *M. tuberculosis kdpE* mutant shown normal growth in human macrophages (S. Walters *et al.*, unpublished results), and another *kdpE* mutant exhibited a hyper-virulent phenotype in SCID mice[45].

## Other two component system regulators

Several other mycobacterial two-component system regulators have been mutated, but less is known about their function or the genes they control, compared to those already discussed above. Parish *et al.* created mutants in the *trcXY*, *narLS*, and the sole histidine kinase *Rv3220*, and tested the mutants' ability to grow in SCID mice. Both, *narLS* and the *Rv3220* mutants displayed a similar phenotype as wild type, while the *trcXY* mutant killed these mice quicker than wild type, adding it to the list of hyper-virulence mutants identified by this group.

Compared to other bacteria, *M. tuberculosis* contains relatively few two-component systems. For example, there are 62 two-component systems in *E. coli*[58], 27 in *Streptococcus pneumoniae*[59], 70 in *B. subtilis*[60], 74 in *Streptomyces coelicolor*[61], but only 11 paired systems in *M. tuberculosis*[30]. Since the primary residence of *M. tuberculosis* in the human lung, perhaps a less adaptive signaling repertoire may be required to survive compared to organisms like *S. coelicolor* that inhabit many different environmental niches. For *M. tuberculosis*, possibly only a few systems are required for a successful infection and through cross communication and unidentified accessory proteins, the different systems can coordinate their responses and further increase the complexity of the regulation of their signaling.

The next step in this exciting field of research is just beginning with the determination of the ligands stimulating the phosphotransfer reaction and the regulated genes of the *M. tuberculosis* two-component systems. Only recently, in a few examples, has there been identification of what the individual histidine kinases are sensing, and the genes the response regulators are controlling[44,53,54]. Identification of the activation signals of the two-component system involved in virulence, such as the PhoP–PhoR, MprA-MprB, and the PrrA–PrrB systems, will provide valuable information about the host environment, leading to a better understanding of the host–pathogen relationship. Additionally, discovery of the genes regulated by the response regulators of such systems are equally interesting, since they are the bacterial response to these signals. From examples in other bacterial systems, such as the *S. typhimurium* PhoP–PhoQ system, many of the genes regulated by response regulators involved in virulence are themselves virulence factors, acting in concert to create the proper bacterial response to the host. Therefore, with the identification of the *M. tuberculosis* systems involved in virulence and their target genes, many addition virulence factors in *M. tuberculosis* will be revealed paving the way for future exciting research.

## Ser/Thr kinases

Many signals that indicate the inner status or the environmental surroundings of bacteria are integrated in the cell by a network system of phosphorylation and desphosphorylation of proteins. Besides the 13 histidine kinases that are part of the two component systems, the *M. tuberculosis* genome is annotated to contain 11 eukaryotic-type serine/threonine kinases (named PknA, B, D, E, F, G, H, I, J, K and L), 36 phosphatases, eight phosphotransferases and at least one kinase inhibitor that are possibly components of the bacterial signaling network. Analysis of the eukaryotic like kinases in other microorganisms indicated that these kinases are components of signaling pathways involved in secondary metabolism and morphogenesis[62,63]. Moreover, it has been shown that they are also important virulence determinants and fundamental components of the bacterial stress response[64]. In *M. tuberculosis*, the ser/thr kinases are a largely unexplored family. By biochemical and structural analysis it was determined the location and the activity of several of these kinases, but there is not much information about possible substrates or pathways.

Expression of *pknA* in *E. coli* induced the elongation of cells indicating that probably this kinase is possibly involved in the control of morphological changes of the bacterium[65]. The crystal structure of PknB has been resolved and it has been determined that it is a transmembrane kinase. Since *pknB* is located in the same operon as *pknA*, it has been suggested that this kinase is also involved in the process of cell elongation. However, the over-expression of *pknA* alone in *E. coli* is sufficient to induce cell elongation in this microorganism. A gene encoding a ser/thr phosphatase is also part of the same operon with *pknA* and *pknB*. The kinases PknF and PknG have been also characterized[66]. PknF is a trans-membrane protein and it has been speculated that the carboxyl terminal region of this kinase might be involved in environment sensing[66]. PknG is cytosolic and has a structure similar to the *Yersinia* YopO. The *Yersinia* YopO is secreted, and it can mediate morphological alterations of infected cultured HeLa cells[67]. However, the effect, if any, of *M. tuberculosis* PknG on the signaling machinery of eukaryotic cells has not been determined.

For the rest of the ser/thr kinases, no evidence exists concerning their possible functions and only there are speculations according to the location of each gene in the

genome of *M. tuberculosis*[68]. *pknD* resides near other genes that code for the components of a phosphate-uptake system. The gene that encodes PknK is located near a putative transcriptional regulator and the carboxy-terminal region of PknK is itself homologous to some members of the LuxR family of transcriptional regulators[68]. It was suggested that PknG and PknH are involved in the process of glutamine uptake induced under nitrogen. *pknI* is in the same operon that the genes that encode for a probable D-amino acid hydrolase (*Rv2913c*) and a signal recognition particle (*ffh*). It has been suggested that these proteins are involved in cell division[68]. *pknL* is in the same operon as a putative transcriptional regulator. *pknF*, *pknE* and *pknH* are in or near operons that code for ABC transporters. *embR* is part of the same operon with *pknH*. EmbR is a transcriptional regulator that regulates *embB* and *embA*, two cell wall arabinosyltransferases that are target of the drug ethambutal[69]. Future studies of this family of eukariotic kinases should elucidate their role in the signaling pathways that couple external and internal signals with the transcriptional machinery of *M. tuberculosis*.

## Sigma factors

Among the most important bacterial transcription activators are the sigma factors. By binding to the RNA polymerase, sigma factors give to the core enzyme, containing the catalytic subunits the specificity for a particular promoter. This specific interaction between enzyme and promoter region provides a means of regulating gene expression in response to various environmental conditions as new sigma factors bound to RNA polymerase allow different groups of genes to be expressed[70].

Analysis of the *M. tuberculosis* genome sequence indicates that this pathogen has 13 sigma factors[30] that are members of the Sigma70 family because of the similarity that they share with the Sig70 from *E. coli*. According to the phylogenetic relationship the Sig70 family is divided into 4 or 5 groups[71]. Group 1 includes principal and essential sigma factors. In *M. tuberculosis* the group 1 is represented by SigA that most probably controls housekeeping genes[72,73]. SigB is the representative member of group 2 that contains sigma factors closely related to the primary sigma factor but are non-essential[70]. Expression of *sigB* in *M. tuberculosis* can be up-regulated under some stress conditions (see below). Group 3 comprises alternative sigma factors that are more distantly related and respond to specific signals. In *M. tuberculosis*, SigF represents this group and shows similarity to the alternative sigma factors SigF of *S. coelicolor* and sigB and SigF of *B. subtilis*. In these bacteria, SigF is involved in sporulation[74,75] and in *B. subtilis*, SigB controls responses to general stress[76]. The other 10 sigma factors are highly diverged from the Sigma70 and belong to the group number 4. These sigma factors are designated as the extracytoplasmic function (ECF) family that is involved in responding to external conditions[77].

Little information is available concerning the functions and regulation for most *M. tuberculosis* sigma factors. Manganelli *et al.* using RT–PCR with molecular beacons have studied the expression of 10 different sigma factors (*sigA*, *sigB*, *sigC*, *sigD*,*sigE*, *sigF*, *sigG*, *sigH*, *sigI*, *sigM*) during exponential phase and under several stress conditions[78]. In this work it was shown that the genes for all sigma factors were expressed during the exponential phase of bacterial growth, with *sigC* showing the highest levels of expression. When the bacteria were treated with non-toxic levels of sodium dodecyl sulphate (SDS) *sigC* and *sigM* were down-regulated and *sigB* and *sigE* up-regulated. Heat shock treatment down-regulated *sigC* and *sigG* but up-regulated *sigB*, *sigE* and *sigH* transcription. Mild cold shock or room temperature induced the expression of *sigI* and down-regulated *sigG*. Incubation of the bacteria with water did not affect *sigI* expression but decreased mRNA levels of the rest. Low aeration conditions (exponential growth cultures incubated 24 h at 37°C without agitation) or stationary phase (bacteria collected after 70 h at OD > 2) decreased mRNA levels of almost all sigma factors. In stationary phase only *sigB*, *sigE* and *sigF* remained constant and under low aeration, *sigE* and *sigF* remained constant but the transcription of *sigB* was significantly up-regulated. Hu *et al.* demonstrated that the expression of *sigI* and *sigJ* was prominently up-regulated in stationary phase when the bacteria were maintained for 100 days in standing cultures. In these conditions levels of *sigG*, *sigH*, *sigD*, *sigF*, *sigL* and *sigK* did not change but *sigA*, *sigB*, *sigC*, *sigE* and *sigM* were down-regulated[79]. These results may suggest a predominant role of SigI and SigJ in stationary growth phase conditions. Graham *et al.* had compared the expression *in vitro* and *ex-vivo* of the different sigma factors of *M. tuberculosis* using the selective capture of transcribed sequences method (SCOTS). These workers found that *sigA*, *sigB*, *sigD*, *sigE*, *sigF*, *sigG*, and *sigH* but not *sigC* where expressed when bacteria were growth in broth culture. When *M. tuberculosis* were growing inside macrophages, the expression of *sigA*, *sigB*, *sigE*, *sigF* and *sigH* [39] could be detected.

## SigA

The *sigA* genes of *M. smegmatis* and *M. tuberculosis* were thought to encode the principal sigma factors in these organisms by their primary amino acid sequence similarity to SigA, the principal factor of *Streptomyces* spp as well as their similarity to each other[80,81]. Moreover, *sigA* was shown to be essential in *M. smegmatis*[82] and in *M. tuberculosis* (J. Timm and I. Smith, unpublished results). As expected the promoter sequences recognized by RNA polymerase containing SigA are similar to those recognized by *E. coli* and *B. subtilis* RNA polymerases containing the principal sigma factor in these organisms[81].

SigA is also involved in virulence as *M. bovis* strain TMC403 that is attenuated in a guinea pig model contains a missense mutation (R515H) in the *sigA* (*rpoV*) gene[83]. This mutation is adjacent to region 4.2 that interacts with the promoter consensus-35 region. Mutations affecting transcription of specific genes have been described in the same region adjacent to region 4.2 in the major sigma factors of *E. coli* and *B. subtilis*[84,85]. These mutations, which have no effect on *in vitro* growth are postulated to prevent interaction between the major sigma factor and several positive activators. The phenotype of the *M. bovis sigA* R515H mutation suggests that there is an activator that interacts with SigA to transcribe *M. bovis* genes that are essential for pathogenicity. WhiB3, discussed below, seems to play this role[86].

## SigF

SigF, is an alternative sigma factor member of the group 3, and according to experiments performed in *M. bovis* BCG, the expression of this sigma factor is highly induced in stationary phase (when bacteria were grown with shaking until OD > 2) and also under nitrogen depletion[87]. A null mutant of *sigF* in *M. tuberculosis* was not attenuated in any of these conditions or when infecting macrophages. However, the *sigF* mutant was more sensitive to the treatment with the antibiotics rifampin and rifapentine and it was attenuated in an experimental model of infection in mice[88]. Genes under SigF regulation have been determined by DNA microarray (W. Bishai, pers. commun.). Previously, it has been reported that the expression of the 16-kilodalton alpha-crystallin protein (acr) in early stationary phase is under SigF regulation[89]. In the reports discussed above there is some discrepancy about the conditions under which some sigma factors are transcribed, particularly, with reference to the expression of *sigF* and *sigB* in stationary phase and low aeration. It is difficult to compare the different studies since definition of these conditions varies from one work to the other. Another complication is that the importance of changes in transcriptional levels of the sigma factors may be relative since other modes of regulation may also be relevant, i.e. post translational control, as discussed below.

## SigE, SigH and SigB

Recently, the response of SigE, SigH and SigB to different stresses has been extensively studied[90–92]. The availability of mutant strains and the use of DNA microarray technique have revealed that there is a transcriptional regulation circuit generated by these sigma factors. As it was described above, the expression of SigE is up-regulated under heat shock, cell wall stress and oxidative stress. A *sigE* mutant is more sensitive to SDS, oxidative stress and heat shock[78]. Using molecular beacons and DNA microarrays, Manganelli *et al.* have defined the SigE regulon under exponential growth and cell wall stress[90]. These studies demonstrated that under non-stress conditions the expression of 38 genes is dependent on SigE. Interestingly, only one gene, *sigB* that required SigE for its expression in exponential growth, had an EFC-sigma-like promoter sequence. The results of this array indicate that during exponential growth SigE, there are indirectly regulated housekeeping genes. After SDS treatment, 62 genes were induced in the wild type strain. The high expression of 23 genes in 13 transcriptional units requires SigE. A sequence that resembles an ECF sigma promotor preceded nine of these sequences. SigE differentially regulates genes involved in mycolic acid synthesis and fatty acid metabolism, as well as transcriptional regulators and there are a many other genes annotated as unknown. The role of these genes in the virulence of the bacteria remains to be elucidated by disrupting them and evaluating the virulence of the mutant strain.

*sigB* is also under SigE regulation during cell wall stress[90]. Recently, using DNA microarrays we compared gene expression of a *sigB* mutant with the wild type strain during SDS treatment. We demonstrated that only two genes, out of the 20 that are under SigE control after SDS treatment also require SigB for their maximum induction (P. Fontan *et al.*, unpublished results). These two genes are annotated to code for a small heat shock protein and a possible transcriptional regulator. The relevance of these genes for the virulence of the *M. tuberculosis* remains to be determined. SigH is the third member of the circuit. This sigma factor is the ortholog of SigR from *S. coelicolor*[93] and controls the expression of genes needed for bacteria to tolerate heat shock or oxidative stress[94]. A *sigH* mutant of *M. tuberculosis* was sensitive to heat shock, and to different oxidative stresses like hydrogen peroxide, organic peroxide and diamide (thiol oxidant)[91,92]. Microarray analysis indicates that SigH is not relevant for expression of genes during bacterial exponential growth, but SigH controls the transcription of 39 genes when bacteria are suffering redox stress induced by diamide, including *sigB* and *sigH*[91]. The induction of thioredoxin and thioredoxin reductase in a SigH dependant manner suggests that this is the system that mycobacteria used to maintain the redox homeostasis. The induction of *sigE* expression by diamide and the control of this expression by SigH have been described in two different reports. However, no *sigH* consensus promotor has been found upstream of *sigE* (Fontan *et al.*, unpublished results).

Interestingly analysis of the promotor region of genes under the control of SigE and SigH suggests that both sigma factors recognize very similar consensus sequence[91]. Thus, the same promoter is used by SigE under stress and non-stress conditions and by SigE under stress to induce the transcription of *sigB*. It is intriguing how promoter discrimination occurs between these two sigma factors. It can be speculated that a subtle changes in the sequence

are responsible for initiation of the transcription using one or another sigma factor. Another possibility is that each sigma factor is differentially available depending on the interaction with their specific antagonists, the anti-sigma factors.

The expression of sigma factors can be regulated at the level of synthesis, proteolysis or by the interaction with their anti-sigma factors[77]. In *M. tuberculosis* only the regulation of SigF has been carefully analysed[95]. Beaucher *et al.* have shown that SigF activity is regulated by reversible interaction with UsfX. This anti-sigma factor is the product of the *usfX* gene positioned directly upstream of the *sigF* gene. These authors also identified two antagonists of UsfX, the anti–anti sigma factors and demonstrated that one of these, RsfA is regulated by redox potential and they suggest that the second, RsfB may be regulated by phosphorylation. The postulated model for SigF and its regulators indicates that SigF remains bound to UsfX until bacteria are subject to reducing conditions, like the possible environment inside the granuloma. When the redox conditions changes, RsfA becomes active and binds UsfX liberating SigF. The role of RsfB is not clear and the kinase involved in the putative phosphorylation of this anti–anti-sigma factor has not been described yet. Three other proteins of *M. tuberculosis* are similar at the level of primary sequence to RsrA, the anti-sigma factor of SigR in *S. coelicolor*. The genes coding for these proteins are downstream of *sigH*, *sigE* and *sigL*. In each case the protein encoded by the putative anti-sigma factor specifically binds to and reversible inactivates the cognate sigma factor (S. Rodrigue *et al.*, unpublished results). Two more genes, *Rv1904* and *Rv2638* have been annotated as similar to an anti–anti-sigma factor of SigB in *S. coelicolor*, but the correspondent anti-sigma and sigma factors have not been identified yet.

The contribution of these global transcriptional regulators to the bacterial pathogenicity has been demonstrated *in vitro* and *in vivo*. The *sigE* mutant was impaired for growth in macrophages[90] and in mice (R. Manganelli *et al.*, unpublished results). *M. tuberculosis* strains carrying mutated *sigF* or *sigH* genes did not show any attenuation when used to infect macrophages but were demonstrated to be attenuated in different models of mice infection[88,96]. Animals infected with either *sigF* or *sigH* mutants did not show a reduced number of bacterial colony forming units (cfu) in the lung when compared with the cfus in the lungs of animal infected with the wild type strain. However, animals infected with a *sigH* mutant showed reduced histopathological signs of lung damage and the time-to death analysis of mice infected with *sigH* or *sigF* mutants was greater when compared with animals infected with the parental strain. These observations suggest that some genes of the regulons under the control of these sigma factors are important bacterial virulence determinants although the survival of the pathogen inside the host is not affected by the absence of these proteins.

