Guardians of the mycobacterial genome: A review on DNA repair systems in *Mycobacterium tuberculosis*

Amandeep Singh*

**Abstract**

The genomic integrity of *Mycobacterium tuberculosis* is continuously threatened by the harsh survival conditions inside host macrophages, due to immune and antibiotic stresses. Faithful genome maintenance and repair must be accomplished under stress for the bacillus to survive in the host, necessitating a robust DNA repair system. The importance of DNA repair systems in pathogenesis is well established. Previous examination of the *M. tuberculosis* genome revealed homologues of almost all the major DNA repair systems, i.e. nucleotide excision repair (NER), base excision repair (BER), homologous recombination (HR) and non-homologous end joining (NHEJ). However, recent developments in the field have pointed to the presence of novel proteins and pathways in mycobacteria. Homologues of archeal mismatch repair proteins were recently reported in mycobacteria, a pathway previously thought to be absent. RecBCD, the major nuclease-helicase enzymes involved in HR in *E. coli*, were implicated in the single-strand annealing (SSA) pathway. Novel roles of archeo-eukaryotic primase (AEP) polymerases, previously thought to be exclusive to NHEJ, have been reported in BER. Many new proteins with a probable role in DNA repair have also been discovered. It is now realized that the DNA repair systems in *M. tuberculosis* are highly evolved and have redundant backup mechanisms to mend the damage. This review is an attempt to summarize our current understanding of the DNA repair systems in *M. tuberculosis*.

**INTRODUCTION**

Tuberculosis still remains among the foremost global health problems, leading to millions of deaths annually. Prevalence of tuberculosis through decades of human history is evident from the success of the pathogen, *Mycobacterium tuberculosis*, in the course of survival and evolution. Pathogenic bacteria face a variety of hostile conditions posed by host defense mechanisms and antibiotic treatments. *M. tuberculosis*, which resides in host macrophages, is exposed to frequent DNA damaging assaults by a variety of endogenous and exogenous factors. Genomic integrity is pivotal for survival and proliferation of all organisms. Consequently, the presence of strong DNA repair systems, in such pathogens, is necessary to ensure an efficient error-free transmission of genetic material.

In addition to its biological significance, it has now been realized that DNA repair has a role in genome diversification and consequently the development of drug resistance in mycobacteria [1, 2]. Mis-sense alterations in genes of three anti-mutator proteins have been attributed to the increased risk of drug resistance in *M. tuberculosis* W-Beijing strains [3]. Later, it was shown that DNA repair genes in strains of the *M. tuberculosis* complex exhibit higher polymorphism than the house-keeping genes [4]. *M. tuberculosis* can stay dormant for years inside the host in persistent or latent stage, leading to the development of active tuberculosis later in life [5]. Most of the current regime therapy drugs against tuberculosis are inactive against the persistence stage bacteria. Many DNA repair proteins are implicated to be essential in the persistent or latent stage [6, 7]. Consequently, proteins involved in DNA repair can also turn out to be co-targets or adjuvant targets in drug design [8].

Reactive oxygen species (ROS) and reactive nitrogen intermediates (RNI), primarily generated by macrophages as host antimicrobial response, can permeate cell membranes and damage the DNA [9, 10]. Reflecting the range of stresses, a variety of DNA damage can occur: base modifications, such as oxidation or alkylation of bases, covalent linking of two bases or elimination of a base leading to creation of an abasic site. DNA lesions, single- or double-stranded breaks, arising as a consequence of damage or processing defects of cellular machinery, pose the most serious threat to the viability of organisms [11].

Multiple DNA repair mechanisms have evolved in the genus mycobacteria and are employed depending on the
nature of DNA damage. Among the major pathways, base modifications are repaired by base excision repair (BER) or nucleotide excision repair (NER), while breaks are repaired by homologous recombination (HR), non-homologous end joining (NHEJ) or single-strand annealing (SSA) [12–14]. The canonical components of mismatch repair (MMR) were found to be absent in mycobacteria [14]. The ribonucleotides incorporated during replication are removed by ribonucleotide excision repair (RER). Owing to the technical difficulties in culturing mycobacterial species, the study of DNA repair systems in mycobacteria is relatively unexplored and has been primarily understood on the basis of homology rather than functional studies [15, 16].

Proteins involved in SOS response, BER and recombination in mycobacteria, form a significant component of our long-range programme on the characterization of mycobacterial proteins [17–22]. Our recent analysis of 43 mycobacterial genomes suggested noteworthy differences in the recombination repair machinery within mycobacteria, and in comparison to other organisms [23]. A number of new proteins with probable involvement in DNA repair were also identified. A large number of outstanding reviews with comprehensive information on DNA repair in mycobacteria are available [12–16, 24, 25]. However, many recent publications have challenged the dogmatic views of DNA repair systems in mycobacteria [26–37]. In this review, all previous information has been incorporated in brief, with a major focus on the new developments in the field. A list of proteins discussed in this review is given in Table 1. The role of DNA repair proteins in infection and pathogenicity of M. tuberculosis has been discussed extensively. The aim of the review is to try and convince the reader that M. tuberculosis has a remarkably evolved and highly redundant DNA repair system, providing it with robust survivability in the harsh environments faced inside the host cell.

**BASE EXCISION REPAIR**

The BER pathway is primarily responsible for the repair of nucleotides damaged following alkylation, deamination or oxidation. A range of damaged nucleotides, including 7-methylG, 3-methylA, 7,8-dihydro-8-oxoguanine (8-oxoG), 5-hydroxyC and 4,6-diamino-5-formamidopyrimidine (FapyA), among others are repaired by the BER pathway proteins [38]. To begin with, the damaged nucleotides are recognized and excised by specific DNA glycosylases leaving an abasic (AP) site in the DNA backbone. AP sites are processed by AP (apurinic/apyrimidinic) lyases (e.g. ExoIII), which hydrolyse the sugar-phosphate bonds leaving a 3’-hydroxyl and 5’-deoxyribose phosphate (dRP). These sites are further processed by a deoxyribose-phosphodiesterase (dRPase, e.g. Fpg, RecF), followed by the action of a DNA polymerase to fill the gap and DNA ligase to seal the nicks [39]. The steps of the pathway are illustrated in Fig. 1.

**GO pathway**

**Base excision**

Guanine is highly susceptible to oxidative damage, owing to its low redox potential. Consequently, the high GC content of mycobacterial genomes increases the chances of oxidative damage [40]. The GO repair system, specific to prevention or repair of oxidized guanine residues, forms the major component of the BER pathway [25, 41]. The presence of 8-oxoG, the most frequent and stable base lesions in DNA, leads to incorporation of A (8oxoG: A) during replication. This results in C to A (or G to T) mutation in the newly synthesized strand. Typically, this pathway involves the interplay of two DNA glycosylases, namely Fapy DNA glycosylase (Fpg or MutM) and adenine glycosylase (MutY). The Fpg or MutM removes 8-oxoG from DNA, while MutY removes adenines mistakenly incorporated against 8-oxoG. Additionally, an 8-oxoG triphosphatase (MutT) that degrades 8-oxo-dGTP is also present, thereby minimizing the chance of its misincorporation into DNA [13, 42].