## Accessory transcriptional factors

### *WhiB family*

WhiB is a transcriptional regulator that has been shown to be essential in *S. coelicolor* for the sporulation of aerial hyphae. Interestingly, this family of transcriptional regulators is present in all actinomycetes but not in other organisms[97]. WhiB has seven homologs in *M. tuberculosis*, and this family is present also in *M. leprae* and *M. smegmatis*. The function of these regulators in mycobacteria is intriguing since these microorganisms do not sporulate. One speculation is that in mycobacteria the *WhiB* genes can encode proteins related to the dormancy state of the bacteria. However, a *whiB3* null mutant in *M. smegmatis* did not show any defect when it was evaluated in the Wayne model for dormancy state[98]. It was demonstrated in *M. smegmatis* that *whiB2* is an essential gene[82]. By constructing an *S. smegmatis* strain with a disruption in the *whiB2* gene but with an extra-chromosomal copy of this gene under an inducible promotor, Gomez *et al.* demonstrated that WhiB2 has a role in cell division and septation[82]. In *M. tuberculosis*, the expression of this gene was observed to be up-regulated when bacteria infected macrophages. As it was described above, a point mutation, a change in Arg-515 to His, in the 4.2 region of SigA was shown to be responsible for the loss of virulence of *M. bovis*[83]. Steyn *et al.* have demonstrated that in *M. tuberculosis* WhiB3 interacts with the 4.2 region of SigA and that the lost of virulence of the *M. bovis* mutant strain is due to the lack of interaction of WhiB3 with SigA[86]. Strains with a mutated *whiB3* gene have been constructed in *M. bovis* and *M. tuberculosis*[86]. The mutant strain in *M. bovis* was attenuated in a guinea pig infection model. Evaluation of the *M. tuberculosis* mutant strain in different animal models of infection did not show any difference in bacterial load in different organs. Interestingly, mice infected with the mutant strain showed less tissue damage specifically in the lungs and the survival of the animal was significantly increased when compared with animals infected with the wild type strain.

## Transcriptional repressors

In the *M. tuberculosis* genome there are 31 proteins annotated as transcriptional repressors. For many of these proteins the only information available is the homology at the sequence level with transcriptional repressors characterized in other microorganisms. However some of these regulators have been well characterized by several laboratories and their roles in the regulation of the global

gene expression has been studied during different environmental stresses.

### *IdeR*

IdeR is the main regulator of iron metabolism in *M. tuberculosis*[99]. This transcriptional regulator was identified first as a repressor but recent work from our laboratory has demonstrated that IdeR is also an activator. Under high intracellular iron condition, IdeR remains bound to the iron box present in the promoter region of genes under its regulation. Thus, under high intracellular levels of iron, IdeR simultaneously represses the *mbt* operon involved in the synthesis of the siderophore mycobactin and activates the genes *bfrA* and *bfrB* that encode iron storage proteins[100,101]. The functional mechanisms of IdeR, as well as the complete regulon under IdeR control have been reviewed recently by Rodriguez and Smith[102].

### *LexA and RecA*

LexA and RecA are two regulatory proteins involved in the bacterial response to DNA damage[103]. When the DNA damage occurs, RecA binds to the single stranded DNA region and induces the autocatalytic cleavage of LexA, then LexA can no longer remain bound to the so-called SOS boxes found upstream genes of the SOS regulon. In a recent work, Davis *et al.* have defined the LexA binding motif in *M. tuberculosis*[104]. In the same work by using microarray analysis these authors demonstrated that the regulon under LexA control comprises 15 genes. Ten of these genes were not known to be induced by DNA damage. Interestingly, some genes known to be under LexA control in *E. coli* were not induced with mitomycin C (an inducer of DNA damage) in *M. tuberculosis* and some others induced genes did not show a LexA-binding motif in their promoters. This last observation indicated that LexA activation is not the only regulatory mechanism of the SOS response in this pathogen.

### *HspR and HrcA*

In *M. tuberculosis*, proteins involved in the heat shock response are negatively controlled by two transcriptional repressors, HspR and HrcA[105]. HspR is homologous to the repressor that controls the Hsp70 operon in *Streptomyces*[106]. HspR in *M. tuberculosis* controls the expression of the Hsp70 operon and in some extent the expression of GroES[105]. A mutant strain of *M. tuberculosis* with a deletion of *hspR* was attenuated when it was used to infect mice[107]. The authors suggest that this phenotype is probably due to an enhanced response of the host's immune response caused by over expression of the Hsp70. HrcA in *B. subtilis* controls both *hsp60* (*groEL*) and *hsp70* heat shock systems[108]. The homologous protein in *M. tuberculosis* mainly regulates the expression of the *hsp60* family of heat shock proteins (*groEL1*, *groEL2* and *groES*). Interestingly, the heat shock response in *M. tuberculosis* is under both, positive and negative regulation. At least three sigma factors, SigH, SigE and SigB are involved in the positive control of this response. SigH seams to be the main regulator controlling the expression of the *hsp70* operon and *clpB*[92]. SigH also controls *sigE* and *sigB* expression under heat shock stress conditions[78]. Preliminary results from our laboratory indicate that under heat shock stress SigB partially controls the expression of the *hsp70* operon, *acr2* and the *groEL/groES* chaperons (P. Fontan *et al.*, unpublished results). Since, SigE controls *acr2* expression when bacteria is treated with diamide, it will be interesting to determine if the sigE also controls the expression of this chaperon or others under heat shock stress.

### *RelA*

As discussed elsewhere in the introduction, *M. tuberculosis* is able to persist in a slow growing or non-growing 'latent' state for long periods in infected hosts. The bacterium must be able to shut down or down regulate the synthesis of those factors that are required for exponential growth, e.g. ribosomes, tRNAs, etc. At the same time, new components must be synthesized to allow bacterial survival under these conditions, e.g. there are marked changes in the cell wall when *M. tuberculosis* is maintained for long period under limited oxygen conditions[109], etc. The primary *in vitro* model used to study the phenomenon of latency has been the Wayne model in which bacteria are exposed to limited oxygen levels[110]. DosR (DevR), the two-component response regulator discussed previously, is induced under these conditions and is required for the up-regulation of other genes induced during anoxia[53,111]. Global *M. tuberculosis* gene expression with DNA arrays has been analysed after exposure to partial or complete nutrient starvation[112], as it is thought that these conditions may mimic the environment to which the bacterium is exposed at later stages of infection. In the course of this work, it was found that genes encoding ribosomal proteins, enzymes involved in intermediary metabolism, energy production and lipid biosynthesis, as well as other genes were all repressed. Significantly, *relA*, encoding (p)ppGpp synthase and some other regulatory proteins were induced by nutrient starvation. The RelA result was not unexpected since much work in other prokaryotes has shown that the regulatory nucleotide (p)ppGpp, the product of the enzymatic function, of this protein is to down-regulate the expression of genes for stable RNAs, ribosomal proteins and lipid biosynthetic enzymes among others, during nutrient starvation, while it activates other genes[113]. The mechanism by which (p)ppGpp selectively inhibits gene expression, while it activates others is unknown, but there is

some evidence that it binds to the β–subunit of RNA polymerase (reviewed in ref. 114).

The importance of (p)ppGpp for the late growth regulation in mycobacterial species had previously been observed, using conditions of carbon limitation and overproduction of the regulatory nucleotide[115]. Recent experiments with a *relA* mutant of *M. tuberculosis*[116] have confirmed the importance of (p)ppGpp in allowing *M. tuberculosis* to survive during stationary growth caused by nutrient limitation or during mouse infections[117]. The mutant shows decreased survival during stationary growth and in mouse lungs and spleens. DNA array analyses also indicated that during the normal limitation of growth caused by nutrient limitation, RelA and presumably its enzymatic product are responsible for down regulating genes for the translational apparatus. At the same time several other genes were higher in the *relA* mutant compared to the wild type strain during nutrient starvation[117]. These exciting results provide many new potential targets for possible antitubercular therapies.

## Conclusions

The relatively recent availability of tools to analyse *M. tuberculosis* at the molecular level had made possible a great advance in the understanding how this microorganism uses its cellular signaling and transcriptional machinery to survive different stresses. As we describe here, there is growing evidence for complex connections between the factors from the different regulatory circuits. These connections are certainly relevant under multiple simultaneous stress conditions, which is more related to the *in vivo* situation than to the carefully controlled single-stress situations usually studied in the laboratory. Now, taking advantage of the existing *in vitro* data and the available technology like DNA microarrays it will be possible to study the orchestrated response of *M. tuberculosis* during host infection. Hopefully, the identification of the regulatory components that contribute to the success of *M. tuberculosis* as a human pathogen will lead to the development of new drugs and vaccines to combat and prevent tuberculosis.


1. Letvin, N. L., Bloom, B. R. and Hoffman, S. L., Prospects for vaccines to protect against AIDS, tuberculosis, and malaria. *JAMA*, 2001, **5**, 606–611.
2. Corbett, E. L., Watt, C. J., Walker, N., Maher, D., Williams, B. G., Raviglione, M. C. and Dye, C., The growing burden of tuberculosis: global trends and interactions with the HIV epidemic. *Arch. Intern. Med.*, 2003, **9**, 1009–1021.
3. Kaufmann, S. H., Is the development of a new tuberculosis vaccine possible? *Nature Med.*, 2000, **9**, 955–960.
4. McKinney, J. D., *In vivo veritas*: the search for TB drug targets goes live. *Nature Med.*, 2000, **12**, 1330–1333.
5. Fenton, M. J., Macrophages and tuberculosis. *Curr. Opin. Hematol.*, 1998, **1**, 72–78.
6. Russell, D. G., *Mycobacterium tuberculosis*: here today, and here tomorrow. *Nat. Rev. Mol. Cell. Biol.*, 2001, **8**, 569–577.
7. Nixon, B. T., Ronson, C. W. and Ausubel, F. M., Two-component regulatory systems responsive to environmental stimuli share strongly conserved domains with the nitrogen assimilation regulatory genes ntrB and ntrC. *Proc. Natl. Acad. Sci. USA,* 1986, **20**, 7850–7854.
8. Parkinson, J. S. and Kofoid, E. C., Communication modules in bacterial signaling proteins. *Annu. Rev. Genet.*, 1992, **26**, 71–112.
9. Kofoid, E. C. and Parkinson, J. S., Transmitter and receiver modules in bacterial signaling proteins. *Proc. Natl. Acad. Sci. USA,* 1988, **14**, 4981–4985.
10. Lukat, G. S., Lee, B. H., Mottonen, J. M., Stock, A. M. and Stock, J. B., Roles of the highly conserved aspartate and lysine residues in the response regulator of bacterial chemotaxis. *J. Biol. Chem.*, 1991, **13**, 8348–8354.
11. Stock, A. M., Robinson, V. L. and Goudreau, P. N., Two-component signal transduction. *Annu. Rev. Biochem.*, 2000, **69**, 183–215.
12. Perraud, A. L., Weiss, V. and Gross, R., Signalling pathways in two-component phosphorelay systems. *Trends Microbiol.*, 1999, **3**, 115–120.
13. Rodrigue, A., Quentin, Y., Lazdunski, A., Mejean, V. and Foglino, M., Two-component systems in *Pseudomonas aeruginosa*: why so many? *Trends Microbiol.*, 2000, **11**, 498–504.
14. McCleary, W. R. and Stock, J. B., Acetyl phosphate and the activation of two-component response regulators. *J. Biol. Chem.*, 1994, **50**, 31567–31572.
15. Galan, J. E. and Curtiss, R., Virulence and vaccine potential of *phoP* mutants of *Salmonella typhimurium*. *Microb. Pathog.*, 1989, **6**, 433–443.
16. Miller, S. I., Kukral, A. M. and Mekalanos, J. J., A two-component regulatory system (*phoP phoQ*) controls *Salmonella typhimurium* virulence. *Proc. Natl. Acad. Sci. USA,* 1989, **13**, 5054–5058.
17. Lupas, A. and Stock, J., Phosphorylation of an N-terminal regulatory domain activates the CheB methylesterase in bacterial chemotaxis. *J. Biol. Chem.*, 1989, **29**, 17337–17342.
18. Hobbs, M., Collie, E. S., Free, P. D., Livingston, S. P. and Mattick, J. S., PilS and PilR, a two-component transcriptional regulatory system controlling expression of type 4 fimbriae in *Pseudomonas aeruginosa*. *Mol. Microbiol.*, 1993, **5**, 669–682.
19. Ritchings, B. W., Almira, E. C., Lory, S. and Ramphal, R., Cloning and phenotypic characterization of *fleS* and *fleR*, new response regulators of *Pseudomonas aeruginosa* which regulate motility and adhesion to mucin. *Infect. Immun.*, 1995, **12**, 4868–4876.
20. Slauch, J. M., Garrett, S., Jackson, D. E. and Silhavy, T. J., EnvZ functions through OmpR to control porin gene expression in *Escherichia coli* K-12. *J. Bacteriol.*, 1988, **1**, 439–441.
21. Taylor, R. K., Hall, M. N., Enquist, L. and Silhavy, T. J., Identification of OmpR: a positive regulatory protein controlling expression of the major outer membrane matrix porin proteins of *Escherichia coli* K-12. *J. Bacteriol.*, 1981, **1**, 255–258.
22. Wanner, B. L. and Chang, B. D., The *phoBR* operon in *Escherichia coli* K-12. *J. Bacteriol.*, 1987, **12**, 5569–5574.
23. Yagil, E., Bracha, M. and Lifshitz, Y., The regulatory nature of the *phoB* gene for alkaline phosphatase synthesis in *Escherichia coli*. *Mol. Gen. Genet.*, 1975, **1**, 11–16.
24. Tommassen, J., de Geus, P., Lugtenberg, B., Hackett, J. and Reeves, P., Regulation of the pho regulon of *Escherichia coli* K-12. Cloning of the regulatory genes *phoB* and *phoR* and identification of their gene products. *J. Mol. Biol.*, 1982, **2**, 265–274.
25. Makino, K., Shinagawa, H. and Nakata, A., Regulation of the phosphate regulon of *Escherichia coli* K-12: regulation and role of the regulatory gene phoR. *J. Mol. Biol.*, 1985, **2**, 231–240.
26. Hoch, J. A., The phosphorelay signal transduction pathway in the initiation of *Bacillus subtilis* sporulation. *J. Cell. Biochem.*, 1993, **1**, 55–61.
27. Grossman, A. D. and Losick, R., Extracellular control of spore formation in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA,* 1988, **12**, 4369–4373.