Unlike *E. coli*, mycobacterial Fpg excises 8-oxoG preferably when paired against C, G or T but not A [43, 44]. Additionally, another Fpg homologue (Fpg2) was also found to be present in M. tuberculosis. Fpg2 was characterized as a non-functional protein, owing to the lack of a conserved proline residue in the catalytic centre [44, 45]. Besides the Fpg or MutY homologues, *M. tuberculosis* also possesses two orthologues of endonucleaseVIII/Nei proteins, belonging to the Fpg/Nei family of DNA glycosylases [46]. *M. tuberculosis* Nei1 showed specificity for oxidized pyrimidines, uracil, and very little excision activity for 8-oxoG, while the role of Nei2 is yet to be established [36]. *M. tuberculosis* was also found to have a single homologue of the endonucleaseIII/Nth gene, involved in the removal of damaged nucleotides [46].

In *M. smegmatis*, a non-pathogenic mycobacterium used as a model system, the Nth and Nei triple deletion mutant (Δnei2 Δnei1 Δnth) shows an exaggerated decline in survival rate and increased rates of mutation. The individual deletion mutants were not able to confer a mutator phenotype [36]. In an earlier study, it was shown that Δfpg and ΔmutY mutants do not affect the survival rate, or show any appreciable changes in mutation rates [43, 47]. However, recently, a synergistic effect of mutY and fpg loss was observed in *M. smegmatis*, leading to four–eight fold higher mutation rates. Combination of Fpg and MutY was found to be crucial in preventing C (G) to A (T) mutations [27]. The fact that single null mutants do not significantly affect the mutation rates and survival makes the apparent redundancy in the BER pathway self-evident.

**Nucleotide pool sanitization**

The other portion of this pathway involves nucleotide pool sanitization enzymes of the MutT family, which hydrolyse nucleotide-triphosphates to nucleotide-monophosphates (e.g. 8-oxo-dGTP to 8-oxo-dGMP+PPi) [48]. The *M. tuberculosis* genome revealed the presence of four MutT homologues, MutT1, MutT2, MutT3 and MutT4 [46]. However,
Table 1. Summary of proteins involved in DNA repair in mycobacteria

<table>
<thead>
<tr>
<th>Protein</th>
<th>Locus</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Base excision repair</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MutM (Fpg)</td>
<td>Rv2924c</td>
<td>Excises 8-oxoG paired against C, additional activity: AP lyase, dRPase</td>
</tr>
<tr>
<td>Fpg2</td>
<td>Rv0944</td>
<td>Non-functional or no identified activity yet</td>
</tr>
<tr>
<td>MutY</td>
<td>Rv3589</td>
<td>Excises A paired against 8-oxoG, additional activity: AP lyase</td>
</tr>
<tr>
<td>Ung</td>
<td>Rv2976c</td>
<td>Excises Uracil and Uracil derivatives from ss- and dsDNA</td>
</tr>
<tr>
<td>UdgB</td>
<td>Rv1259</td>
<td>Excises Uracil, hypoxanthine and a range of oxidized pyrimidines from dsDNA, thermo-tolerant protein</td>
</tr>
<tr>
<td>TagA</td>
<td>Rv1210</td>
<td>Probably excises alkylated bases (by homology)</td>
</tr>
<tr>
<td>AlkA</td>
<td>Rv1317c</td>
<td>Lacks DNA glycosylase activity, but has methyl-transferase activity</td>
</tr>
<tr>
<td>Nei1</td>
<td>Rv2464c</td>
<td>Excises thymine-glycol and 5,6-dihydrouracil from dsDNA, weak activity for FapyA and FapyG, additional activity: AP lyase</td>
</tr>
<tr>
<td>Nei2</td>
<td>Rv3297</td>
<td>Excises dihydrouracil residues in ss- and dsDNA and has strong AP lyase activity</td>
</tr>
<tr>
<td>Nth</td>
<td>Rv3674c</td>
<td>Excises dihydrouracil (DHU), 5-hydroxyU, 5-hydroxyC and methyl-hydantoin (MeHyd) as well as FapyA, FapyG, and 8-oxoA</td>
</tr>
<tr>
<td><strong>AP endonucleases</strong></td>
<td></td>
<td>AP endonuclease with preferential selectivity for sites against C</td>
</tr>
<tr>
<td>End (Nfo)</td>
<td>Rv0670</td>
<td>DNA-dependent ATPase activity, dimer binds to DNA and directs UvrB binding</td>
</tr>
<tr>
<td>XhrA</td>
<td>Rv0427c</td>
<td>DNA-dependent ATPase activity, dimer binds to DNA and directs UvrB binding</td>
</tr>
<tr>
<td><strong>Nucleotide pool sanitization enzymes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MutT2</td>
<td>Rv1160</td>
<td>Hydrolyses dCTP, 5-methylCTP, 8-oxoGTP</td>
</tr>
<tr>
<td>MutT3</td>
<td>Rv0413</td>
<td>Hydrolyses dATP, not characterized thoroughly</td>
</tr>
<tr>
<td>MutT4</td>
<td>Rv3908</td>
<td>Not characterized thoroughly</td>
</tr>
<tr>
<td>Dut</td>
<td>Rv2697c</td>
<td>Displays both dUTPase and dCTPase activities</td>
</tr>
<tr>
<td>MazG</td>
<td>Rv1021</td>
<td>Hydrolyses 5-OH-dCTP</td>
</tr>
<tr>
<td><strong>Nucleotide excision repair</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UvrA</td>
<td>Rv1638</td>
<td>DNA-dependent ATPase activity, dimer binds to DNA and directs UvrB binding</td>
</tr>
<tr>
<td>UvrB</td>
<td>Rv1633</td>
<td>In complex with UvrA dimer, scans DNA for damage</td>
</tr>
<tr>
<td>UvrC</td>
<td>Rv1420</td>
<td>Nucleotide cleavage activity directed by UvrA,B</td>
</tr>
<tr>
<td>Mfd</td>
<td>Rv1020</td>
<td>Transcription-coupled DNA repair</td>
</tr>
<tr>
<td>Cho</td>
<td>Rv2191</td>
<td>Not characterized yet</td>
</tr>
<tr>
<td>UvrD1</td>
<td>Rv0949</td>
<td>Helicase activity, displacement of DNA cleaved by UvrABC</td>
</tr>
<tr>
<td>UvrD2</td>
<td>Rv3198c</td>
<td>Non-essential helicase activity, yet unidentified roles</td>
</tr>
<tr>
<td><strong>DNA polymerases</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PolA</td>
<td>Rv1629</td>
<td>Canonical DNA polymerase I</td>
</tr>
<tr>
<td>DnaE1</td>
<td>Rv1547</td>
<td>Replicative polymerase</td>
</tr>
<tr>
<td>DnaE2</td>
<td>Rv3570c</td>
<td>Essential for damage-induced mutagenesis, role in trans-lesion synthesis</td>
</tr>
<tr>
<td>DnB1</td>
<td>Rv1537</td>
<td>Essential for damage-induced mutagenesis, role in trans-lesion synthesis</td>
</tr>
<tr>
<td>DnB2</td>
<td>Rv3056</td>
<td>Mutagenic polymerase activity, probable role in trans-lesion synthesis</td>
</tr>
<tr>
<td>PolD1/Prim-PolC</td>
<td>Rv3730c</td>
<td>AEP superfamily polymerase, role in BER along with LigC</td>
</tr>
<tr>
<td>PolD2</td>
<td>Rv0296c</td>
<td>AEP superfamily polymerase, yet unidentified roles</td>
</tr>
<tr>
<td><strong>DNA ligases</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LigA</td>
<td>Rv3104c</td>
<td>Essential NAD+-dependent DNA ligase</td>
</tr>
<tr>
<td>LigB</td>
<td>Rv3062</td>
<td>Yet unidentified roles</td>
</tr>
<tr>
<td>LigC</td>
<td>Rv3731</td>
<td>ATP-dependent ligase, role in BER, backup role in NHEJ</td>
</tr>
<tr>
<td><strong>Ribonucleotide excision</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNaseH1</td>
<td>Rv2228c</td>
<td>Incises a minimum 4–5 ribonucleotide long tract from DNA-RNA duplex</td>
</tr>
<tr>
<td>RNaseH2</td>
<td>Rv2902</td>
<td>RER, able to incise single ribonucleotide from DNA-RNA duplex</td>
</tr>
<tr>
<td><strong>Mismatch repair</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MutSL</td>
<td>–</td>
<td>Absent in mycobacteria</td>
</tr>
<tr>
<td>NucS</td>
<td>Rv1321</td>
<td>Possible role in mismatch repair, to be explored in more depth</td>
</tr>
<tr>
<td><strong>Homologous recombination: end resection and RecA loading</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AdnA</td>
<td>Rv3202c</td>
<td>End resection helicase-nuclease, RecA loading dependent on RecR as a mediator</td>
</tr>
<tr>
<td>AdnB</td>
<td>Rv3201c</td>
<td>End resection helicase-nuclease, RecA loading dependent on RecR as a mediator</td>
</tr>
<tr>
<td>RecF</td>
<td>Rv0003</td>
<td>Binds to ssDNA, accessory role in RecOR pathway</td>
</tr>
</tbody>
</table>
an exact functional homologue of *E. coli* MutT is not present. *E. coli* MutT is a single domain protein, while an additional domain with a yet unidentified role is present in mycobacterial MutT1. *M. tuberculosis* MutT1 hydrolyses 8-oxo-dGTP to 8-oxo-dGDP, which is further processed by an ADPRase (Rv1700) to form 8-oxo-dGMP [30]. Initially, the processing of 8-oxo-dGTP was seen as a two-step process. However, in a recent study in our laboratory, we observed that on increasing the concentration of protein in solution, Ms-MutT1 could directly hydrolyse 8-oxo-dGTP to 8-oxo-dGMP [17]. The exact consequence of the same is yet to be analysed. Additionally, Ms-MutT1 can hydrolyse di-adenosine polyphosphates like Ap4A, Ap5A or Ap6A [35]. This activity has not been demonstrated for *E. coli* MutT. It is known that the *M. tuberculosis mutT1* mutant has increased mutation rates, but at a much lower rate as compared to other bacteria [49].