28. Mathiopoulos, C. and Sonenshein, A. L., Identification of *Bacillus subtilis* genes expressed early during sporulation. *Mol. Microbiol.*, 1989, **8**, 1071–1081.
29. Trach, K. *et al.*, Control of the initiation of sporulation in *Bacillus subtilis* by a phosphorelay. *Res. Microbiol.*, 1991, **7–8**, 815–823.
30. Cole, S. T. *et al.*, Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, 1998, **6685**, 537–544.
31. Zahrt, T. C. and Deretic, V., *Mycobacterium tuberculosis* signal transduction system required for persistent infections. *Proc. Natl. Acad. Sci. USA,* 2001, **22**, 12706–12711.
32. Zahrt, T. C. and Deretic, V., An essential two-component signal transduction system in *Mycobacterium tuberculosis*. *J. Bacteriol.*, 2000, **13**, 3832–3838.
33. Groisman, E. A., Chiao, E., Lipps, C. J. and Heffron, F., *Salmonella typhimurium* phoP virulence gene is a transcriptional regulator. *Proc. Natl. Acad. Sci. USA,* 1989, **18**, 7077–7081.
34. Vescovi, E. G., Ayala, Y. M., Di Cera, E. and Groisman, E. A., Characterization of the bacterial sensor protein PhoQ. Evidence for distinct binding sites for $Mg^{2+}$ and $Ca^{2+}$. *J. Biol. Chem.*, 1997, **3**, 1440–1443.
35. Groisman, E. A., The pleiotropic two-component regulatory system PhoP–PhoQ. *J. Bacteriol.*, 2001, **6**, 1835–1842.
36. Perez, E., Samper, S., Bordas, Y., Guilhot, C., Gicquel, B. and Martin, C., An essential role for *phoP* in *Mycobacterium tuberculosis* virulence. *Mol. Microbiol.*, 2001, **1**, 179–187.
37. Ludwiczak, P., Brando, T., Monsarrat, B. and Puzo, G., Structural characterization of *Mycobacterium tuberculosis* lipoarabinomannans by the combination of capillary electrophoresis and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Anal. Chem.*, 2001, **10**, 2323–2330.
38. Ewann, F. *et al.*, Transient requirement of the PrrA–PrrB two-component system for early intracellular multiplication of *Mycobacterium tuberculosis*. *Infect. Immun.*, 2002, **5**, 2256–2263.
39. Graham, J. E. and Clark-Curtiss, J. E., Identification of *Mycobacterium tuberculosis* RNAs synthesized in response to phagocytosis by human macrophages by selective capture of transcribed sequences (SCOTS). *Proc. Natl. Acad. Sci. USA,* 1999, **20**, 11554–11559.
40. Supply, P., Magdalena, J., Himpens, S. and Locht, C., Identification of novel intergenic repetitive units in a mycobacterial two-component system operon. *Mol. Microbiol.*, 1997, **5**, 991–1003.
41. Himpens, S., Locht, C. and Supply, P., Molecular characterization of the mycobacterial SenX3–RegX3 two-component system: evidence for autoregulation. *Microbiology*, **146**, 2000, 3091–3098.
42. Haydel, S. E., Dunlap, N. E. and Benjamin, W. H. Jr., *In vitro* evidence of two-component system phosphorylation between the *Mycobacterium tuberculosis* TrcR/TrcS proteins. *Microb. Pathog.*, 1999, **4**, 195–206.
43. Haydel, S. E., Benjamin, W. H. Jr., Dunlap, N. E. and Clark-Curtiss, J. E., Expression, autoregulation, and DNA binding properties of the *Mycobacterium tuberculosis* TrcR response regulator. *J. Bacteriol.*, 2002, **8**, 2192–2203.
44. Wernisch, L. *et al.*, Analysis of whole-genome microarray replicates using mixed models. *Bioinformatics*, 2003, **1**, 53–61.
45. Parish, T., Smith, D. A., Kendall, S., Casali, N., Bancroft, G. J. and Stoker, N. G., Deletion of two-component regulatory systems increases the virulence of *Mycobacterium tuberculosis*. *Infect. Immun.*, 2003, **3**, 1134–1140.
46. Meyer, M., Dimroth, P. and Bott, M., *In vitro* binding of the response regulator CitB and of its carboxy-terminal domain to A + T-rich DNA target sequences in the control region of the divergent *citC* and *citS* operons of *Klebsiella pneumoniae*. *J. Mol. Biol.*, 1997, **5**, 719–731.
47. Aiba, H., Nakasai, F., Mizushima, S. and Mizuno, T., Phosphorylation of a bacterial activator protein, OmpR, by a protein kinase, EnvZ, results in stimulation of its DNA-binding ability. *J. Biochem. (Tokyo)*, 1989, **1**, 5–7.
48. Makino, K., Shinagawa, H., Amemura, M., Kawamoto, T., Yamada, M. and Nakata, A., Signal transduction in the phosphate regulon of *Escherichia coli* involves phosphotransfer between PhoR and PhoB proteins. *J. Mol. Biol.*, 1989, **3**, 551–559.
49. Forst, S., Delgado, J. and Inouye, M., Phosphorylation of OmpR by the osmosensor EnvZ modulates expression of the *ompF* and *ompC* genes in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA,* 1989, **16**, 6052–6056.
50. Boucher, P. E., Menozzi, F. D. and Locht, C., The modular architecture of bacterial response regulators. Insights into the activation mechanism of the BvgA transactivator of *Bordetella pertussis*. *J. Mol. Biol.*, 1994, **3**, 363–377.
51. Lynch, A. S. and Lin, E. C., Transcriptional control mediated by the ArcA two-component response regulator protein of *Escherichia coli*: characterization of DNA binding at target promoters. *J. Bacteriol.*, 1996, **21**, 6238–6249.
52. Dahl, J. L., Wei, B. Y. and Kadner, R. J., Protein phosphorylation affects binding of the *Escherichia coli* transcription activator UhpA to the *uhpT* promoter. *J. Biol. Chem.*, 1997, **3**, 1910–1919.
53. Park, H. D. *et al.*, Rv3133c/dosR is a transcription factor that mediates the hypoxic response of *Mycobacterium tuberculosis*. *Mol. Microbiol.*, 2003, **3**, 833–843.
54. Steyn, A. J., Joseph, J. and Bloom, B. R., Interaction of the sensor module of *Mycobacterium tuberculosis* H37Rv KdpD with members of the Lpr family. *Mol. Microbiol.*, 2003, **4**, 1075–1089.
55. Epstein, W., Buurman, E., McLaggan, D. and Naprstek, J., Multiple mechanisms, roles and controls of $K^+$ transport in *Escherichia coli*. *Biochem. Soc. Trans.*, 1993, **4**, 1006–1010.
56. Altendorf, K., Voelkner, P. and Puppe, W., The sensor kinase KdpD and the response regulator KdpE control expression of the *kdpFABC* operon in *Escherichia coli*. *Res. Microbiol.*, 1994, **5–6**, 374–381.
57. Walderhaug, M. O., Polarek, J. W., Voelkner, P., Daniel, J. M., Hesse, J. E., Altendorf, K. and Epstein, W., KdpD and KdpE, proteins that control expression of the *kdpABC* operon, are members of the two-component sensor-effector class of regulators. *J. Bacteriol.*, 1992, **7**, 2152–2159.
58. Mizuno, T., Compilation of all genes encoding two-component phosphotransfer signal transducers in the genome of *Escherichia coli*. *DNA Res.*, 1997, **2**, 161–168.
59. Throup, J. P. *et al.*, A genomic analysis of two-component signal transduction in *Streptococcus pneumoniae*. *Mol. Microbiol.*, 2000, **3**, 566–576.
60. Fabret, C., Feher, V. A. and Hoch, J. A., Two-component signal transduction in *Bacillus subtilis*: how one organism sees its world. *J. Bacteriol.*, 1999, **7**, 1975–1983.
61. Bentley, S. D. *et al.*, Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature*, 2002, **6885**, 141–147.
62. Matsumoto, A., Hong, S. K., Ishizuka, H., Horinouchi, S. and Beppu, T., Phosphorylation of the AfsR protein involved in secondary metabolism in *Streptomyces* species by a eukaryotic-type protein kinase. *Gene*, 1994, **1**, 47–56.
63. Hanlon, W. A., Inouye, M. and Inouye, S., Pkn9, a Ser/Thr protein kinase involved in the development of *Myxococcus xanthus*. *Mol. Microbiol.*, 1997, **3**, 459–471.
64. Rajagopal, L., Clancy, A. and Rubens, C. E., A eukaryotic type serine/threonine kinase and phosphatase in *Streptococcus agalactiae* reversibly phosphorylate an inorganic pyrophosphatase and affect growth, cell segregation, and virulence. *J. Biol. Chem.*, 2003, **16**, 14429–14441.
65. Chaba, R., Raje, M. and Chakraborti, P. K., Evidence that a eukaryotic-type serine/threonine protein kinase from *Mycobacterium tuberculosis* regulates morphological changes associated with cell division. *Eur. J. Biochem.*, 2002, **4**, 1078–1085.
66. Koul, A., Choidas, A., Tyagi, A. K., Drlica, K., Singh, Y. and Ullrich, A., Serine/threonine protein kinases PknF and PknG of

*Mycobacterium tuberculosis*: characterization and localization. *Microbiology*, 2001, **147**, 2307–2314.
67. Hakansson, S., Galyov, E. E., Rosqvist, R. and Wolf-Watz, H., The Yersinia YpkA Ser/Thr kinase is translocated and subsequently targeted to the inner surface of the HeLa cell plasma membrane. *Mol. Microbiol.*, 1996, **3**, 593–603.
68. Av-Gay, Y. and Everett, M., The eukaryotic-like Ser/Thr protein kinases of *Mycobacterium tuberculosis*. *Trends Microbiol.*, 2000, **5**, 238–244.
69. Belanger, A. E., Besra, G. S., Ford, M. E., Mikusova, K., Belisle, J. T., Brennan, P. J. and Inamine, J. M., The *embAB* genes of *Mycobacterium avium* encode an arabinosyl transferase involved in cell wall arabinan biosynthesis that is the target for the antimycobacterial drug ethambutol. *Proc. Natl. Acad. Sci. USA,* 1996, **21**, 11919–11924.
70. Gomez, J. E., Chen, J. M. and Bishai, W. R., Sigma factors of *Mycobacterium tuberculosis*. *Tuber. Lung. Dis.*, 1997, **3–4**, 175–183.
71. Paget, M. S. and Helmann, J. D., The sigma70 family of sigma factors. *Genome Biol.*, 2003, **1**, 203.
72. Gomez, M., Doukhan, L., Nair, G. and Smith, I., *sigA* is an essential gene in *Mycobacterium smegmatis*. *Mol. Microbiol.*, 1998, **2**, 617–628.
73. Hu, Y. and Coates, A. R., Transcription of two sigma 70 homologue genes, *sigA* and *sigB*, in stationary-phase *Mycobacterium tuberculosis*. *J. Bacteriol.*, 1999, **2**, 469–476.
74. Potuckova, L., Kelemen, G. H., Findlay, K. C., Lonetto, M. A., Buttner, M. J. and Kormanec, J., A new RNA polymerase sigma factor, sigma F, is required for the late stages of morphological differentiation in *Streptomyces* spp. *Mol. Microbiol.*, 1995, **1**, 37–48.
75. Haldenwang, W. G., The sigma factors of *Bacillus subtilis*. *Microbiol. Rev.*, 1995, **1**, 1–30.
76. Boylan, S. A., Redfield, A. R., Brody, M. S. and Price, C. W., Stress-induced activation of the sigma B transcription factor of *Bacillus subtilis*. *J. Bacteriol.*, 1993, **24**, 7931–7937.
77. Helmann, J. D., The extracytoplasmic function (ECF) sigma factors. *Adv. Microb. Physiol.*, 2002, **46**, 47–110.
78. Manganelli, R., Dubnau, E., Tyagi, S., Kramer, F. R. and Smith, I., Differential expression of 10 sigma factor genes in *Mycobacterium tuberculosis*. *Mol. Microbiol.*, 1999, **2**, 715–724.
79. Hu, Y. and Coates, A. R., Increased levels of *sigJ* mRNA in late stationary phase cultures of *Mycobacterium tuberculosis* detected by DNA array hybridisation. *FEMS Microbiol. Lett.*, 2001, **1**, 59–65.
80. Doukhan, L. *et al.*, Genomic organization of the mycobacterial sigma gene cluster. *Gene*, 1995, **1**, 67–70.
81. Predich, M., Doukhan, L., Nair, G. and Smith, I., Characterization of RNA polymerase and two sigma-factor genes from *Mycobacterium smegmatis*. *Mol. Microbiol.*, 1995, **2**, 355–366.
82. Gomez, J. E. and Bishai, W. R., *whmD* is an essential mycobacterial gene required for proper septation and cell division. *Proc. Natl. Acad. Sci. USA,* 2000, **15**, 8554–8559.
83. Collins, D. M., Kawakami, R. P., de Lisle, G. W., Pascopella, L., Bloom, B. R. and Jacobs, W. R., Jr., Mutation of the principal sigma factor causes loss of virulence in a strain of the *Mycobacterium tuberculosis* complex. *Proc. Natl. Acad. Sci. USA,* 1995, **17**, 8036–8040.
84. Siegele, D. A., Hu, J. C., Walter, W. A. and Gross, C. A., Altered promoter recognition by mutant forms of the sigma 70 subunit of *Escherichia coli* RNA polymerase. *J. Mol. Biol.*, 1989, **4**, 591–603.
85. Zuber, P., Healy, J., Carter, H. L., Cutting, S., Moran, C. P. Jr. and Losick, R., Mutation changing the specificity of an RNA polymerase sigma factor. *J. Mol. Biol.*, 1989, **4**, 605–614.
86. Steyn, A. J., Collins, D. M., Hondalus, M. K., Jacobs, W. R. Jr., Kawakami, R. P. and Bloom, B. R., *Mycobacterium tuberculosis* WhiB3 interacts with RpoV to affect host survival but is dispensable for *in vivo* growth. *Proc. Natl. Acad. Sci. USA,* 2002, **5**, 3147–3152.
87. DeMaio, J., Zhang, Y., Ko, C., Young, D. B. and Bishai, W. R., A stationary-phase stress-response sigma factor from *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. USA,* 1996, **7**, 2790–2794.
88. Chen, P., Ruiz, R. E., Li, Q., Silver, R. F. and Bishai, W. R., Construction and characterization of a *Mycobacterium tuberculosis* mutant lacking the alternate sigma factor gene, *sigF*. *Infect. Immun.*, 2000, **10**, 5575–5580.
89. Manabe, Y. C., Chen, J. M., Ko, C. G., Chen, P. and Bishai, W. R., 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.*, 1999, **24**, 7629–7633.
90. Manganelli, R., Voskuil, M. I., Schoolnik, G. K. and Smith, I., The *Mycobacterium tuberculosis* ECF sigma factor sigmaE: role in global gene expression and survival in macrophages. *Mol. Microbiol.*, 2001, **2**, 423–437.
91. Manganelli, R., Voskuil, M. I., Schoolnik, G. K., Dubnau, E., Gomez, M. and Smith, I., Role of the extracytoplasmic-function sigma factor sigma(H) in *Mycobacterium tuberculosis* global gene expression. *Mol. Microbiol.*, 2002, **2**, 365–374.
92. Raman, S., Song, T., Puyang, X., Bardarov, S., Jacobs, W. R. Jr. and Husson, R. N., The alternative sigma factor SigH regulates major components of oxidative and heat stress responses in *Mycobacterium tuberculosis*. *J. Bacteriol.*, 2001, **20**, 6119–6125.
93. Paget, M. S., Molle, V., Cohen, G., Aharonowitz, Y. and Buttner, M. J., Defining the disulphide stress response in *Streptomyces coelicolor* A3(2): identification of the sigmaR regulon. *Mol. Microbiol.*, 2001, **4**, 1007–1020.
94. Fernandes, N. D., Wu, Q. L., Kong, D., Puyang, X., Garg, S. and Husson, R. N., A mycobacterial extracytoplasmic sigma factor involved in survival following heat shock and oxidative stress. *J. Bacteriol.*, 1999, **14**, 4266–4274.
95. Beaucher, J., Rodrigue, S., Jacques, P. E., Smith, I., Brzezinski, R. and Gaudreau, L., Novel *Mycobacterium tuberculosis* anti-sigma factor antagonists control sigmaF activity by distinct mechanisms. *Mol. Microbiol.*, 2002, **6**, 1527–1540.
96. Kaushal, D. *et al.*, Reduced immunopathology and mortality despite tissue persistence in a *Mycobacterium tuberculosis* mutant lacking alternative sigma factor, SigH. *Proc. Natl. Acad. Sci. USA,* 2002, **12**, 8330–8335.
97. Soliveri, J. A., Gomez, J., Bishai, W. R. and Chater, K. F., Multiple paralogous genes related to the *Streptomyces coelicolor* developmental regulatory gene *whiB* are present in *Streptomyces* and other actinomycetes. *Microbiology*, 2000, **146**, 333–343.
98. Hutter, B. and Dick, T., Molecular genetic characterisation of *whiB3*, a mycobacterial homologue of a *Streptomyces* sporulation factor. *Res. Microbiol.*, 1999, **5**, 295–301.
99. Rodriguez, G. M., Voskuil, M. I., Gold, B., Schoolnik, G. K. and Smith, I., *ideR*, An essential gene in *Mycobacterium tuberculosis*: role of IdeR in iron-dependent gene expression, iron metabolism, and oxidative stress response. *Infect. Immun.*, 2002, **7**, 3371–3381.
100. Gold, B., Rodriguez, G. M., Marras, S. A., Pentecost, M. and Smith, I., The *Mycobacterium tuberculosis* IdeR is a dual functional regulator that controls transcription of genes involved in iron acquisition, iron storage and survival in macrophages. *Mol. Microbiol.*, 2001, **3**, 851–865.
101. Rodriguez, G. M., Gold, B., Gomez, M., Dussurget, O. and Smith, I., Identification and characterization of two divergently transcribed iron regulated genes in *Mycobacterium tuberculosis*. *Tuber. Lung. Dis.*, 1999, **5**, 287–298.
102. Rodriguez, G. M. and Smith, I., Mechanisms of iron regulation in mycobacteria: role in physiology and virulence. *Mol. Microbiol.*, 2003, **6**, 1485–1494.
103. d'Ari, R., The SOS system. *Biochimie*, 1985, **3–4**, 343–347.

104. Davis, E. O., Dullaghan, E. M. and Rand, L., Definition of the mycobacterial SOS box and use to identify LexA-regulated genes in *Mycobacterium tuberculosis*. *J. Bacteriol.*, 2002, **12**, 3287–3295.
105. Stewart, G. R. *et al.*, Dissection of the heat-shock response in *Mycobacterium tuberculosis* using mutants and microarrays. *Microbiology*, 2002, **148**, 3129–3138.
106. Bucca, G., Brassington, A. M., Schonfeld, H. J. and Smith, C. P., The HspR regulon of *Streptomyces coelicolor*: a role for the DnaK chaperone as a transcriptional co-repressordagger. *Mol. Microbiol.*, 2000, **5**, 1093–1103.
107. Stewart, G. R. *et al.*, Overexpression of heat-shock proteins reduces survival of *Mycobacterium tuberculosis* in the chronic phase of infection. *Nature Med.*, 2001, **6**, 732–737.
108. Hecker, M., Schumann, W. and Volker, U., Heat-shock and general stress response in *Bacillus subtilis*. *Mol. Microbiol.*, 1996, **3**, 417–428.
109. Cunningham, A. F. and Spreadbury, C. L., Mycobacterial stationary phase induced by low oxygen tension: cell wall thickening and localization of the 16-kilodalton alpha-crystallin homolog. *J. Bacteriol.*, 1998, **4**, 801–808.
110. Wayne, L. G. and Hayes, L. G., An in vitro model for sequential study of shiftdown of *Mycobacterium tuberculosis* through two stages of nonreplicating persistence. *Infect. Immun.*, 1996, **6**, 2062–2069.
111. Sherman, D. R., Voskuil, M., Schnappinger, D., Liao, R., Harrell, M. I. and Schoolnik, G. K., Regulation of the *Mycobacterium tuberculosis* hypoxic response gene encoding alpha-crystallin. *Proc. Natl. Acad. Sci. USA,* 2001, **13**, 7534–7539.
112. Betts, J. C., Lukey, P. T., Robb, L. C., McAdam, R. A. and Duncan, K., Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling. *Mol. Microbiol.*, 2002, **3**, 717–731.
113. Gentry, D. R. and Cashel, M., Mutational analysis of the *Escherichia coli spoT* gene identifies distinct but overlapping regions involved in ppGpp synthesis and degradation. *Mol. Microbiol.*, 1996, **6**, 1373–1384.
114. Chatterji, D. and Ojha, A. K., Revisiting the stringent response, ppGpp and starvation signaling. *Curr. Opin. Microbiol.*, 2001, **2**, 160–165.
115. Ojha, A. K., Mukherjee, T. K. and Chatterji, D., High intracellular level of guanosine tetraphosphate in *Mycobacterium smegmatis* changes the morphology of the bacterium. *Infect. Immun.*, 2000, **7**, 4084–4091.
116. Primm, T. P., Andersen, S. J., Mizrahi, V., Avarbock, D., Rubin, H. and Barry, C. E., The stringent response of *Mycobacterium tuberculosis* is required for long-term survival. *J. Bacteriol.*, 2000, **17**, 4889–4898.
117. Dahl, J. L. *et al.*, The role of RelMtb-mediated adaptation to stationary phase in long-term persistence of *Mycobacterium tuberculosis* in mice. *Proc. Natl. Acad. Sci. USA,* 2003, (in press).

ACKNOWLEDGEMENTS. We thank all the members of the Smith laboratory for fruitful discussions. Work of the authors' mentioned in this review is supported by NIH grants AI-44856 and HL-68513 (awarded to I.S.). The literature survey for this article was completed in August 2003.

# Regulation of DNA topology in mycobacteria

**V. Nagaraja**

Department of Microbiology and Cell Biology, Indian Institute of Science, Bangalore 560 012 and Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore 560 064, India

**DNA topoisomerases catalyse essential DNA transaction processes in order to attain the balanced topology of the genome. Contrasting activities of DNA topoisomerase I and DNA gyrase result in the maintenance of topological homeostasis. The regulation of expression of different topoisomerases ensure steady state optimum level expression of the enzymes. Many aspects of their organization and regulation seem to be different in mycobacteria when compared with that of *Escherichia coli*. Here we present several aspects of the regulation of mycobacterial topoisomerases and discuss the significance.**

## Introduction

DNA topoisomerases have evolved to catalyse the topological alterations in DNA in order to ensure that DNA-transaction processes are completed without topological interruptions. Hence the enzymes play essential roles during replication, transcription, recombination, repair and chromosome segregation[1]. The genomes of bacteria are normally maintained in negative supercoiled state. Since all processes that involve DNA as a substrate either need to melt, bend or distort DNA; negative supercoiling modulates these cellular processes. For example, supercoiling influences recombination at two levels independently. First, supercoiling enhances recombination because the plectonemic winding of DNA facilitates the juxtapositioning of recombination sites and limits the extent of diffusion required for the sites to collide[2]. Second, the extent of supercoiling of the substrate determines the complexity of the product(s) since recombination converts the supercoil nodes into nodes of catenation or knots, depending on the relative orientation of the sites[3]. In a complementary manner, various DNA transactions alter the topology of DNA. The most obvious of these being the generation of catenated daughter duplexes after replication and activities of DNA tracking machineries. In eubacteria, the principal enzymes that influence the vital processes are topoisomerase I and DNA gyrase with substantial contribution from topoisomerase IV when present (see later section). Hence, the regulation of their expression and activities is an important determinant in the maintenance of balanced topological state and the global supercoiling of DNA is thus dependent on the balance of activities of various topoisomerases.

*For correspondence. (e-mail: vraj@mcbl.iisc.ernet.in)

Amongst the four topoisomerases found in *E. coli*, topoisomerase I and III belong to type IA group while DNA gyrase and topoisomerase IV are type II enzymes[4]. The key enzyme in all bacteria catalysing the formation of negatively supercoiled DNA in an ATP dependent reaction is DNA gyrase. The enzyme, encoded by *gyr*A and *gyr*B, is a heterotetrameric protein[5]. By virtue of its indispensability, the enzyme has been and continues to be a favourite drug target. As a consequence, several inhibitors and poisons, both natural and synthetic, have been characterized[6]. The second major player in influencing global topology is DNA topoisomerase I. In contrast to DNA gyrase, it comprises of a single polypeptide, encoded by *top*A gene. The enzyme catalyses the conversion of negatively supercoiled DNA into relaxed form in an ATP independent reaction[4].

Unlike *E. coli*, where four topoisomerases have been characterized, mycobacteria and many other organisms do not encode the full complement of topoisomerases. This was evident during our efforts to clone the genes for DNA gyrase from both *M. smegmatis* and *M. tuberculosis*[7,8]. Efforts to clone genes for other topoisomerases such as topoisomerase IV were unsuccessful hinting at the possibility of absence of these genes in *M. tuberculosis*. Genome sequencing efforts subsequently confirmed the presence of only single topoisomerase I and DNA gyrase in *M. tuberculosis*, while some other species such as *M. smegmatis*, *M. bovis* appear to have genes encoding for additional topoisomerases[9–11]. Amongst the two type IA enzymes found in *E. coli* and other bacteria, only topoisomerase I is present in mycobacteria[12].

Figure 1 depicts the organization of genes encoding DNA gyrase in *E. coli*, *M. smegmatis* and *M. tuberculosis*. Notably, *gyr*B and *gyr*A in *E. coli* are located far apart in the circular chromosome but present next to each other in both the species of mycobacteria. Furthermore, significant additional differences are observed in their primary sequences[13]. The genetic linkage between the *gyr* genes seems to correlate with the size of the *gyr*B gene. Species in which the genes are present far apart have 165 amino acids extra in the C-terminal half of *Gyr*B[14] and this insertion appears to be involved in DNA binding[15]. Our work over the last decade has revealed some of the distinctive characteristics of mycobacterial topoisomerase organization, function and regulation. In the following sections, the salient features of topoisomerase regulation in mycobacteria are presented and compared with that of *E. coli*.

## Topology and transcription

Topological organization of DNA is known to have important influence in regulation of gene expression. Movement of RNA polymerase along the helical axis results in an increase in twist ahead of the tracking machinery and decrease behind, introducing positive and negative supercoils respectively. This is termed as twin domain of supercoiling[16]. As a result, DNA gyrase and topoisomerase I have to function ahead and behind the transcription bubble to remove positive and negative supercoils respectively (Figure 2). The twin domain model of supercoiling has several biological implications. (i) DNA transactions may prove to be a major determinant of local DNA topology; (ii) transcription of adjacent genes could significantly influence expression of a particular gene; (iii) Most importantly, for the first time, there appeared to be a necessity to expect efficiency in topoisomerases. Supercoiling influences transcription of many genes in the cell[17–19], modulating by several ways. Directly, it can realign promoter elements or facilitate open complex formation. Indirectly, it can stabilize loops, bends or other non-B-DNA structures in DNA.

In majority of the promoters, negative supercoiling facilitates isomerization of closed complex to open complex. However, failure to remove negative supercoils generated behind the transcription elongation complex would lead to the accumulation of R-loops and as a consequence, inhibition of transcription[20]. In addition, study of regulation of topoisomerase expression in mycobacteria is important especially since the genome lacks full complement of topoisomerases. Furthermore, in many pathogenic bacteria, expression of virulence genes is dependent on topological status of the genome[21,22]. As sensor of supercoils, the topoisomerases influence the specific gene expression.

## Transcription of topoisomerases and regulation of topology

The net supercoiling of intracellular DNA is maintained by the relaxation activities of DNA topoisomerases I and IV opposing the supercoiling activity of DNA gyrase. Thus, by modulating the expression of any one of these genes, the cell can bring about rapid changes in supercoiling as well as compensate for sudden changes in supercoiling. As the sole supercoiling activity in the cell, DNA gyrase faces the daunting task of opposing the relaxation activities of both topoisomerases I and IV[23] and regulates its expression by a unique mechanism termed relaxation-stimulated transcription (RST). In general, transcription of most genes is induced by increased negative supercoiling. In contrast, negative supercoiling represses transcription of the gyrase genes in *E. coli*[24]. Increased gyrase levels lead to an increase in supercoiling, which, in turn, represses the expression of gyrase and allows other topoisomerases to bring the topology of the DNA back to its optimum state. Following observations led to the discovery of RST in *E. coli*. Cell-free transcription showed that transcription was dependent directly on the DNA topology, being maximal on a relaxed template[24]. Deletion analysis of the promoter regions of both *gyr*A and *gyr*B genes defined a short region around the transcriptional start site, including the –10 region, that is necessary and sufficient for conferring RST to a reporter gene[25,26]. Extensive mutagenesis of the

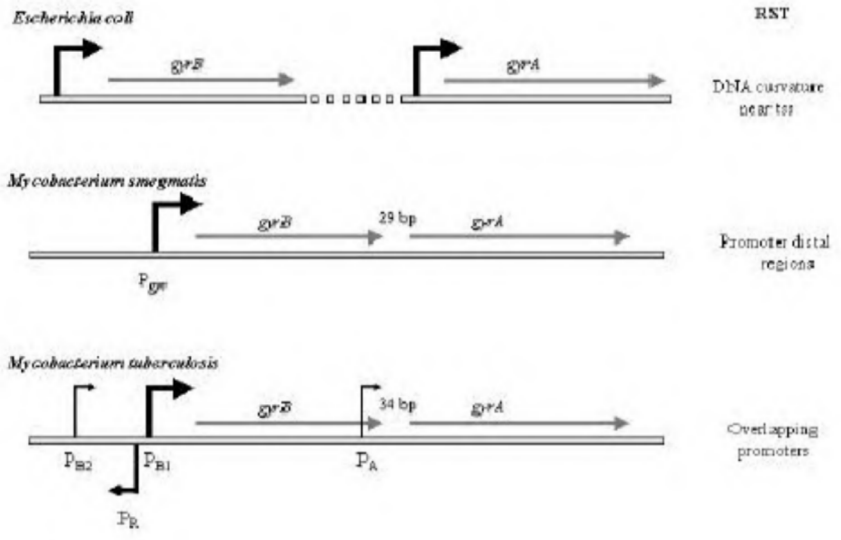


**Figure 1.** Transcriptional organization and regulation of *gyr* genes in *E. coli* and two species of mycobacteria. In *E. coli*, the genes are transcribed independently. A strong promoter drives the transcription of *gyr*BA operon in both *M. smegmatis* and *M. tuberculosis*. The vertical arrows and their thickness depict the promoters and their strength. tss; transcription start site.

*gyr*A promoter showed that the –10 region is responsible for both promoter strength and supercoil-sensitive behaviour[27]. Paradoxically, the promoter region harbours a sequence that matches the *E. coli* consensus for extended –10 promoters[28]. Since most extended –10 promoters do not show RST, it appears unlikely that the sequence of the –10 region alone is responsible for RST. Our recent analysis suggests that DNA curvature around the transcription start point plays a role in RST in *E. coli*[29].

The *gyr* operon in *M. smegmatis* is induced by novobiocin at the transcriptional level. However, unlike *E. coli*, minimal promoter of *gyr* operon do not confer RST in this case[30]. The presence of a strong CHPS (cruciform/hairpin potential sequence) with an 8 base pair stem and a 4 base loop in the 5′ untranslated region suggested a potential mechanism for RST in *M. smegmatis* and found to have a positive effect on promoter activity but is not sufficient for the operon to respond to novobiocin. The induction of the genomic copy and the results with the constructs harbouring upstream and downstream DNA sequences show the essential role played by promoter distal elements. DNA elements that are present 600 bp downstream of the promoter are necessary for RST to occur in the plasmid context[30]. This suggests the involvement of long range interactions and formation of repressor loops which could either prevent the binding of the polymerase to the promoter or prevent its release. Since such repressor loops are stabilized by negative supercoiling[31–33], the repression would occur in a supercoil sensitive manner. However, the downstream element along with the minimal promoter region is not sufficient to respond to novobiocin. Sequential deletions of both upstream and downstream distal regions reveal interesting novel features of the regulation. Therefore, RST appears to operate in *M. smegmatis* by some sort of de-repression rather than direct induction and a distal element has a negative influence on transcription[30].

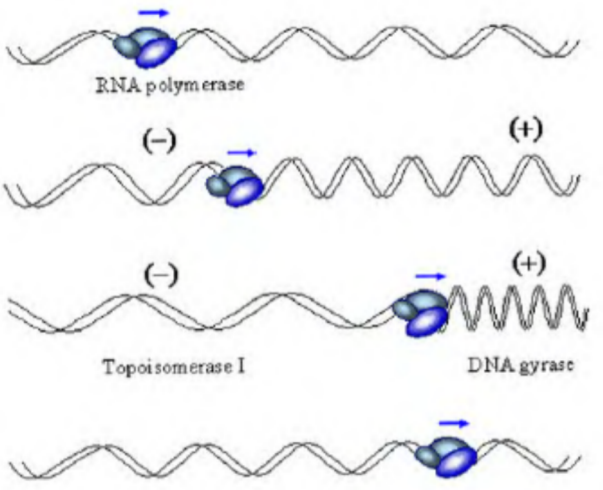


**Figure 2.** Twin domain of supercoiling. During transcription, RNA polymerase movement results in positive (+) in front and negative (–) supercoils behind the tracking machine. Action of the topoisomerases relieve the excessive supercoiling ensuring the further movement of RNA polymerase.

A comparison of the expression of DNA gyrase in *M. smegmatis* and *M. tuberculosis* reveals an amalgamation of conserved and divergent features (Figure 1). In addition to conserved genomic arrangement and dicistronic organization, the primary promoter in *M. tuberculosis*, $P_{B1}$, is located upstream of the *gyrB* gene at a position similar to that of the *M. smegmatis gyr* promoter[34]. Furthermore, the promoter region of $P_{B1}$ shows extensive conservation with $P_{gyr}$, the promoter driving the *gyr* genes of *M. smegmatis*, indicating the evolutionarily relationship. Apart from the primary promoter, the *gyr* locus in *M. tuberculosis* employs at least three other promoters[34]. These additional promoters are weak and appear to play a regulatory role. $P_A$, the internal promoter for *gyr*A is 70-fold weaker than $P_{B1}$ in exponentially growing *M. tuberculosis,* possibly employs an *M. tuberculosis*-specific sigma factor. Moreover, $P_A$ may be induced under specific conditions which require the production of excess GyrA. Induction of GyrA alone in *E. coli* in response to treatment with GyrA inhibitors has been demonstrated earlier[35]. The other weak promoter, $P_R$, is divergently oriented and almost completely overlaps $P_{B1}$. Therefore, the binding of RNA polymerase to one of them would prevent binding in the opposite orientation. There are no identifiable coding sequences upstream of *gyrB* that $P_R$ could be involved in transcribing, suggesting the function of $P_R$ to be regulatory. Overlapping, mutually exclusive promoters are one of the mechanisms for regulating gene expression[36]. Recruitment of the polymerase to $P_R$ would decrease expression of DNA gyrase by reducing transcription initiation at $P_{B1}$. In the converse scenario, as in relaxation of the template, $P_R$ is repressed and $P_{B1}$ gets induced to almost the same extent.

These studies highlight the importance of regulation of constituitively expressed, housekeeping, essential functions. While RST is a convenient mechanism to attain steady state levels of the enzyme, the complete molecular details of its operation vary and are not yet understood. Analysis of the promoter region of *M. smegmatis* and *M. tuberculosis* reveals a distinct lack of any axial distortion upstream of the +1 start site unlike *E. coli*[29]. To further substantiate the demarcation, all known *gyr*ase promoters were analysed for the presence of curvature in the vicintiy of the –10 region (+5 nucleotides). It is noteworthy that roughly half of these show a significant curvature in this region while others do not[29]. Interestingly, both the position and the extent of curvature are conserved between the *E. coli* and *Klebsiella pnemoniae gyr*A promoters[29]. How a promoter distal (downstream) element located within the ORF contributes for RST in *M. smegmatis* is not clear. Although it appears that trans factor/s could be involved in promoting the long range interactions, the molecular mechanism is still elusive. *M. tuberculosis*

genome unlike that of *M. smegmatis*, has its own variant mechanism of RST. The RST response in case of *M. tuberculosis* is extremely slow. The gyrases from both the species are very similar (about 90% identity) and hence the difference in the response is not likely due to catalytic properties. On the other hand, rate of transcription is varied between the species. *M. tuberculosis* RNA polymerase is at least 3–5 times slower than *M. smegmatis* enzyme[37]. Lower transcriptional rates in conjunction with yet unknown features may contribute for the difference in the response.

In contrast to *gyrB* and *gyrA* genes which are transcribed by single promoters, topoisomerase I expression appears to be regulated by multiple promoters in *E. coli*[35,38]. Two of them seem to be dependent on sigma-70, the major sigma factor of *E. coli* RNA polymerase. Amongst the others, one promoter is active during heat shock response while the other directs transcription at stationary phase of the growth[38–40]. As a consequence, the overall topoisomerase I expression is the net result of combined activation of various promoters[38]. Although the underlying mechanisms of regulation of *top*A and *gyr* genes by supercoiling are not completely understood, these represent an efficient homeostatic mechanism for the maintenance of supercoiling within physiological limits. For example, when global supercoiling goes down, *top*A is repressed while *gyr* genes are induced, compensating for the deficit in supercoiling. Instead of transfactor/s, DNA topology, the substrate/product for topoisomerases, directly modulates the levels of the enzymes in a manner reminiscent of product mediated inhibition in metabolic pathways.

## Post-transcriptional regulation: mRNA stability

The half-life of the bulk of the mRNA in *E. coli* is 2.4 min at 37°C[41]. This short half-life could reflect the fastgrowing nature of *E. coli*, possibly facilitating rapid adaptation to environmental changes[42]. Thus, one would expect mycobacteria and other slow-growing organisms would have more stable messages. In addition, the regulation of degradation of these messages would be different. Based on this hypothesis, stability of the DNA *gyr*ase mRNA in *M. smegmatis* was analysed[43]. A secondary structure near the 5′ end of mRNA that protects the message against degradation was identified (Figure 3). The stabilization effect is significantly pronounced in nutrient-deprived conditions. In addition to the transcriptional regulation discussed earlier, the nutrient-dependent stabilization of the gyrase message, represents a second, hitherto unexplored, level of regulation of the *gyr* genes in any organism. While, in general, stabilization of a housekeeping message would be important for slow-growing organisms like mycobacteria, in the specific case of DNA gyrase, it probably has additional significance due to the operonic arrangement and for all known biological functions both proteins are required in equimolar amounts. Since genes present downstream in an operon are usually underrepresented at the protein level, it would be useful for the organism to evolve methods to prevent this discrepancy. The mycobacterial *gyr* operon attempts to circumvent this problem by subtle changes in its primary sequence (see later section). In such a context, the presence of a stabilizing secondary structure is probably an additional mechanism to ensure that the downstream message is maintained long enough to be translated efficiently. Thus, *M. smegmatis* appears to rely on two distinct sensors: a promoter-proximal sensor for nutrient levels and a promoter-distal sensor for DNA topology. Although not experimentally verified for its function, similar secondary structure is found upstream of *gyr*BA operon in *M. tuberculosis*.