In another study, MutT2 was found to hydrolyse dCTP, 5-methyl-dCTP and 8-oxo-dGTP directly to their respective mono-phosphate products [50]. This functional redundancy could possibly be the reason for the less severe MutT1 loss in *M. tuberculosis*. *M. smegmatis* MutT2 is also able to complement the loss of MutT in *E. coli* [50]. The sequence of *M. tuberculosis* MutT2 is more similar to *E. coli* Orf135 protein, which shows similar substrate specificity. However, unlike Ec-Orf135, *M. smegmatis* MutT2 cannot hydrolyse 2-OH-dATP (our unpublished results). Additionally, an unrelated pyrophosphorylase, MazG that hydrolyses 5-OH-dCTP is also present in mycobacteria [51, 52]. A mazG null mutant showed attenuated virulence in the mouse model of infection [53]. Together, these enzymes are involved in regulating the pool of mutagenic nucleotides. A debatable point is the hydrolysis of normal undamaged nucleotide dCTP by MutT2. A possible explanation of this activity might be to maintain optimal levels of these nucleotides in the pool. Excessive presence of these nucleotides might affect the efficiency of DNA repair and replication machinery. Yet again, the multiple processing pathways for 8-oxo-dGTP and a broad substrate specificity is evidence of a notable redundancy in this pathogen.

### Uracil repair

Apart from high guanine oxidation rates, the mycobacterial genome is also susceptible to the deamination of cytosine, leading to the formation of uracil. In addition to glycosylases that remove already incorporated uracil from DNA, dUTPase enzymes minimize the probability of uracil

---

**Table 1. cont.**

<table>
<thead>
<tr>
<th>Protein</th>
<th>Locus</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>RecO</td>
<td>Rv2362c</td>
<td>Annealing of ssDNA and interaction with RecR to mediate RecA loading, role in SSA</td>
</tr>
<tr>
<td>RecR</td>
<td>Rv3715c</td>
<td>DNA binding activity</td>
</tr>
<tr>
<td>RecJ</td>
<td>–</td>
<td>Exonuclease in <em>E. coli</em>, absent in mycobacteria</td>
</tr>
<tr>
<td>RecQ</td>
<td>–</td>
<td>Helicase in <em>E. coli</em>, absent in mycobacteria</td>
</tr>
<tr>
<td>RecA</td>
<td>Rv2737c</td>
<td>Catalyses strand exchange</td>
</tr>
<tr>
<td>SSBa</td>
<td>Rv0054</td>
<td>Canonical ssDNA binding protein role in the cell, role in RecA loading</td>
</tr>
<tr>
<td>SSBb</td>
<td>Rv2478c</td>
<td>ssDNA binding, probable role in recombination during stress</td>
</tr>
<tr>
<td>RuvA</td>
<td>Rv2593c</td>
<td>Holliday junction binding protein, binds HJ during branch migration by RuvB</td>
</tr>
<tr>
<td>RuvB</td>
<td>Rv2592c</td>
<td>Holliday junction branch migration helicase</td>
</tr>
<tr>
<td>RecG</td>
<td>Rv2973c</td>
<td>Holliday junction branch migration, among many other roles in the cell</td>
</tr>
<tr>
<td>RuvC</td>
<td>Rv2594c</td>
<td>Holliday junction resolvase</td>
</tr>
<tr>
<td>RuvX</td>
<td>Rv2554c</td>
<td>Holliday junction resolvase</td>
</tr>
<tr>
<td>RecB</td>
<td>Rv0630c</td>
<td>End resection helicase-nuclease complex</td>
</tr>
<tr>
<td>RecC</td>
<td>Rv0631c</td>
<td></td>
</tr>
<tr>
<td>RecD</td>
<td>Rv0629c</td>
<td></td>
</tr>
<tr>
<td>Ku</td>
<td>Rv0937c</td>
<td>DNA bridging activity binds to broken ends</td>
</tr>
<tr>
<td>LigD</td>
<td>Rv0938</td>
<td>ATP-dependent ligase</td>
</tr>
<tr>
<td>RecN</td>
<td>Rv1696</td>
<td>Involved in forming the break repair centre</td>
</tr>
<tr>
<td>RecX</td>
<td>Rv2736c</td>
<td>Controls expression of RecA</td>
</tr>
<tr>
<td>RadA</td>
<td>Rv3585</td>
<td>Potential role in radiation damage repair</td>
</tr>
<tr>
<td>RecG&lt;sub&gt;end&lt;/sub&gt;</td>
<td>Rv2694c</td>
<td>Binds HJ in vitro, potential role in DNA repair</td>
</tr>
<tr>
<td></td>
<td>Rv2119</td>
<td>Not characterized yet</td>
</tr>
</tbody>
</table>

**Homologous recombination: strand exchange**

**Homologous recombination: resolution**

**Single-strand annealing pathway**

**Non-homologous end joining**

**Other proteins**

---

Singh, *Microbiology* 2017;163:1740–1758
incorporation [54]. M. tuberculosis possesses a canonical dUTPase and an additional archeal-type dUTPase, which also deaminates dCTP. However, once incorporated into the genome, uracil is excised by proteins of the uracil DNA glycosylase (UDG) family. M. tuberculosis Ung, a family 1 UDG, can excise uracil from both single- and double-stranded DNA and is essential for full virulence in a murine model of infection [53, 55, 56]. M. tuberculosis Ung was also found to bind a wide range of uracil derivatives known to have inhibitory effects to various extents [18]. However, the functional importance of these interactions is yet to be explored. It may play a role in the feedback control system of this enzyme.