Enhanced stability of mRNA upon starvation has been reported in many organisms[42,44,45] which would allow the cells to utilize the already synthesised messages to their fullest and conserve resources when they are scarce. This is arguably more important for organisms like mycobacteria that grow slowly even under nutrient-rich conditions. In these organisms, the lower rate of transcription elongation is probably compensated for by enhanced stability of the message. Secondary structures at or near the 5′ end of the mRNA in *E. coli* are believed to function by preventing access of RNase E[46]. Furthermore, cleavage by RNase E appears to be the primary rate

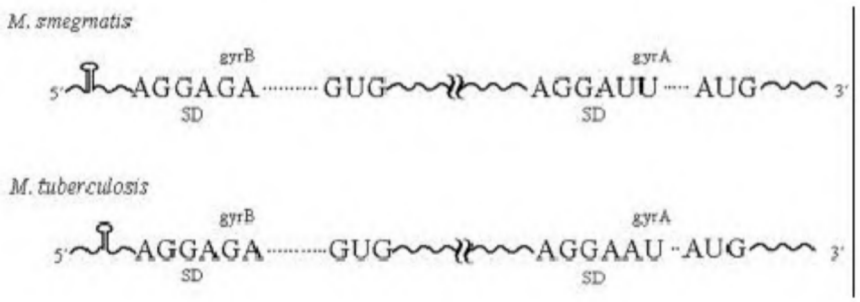


**Figure 3.** Translational regulation of *gyr* mRNA. The stabilizing stem loop structure is shown at 5′ end of the *gyr*BA mRNA. The SD and start codon sequences of the individual cistrons are shown which suggest translational optimization. The dots between SD sequence and start codon represent the actual number of nucleotides.

determining step in the degradation of most messages in *E. coli*[47]. Genomes of *M. tuberculosis*[10] and *Mycobacterium leprae*[48] encode for a homologue of this enzyme. Genome-wide analysis of the distribution of secondary structures indicates that genes in slow-growing organisms like *M. tuberculosis* are more likely to have a strong secondary structure ahead of them than those in fast-growing organisms like *E. coli*[49]. This probably protects a majority of the messages against the degradative activities. Such a strategy would make economic sense for a slow growing organism that does not necessarily need to respond to environmental changes rapidly[29,34,50].

## Post-translational regulation

The *gyr*BA dicistron in both *M. smegmatis* and *M. tuberculosis* exhibits additional interesting regulatory features (Figure 3). The *gyr*B has a near perfect Shine-Dalgarno (SD) sequence upstream of a weak start codon while *gyr*A has relatively weak SD sequence and an efficient start codon. Such an arrangement suggests a translational regulation that could facilitate the production of equimolar amounts of the two subunits that constitute the heterotetrameric functional holoenzyme.

Notably, there is a remarkable difference in the activities of DNA gyrase from *E. coli* and mycobacteria. *M. smegmatis* DNA gyrase has 3–5 fold weaker ATP hydrolyzing activity compared to that of *E. coli*[51]. However, the enzyme is a potent decatenase suggesting a more important role during segregation of daugther chromosomes. Since *M. tuberculosis* genome encodes only for topoisomerase I and DNA gyrase[10], the DNA gyrase is likely to possess strong decatenase activity to take care of added responsibility of daughter genome segregation.

Additional post translational measures seem to play a role in the expression of functional gyrase in mycobacteria. The mycobacterial recombinant DNA gyrase expressed in *E. coli* do not show high specific activity when analysed for supercoiling or ATPase activities in contrast to the enzymes isolated from wild type cells or overexpressed in mycobacteria (unpublished). This suggests a role for post-translational modification of the enzyme. Furthermore, the presence of two *gyr*B in *M. smegmatis* raises interesting possibilities regarding the intracellular functions[11]. The GyrB encoded by *gyr*BA operon is known to be associated with GyrA in a tetrameric holoenzyme that carries out the DNA supercoiling reaction[51]. From our comparative analysis it appears that the additional *gyr*B, termed as orphan *gyrB* is a functional allele and hence retained in the genome at a different location. Considering the difference in the growth rates of *M. smegmatis* and *M. tuberculosis*, the orphan GyrB could be contributing to the higher levels of enzymatic activity required during exponential growth phase. Alternatively, it is expressed differentially under certain conditions as an immediate requirement for cellular function. Another point to be noted is that GyrB is intrinsically less stable than GyrA in *E. coli*[52] and also in *M. smegmatis* (unpublished results).

## Conclusions

Diverse topoisomerases influence the topological state of the genome. Although topoisomerases are essential housekeeping functions, the fine tuning of their expression is important in order to maintain the balanced topological state. Our analysis of regulation of gyrase expression between two species of mycobacteria has revealed an amalgamation of several concepts with important species specific differences. The conserved features include dicistronic organization, mycobacteria specific promoters, RNA stability etc. However, autoregulation of transcription appears to have species specific variation. In *M. smegmatis* promoter distal downstream elements and possibly transfactors have a role in RST while in *M. tuberculosis*, overlapping mutually exclusive divergently organized promoters regulate the process. The organization as an operon in order to assemble heterotetrameric enzyme rapidly and extraordinary stability of the dicistronic mRNA are some of the measures taken by these group of bacteria to compensate for slower growth rates. The studies on regulation of topoisomerase I expression now underway would reveal other facets of regulation which contribute to attain cellular homeostasis in DNA topology.


1. Gellert, M., DNA topoisomerases. *Annu. Rev. Biochem.*, 1981, **50**, 879–910.
2. Droge, P. Protein tracking-induced supercoiling of DNA: a tool to regulate DNA transcription *in vivo*? *Bioassays*, 1994, **16**, 91–99.
3. Bliska, J. B. and Cozzarelli, N. R., Use of site-specific recombination as a probe of DNA structure and metabolism *in vivo*. *J. Mol. Biol.*, 1987, **194**, 205–218.
4. Champoux, J. J., DNA topoisomerases: structure, function and mechanism. *Annu. Rev. Biochem.*, 2001, **70**, 369–413.
5. Reece, R. J. and Maxwell, A., DNA gyrase: structure and function. *Crit. Rev. Biochem. Mol. Biol.*, 1991, **26**, 335–375.
6. Maxwell, A., DNA gyrase as a drug target. *Trends Microbiol.*, 1997, **5**, 102–109.
7. Madhusudan, K., Ramesh, V. and Nagaraja, V., Molecular cloning of *gyr*A and *gyr*B genes of *Mycobacterium tuberculosis*: analysis of nucleotide sequence. *Biochem. Mol. Biol. Int.*, 1994, **33**, 651–660.
8. Madhusudan, K. and Nagaraja, V. *Mycobacterium smegmatis* DNA gyrase: cloning and overexpression in *Escherichia coli*. *Microbiology*, 1995, **141**, 3029–3037.
9. Brosch, R., Gordon, S. V., Buchrieser, C., Pym, A. S., Garnier, T. and Cole, S. T., Comparative genomics uncovers large tandem chromosomal duplications in *Mycobacterium bovis* BC Pasteur. *Yeast*, 2000, **17**, 111–123.
10. Cole, S. T. *et al.*, Deciphering the biology of *Mycboacterium tuberculosis* from the complete genome sequence. *Nature*, 1998, **393**, 537–544.
11. Jain, P. and Nagaraja, V., An orphan *gyr*B in the *Mycobacterium smegmatis* genome uncovered by comparative genomics. *J. Genet.*, 2002, **81**, 105–110.
12. Nagaraja, V., Sikder, D. and Jain, P., DNA topoisomerase I from mycobacteria – a potential drug target. *Curr. Pharm. Des.*, 2002, **8**, 1995–2007.

13. Madhusudan, K. and Nagaraja, V., Alignment and phylogenetic analysis of type II DNA topoisomerases., *J. Biosci.*, 1996, **21**, 613–629.
14. Huang, W. M., Type II DNA topoisomerase genes. *Adv. Pharmacol.*, 1994, **29**, 201–225.
15. Chatterji, M., Unniraman, S., Maxwell, A. and Nagaraja, V., The additional 165 amino acids in the B protein of *Escherichia coli* DNA gyrase have an important role in DNA binding. *J. Biol. Chem.*, 2000, **275**, 22888–22894.
16. Liu, L. F. and Wang, J. C., Supercoiling of the DNA template during transcription. *Proc. Natl. Acad. Sci. USA*, 1987, **84**, 7024–7027.
17. Kreuzer, K. N. and Cozzarelli, N. R., *Escherichia coli* mutants thermosensitive for deoxyribonucleic acid gyrase subunit A: effects on deoxyribonucleic acid replication, transcription, and bacteriophage growth. *J. Bacteriol.*, 1979, **140**, 424–435.
18. Sanzey, B., Modulation of gene expression by drugs affecting deoxyribonucleic acid gyrase. *J. Bacteriol.*, 1979, **138**, 40–47.
19. Yang, H. L., Heller, K., Gellert, M. and Zubay, G., Differential sensitivity of gene expression by drugs affecting deoxyribonucleic acid gyrase. *Proc. Natl. Acad. USA*, 1979, **76**, 3304–3308.
20. Phoenix, P., Raymond, M. A., Masse, E. and Drolet, M., Role of DNA topoisomerases in the regulation of R-loop formation *in vitro*. *J. Biol. Chem.*, 1997, **272**, 1473–1479.
21. Dorman, C. J., Fleming Lecture: DNA topology and the global control of bacterial gene expression: implications for the regulation of virulence gene expression. *Microbiology*, 1995, **141**, 1271–1280.
22. Wang, J. C. and Lynch, A. S., Transcription and DNA supercoiling. *Curr. Opin. Genet. Dev.*, 1993, **3**, 764–768.
23. Zechiedrich, E. L., Khodursky, A. B., Bachellier, S., Schneider, R., Chen, D., Lilley, D. M. and Cozzarelli, N. R., Roles of topoisomerases in maintaining steady-state DNA supercoiling in *Escherichia coli*. *J. Biol. Chem.*, 2000, **275**, 8103–8113.
24. Menzel, R. and Gellert, M., Regulation of the genes for *E. coli* DNA gyrase: homeostatic control of DNA supercoiling. *Cell*, 1983, **34**, 105–113.
25. Menzel, R. and Gellert, M., Fusions of the *Escherichia coli gyr*A and *gyr*B control regions to the galactokinase gene are inducible by coumermycin treatment. *J. Bacteriol.*, 1987, **169**, 1272–1278.
26. Menzel, R. and Gellert, M., Modulation of transcription by DNA supercoiling: a deletion analysis of the *Escherichia coli gyr*A and *gyr*B promoters. *Proc. Natl. Acad. Sci. USA*, 1987, **84**, 4185–4189.
27. Straney, R., Krah, R. and Menzel, R., Mutations in the – 10 TATAAT sequence of the *gyr*A promoter affect both promoter strength and sensitivity to DNA supercoiling. *J. Bacteriol.*, 1994, **176**, 5999–6006.
28. Keilty, S. and Rosenberg, M., Constitutive function of a positively regulated promoter reveals new sequences essential for activity. *J. Biol. Chem.*, 1987, **262**, 6389–6395.
29. Unniraman, S. and Nagaraja, V. Axial distortion as a sensor of supercoil changes: a molecular model for the homeostatic regulation of DNA gyrase. *J. Genet.*, 2001, **80**, 119–124.
30. Unniraman, S. and Nagaraja, V., Regulation of DNA gyrase operon in *Mycobacterium smegmatis*: a distinct mechanism of relaxation stimulated transcription. *Genes Cells*, 1999, **4**, 697–706.
31. Borowiec, J. A., Zhang, L., Sasse-Dwight, S. and Gralla, J. D., DNA supercoiling promotes formation of a bent repression loop in lac DNA. *J. Mol. Biol.*, 1987, **196**, 101–111.
32. Choy, H. E., Park, S. W., Parrack, P. and Adhya, S., Transcription regulation by inflexibility of promoter DNA in a looped complex. *Proc. Natl. Acad. Sci. USA*, 1995, **92**, 7327–7331.
33. Lewis, D. E., Geanacopoulos, M. and Adhya, S., Role of HU and DNA supercoiling in transcription repression, specialized nucleoprotein repression complex at gal promoters in *Escherichia coli*. *Mol. Microbiol.*, 1999, **31**, 451–461.
34. Unniraman, S., Chatterji, M. and Nagaraja, V., DNA gyrase genes in *Mycobacterium tuberculosis*: a single operon driven by multiple promoters. *J. Bacteriol.*, 2002, **184**, 5499–5456.
35. Neumann, S. and Quinones, A., Discoordinate gene expression of *gyr*A and *gyr*B in response to DNA gyrase inhibition in *Escherichia coli*. *J. Basic Microbiol.*, 1997, **37**, 53–59.
36. Nagaraja, V., Control of transcription initiation. *J. Biosci.*, 1993, **18**, 13–25.
37. Harshey, R. M. and Ramakrishnan, T., Rate of RNA chain growth in *Mycobacterium tuberculosis* $H_{37}R_v$. *J. Bacteriol.*, 1977, **129**, 616–622.
38. Rui, S. and Tse-Dinh, Y.-C., Topoisomerase function during bacterial responses to environmental challenge. *Front. Biosci.*, 2003, **8**, 256–263.
39. Leslie, S. A., Jovanovich, S. B., Tse-Dinh, Y.-C. and Burgess, R. R., Identification of a heat shock promoter in the *top A* gene of *Escherichia coli*. *J. Bacteriol.*, 1990, **172**, 6871–6874.
40. Qi, H., Menzel, R. and Tse-Dinh, Y.-C., Regulation of *Escherichia coli top A* gene transcription: involvement of a sigma-S dependent promoter. *J. Mol. Biol.*, 1997, **267**, 481–489.
41. Regnier, P. and Arraiano, C. M., Degradation of mRNA in bacteria: emergence of ubiquitous features. *BioEssays*, 2000, **22**, 235–244.
42. Takayama, K. and Kjelleberg, S., The role of RNA stability during bacterial stress responses and starvation. *Environ. Microbiol.*, 2000, **2**, 355–365.
43. Unniraman, S., Chatterji, M. and Nagaraja, V., A hairpin near the 5′ end stabilizes the DNA gyrase mRNA in *Mycobacterium smegmatis*. *Nucl. Acids Res.*, 2002, **30**, 5376–5381.
44. Condon, C., Putzer, H. and Grunberg-Manago, M., Processing of the leader mRNA plays a major role in the induction of *thr*S expression following threonine starvation in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 6992–6997.
45. Thorne, S. H. and Willimans, H. D., Adapatation to nutrient starvation in *Rhizobium leguminosarum by phaseoli*: analysis of survival, stress resistance and changes in macromolecular synthesis during entry to and exit from stationary phase. *J. Bacteriol.*, 1997, **179**, 6894–6901.
46. Hansen, M. J., Chen, L. H., Fejzo, M. L. and Belasco, J. G., The *omp*A 5′ untranslated region impedes a major pathway for mRNA degradation in *Escherichia coli*. *Mol. Microbiol.*, 1994, **12**, 707–716.
47. Jain, C., Deana, A. and Belasco, J. G., Consequences of RNase E scarcity in *Escherichia coli*. *Mol. Microbiol.*, 2002, **43**, 1053–1064.
48. Cole, S. T. *et al.*, Massive gene decay in leprosy bacillus. *Nature*, 2001, **409**, 1007–1011.
49. Unniraman, S., Prakash, R. and Nagaraja, V., Alternate paradigm for intrinsic transcription termination in eubacteria. *J. Biol. Chem.*, 2001, **276**, 41850–41855.
50. Papavinasasundaram, K. G. *et al.,* Slow induction of RecA by DNA damage in *Mycobacterium tuberculosis*. *Microbiology*, 2001, **147**, 3271–3279.
51. Manjunatha, U. H., Dalal, M., Chatterji, M., Radha, D. R., Visweswariah, S. S. and Nagaraja, V., Functional characterization of mycobacterial DNA gyrase: an efficient decatenase. *Nucl. Acids Res.*, 2002, **30**, 2144–2153.
52. Higgins, N. P., Peebles, C. L., Sugino, A. and Cozzarelli, N. R., Purification of subunits of *Escherichia coli* DNA gyrase and reconstitution of enzymatic activity. *Proc. Natl. Acad. Sci. USA*, 1978, **75**, 1773–1777.

ACKNOWLEDGEMENTS. I would like to thank my colleagues especially S. Unniraman, U. H. Manjunatha, D. R. Radha, M. Chatterjee and P. Jain who carried out most of the experiments. H. V. Jayashree and J. Robert have helped in the preparation of the manuscript. Research in my laboratory for this work has been supported by extramural grants from Council for Scientific and Industrial Research, Department of Science and Technology, Indian Council of Medical Research, Government of India. Part of the infrastructural support has come from ICMR Centre for Advanced Study in Molecular Medical Microbiology to the department of Microbiology and Cell Biology, Indian Institute of Science.