Apart from Ung, M. tuberculosis also contains a thermo-tolerant uracil DNA glycosylase UdgB, that acts only on double-stranded DNA [57]. In addition to uracil, it can hydrolyse a broad range of substrates, including ethenocytosine and hypoxanthine. It was shown that UdgB is regulated by its final products and is likely to act as a backup when uracil levels are unbalanced [58]. Intriguingly, it was also observed that UdgB binds to AP sites [57]. It may play a possible role in shielding AP sites until downstream processing by other BER enzymes. A novel UDG, named UdgX, has recently been identified in M. smegmatis, that binds strongly to the uracil in DNA and marks it for repair by RecA [26]. Although this enzyme is present in many mycobacterial species, it was found to be absent in M. tuberculosis.

Alkylation repair

In E. coli, products of tagA and alkA genes encode 3-methyladenine DNA glycosylase I and II, respectively. TagA is highly specific to 3-methyl purines, while AlkA recognizes a wide range of methylated bases [59]. The ada gene product along with alkA gene product controls adaptive response to
In M. tuberculosis, ada and alkA encode fused proteins and are found within an operon containing ogt [46]. M. tuberculosis mutants lacking this operon are mildly sensitive to alkylating agents but not critical for virulence. No major biochemical work has been reported for this enzyme. Another DNA glycosylase Mpg, a 3-methylpurine DNA glycosylase, was also found to be present in M. tuberculosis [46, 61]. This functional redundancy provided by TagA and Mpg probably forms a backup of an extensive network to repair alkylation damage.

AP endonucleases

The AP sites arising from the action of DNA glycosylases can result in stalling of the replication fork [62]. These lesions are repaired by endonuclease IV (End or Nfo) and exonuclease III (XthA) proteins. In M. tuberculosis, End plays the major role in the processing of AP sites and is more efficient than XthA, contrary to what is observed in E. coli [63]. M. tuberculosis end and xthA deletion mutants have increased sensitivity towards oxidative stress and show attenuation during virulence [53, 64]. Consistent with this, xthA was found to be a pseudogene of M. leprae, a pathogenic mycobacteria with highly reduced genome size [24]. Unlike E. coli AP endonucleases that have no substrate preference, M. tuberculosis AP endonucleases favour processing of AP sites formed against cytosine residues [63, 65]. Thus, the presence of two endonucleases with a preferential selectivity of sites against C, indicate the additional bias of the BER system towards the repair of 8-oxoG to counteract the risks due to high GC content. The gaps resulting due to the action of AP endonucleases are further processed by DNA polymerases and ligases, which are discussed in detail in the following segments.

Nucleotide excision repair

NER is a low specificity repair system, in which the whole damaged nucleotide is excised, rather than just the base. Thus, it serves as an alternative pathway for a wider range of substrates and is essentially conserved across evolution [66, 67]. In most bacteria, repair is initiated by the action of UvABC excinuclease [68]. Recognition of damage happens by a ternary complex of (UvrA)_{2}UvrB in an ATP-dependent manner, following which the endonuclease UvrC is recruited (Fig. 1) [69]. UvrC cleaves the seventh or eighth nucleotide upstream and the fourth or fifth nucleotide downstream of the damage, leading to excision of about 12 or 13 nucleotides. A DNA helicase UvrD disassembles the excised DNA and DNA polymerase (Pol I) fills the gap, followed by the action of DNA ligase. Additionally, transcription-repair coupling factor (Mfd) that recognizes damaged DNA in a stalled transcription unit, recruits a UvrA:B ternary complex to initiate a transcription-coupled repair [70, 71].

Uvr ABC excinuclease

Most mycobacterial genomes contain a copy of all uvr genes [46]. M. tuberculosis uvr genes are up-regulated and are imperative for infection [72, 73]. Only a modest effect on pathogenicity was observed for a uvrA mutant in mouse models. However, uvrB null strains were found to exhibit severe sensitivity towards multiple clastogens and were attenuated for infection in the mice and non-human primate model [74, 75]. Mutations in uvrB have also been linked to increased drug resistance in a strain isolated from a patient [76]. To compare significance of NER and BER, the relative fitness of uvrB null mutant of M. smegmatis was compared to that of ung and fpg null strains. The strain deficient in NER showed relatively higher sensitivity towards a multitude of clastogens, as compared to BER mutants, again indicating a higher importance of this broad substrate recognizing system [77].

A UvrC homologue (Cho) identified in E. coli has high sequence similarity to the N-terminal domain and shows similar activity [78]. A similar conserved hypothetical protein Rv2191 has also been identified in M. tuberculosis. An additional domain similar to the proofreading domain of DNA Pol III is present at the C-terminus of this protein. Expression levels of Rv2191 are elevated upon DNA damage in vitro and in a rabbit model of infection [79]. This might form an alternative mechanism for repair in which exonuclease activity is performed after incision by the same protein.

Transcription repair coupling factor Mfd

Damage to DNA usually leads to stalling of RNA polymerase and hence transcription. Mfd recognizes the damaged DNA at a stalled transcription unit and displaces the machinery for recruitment of UvrABC excinuclease [80]. Unlike its monomeric counterpart in E. coli, Mfd from M. tuberculosis (Rv1020) exists in the population distributed either as a monomer or a hexamer. This probably indicates a different mechanism of action [81]. Expression of the mfd gene was found up-regulated in macrophages as well as the SCID mice model [72, 82].

DNA helicases – UvrD

Superfamily I helicase UvrD unwinds the excised nucleotide region and facilitates disassembly of UvrC post-incision. Although most bacteria carry only one uvrD gene, M. tuberculosis has two putative genes, uvrD1 and uvrD2 [46]. The expression of both these genes is induced by DNA damage in vitro and is also up-regulated in macrophages [79]. Knockout of uvrD1 renders the bacteria sensitive to DNA damaging agents, but is non-essential for survival. A role of UvrD1 in establishing persistence in a mouse model of infection has been ascertained [83]. However, UvrD2 is essential for M. smegmatis survival and is predicted to be essential for M. tuberculosis growth in vitro [84, 85].