# Homologous recombination in mycobacteria

**K. Muniyappa*, N. Ganesh, N. Guhan, Pawan Singh, G. P. Manjunath, S. Datta[¶], Nagasuma R. Chandra[†] and M. Vijayan[¶]**

Department of Biochemistry, [¶]Molecular Biophysics Unit, and [†]Bioinformatics Center, Indian Institute of Science, Bangalore 560 012, India

**In recent years, considerable effort and resources have been expended to develop targeted gene delivery methods, and generation of auxotrophic mutants of mycobacteria. The results of these studies suggest that mycobacteria exhibit a wide range of recombination rates, which vary from loci to loci. Here we review the methods developed for allele exchange and targeted gene disruption as well as the mechanistic aspects of homologous recombination in mycobacteria. The results of whole genome, functional and structural analyses of *Mycobacterium tuberculosis* and *Mycobacterium smegmatis* RecA and SSB proteins provide insights into variations of the prototypic *Escherichia coli* paradigm. This variation of a common theme might allow mycobacteria to function in their natural but complex physiological environments.**

STUDIES of *Mycobacterium tuberculosis* are hindered by its long generation time (12 h), severe clumping of the bacilli and the safety risk involved with handling live cultures. Our understanding of the mechanisms of pathogenesis caused by the tubercle bacillus is inadequate, and the factors responsible for virulence are poorly understood. Although much research has focused on immunology, biochemistry, and microbiology of this pathogen, investigations into molecular interactions between specific gene products have not been possible because of the lack of defined mutants with specific phenotypes. It is believed that transfer of DNA into tubercle bacilli, either by allele replacement or transposon mutagenesis would provide insights into understanding of the role(s) of virulence determinants as well as mechanisms of pathogenesis. Therefore, understanding of the mechanistic aspects of homologous recombination may help molecular genetic manipulation of mycobacteria as well as knowledge needed to develop strategies to control tuberculosis.

## Introduction of foreign DNA into mycobacteria

Introduction of foreign DNA by transduction or conjugation has greatly facilitated the generation of mutant strains and the functional analysis of the genomes of *Escherichia coli* and *Salmonella typhimurium*[1,2]. Similarly, introduction of foreign DNA into mycobacterial strains via a genetic route has relied on the processes of transformation or transduction. Various plasmids, derived from mycobacteriophages, such as TM4, L1, and D29, have proven useful for the development of transformation systems for mycobacteria[3]. The transfer of DNA by transduction by a virus was first demonstrated for *M. smegmatis*[4,5]. More recently, a single-step and relatively efficient allele exchange method was developed using a shuttle plasmid integrated into a specialized transducing mycobacteriophage TM4. This method was used to construct seven isogenic auxotrophic mutant strains of *M. smegmatis*, three substrains of *M. bovis* BCG and three strains of *M. tuberculosis*[6]. A number of investigators have ascertained the potential utility of this method for targeted gene disruptions at several loci in *M. tuberculosis*[7–10]. Mycobacteriophages have been used as vectors to generate luciferase reporter phages for the rapid detection of pathogenic species of mycobacteria and the assessment of their drug susceptibilities.

Bacterial conjugation is a process by which DNA is transferred from a donor to recipient cell through cell-to-cell contact mediated by energy-driven transport. The process is conceptualized as two sub-processes: DNA preparation, and mating bridge formation. Studies of conjugation in *E. coli* have played a crucial role in the development of bacterial genetics, and led to the isolation of the first recombination-deficient (*rec*) mutants. In *E. coli* the functions required for conjugation are mainly encoded by the F factor, which act at a unique *cis*-acting site to initiate and complete DNA transfer. By contrast, in the naturally occurring conjugation system of *M. smegmatis*, DNA transfer is chromosomally encoded[11]. In addition, unlike conventional plasmid transfer, recipient recombination functions are required to allow this plasmid, and derivatives of it, to re-circularize through a process similar to gap repair. Extended DNA homology with the recipient chromosome and the F factor is required to facilitate repair, resulting in acquisition of recipient chromosomal DNA by the plasmid. Together, these results show that DNA transfer in *M. smegmatis* occurs by a mechanism different from that of prototypical plasmid transfer systems[11]. Plasmid-mediated conjugative gene transfer has not been demonstrated in strains belonging to *M. tuberculosis* complex.

## Gene transfer in mycobacteria

In recent years, considerable effort and resources have been expended to the development of methods for targeted

*For correspondence. (e-mail: kmbc@biochem.iisc.ernet.in)

gene delivery and mutation in mycobacteria. The methods are mechanistically similar to those developed for *E. coli* or yeast. Such approaches indicate that generation of genetically defined isogenic strains containing single or multiple mutations has been hampered by the lack of suitable tools. *M. tuberculosis* and *M. smegmatis* genomes exhibit a wide range of recombination rates as reflected in the efficiency of allele exchange, which is known to vary from loci to loci. It has been technically difficult to generate defined auxotrophic mutants of *M. tuberculosis* at high frequency in a routine manner. Mutagenesis of mycobacteria has been performed by random or targeted strategies. In organisms in which gene targeting has been observed at high efficiency, DNA molecules with broken ends have been shown to be more recombinogenic than covalently closed circular DNA. However, the stimulatory role of double-strand breaks in mycobacteria is poorly understood. Historically, the first recombinant DNA vectors developed for mycobacteria include shuttle plasmid vectors and chimeric DNA molecules that replicate in *E. coli* as plasmids and in mycobacteria as phages[12]. These are integrated into the bacterial genome, by recombination, so that encoded resistance genes may be maintained over time if the plasmid cannot replicate independently within that cell. The early studies of successful isolation of auxotrophic mutants for the *M. tuberculosis* complex strains used insertional mutagenesis systems, which resulted in illegitimate recombination[13], transposon mutagenesis[14] or allele exchange[15]. Over the years, a variety of alternative gene transfer strategies have been developed to achieve high frequency of allele exchange in *M. smegmatis*[16–20]. On the other hand, similar studies in *M. tuberculosis* involving random shuttle mutagenesis using transposons displayed low frequency of mutations at allelic sites[13,21,22]. The difficulties encountered in these studies led to the conclusion that slow-growing species of mycobacteria promote a high frequency of illegitimate recombination[13,23,24]. Why is this the case? The probable answer stems from the fact that the methods used for detection of very rare allelic exchange events are hindered by low transformation efficiencies and high frequencies of illegitimate recombination, especially in the slow-growing species of pathogenic mycobacteria.

Advances in the construction of gene targeting vectors together with the improvement in the delivery systems have led to increased efficiency of generation of 'knockout' mutants of *M. tuberculosis* and *M. bovis* BCG. These investigations involved short[25,26] or long linear DNA fragments[15] as substrates. Several groups have demonstrated the use of 'suicide' plasmid vectors (using a non-temperature-sensitive plasmid) to achieve insertional mutagenesis in both fast- and slow-growing species of mycobacteria[18–20,21,27–33]. A two-step selection method using selectable and counterselectable markers, positioned on either replicating or non-replicating plasmids, has also been successfully used in *M. smegmatis*[20,32], *M. bovis* BCG and *M. tuberculosis*[30,34–36]. Interestingly, in the case of 'suicide' plasmid vectors, pretreatment of DNA with UV light or alkali enhanced homologous recombination (HR), and abolished illegitimate recombination in the recipient cells. The suicide vector approach is dependent upon the delivery of the gene targeting vectors by electroporation. Therefore, the HR frequencies are very close to the efficiency at which plasmids can be electroporated into slow-growing mycobacteria, whereas the suicide vector approach is limited to those cases where high transformation efficiencies can be obtained. Consequently, it is surmised that this electroporation limitation, rather inefficient HR, may be the reason for difficulties encountered in allele exchange experiments in slow-growing species of mycobacteria[19].

An alternative strategy for gene targeting involves the use of replicating vectors. The method offers the advantage that high density of recombinant vectors increase the frequency of allele exchange, compared with that obtained with the suicide vectors. These vectors have greatly improved reproducibility of allele exchange in the slow-growing species of mycobacteria[37]. The possible reasons are (i) the availability of increased time for HR and (ii) DNA replication and recombination occur concurrently in the cell. A further increase in the efficiency of isolation of allelic replacements has been achieved by combining a counter-selection method with vectors bearing temperature-sensitive origin of replication[38]. Recently, a promising method has been developed for making targeted gene knockouts in *M. smegmatis* and *M. bovis* BCGs based on two-plasmid incompatibility system. This method uses a pair of replicating plasmids carrying a mutated allele of a targeted gene or a transposon, and has the advantage by providing prolonged time for HR[37]. When used for the generation of *M. smegmatis pyrF* mutant alleles, high frequency of recombinants was obtained by this method.

## Analysis of *M. tuberculosis* and *M. leprae* genomes for *rec* genes

The *M. tuberculosis* genome is 4.4 Mb long, which is exceedingly rich in genes for lipid biosynthesis and degradation[38]. In parallel, the 3.3 Mb genome sequence of *M. leprae* has been determined[39]. *M. tuberculosis* genome can potentially encode 3924 genes, while the *M. leprae* encodes 1604 proteins and contains 1116 pseudogenes, compared to six in *M. tuberculosis*[38,39]. Comparison of the genome sequence of *M. leprae* with that of *M. tuberculosis* indicates that the former has undergone massive gene decay, losing large number of genes since its divergence from a common mycobacterial ancestor[40,41]. It is possible that its genes were rendered inactive once their functions were no longer essential for survival, and this was followed by genome shrinkage through rear-

rangements and/or deletions. It has been proposed that downsizing of *M. leprae* genome, and mutations in several metabolic genes, may account for its exceptionally slow growth as well as its failure to grow *in vitro*. It seems to have completely dispensed with or substantively reduced certain metabolic pathways, including oxidative and anaerobic respiratory chains. The enzymes for breaking down host-derived lipids, a means by which many mycobacterial pathogens derive their energy, are also drastically reduced in *M. leprae*. By contrast, most anabolic pathways seem to be intact, indicating that *M. leprae* depends on these pathways to survive in the nutrient-poor microenvironment of phagosomes[38–41]. The availability of the mycobacterial genome sequences and the ability to generate transposon mutants, targeted gene disruptions, and complementation analyses provide an unprecedented opportunity for the elucidation of the functions of mycobacterial genes.

Comparative analysis of the genomes of *M. tuberculosis* and *M. leprae* has revealed a considerable decay or deletion of genes involved in recombination, especially of those encoding for alternate pathways of HR[40–42]. In *E. coli*, at least four alternate pathways exist for HR, each featuring the action of a distinct exonuclease and/or helicase[43–44]. These are required to generate 3′ invasive ends for polymerization of RecA to initiate recombination. These include RecBCD, RecE/RecT or RecJ/RecQ. Most notably, the *M. tuberculosis* genome is devoid of homologues of *E. coli recE*, *recT*, *recQ*, *recJ*, *recO* and *rusA*[42]. RecQ helicase has been shown to disrupt illegitimate recombination in *E. coli*, and its absence could be one of the reasons for higher frequency of illegitimate recombination in *M. tuberculosis*. Intriguingly, *M. tuberculosis recB*, *recC* and *recD* genes resemble those of Gram-negative species rather than analogues of *addA addB* that exist in Gram-positive bacteria[45]. *M. leprae* genome possess neither of these systems, however, it contains an archaeal-type exonuclease and helicase similar to the *recB* family of exonuclease/helicase. Mutations are also found in *M. leprae* genes involved in DNA repair (the *mutT*, *dnaQ*, *alkA*, *dinX*, and *dinP* genes)[45].

In *E. coli*, early steps of HR involve RecA, RecBCD enzyme, and the recombination hotspot called Chi (χ) site. *E. coli* χ sites are G-rich (5′-GCTGGTGG-3′) asymmetric *cis*-acting regulatory sequences that modify the activities of the RecBCD enzyme, thereby leading to the generation of single-stranded DNA. This results in preferential loading of RecA onto the χ-containing DNA strand. The RecA nucleoprotein filament then invades homologous double-stranded DNA to produce a D-loop structure. Although *recB*, *recC*, *recD* genes and putative χ-like sites have been identified in the *M. tuberculosis* genome[42,46], and are likely to exist in other mycobacteria, it has not been shown whether they constitute recombination hotspots in any of the mycobacterial species.

## Organization and expression of mycobacterial *recA*

The biochemical activities of many of the factors involved in HR in mycobacteria are poorly understood. However, two components of the pathway of HR in mycobacteria, RecA and SSB, have been studied in considerable detail. One of the primary functions of eubacterial *recA* is its role in recombinational DNA repair[43,44]. Recombination between similar DNA sequences contributes significantly to genome plasticity; post-replicative mismatch repair, and restricts recombination between homologous sequences. RecA is both ubiquitous and well conserved among a range of organisms. Unlike *M. smegmatis*, pathogenic species of mycobacteria display relatively low levels of HR[47]. In contrast to *M. smegmatis recA*, the *M. tuberculosis* and *M. leprae recA* contain an in-frame open reading frame encoding an intein[48–49]. RecA intein is removed from the precursor RecA by an autocatalytic protein splicing reaction, and active RecA is generated by ligation of amino- and carboxyl-terminal fragments mediated by intein (Figure 1). This post-translational processing is required for RecA activity: a mutant gene that no longer undergoes protein splicing fails to complement *E. coli recA* mutants, whereas the wild-type gene can[49]. Therefore, it is possible that this unusual arrangement for the production of mature, active RecA protein might affect its function in *M. tuberculosis*, either by regulating the splicing reaction or by subsequent interaction of the intein with RecA. The biological significance of the presence of intein in the *recA* gene in pathogenic mycobacteria is the subject of the on-going debate[47,50–52]. Given the fact that inteins are mostly found in *recA* of pathogenic mycobacteria, the advantage is unclear. Although the significance of *recA* intervening sequence is unknown, it has been shown that *M. leprae* or *M. tuberculosis recA* complements *M. smegmatis* Δ*recA* strains for recombination and UV repair[50–52].

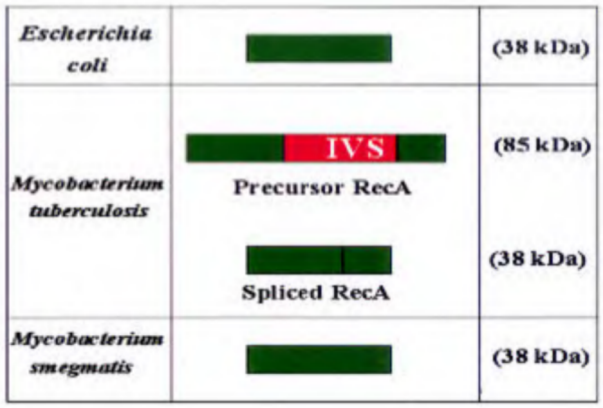


**Figure 1.** Structural organization of *E. coli*, *M. tuberculosis* and *M. smegmatis* RecA proteins. IVS, intervening sequence.

The characterization of *M. tuberculosis* RecA intein revealed that it is a unique member of the LAGLIDADG family of homing endonucleases[53]. *M. tuberculosis* RecA intein displayed very novel characteristics: In the presence of $Mn^{2+}$ and ATP, it was able to cleave cognate site in the inteinless *recA* allele at 24 and 33/43 bases upstream of the intein insertion site, in the upper and lower strands respectively[54]. This property is consistent with the class of homing endonucleases that tolerate some sequence degeneracy within their recognition sequences[53]. Recent studies also provided a great deal of insight into the catalysis of DNA cleavage as well as ATPase activity of RecA intein[55]. Intriguingly, RecA intein displayed robust site-specific endonuclease activity with non-cognate DNA in the presence of $Mg^{2+}$ generating DNA fragments with blunt end or 1–2 base overhangs[56]. The latter activity has been implicated in the movement of RecA intein DNA sequence from one chromosome location to another in natural populations[56]. It is unknown whether *M. leprae* RecA intein possess similar activities.

In *E. coli*, *recA* is a part of the SOS response system[43]. SOS response is activated by agents or processes related to DNA metabolism that generate single-stranded DNA. The SOS box is the target for LexA binding, which exists upstream of all genes expressed in the SOS regulon. Under normal growth conditions, low levels of RecA and LexA exist in the cells[43]. LexA functions as a repressor to inhibit *recA*, *lexA* and many other repair operons involved in the SOS response. When RecA is activated by DNA damage, it promotes autocatalytic cleavage of LexA. In *M. tuberculosis*, in addition to *recA*, a number of DNA-damage induced genes are regulated by LexA-dependent mechanism[57,58]. However, few genes induced in response to DNA damage are not regulated by LexA binding, but by an alternate mechanism of gene regulation[57]. Recently, it has been shown that *M. tuberculosis recA* is expressed from two promoters: one is LexA-regulated, and the second remains DNA damage inducible in the absence of RecA or when LexA binding is prevented[59]. These findings indicate that the mycobacterial DNA repair system is different in many aspects compared to the prototypic model species, e.g. *E. coli* or *B. subtilis*. *recA* deletion mutant of *M. smegmatis* strain (HS42) exhibited enhanced sensitivity to UV irradiation and failed to display HR[34]. The deficiencies in UV survival and recombination were complemented by introduction of the cloned *M. smegmatis recA* gene[34].

## Recombination activities of mycobacterial RecA and SSB proteins

Recombination is central to the identification of genes and to the understanding of the biology of any organism. Using *E. coli* as a model, the process of HR has been separated into four kinetically distinguishable phases: presynapsis, synapsis, strand exchange and resolution[43,44]. *Presynapsis* involves cooperative binding of RecA protein on single-stranded DNA forming a helical nucleoprotein filament; *synapsis*, the homologous alignment of nucleoprotein filament comprised of RecA–ssDNA with naked duplex DNA; and unidirectional *strand exchange*, which creates long stretches of heteroduplex DNA. Finally, the heteroduplex DNA is expanded by RuvAB motor proteins and resolved by the RuvC endonuclease. *E. coli* RecA is the central component in these processes, and, because its functions are conserved from bacteriophage to humans, its study has provided a paradigm for understanding the biologically important process of HR. This complex process requires the action of $> 20$ gene products. In *E. coli*, the proteins that carry out all of the steps of HR have been purified and characterized *in vitro*. These studies are quite advanced in the case of *E. coli*, and portions of the HR pathway are being reconstituted *in vitro*. RecA-like proteins constitute a group of DNA strand transfer proteins ubiquitous in eubacteria, eukarya and archaea. However, the functional relationship among RecA-like proteins is poorly understood.

To understand the basis for inefficient allele exchange in mycobacteria compared to *E. coli*, to obtain gene targeting in mycobacteria with reasonable efficiency, and to understand the differences in allele exchange between *M. tuberculosis* and *M. smegmatis*, what is needed is greater insight into the molecular mechanism of HR in mycobacteria, and detailed characterization of the biochemical activities of the components of HR in these species. It is possible that the endogenous DNA repair and recombination machinery in mycobacteria is different from that of *E. coli*. To this end, *M. tuberculosis* RecA (38 kDa), but not its precursor (85 kDa), displayed the hallmark features of *E. coli* RecA, including binding to single-stranded DNA, ssDNA-dependent ATP hydrolysis, formation of D-loops, homologous pairing between single-stranded DNA with duplex DNA, and strand exchange[60,61]. There were, however, striking qualitative and quantitative differences in the activities and pattern of strand exchange promoted by *E. coli and M. tuberculosis* RecA on one hand, and between *M. smegmatis* and *M. tuberculosis* RecA on the other[60–62]. These include rates of ssDNA-dependent ATP hydrolysis, conditions and cofactors required for the display of maximum homologous pairing and strand exchange. Mycobacterial RecA proteins promoted much slower rates of ATP hydrolysis than the rates of the reactions catalyzed by *E. coli* RecA in side-by-side comparisons. Results of molecular modelling and the crystal structure of *M. tuberculosis* RecA indicated that the reduced affinity for ATP and relative catalytic inefficiency of *M. tuberculosis* RecA is related to the expansion of the P-loop region, compared to its homologue in *E. coli*[60].

Another set of observations revealed significant differences in the pattern of homologous pairing and strand

exchange promoted by *M. tuberculosis* and *M. smegmatis* RecA proteins. *M. tuberculosis* RecA was able to effect maximum strand exchange in the alkaline pH range, whereas *M. smegmatis* RecA was around neutral pH[61–63]. Although the rates and the pH profiles of dATP hydrolysis catalysed by *M. tuberculosis* and *M. smegmatis* RecA were similar, only the latter was able to couple dATP hydrolysis to strand exchange. A number of studies have shown that single-stranded DNA binding proteins (SSB) serve as accessory factors in HR[43,44,64]. *M. smegmatis* SSB (165 aa) shares 84% identity and 89% similarity with the *M. tuberculosis SSB* (164 aa)[65]. While *E. coli* RecA promoted substantial strand exchange in the absence of SSB, mycobacterial RecA proteins were completely unable to do so in the absence of SSB[61,63]. This finding confirmed the absolute requirement of SSB for a HR in *M. smegmatis*. Significantly, unlike *E. coli* SSB, mycobacterial SSB proteins physically interacted with their cognate RecA proteins with high affinity. Further, DNA size played an important role on the ability of mycobacterial RecA proteins to synthesize extended lengths of heteroduplex DNA. For example, with duplex DNA length of $< 2$ kb, the efficiency of strand exchange was indistinguishable from that of the prototype *E. coli* RecA, whereas it decreased with increase in the size of duplex DNA (Figure 2)[61,63]. *E. coli* RecA was able to effect complete strand exchange between linear duplex DNA (6.4 kb) and ssDNA (6407 nucleotides) to generate gapped or nicked circular duplex DNA. In contrast to this, *M. tuberculosis* RecA generated substantial amounts of intermediates and networks of DNA as the length of linear duplex DNA was increased from 1 kb to 6.4 kb. The direct correlation between the length of duplex DNA and accumulation of DNA intermediates and networks of DNA indicate that the ability of *M. tuberculosis* RecA to generate extended stretches of heteroduplex DNA is limited. In addition, strand exchange promoted by *M. tuberculosis* and *M. smegmatis* RecA displayed distinctly different pH profiles, suggesting functional diversity between RecA from pathogenic and non-pathogenic species of mycobacteria[62,63].

## Structure of mycobacterial RecA proteins

In the absence of DNA, the crystal structure of *E. coli* RecA revealed a central core domain and two smaller domains at the amino and carboxyl termini. The core which is made up of twisted eight stranded β-sheet flanked by four α-helices contains domains for DNA-binding, designated as L1 and L2 loops, and P-loop containing the nucleotide triphosphate-fold[66,67]. In the *E. coli* RecA crystal structure, the monomers are packed so as to form a right-handed helical filament with 6 monomers

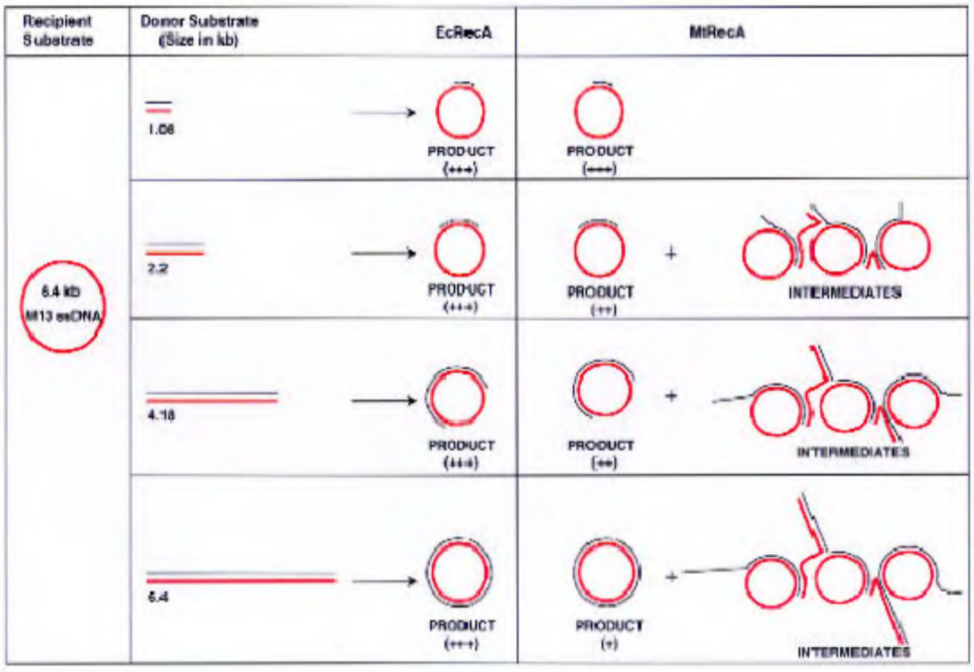


**Figure 2.** Model for the effect of the length of duplex DNA on strand exchange promoted by *E. coli* or *M. tuberculosis* RecA. Reactions were performed with linear duplex DNA (donor; 1–6.407 kb) and nucleoprotein filaments of RecA-circular single-stranded DNA (recipient; 6407 nucleotide residues) as described[61]. 'Plus' symbols correspond to the extent of strand exchange: '+++' denotes maximum strand exchange; '++' half-maximal; and '+' one-third of the maximum value. EcRecA, *E. coli* RecA; MtRecA, *M. tuberculosis* RecA.

per turn[66,67]. The overall structures of *M. tuberculosis* and *M. smegmatis* RecA are nearly identical to the *E. coli* RecA structure with r.m.s. deviations of 0.6 to 1.2 Å[68,69]. In comparison, the molecular surface of the *M. tuberculosis* and *M. smegmatis* RecA filaments possess negative electrostatic potential, compared to *E. coli* RecA (Figure 3). The ligand-bound structures of *M. tuberculosis* RecA revealed subtle variations in nucleotide conformations. Furthermore, the neighbouring filaments in the bundle are involved in several hydrogen bonds in *E. coli* RecA, whereas are hardly any in the case of *M. tuberculosis* RecA. As a consequence, the association of filaments of mycobacterial RecA into bundles is much weaker. The DNA-binding loops, L1 and L2, were undefined in the *E. coli* RecA crystal structure. On the other hand, the conformation and orientation of L1 and L2 loops in the mycobacterial RecA structures were defined, and appear to be different[70]. More importantly, the nucleotide binding by the *M. smegmatis* RecA was accompanied by the movement of Gln196 in the L2 loop, which has been implicated in the propagation of the signal induced by the binding of nucleotide cofactor to the DNA-binding loops[69].

## Regulation of recombination

Recombination is vital for various cellular processes related to DNA metabolism, but it must be tightly controlled. The insight into regulation of HR has been derived from studies on *recX* in eubacteria. In *M. smegmatis*, *M. tuberculosis*, *Pseudomonas aeruginosa*, *Streptomyces lividans*, or *Thiobacillus ferrooxidans*, the ORFs of *recA* and *recX* overlap and the two genes are co-transcribed[71–76]. It is known that overexpression of *recA* in *recX* mutants of *S. lividans*, *M. smegmatis*, or *P. aeruginosa*, but not mutant RecA, lead to induction of deleterious effects[72–73]. However, the molecular mechanisms by which *recX* attenuates the deleterious effects induced by *recA* overexpression have remained unknown. Using *M. tuberculosis* as a model, it has been shown that *M. tuberculosis* RecX binds directly to *M. tuberculosis* RecA as well as *M. smegmatis* and *E. coli* RecA proteins *in vivo* and *in vitro*, but not SSB[77]. The direct association of RecX with RecA failed to regulate the specificity or extent of binding of RecA either to DNA or ATP, ligands that are central to activation of its functions. Significantly, RecX severely impeded ATP hydrolysis and the generation of heteroduplex DNA promoted by homologous as well as heterologous RecA proteins[77]. These findings reveal a novel mode of negative regulation of RecA, and imply that RecX might act as an anti-recombinase to repress inappropriate recombinational repair events during normal DNA metabolism (Figure 4). Consistent with these observations, *E. coli* RecX was shown to inhibit strand exchange as well as ATPase activities of its cognate RecA[78], indicating that negative regulation of HR might be a general phenomenon.

## Perspectives

Understanding of the biology of tubercle bacillus requires inputs from comparative analysis of non-pathogenic species of mycobacteria as well. From the earliest studies of HR, it was recognized that *E. coli* is the best model for

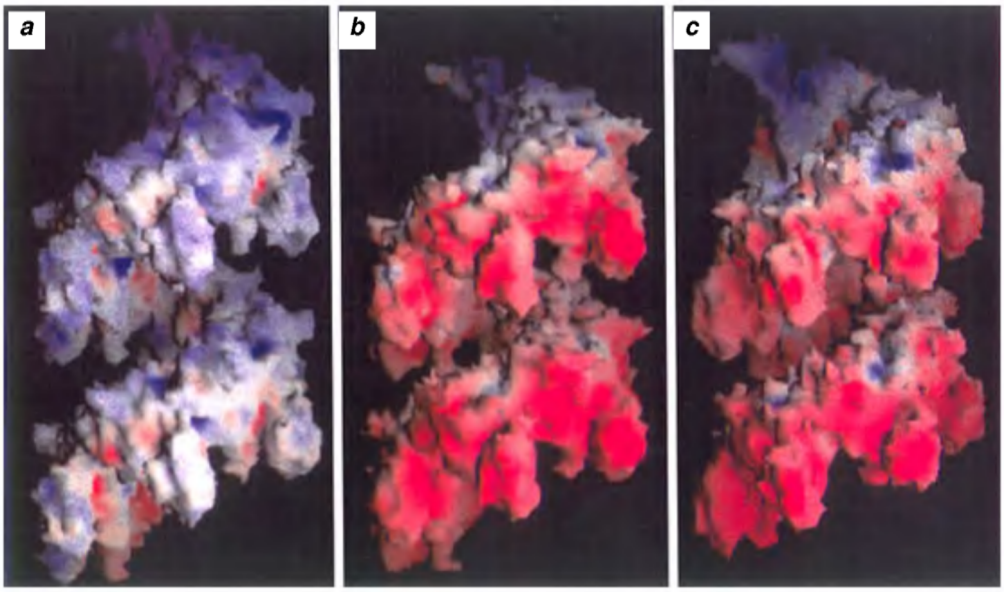


**Figure 3.** Molecular surface representation of the RecA filament: ***a***, *E. coli*; ***b***, *M. tuberculosis*; ***c***, *M. smegmatis*. The surfaces are colour-coded according to the electrostatic potential: Red (negative values), blue (positive values), and white (neutral values).

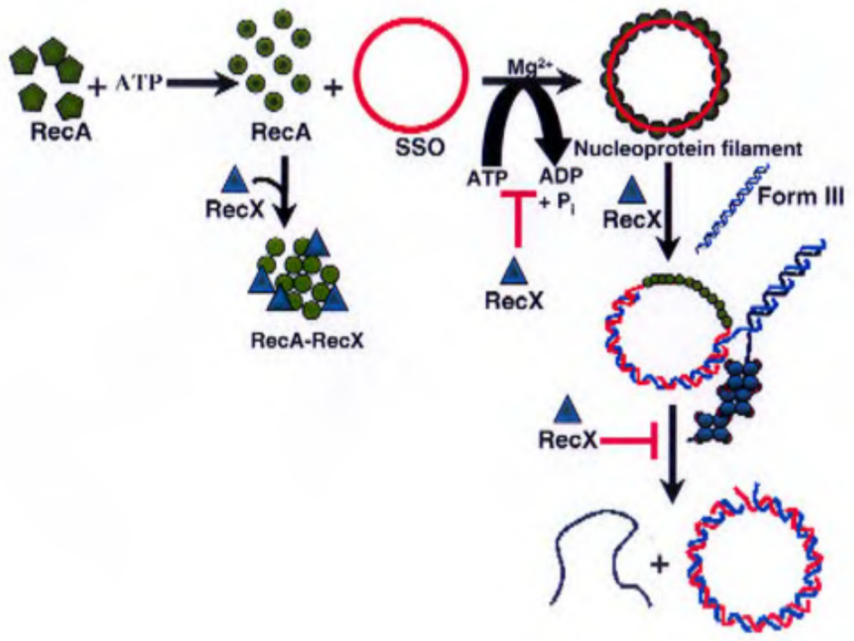


**Figure 4.** Schematic representation of three-strand exchange promoted by the concerted action of RecA and SSB and inhibition by RecX. Both in *in vivo* and *in vitro* conditions, RecX (triangle) has been shown to interact with RecA (circle)[77]. The displaced single-stranded DNA is sequestered by tetramers of SSB. The site of action of RecX is indicated by the T symbol. Form III, linear duplex DNA.

elucidation of the mechanism of HR at the molecular level. The results of whole genome, functional and structural analyses of *M. tuberculosis* and *M. smegmatis* RecA and SSB proteins provide insights into variations of the prototypic *E. coli* paradigm. This variation of a common theme might allow mycobacteria to function in their natural but complex physiological environments. However, further functional and structural studies will be required in understanding the activities of the recombination machinery as well as the mechanistic aspects of HR in mycobacteria.


1. Zinder, N. D. and Lederberg, J., *J. Bacteriol.*, 1952, **64**, 679–699.
2. Lennox, E. S., *Virology*, 1955, **1**, 190–206.
3. Snapper, S. B., Lugosi, L., Jekkel, A., Melton, R. E., Kieser, T., Bloom, B. R. and Jacobs, W. R., Jr., *Proc. Natl. Acad. Sci. USA*, 1988, **85**, 6987–6991.
4. Raj, C. V. and Ramakrishnan, T., *Nature*, 1970, **228**, 280–281.
5. Tokunaga, T., Mizuguchi, Y. and Suga, K., *J. Bacteriol.*, 1973, **113**, 1104–1111.
6. Bardarov, S. *et al.*, *Microbiology*, 2002, **148**, 3007–3017.
7. Glickman, M. S., Cox, J. S. and Jacobs, W. R. Jr., *Mol. Cell*, 2000, **5**, 717–727.
8. Raman, S., Song, T., Puyang, X., Bardarov, S., Jacobs, W. R., Jr. and Husson, R. N., *J. Bacteriol.*, 2001, **183**, 6199–6125.
9. Sirakova, T. D., Thirumala, A. K., Dubey, V. S., Sprecher, H., Kolattukudy, P. E., *J. Biol. Chem.*, 2001, **2**, 16833–16839.
10. Steyn, A. J., Collins, D. M., Hondalus, M. K., Jacobs, W. R. Jr., Kawakami, R. P. and Bloom, B. R., *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 3147–3152.
11. Parsons, L. M., Jankowski, C. and Derbyshire, K. M., *Mol. Microbiol.*, 1998, **28,** 571–582.
12. Jacobs, Jr., W. R., Tuckman, M. and Bloom, B. R., *Nature*, 1987, **327**, 532–534.
13. Kalpana, G. V., Bloom, B. R. and Jacobs, W. R. Jr., *Proc. Natl. Acad. Sci. USA*, 1991, **88**, 5433–5437.
14. McAdam, R. A. *et al.*, *Infect. Immun.*, 1995, **63**, 1004–1012.
15. Balasubramanian, V. *et al.*, *J. Bacteriol.*, 1996, **178**, 273–279.
16. Boshoff, H. I. and Mizrahi, V., *J. Bacteriol.*, 2000, **182**, 5479–5485.
17. Braunstein, M., Brown, A. M., Kurtz, S. and Jacobs, W. R., Jr., *J. Bactriol.*, 2001, **183**, 6979–6990.
18. Pavelka, M. S. Jr. and Jacobs, W. R., *J. Bacteriol.*, 1999, **181**, 4780–4789.
19. Frischkorn, K., Sander, P., Scholz, M., Teschner, K., Prammananan, T. and Bottger, E. C., *Mol. Microbiol.*, 1998, **29**, 1203–1214.
20. Knipfer, N., Seth, A. and Sharder, T. E., *Plasmid*, 1997, **37**, 129–140.
21. Sander, P., Meier, A. and Bottger, E. C., *Mol. Microbiol.*, 1995, **16**, 991–1000.
22. Gordhan, B. G., Andersen, S. J., De Meyer, A. R. and Mizrahi, V., *Gene*, 1996, **178**, 125–130.
23. Aldovini, A., Husson, R. N. and Young, R. A., *J. Bacteriol.*, 1993, **175**, 7282–7289.
24. McFadden, J., *Mol. Microbiol.*, 1996, **21**, 205–211.
25. Azad, A. K., Sirakova, T. D., Rogers, L. M. and Kolattukudy, P. E., *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 4787–4792.
26. Reyrat, J.-M., Berthet, F.-X. and Gicquel, B., *Proc. Natl. Acad. Sci. USA*, 1995, **92**, 8768–8772.
27. Husson, R. N., James, B. E. and Young, R. A., *J. Bacteriol.*, 1990, **172**, 519–524.
28. Berthet, F. X. *et al.*, *Science*, 1998, **282**, 759–762.
29. Fitzmaurice, A. M. and Kolattukudy, P. E., *J. Biol. Chem.*, 1998, **273**, 8033–8039.