Intriguingly, the 3’ to 5’ helicase activity of UvrD1 was found to be stimulated by the interaction of its C-terminus...
with Ku, a component of the NHEJ repair system [86]. While UvrD2 has intact helicase activity, no such interaction with Ku is reported. This is probably a consequence of the differential domain organizations, as UvrD2 has an unusual superfamily II like C-terminal domain [84]. Plasmid expressed mutants of UvrD2 lacking helicase activity were able to permit a deletion of uvrD2 in M. tuberculosis, implying that the essentiality of this gene is not by virtue of its helicase activity [85]. The clear role of UvrD2 is yet to be elucidated. In addition to NER, mycobacterial UvrD1 and UvrA have been implicated to act together to inhibit DNA strand exchange by RecA [74]. Also, UvrD1 and UvrB act together to ensure fidelity of HR in the absence of canonical MMR [87]. This cross-talk between pathways and two non-redundant UvrD proteins provide an additional level of control in DNA damage response.

DNA polymerases

The gaps generated after the action of AP endonucleases and UvrD in BER and NER respectively, are filled by DNA polymerases. Both, NER and BER, pathways converge at this point. Three major DNA polymerases Pol I, Pol II and Pol III are known to be present in prokaryotes. However, mycobacteria lack the components of high fidelity DNA Pol II [46]. Predominantly, these gaps are filled by DNA-dependent A-family polymerase I, Pol A. The active polymerase component of the C-family Pol III is provided by dnaE gene product (a subunit of DNA Pol III). The deficiency of Pol A can be restored by the gene encoding DnaE. The other components of DNA Pol III, i.e. DnaZX, DnaQ and DnaN, required for efficient loading, are present in mycobacteria (reviewed in [88]). Deletion of dnaE1 was found to be lethal and could not be generated, suggestive of it being a replicative polymerase. Consistent with this, it was shown recently that DnaE1 itself encodes an intrinsic proofreading activity within its PH-domain and that canonical proofreading ε exonuclease (DnaQ) is completely dispensable in M. tuberculosis [89]. Nevertheless, the presence of ε exonuclease in mycobacteria is evident of a backup proofreading mechanism.

While E. coli has only one dnaE gene, most mycobacterial genomes contain two C-family DNA polymerase genes dnaE1 and dnaE2 [46]. The deletion of the M. tuberculosis dnaE2 gene results in increased sensitivity to damage and eliminates damage-induced mutagenesis in vitro [90]. Two accessory proteins, ImuA (Rv3395c) and ImuB (Rv3394c), are essential for damage-induced mutagenesis by DnaE2 [91]. This indicated the role of DnaE2 in trans-lesion synthesis (TLS), a procedure where DNA polymerases bypass lesions thereby integrating mutations in the genome [92]. Consequently, DnaE2 has been implicated in the emergence of drug resistance in M. tuberculosis [90]. It is also suggested that up-regulation of dnaE2 during DNA damage, leads to an increase in adaptability by increasing the error-prone synthesis of DNA [93].

Along with this, the M. tuberculosis genome encodes for two Y-family polymerases, DinB1 (dinX) and DinB2 (dinP) [94]. DinB1 was implicated as a DNA-dependent polymerase, while DinB2 was shown to be a low fidelity polymerase with a preference for ribonucleotides [95]. Also, DinB2 is capable of incorporation of oxo-rGTP and 8-oxo-dGMP as well as incorporation of rNTPs against 8-oxodG lesions [96, 97]. Although the mutagenic polymerase activity of DinB2 has been established, the biological role of these proteins is not yet understood clearly. A duplication of the dinB2 gene has been observed in some mycobacterial genomes, including that of M. smegmatis mc²155. DinB2 was noted to aid survival under mycobacteriophage-induced dTTP-limiting conditions [98]. The expression of these genes is not dependent on known mechanisms induced by DNA damage [79, 99]. However, differential expression of dinB1 in pulmonary tuberculosis and dinB2 following exposure to novobiocin, suggest a distinct role of these genes in M. tuberculosis [72, 99].

The M. tuberculosis genome also encodes for three polymerases of the archeo-eukaryotic primase (AEP) superfamily, LigD-Pol/Prim-Pol/D/PolDom (polymerase domain of NHEJ enzyme LigD) and two standalone polymerases PolD1 and PolD2, similar to the LigD-Pol domain [100, 101]. Members of the AEP family are functionally diverse enzymes, capable of a range of enzymatic activities such as template-dependent RNA/DNA polymerase, strand displacement, terminal transferase and gap filling during single-stranded breaks [102]. These polymerases are error-prone and have a notable preference for incorporation of ribonucleotides [103]. The role of Prim-PolD (LigD-Pol) has been established in NHEJ and has been discussed later in detail. PolD1 and PolD2 were thought to be backup polymerases acting in NHEJ, although it is not proven. However, PolD1/Prim-PolC has very recently been identified to work in conjunction with LigC in the BER pathway [37].

DNA ligases

The sealing of nicks resulting from damage or repair is performed by DNA ligases. In most bacteria, this reaction is carried out by a NAD+-dependent DNA ligase, LigA. Mt-LigA function is essential for mycobacterial viability and has been explored as a potential drug target [104, 105]. While E. coli encodes for only one ligase, M. tuberculosis contains three additional ATP-dependent ligases, LigB, LigC and LigD [46]. The genes encoding LigB, LigC and LigD can be deleted without substantial effect on cell growth under laboratory conditions [106]. These ligases have been shown to have a strong propensity for ligation of DNA breaks with a ribonucleotide at the 3' end of the break, which is in strong correlation with the preference of ribonucleotides as substrates by their associated Prim-Pols [107]. Mt-LigD is a multidomain-multifunctional enzyme with modules for ligase (Lig), polymerase (Prim-PolD) and phosphoesterase (PE) domain [101]. Mt-LigD, along with the Ku protein, forms the functional NHEJ enzyme complex (discussed later) [108]. The function of LigB has not been elucidated yet. M. smegmatis mc²155 has two paralogous LigC (LigC1 and LigC2), encoded by adjacent genes. These two
ligases are in close genomic proximity to Prim-PolC (PolD1) and were proposed to act in concert as domains of LigD. LigC1 was able to restore plasmid re-circularization (PolD1) and were proposed to act in concert as domains of rnhA

Loss of 2 and Pol I [112]. In DNA duplex and their replacement with the dNMP. RER is RER usually refers to the removal of single NMPs from excised by NER pathway proteins in the case of RNase H activity loss, however, it is yet to be shown in mycobacteria [118]. RNase H2 enzyme nicks 3’ hydroxyl and 5’ phosphate termini, which is 'nick translated' by Pol I in a classical mechanism (reviewed in [112]). M. tuberculosis RnhB (RNase H type 2, Rv2902) or M. smegmatis RnhB has not been characterized biochemically. However, ΔrnhB M. smegmatis cells presented a profound sensitivity to oxidative stress in stationary phase [119]. Establishing clear roles of RNase enzymes in M. tuberculosis can provide possible insights into the regulation of nucleotide usage by multiple repair enzymes during the stationary phase.

NON-CANONICAL MISMATCH REPAIR

Mismatched bases mostly arise due to error-prone replication and the MMR system detects the incorrect bases in the newly synthesized strand, while discriminating against the parental strand. The MMR pathway plays a critical role in avoiding mutations and also in preventing recombination between non-perfectly identical, i.e. homologous DNA sequences. The absence of recognized MMR homologues MutS and MutL, in mycobacteria, has been a frequent point of discussion considering that the rates and spectra of spontaneous mutations in M. tuberculosis are similar to that in MMR-bearing bacteria [120–122].