30. Parish, T. and Stoker, N. G., *Microbiology*, 2000, **146**, 1969–1975.
31. Pelicic, V., Jackson, M., Reyrat, J. M., Jacobs, W. R., Gicquel, B. and Guilhot, C., *Proc. Natl. Acad. Sci. USA*, 1997, **94**, 10955–10960.
32. Pelicic, V., Reyrat, J. M. and Gicquel, B., *J. Bacteriol.*, 1996, **178**, 1197–1199.
33. Pelicic, V., Reyrat, J. M. and Gicquel, B., *Mol. Microbiol.*, 1996, **20**, 919–925.
34. Papavinasasundaram, K. G., Colston, M. J. and Davis, E. O., *Mol. Microbiol.*, 1998, **30**, 525–534.
35. Hinds, J., Mahenthirlingam, E., Kempsell, K. E., Duncan, K., Stokes, R. W., Parish, T. and Stoker, N. G., *Microbiology*, 1999, **145**, 519–527.
36. Parish, T., Gordhan, B. G., McAdam, R. A., Duncan, K., Mizrahi, V. and Stoker, N. G., *Microbiology*, 1999, **145**, 3497–3503.
37. Pashley, C. A., Parish, T., McAdam, R. A., Duncan, K. and Stoker, N. G., *Appl. Environ. Microbiol.*, 2003, **69**, 517–523.
38. Cole, S. T. *et al.*, *Nature*, 1998, **393**, 537–544.
39. Cole, S. T. *et al.*, *Nature*, 2001, **409**, 1007–1011.
40. Young, D. and Robertson, B., *Curr. Biol.*, 2001, **11**, R381–383.
41. Brosch, R., Gordon, S. V., Eiglmeir, K., Garnier, T. and Cole, S. T., *Res. Microbiol.*, 2000, **151**, 135–142.
42. Muniyappa, K., Vaze, M. B., Ganesh, N., Reddy, M. S., Guhan, N. and Venkatesh, R., *Microbiology*, 2000, **146**, 2093–2095.
43. Kowalczykowski, S. C., Dixon, D. A., Eggleston, A. K., Lauder, S. D. and Rehrauer, W. M., *Microbiol. Rev.*, 1994, **58**, 401–465.
44. Cox, M. M., *Annu. Rev. Genet.*, 2001, **35**, 53–82.
45. Dawes, S. S. and Mizrahi, V., *Lepr. Rev.*, 2001, **72**, 408–414.
46. Kyuma, D., Arakawa, K., Uno, R., Nakayama, Y. and Tomita, M., *Genome Informatics*, 2002, **13**, 533–534.
47. McFadden, J., *Mol. Microbiol.*, 1996, **21**, 205–211.
48. Davis, E. O., Sedgwick, S. G. and Colston, M. J., *J. Bacteriol.*, 1991, **173**, 5653–5662.
49. Davis, E. O., Jenner, P. J., Brooks, P. C., Colston, M. J. and Sedgwick, S. G., *Cell*, 1992, **71**, 201–210.
50. Davis, E. O., Thangaraj, H. S., Brooks, P. C. and Colston, M. J., *EMBO J.*, 1994, **13**, 699–703.
51. Frischkorn, K., Sander, P., Scholz, M., Teschner, K., Prammananan, T. and Böttger, E. C., *Mol. Microbiol.*, 1998, **29**, 1203–1214.
52. Frischkorn, K., Springer. B., Bottger, E. C., Davis, E. O., Colston, M. J. and Sander, P., *J. Bacteriol.*, 2000, **182**, 3590–3592.
53. Guhan, N. and Muniyappa, K., *Crit. Rev. Biochem. Mol. Biol.*, 2003, **38**, 199–248.
54. Guhan, N. and Muniyappa, K., *J. Biol. Chem.*, 2002, **277**, 16257–16264.
55. Guhan, N. and Muniyappa, K., *Nucleic Acids Res.*, 2003, **31**, 4184–4191.
56. Guhan, N. and Muniyappa, K., *J. Biol. Chem.*, 2002, **277**, 40352–40361.
57. Brooks, P. C., Movahedzadeh, F. and Davis, E. O., *J. Bacteriol.*, 2001, **183**, 4459–4467.
58. Durbach, S. I., Anderson, S. J. and Mizrahi, V., *Mol. Microbiol.*, 1997, **26**, 643–653.
59. Davis, E. O., Springer, B., Gopaul, K. K., Papavinasasundaram, K. G., Sander, P. and Bottger, E. C., *Mol. Microbiol.*, 2002, **46**, 791–800.
60. Kumar, R. A., Vaze, M. B., Chandra, N. R., Vijayan, M. and Muniyappa, K., *Biochemistry*, 1996, **35**, 1793–1802.
61. Vaze, M. B. and Muniyappa, K., *Biochemistry*, 1999, **38**, 3175–3186.
62. Ganesh, N. and Muniyappa, K., *Proteins Struct. Funct. Genet.*, 2003, **53**, 6–17.
63. Ganesh, N. and Muniyappa, K., *Biochemistry*, 2003, **42**, 7216–7225.
64. Reddy, M. S., Vaze, M. B., Madhusudan, K. and Muniyappa, K., *Biochemistry*, 2000, **39**, 14250–14262.
65. Reddy, M. S., Guhan, N. and Muniyappa, K., *J. Biol. Chem.*, 2001, **276**, 45959–45968.
66. Story, R. M., Weber, I. T. and Steitz, T. A., *Nature*, 1992, **355**, 318–325.
67. Story, R. M. and Steitz, T. A., *Nature*, 1992, **355**, 374–376.
68. Datta, S., Prabhu, M. M., Vaze, M. B., Ganesh, N., Chandra, N. R., Muniyappa, K. and Vijayan, M., *Nucleic Acids Res.*, 2000, **28**, 4964–4973.
69. Datta, S., Krishna, R., Ganesh, N., Chandra, N. R., Muniyappa, K. and Vijayan, M., *J. Bacteriol.*, 2003, **185**, 4280–4284.
70. Datta, S., Ganesh, N., Chandra, N. R., Muniyappa, K. and Vijayan, M., *Proteins: Struct, Funct.Genet.*, 2003, **50**, 474–485.
71. De Mot, R., Schoofs, G. and Vanderleyden, J., *Nucleic Acids Res.*, **22**, 1994, 1313–1314.
72. Vierling, S., Weber, T., Wohlleben, W. and Muth, G., *J. Bacteriol.*, 2000, **182**, 4005–4011.
73. Papavinasasundaram, K. G, Movahedzadeh, F., Keer, J. T., Stoker, N. G., Colston, M. J. and Davis, E. O., *Mol. Microbiol.*, 1997, **24**, 141–153.
74. Sano, Y., *J. Bacteriol.*, 1993, **175**, 2451–2454.
75. Horn, J. M. and Ohman, D. E., *J. Bacteriol.*, 1988, **170**, 1637–1650.
76. Guilani, N., Bengrine, A., Borne, F., Chippaus, M. and Bonnefoy, V., *Microbiology*, 1997, **143**, 2179–2187.
77. Venkatesh, R., Ganesh, N., Guhan, N., Reddy, M. S., Chandrasekhar, T. and Muniyappa, K., *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 12091–12096.
78. Stohl, E. A., Brockman, J. P., Burkel, K. L., Moromatsu, K., Kowalczykowski, S. C. and Seifert, S. H., *J. Biol. Chem.*, 2003, **278**, 2278–2285.

ACKNOWLEDGEMENTS. This work was supported by a grant from the Indian Council of Medical Research, New Delhi.

# Deletion of the *rel* gene in *Mycobacterium smegmatis* reduces its stationary phase survival without altering the cell-surface associated properties

**Renjith Mathew[†], Anil Kr Ojha[†,#], Anjali A. Karande[‡] and Dipankar Chatterji[†,#,*]**

[†]Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India,
[‡]Department of Biochemistry, Indian Institute of Science, Bangalore 560 012, India
[#]Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore 560 064, India

**Dormant or latent physiology of the mycobacterial species is a subject of current investigation in order to understand the long-term persistence of these organisms inside the host. It is argued that the carbon-starved mycobacteria may serve as a good model for the dormant bacilli. The *relA/spoT* gene is upregulated during carbon starvation in *Mycobacterium tuberculosis* and the deletion of the gene resulted in reduction of long-term persistence in *M. tuberculosis*. Over-expression of the gene in *M. smegmatis* changes the colony morphology of the bacteria. Here we show that knock-out of the *relA/spoT* gene compromises stationary-phase survival of *M. smegmatis*. However, the Δ*relA/spoT* bacteria show similar profile of glycopeptidolipids as the wild-type bacteria under carbon starvation. We have seen here that *M. smegmatis*, a non-pathogenic species, upon carbon starvation exhibits reduced association with murine macrophage cell line RAW 264.7 in comparison to *M. smegmatis* grown in carbon-enriched medium. But the clearance of the bacteria from macrophages takes place in the same window of time in both conditions.**

*MYCOBACTERIUM SMEGMATIS* is a fast-growing, non-pathogenic species of mycobacteria. Due to its short doubling time, *M. smegmatis* has always been argued to be a suitable model for studying the slower-growing pathogenic mycobacteria. But as the species lacks the determinants of mycobacterial virulence, the design of experiments and interpretation of results require careful control. On the other hand, the stationary-phase physiology and the survival of the organism for extended periods of time are interesting phenomena, which need to be addressed separately. Long-term persistence of *M. tuberculosis* was first demonstrated by Corper and Cohn[1], following which the anaerobic culture of the bacterium called the Wayne's model of dormancy, gained popularity[2,3]. However, Nyka[4] first proposed that an *in vitro* starved culture of mycobacteria exhibited similar morphology and hydrophobicity as *in vivo* persisting bacilli. We decided to look at the long-time survival of *M. smegmatis* in the stationary phase under starved condition in order to understand the stress physiology of this organism and with an aim to correlate it with the persisting pathogens. It was noticed that the passage of the culture into the stationary phase or carbon limitation in the medium induced several genes like *sigF*[5] and *relA/spoT*[6]. These genes have important roles in the long-term survival of *M. tuberculosis*. We reported earlier that carbon starvation in *M. smegmatis* alters the cellular morphology[7], concomitant with the synthesis of a new polar glycopeptidolipid (GPL)[8]. All mycobacteria are internalized by macrophages[9,10], but only the pathogenic ones survive and replicate intracellularly[11]. In this article we address two questions:

(i) Whether deletion of *relA/spoT* gene, which is responsible for the maintenance of the level of ppGpp in the cell, in *M. smegmatis* has any relevance to its survival in the stationary phase.
(ii) Whether the mutant *M. smegmatis* also shows similar cell surface-associated properties like the GPL profile and association with macrophages.

## Materials and methods

### *Bacterial strains and growth conditions*

*M. smegmatis* $mc^2155$ was grown in MB7H9 broth (Difco) or agar supplemented with 0.05% Tween-80 (Sigma) and 2 or 0.02% (w/v) glucose for normal and carbon-limiting cultures respectively, at 37°C. Media for growing the *relA/spoT* mutant of *M. smegmatis* had 25 μg/ml hygromycin (Sigma). Amikacin (Torrent Pharma) was used in the cell association and intracellular survival assay, in order to kill the extracellular bacteria following infection of the cells.

### *Eukaryotic cell line*

Murine macrophage cells RAW 264.7 (ATCC number TIB-71) were grown in Dulbecco's modified Eagle's medium high glucose (Sigma) supplemented with 10% foetal bovine serum (Sigma), hereafter referred to as DMEM-10%, under 5% $CO_2$ at 37°C. RAW 264.7 cells were obtained from Maneesha Inamdar, JNCASR, Bangalore.

*For correspondence. (e-mail: dipankar@mbu.iisc.ernet.in)

### *Cell association and intracellular survival assay*

The method was devised similar to the procedure of Ramakrishnan and Falkow[9], developed for *M. marinum* and the murine macrophage cell line J-774.

The RAW 264.7 cells used in the experiments had undergone 19 to 24 passages. Adherent cultures of RAW 264.7 were grown in 25 $cm^2$ tissue-culture flasks (Greiner) for three days. The cells were dislodged and the released cells were recovered by centrifugation, and resuspended in DMEM-10%. The macrophages were then plated in 48-well tissue-culture plates (Costar) at a density of $1 \times 10^5$ cells per well (in 200 µl of medium per well). Cells were grown for a period of 24 h at 37°C under 5% $CO_2$ prior to the infection.

For preparing the bacterial inoculum, $mc^2155$ was grown to the mid-log phase in normal growth medium and to the early stationary phase in carbon-limiting medium. The *relA/spoT* mutant was grown in the carbon-limiting medium to the early stationary phase. Bacteria were harvested by centrifugation at 1300 *g* for 10 min. The pellets were washed and resuspended in DMEM-10% at a concentration of $3 \times 10^7$ cells per ml. Then 200 µl of this bacterial suspension (with an approximate multiplicity of infection [MOI] of 30) was added to the overnight-grown monolayers of RAW 264.7. Phagocytosis of the bacteria was allowed to proceed for 2 h at 37°C under 5% $CO_2$. Then the macrophage monolayers were washed twice with DMEM-10% and treated with DMEM-10% containing 200 µg/ml Amikacin for 1 h at 37°C to kill the residual extracellular bacteria. The monolayers were washed again with DMEM-10% and the washed medium was plated for assessing the presence of any residual extracellular bacteria. The wells for assaying the intracellular survival of bacteria were then incubated in a medium containing 20 µg/ml of Amikacin. At various time points after phagocytosis, cells in triplicates of wells were lysed. For lysing the cells, 200 µl of sterile water was added to the wells after removal of the medium. Ten minutes later, the contents of the wells were vigorously pipetted and the lysate was collected. The lysates were plated on 7H9 agar at appropriate dilutions and *M. smegmatis* colonies appearing after two to three days were counted for the intracellular bacterial counts.

The wells which were assayed immediately after treatment with the medium containing 200 µg/ml Amikacin, served as the standard for measuring the number of phagocytosed bacteria. The time at which these wells were assayed was considered time zero. Lysis was carried out at 12, 24 and 48 h post infection.

### *Targetted replacement of relA/spoT in M. smegmatis with a disrupted copy*

A 400 base pair (bp) internal region of *relA/spoT* from *M. smegmatis* was amplified in polymerase chain reaction using degenerate primers designed from a highly conserved region of *relA/spoT* as observed in other bacteria, including *M. tuberculosis*.

The sequences of the primers are:

F1 = 5′GTGTGCACCGCNTANGCNAAGT3′
R1 = 5′GTGTACCAGTCNNTGCACACCAC3′

A genomic library of *M. smegmatis* constructed in λ-ZAP II (Stratagene) vector at *Eco*RI site (a kind gift by William Bishai, Johns Hopkins University, Baltimore) was screened by the $^{32}P$ labelled 400 bp probe. Three plaques which showed positive signals were lifted from the plate and the recombinant phages from the three plaques were excised into recombinant pBluescript SK phagemid with the DNA of interest according to the earlier published method[12]. All the three recombinant plasmids were found to contain 4.2 kb *Eco*RI fragment and were then subjected to automated DNA sequencing to confirm that the three clones had the same origin. One of three clones, pRelMs, was used for further study. pCK0686 (a kind gift from Prof. William Bishai) was used as a suicide vector for disrupting the *relA/spoT*. The vector has *OriE* and $kan^r$, $suc^s$, $amp^r$ as a marker on the plasmid. Multiple cloning sites flanked another marker $hyg^r$ on its either side and hence this marker was used for making the disruption construct.

The plasmid pRelMs was digested with *Mlu*I, end-filled with DNA polymerase I (klenow fragment), redigested with *Sma*I and self-ligated to generate a pBluescript SK with 584 bp from the start codon of *relA/spoT*. This N-terminal region was released with *Eco*RI–*Not*I and ligated to pET21b at *Eco*RI–*Not*I site. This extra step of subcloning was carried out to gain an *Nde*I site from pET21b (located upstream of *Eco*RI). Then the subcloned fragment was released with *Nde*I–*Not*I and ligated to *Nde*I–*Not*I of pCK0686. The resulting plasmid pCKNT was used for cloning the DNA fragment downstream to the gene. Similarly, the 1.8 kb fragment downstream to the stop codon of *relA/spoT* was cloned on the other side of the $hyg^r$. In order to do this, pRelMs was digested with *Mlu*I, end-filled with DNA polymerase I (Klenow fragment), redigested with *Eco*RV and self-ligated to generate pBlueScript SK with 1.8 kb DNA fragment downstream to the stop codon of *relA/spoT*. The 1.8 kb C-terminal fragment was released with *Eco*RI–*Kpn*I, subcloned into *Eco*RI–*Kpn*I site of pMV261. This extra subcloning step was carried out to gain an appropriate cloning site. Then the downstream fragment was released by *Xho*I and subcloned into the *Xho*I site of pCKNT. Thus the resultant plasmid pΔrelMs had a disrupted *relA/spoT*, which was replaced by $hyg^r$ between the 584th and 2423rd nucleotide sequence.

The disrupted copy of *M. smegmatis relA/spoT* placed on pΔrelMs was electroporated into *M. smegmatis*, $mc^2155$, following the protocol described by Jacobs *et al.*[13], using BIO-RAD electropulsator at 1.5 kV/mm. After 3 h of incubation in antibiotic free medium, the