In a recent report, a homologue of archaean endonuclease NucS has been identified in mycobacteria [28]. NucS from Thermococcus kodakarensis has a nuclease activity on dsDNA substrates with mismatched bases [123]. M. smegmatis nucS null mutants had high mutation rates and increased homologous recombination rates, which are the hallmarks of MMR inactivation. The increased mutation rates could be complemented by a nucS carrying vector. However, purified M. smegmatis NucS was not able to bind dsDNA or show any significant specific cleavage activity, as shown for archaeal NucS. But the direct increase in drug resistance and mutation rates in NucS null strains, suggest an active mutation avoidance mechanism. A possible disparity between functional mechanisms of bacterial and archaeal NucS cannot be ruled out. Probable involvement of other components or proteins along with NucS, or separately, is also yet to be probed. However, this is a first step towards ending the hunt for MMR in mycobacteria.

This report is in stark contrast to the previous understanding, that the loss of MMR has played a direct role in increasing the genetic diversity, and hence the evolution of M. tuberculosis strains [14]. Even though the study on this system is in its preliminary phase, it opens up a new dimension in understanding the hypermutable strains of M. tuberculosis. It would be interesting to know whether these hypermutable variants have diminished NucS activity due to polymorphisms in nucS genes, as seen with other DNA repair proteins.

DOUBLE-STRAND DNA BREAK (DSB) REPAIR

DSBs can be lethal to dividing cells and thus the repair of DSBs is essential [124]. While E. coli has just one pathway,
Mycobacteria exploit three distinct mechanisms, i.e. HR, NHEJ and SSA for repair of DSBs (illustrated in Fig. 2) [32]. Although HR is the most faithful repair pathway among the three, the prerequisite of a second intact copy as a template limits the pathway to only post-replicative stages of the cell cycle [125]. NHEJ can either be error-free (if the ends are sealed directly) or it can be mutagenic (if the ends are modified by nucleases or polymerases) [126]. Since NHEJ acts as the major pathway in non-replicating cells, it may be of significant relevance in persistence and pathogenesis of *M. tuberculosis*. The SSA mechanism of repair comes into play when the DSB is flanked by repeats on both sides [72].

**Homologous recombination**

HR is a ubiquitous process, with a broad range but similar mechanics of the process in evolution. It requires the presence of an unbroken homologous strand to direct the repair and therefore is limited to post-replicative stages of the cell cycle [124]. The process of recombination can be divided into three stages: (a) *end resection* – processing of broken DNA strands to generate a 3’ single-stranded overhang, (b) *strand exchange* – RecA-mediated exchange of an intact strand from the undamaged strand and broken strand of DNA, followed by DNA synthesis, (c) *resolution* – processing of the Holliday junction, formed by virtue of strand exchange, to yield two repaired DNA duplexes [127]. A review on differential expression of DNA repair proteins during multiple stages of infection suggests the high importance of proteins involved in recombination repair [7].

**End resection in mycobacteria**

In *E. coli*, initiating functions of helicase-nuclease is provided by a complex of RecBCD enzymes. RecBCD resects dsDNA until it encounters an 8 bp recognition site, called Chi, following which it acts as a single-stranded 5’ to 3’ exo-nuclease [128]. RecBCD also assists loading of RecA onto...
the generated 3’ ssDNA overhang coated by single-stranded DNA binding (SSB) proteins. In ΔrecBCD E. coli strains, the function is taken over by exonuclease RecJ and helicase RecQ, followed by RecA loading assisted by RecFOR complex [125]. Although mycobacterial genomes encode homologues of RecBCD, their function differs substantially [32]. The M. smegmatis recBCD null mutant was shown to be insensitive to exposure to UV, suggesting the presence of other resection machinery. The mycobacterial RecBCD complex, an efficient reporter system, developed in the Glickman lab, suggested the involvement of the RecBCD complex in the SSA pathway [32].

A novel heterodimeric helicase-nuclease AdnAB was later implicated in mycobacterial HR [129]. Both adnA and adnB genes are found to be up-regulated in H₂O₂-treated cultures and macrophages [7]. AdnA and AdnB each contain a helicase and a nuclease domain. A similar complex (AddAB) with one helicase and two nuclease domains is observed in B. subtilis [130]. The ΔadnAB mutant of M. smegmatis was rendered sensitive to ionizing radiation, however, the effect was not as severe as the ΔrecA mutant. The clastogen sensitivity of the ΔadnAB ΔrecBCD strain was similar to that of only the ΔadnB strain, suggesting that the residual HR was by means of another pathway [129]. However, deletion of recO along with adnB abolished HR events completely. This suggests the presence of a parallel pathway of HR mediated by RecO [131]. Recently, an accessory role of RecF was also implicated in the RecO pathway [132]. Intriguingly, a ΔrecR mutant completely abolished the HR events, suggesting that RecR provides the necessary mediator activity required for RecA loading [132]. Consistent with this, recR is up-regulated in samples isolated from patients being treated with a wide array of drugs [7]. Consequently, mycobacterial HR may proceed via RecFO-RecR or AdnAB-RecR pathways.

The above data is in agreement with the sequence conservation levels of these proteins among mycobacterial genomes. In our previous study, we found that recBCD and adnAB genes show a high level of variation within the M. tuberculosis complex and the studied range of mycobacterial genomes [23]. RecR is among the most highly conserved proteins in the recombination pathway, next only to RecA and RuvB, consistent with its critical roles [23].

Very recently, a nuclease dead but helicase active AdnAB was found to be sufficient for driving HR in M. smegmatis, in the presence of RecOR. However, inactivation of AdnB helicase activity in the nuclease dead AdnAB, lead to the abolishment of HR [29]. This indicates a probable presence of other nucleases. An alternative pathway in which helicases unwind the DNA duplex and the 3’ end of DNA protected by SSB is directly availed for loading of RecA, was also suggested [29]. The homologues of RecJ and RecQ, required for strand processing in E. coli, are absent in mycobacteria [14]. Intriguingly, M. leprae does not encode for AdnAB or RecBCD, with the former being present as pseudogenes and the latter being absent [23]. However, for this alternative pathway to proceed in the case of M. leprae, an additional helicase must come into play.

Nevertheless, a lookout for unidentified resection nucleases and helicases in mycobacteria is still in progress. In our recent genomic search, we identified a standalone nuclease (Rv2119 in M. tuberculosis) in mycobacterial genomes, including M. leprae, with a cas4 nuclease domain architecture similar to nuclease domains of AdnA and AdnB [23]. The expression of this gene is up-regulated during DNA damaging stress [99]. The function of this nuclease in recombination repair is another aspect that can be explored. Nonetheless, at this moment, a supposition of this protein acting as a backup nuclease cannot be eliminated. The plethora of options present for processing of strands before RecA loading again suggests an elegant way to increase the chances of faithful DSB repair over NHEJ or SSA.

**Strand exchange**

**RecA**

Exchange of a damaged strand with an intact strand is central to the process of recombination repair. A multifunctional protein RecA plays a pivotal role in the repair of these strands [133]. Post-binding to 3’ ssDNA overhang, RecA filaments bind and search for homology in the intact duplex DNA, the mechanism of which is unclear so far. When homology is found, RecA catalyses the exchange of strands, thereby initiating the actual process of recombination [133].

M. tuberculosis recA shows an unusual gene structure, with the presence of an intein in the translated ORF [134]. The intein, a homing endonuclease, is later spliced to generate RecA protein [135, 136]. However, it has been shown that the presence of unprocessed intein does not affect the RecA function. In a previous structural examination of RecA from our lab, we found that mycobacterial RecA have a substantially negatively charged surface, while no such patches exist on E. coli RecA filaments [137–139]. This may have direct implications on the RecA-DNA interactions. Among other differences is the ability of the filament to further aggregate into bundles in solution. EcRecA aggregates in bundles while the Ms- and MtRecA do not [140]. The repercussions of these differences in the mechanism are still unexplored.

The expression of M. tuberculosis RecA is controlled by two different promoters, both of which are DNA damage-inducible. One of the promoters is regulated by LexA, while the other is independent of both RecA and LexA [141, 142]. In turn, active RecA filaments induce the autocalytic activity of LexA, which further regulates the 'SOS response' to DNA damage [143, 144]. The clear role of the two-promoter system is not known but possibly helps in scrutinizing the activity of RecA at multiple checkpoints. Recently, NucS was implicated in checking strand exchange in the case of micro-homology, i.e. homologous recombination [28]. In M. tuberculosis, UvrD1 and UvrB are also known to interact with RecA and probably function in dispersing RecA filaments in the case of homologous sequences [87].
Single stranded DNA binding protein
SSBs bind and protect the 3' ssDNA overhang generated by resection enzymes. It also assists RecA polymerization onto ssDNA, by helping dissolution of the secondary structures in the DNA [127, 145]. Unlike E. coli, mycobacterial SSBs are known to physically interact with their cognate RecA [146]. Also, mycobacterial SSBs are more stable than E. coli SSB, owing to a 'clamp' like inter-subunit strand exchange [147].

While E. coli has only one ssb (hereafter referred to as ssbA) gene, mycobacterial genomes consist of a second paralogous ssbB gene. The sequence of SSBB appears to be less conserved than SSBa in mycobacteria, while both are structurally very similar [148]. A recent study on SSBB from M. smegmatis in our lab, showed that the expression levels of the ssbB gene increased by approximately two and seven-fold in UV and hypoxic stress respectively, while simultaneously the levels of ssbA expression declined [149]. The expression data are consistent with microarray data present for MtSSBB in the tuberculosis database (TBDB) [99, 150]. A direct physical interaction of M. smegmatis SSBB and its cognate RecA in solution, mediated by the C-terminal tail of SSBB was also established. These results indicate the possible role of SSBB in assisting DNA repair during the persistence stage of M. tuberculosis. The ssbA gene in M. avium 104 has a frame-shift mutation that renders it non-functional. A probability that SSBB might take over the basic functions of SSBA cannot be ignored. A clear role for SSBB is yet to be established.

Resolution
Strand exchange leads to the formation of a four-stranded 'Holliday junction' (HJ) at both ends of the crossover [151, 152]. The length of the hetero-duplex DNA is extended by branch migration, involving the movement of the junction along DNA, mediated by the RuvAB complex or RecG [151]. The joint DNA molecules are then resolved and eliminated by the HJ-specific endonuclease/resolvase RuvC [153]. RuvC has been implicated in the resolution of intermediate formed by RuvAB, while the search for a nuclease downstream to the action of RecG is still ongoing [154]. In E. coli, an additional HJ resolvase, RusA can act along with RecG or RuvAB to resolve HJ in vitro [155]. However, the rusA gene is not normally expressed in E. coli and its deletion is not detrimental in the background of ruv mutants. Although the role of RecG in conjugal recombination is observed, its role in intra-chromosomal recombination is not clear [154].

Homologues of RuvABC and RecG are found in all mycobacteria, including M. leprae [23]. Similar enzymatic roles of mycobacterial RuvABC and RecG have been established [156–158]. The expression levels of ruvA and ruvC are up-regulated upon UV damage in M. tuberculosis. recG is up-regulated in macrophages in the mouse model of infection, as well as in clinical lung samples [7]. While no RusA homologue is found in mycobacteria, all mycobacteria encode for another HJ nuclease RuvX [46]. RuvX is formed by dimerization of a conserved family of YqgF proteins and is found widely in the bacterial kingdom [159]. Mycobacterial RuvX was found to efficiently catalyse HJ resolution in vitro [33]. However, no HJ resolution activity has been observed for the monomeric E. coli YqgF protein and has instead been implicated in RNA metabolism [160, 161]. M. tuberculosis yqgf/ruvX, along with ruvC, was found to be up-regulated upon UV and MMS damage [33]. Hence, speculation that RuvX may be a backup nuclease cannot be considered as expression levels of both ruvC and yqgf rise in a concerted manner. However, whether or not the mycobacterial Yqgf acts in tandem with RuvAB or if it is the 'not yet identified downstream HJ nuclease' to RecG, is to be established. Fig. 1 summarizes the developments and questions in DSB repair.

The 'wedge' domain in RecG is involved in binding to branched DNA substrates, such as HJ, and occurs along with a C-terminal helicase domain [156]. We identified a standalone wedge domain protein, named RecGwed (Rv2694c), in the previous genomic search [23]. Purified M. smegmatis RecGwed is able to bind to three-way junctions, replication forks and HJ in solution (our unpublished results). This gene is also up-regulated during DNA damage and during respiratory inhibition, as observed in the TBDB. A thorough analysis of functions and interaction partners of RuvX and RecGwed in the cell may yield new insights into the resolution mechanisms in mycobacteria.

Non homologous end joining pathway
In contrast to HR, NHEJ does not require homology and re-anneals the two damaged ends directly. A modest amount of end-processing and addition or removal of a few nucleotides is required to render the ends suitable for ligation. The process is mediated by an end-binding and end-bridging DNA binding protein, Ku and an ATP-dependent DNA ligase [162]. Bacterial orthologues for NHEJ were first identified by the Doherty group [163, 164], who then confirmed the hypothesis with the characterization of NHEJ components of B. subtilis in vivo [165]. Subsequently, M. tuberculosis Ku and ligase LigD were first shown to seal DNA ends in vitro [108]. Mycobacterial NHEJ is the most studied procaryotic NHEJ system (reviewed in [12, 166, 167]).

The Ku protein binds to DNA and is essential for the recruitment of LigD. As discussed earlier, Mt-LigD has three domains: ligase (Lig), polymerase (Pol) and phosphoesterase (PE). The Pol and PE domain together are involved in end modification before Lig domain action can seal the ends [168]. LigD-Pol/Prim-PoD domain can add templated and non-templated nucleotides to the repaired ends [169, 170]. The LigD-Pol/Prim-PoD display a strong preference for ribonucleotides and LigD-Lig has a strong propensity for ligation of nicks with ribonucleotide at 3' position [103, 107]. Together, these domains lead to the incorporation of patches of ribonucleotides forming a DNA–RNA hybrid. It was shown that the LigD-PE domain is required for ribonucleotide resection in a mesophlic archaeon M. paludicola. It also helps in the removal of the unnecessary/extra NMPs and
helps the displaced DNA to realign with the template [171, 172]. A similar role for mycobacterial LigD-PE is yet to be ascertained. NHEJ is the prominent pathway during stationary phase, where pools of dNTPs are depleted and rNTPs are present in abundance [173]. However, the exact role for the preference of ribonucleotides is not known. The ribonucleotides incorporated in the DNA are probably processed by the RNase enzymes/PolA as discussed previously. However, this role is yet to be established.

Homologues of Ku were also identified in genomes of two mycobacteriophages, Omega and Corndog. Phage and mycobacterial Ku homologues are similar in sequences, indicating a probable lateral gene transfer event. Intriguingly, these Ku proteins have an ability to recruit mycobacterial LigD to form a functional NHEJ complex. The NHEJ activity is necessary for circularization of the genome, to support the rolling circle replication model for making progeny and maintaining infectivity [174].

Deletion of Ku and LigD leads to the abolishment of NHEJ in vivo [109]. The absence of NHEJ sensitized M. smegmatis to desiccation during the stationary phase of growth, suggesting the predominant role of NHEJ in DSB repair during stationary phase [173]. It was shown that the loss of LigD-ligase activity did not confer serious defects in NHEJ, and LigC can supply the activity [109, 175]. However, the LigD-Pol inactive and ΔpolD1ΔpolD2 strain still retained the NHEJ fidelity [100], suggesting the possibility of yet unidentified polymerases involved in the pathway. Exploratory studies for the phenotypes of a cell lacking NHEJ components will be required, to provide clear insights into the regulation of LigC and LigD in oxidative damage and NHEJ. Considering that NHEJ may be the predominant choice for repair of DSBs during the persistence stage of M. tuberculosis, makes it an important area for future research.

Single strand annealing pathway

SSA is defined by bi-directional resection of strands from a DSB flanked by repeats, such that complementary single strands are revealed. Annealing of complementary strands, followed by the action of ligase, leads to repair with the loss of the segment between the repeats. This process is independent of RecA and there is no requirement for the presence of a second intact strand. While E. coli RecBCD has been implicated in HR, mycobacterial RecBCD was shown to mediate SSA [58]. Recently, RecO was also shown to be essential for the mycobacterial SSA pathway. RecO can mediate faster annealing of single strands in vitro [27]. However, the exact function of RecO, along with the mechanics of SSA, is yet to be established.

Regulation of pathways

Not much information about regulating the choice of pathways in a replicative stage of mycobacteria is known. SSA and NHEJ are the only two choices during pre-replicative stages of cell cycle, with both being error-prone. However, the choice between the three pathways during post-replicative state is unclear. Mammalian Ku homologues are known to first reach the site of DSB and bind to DNA ends. Since DNA ends are not available for resection, HR cannot proceed from there [176]. Very recently, it was shown that phosphorylation of Ku leads to its dissociation from DNA ends and is involved in regulating the choice of the pathway in mammalian cells [177]. Phosphorylation of tyrosine and arginine in SSB of B. subtilis has been observed [178, 179]. Phosphorylation of many proteins in M. tuberculosis has also been reported [180, 181]. A possible role of post-translational modifications in the regulation of complex DNA repair mechanisms cannot be eliminated. A critical examination of mycobacterial proteins for post-translational modification and interacting partners will be required to understand the choice of pathways a cell makes during DSB damage. This is a major area of research which is yet to be explored.

OTHER PROTEINS

RecN

In bacterial systems, RecN is known to first reach the site of DNA damage and form a discrete focus on nucleoids called the ‘repair centre’ (RC). Other proteins involved in processing and recombination are later assembled at the RC in a distinct sequential order [182]. In yeast, a single nuclear RC can accommodate multiple DSBs, and once formed can accommodate more DSBs later [183]. However, this could be precarious as the concurrence of DNA ends from multiple breaks can lead to an increased likelihood of chromosomal translocation. Thus, this concept remains laden with multiple unanswered questions. Even though RecN is conserved in all mycobacteria, to our current knowledge, no report on the functional characterization of mycobacterial RecN is available.

RecX

RecX is a small regulatory protein that negatively modulates the expression of RecA, the protein central to recombination events. RecX was first shown to down-regulate RecA in Pseudomonas aeruginosa, where it was required for the normal growth in the RecA over-expression strain [184]. Subsequently, it was shown to regulate RecA in M. smegmatis [185]. RecX inhibits the co-protease, ATPase and recombination activity of RecA by binding to the ends of growing RecA filament and leading to net congestion [186, 187]. The recX gene is found downstream of the recA gene in E. coli, while M. tuberculosis recX ORF overlaps that of recA [188–190].

RadA

In E. coli, RadA has a role in radiation damage resistance and is hence so named. Mutations in the RadA gene leads to increased sensitivity to DNA damaging agents [191]. This protein shares similarity to RecA in its middle region, to Lon protease at the C-terminus region and has a Zing finger motif at N-terminus. These features are enough for a role similar to RecA. RadA was suspected and later found to be involved in recombination repair in E. coli [192]. These
bacterial RadA proteins are different from the archaean RadA protein which catalyse branch migration, as RecA is absent. Recently, RadA was proposed to stabilize the strand invasion products and help in recombination events [193]. In *B. subtilis* and *S. pneumoniae*, RadA was shown to be involved in the integration of donor DNA in the nucleoprotein filament [194, 195]. *M. tuberculosis* RadA was found to be up-regulated under hydrogen-peroxide or mitomycin C stress and also in the macrophages of mouse models [7]. However, to our understanding, no functional characterization of this protein has been reported in mycobacteria.

**PERSPECTIVES**

It is clear that *M. tuberculosis* comprises some distinct DNA repair pathways, that are not yet fully understood. The high polymorphism in DNA repair genes probably provides this organism with high adaptability to the incessant stress of antibiotic treatment. Presence of a plethora of error-prone mechanisms, such as DnaE2, SSA and NHEJ among others, may predispose this organisms’ ability to adapt resistance to antibiotics. As observed, multiple backup systems for repair of DNA reflect the robustness and strong capability of this bacteria to proliferate even in harsh conditions. A summary of proteins is given in Table 1.

Many of the DNA repair genes discussed above are essential for *in vivo* growth while knocking out some of these genes leads to attenuation of infection in model organisms. A lot of questions regarding the nature and function of many proteins, as well as involvement of yet unidentified proteins, remain unanswered. Some of these questions are listed in Table 2. Now that we understand that many of the above-discussed proteins are important during persistence, a better understanding of their function is essential to develop effective therapeutic interventions.

### Table 2. Some of the questions left to be addressed in future studies

<table>
<thead>
<tr>
<th>BER</th>
<th>Question still left to be addressed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oxidation</td>
<td>What is the preference of enzymes, during a damage (such as 8-oxoG) that can be repaired by multiple enzymes?</td>
</tr>
<tr>
<td></td>
<td>How is the high polymorphism observed in the genes, like <em>mutT</em> genes, linked to the increased adaptability to antibiotic stress?</td>
</tr>
<tr>
<td></td>
<td>What is the significance of having a plethora of redundant enzymes for oxidative damage?</td>
</tr>
<tr>
<td></td>
<td>What is the role of the second domain in MutT1 enzyme? What cellular role do other MutT proteins play?</td>
</tr>
<tr>
<td>Uracil</td>
<td>What is the significance of UdgB binding to AP sites?</td>
</tr>
<tr>
<td></td>
<td>What is the cellular role of bi-functional archaean-type dUTPase?</td>
</tr>
<tr>
<td>Alkylation</td>
<td>What drives the interplay of enzymes in this extensive network?</td>
</tr>
<tr>
<td>NER</td>
<td>Does Rv2191 play an active role in NER?</td>
</tr>
<tr>
<td></td>
<td>What role of UvrD2, other than helicase, makes it essential for the cell? What are the interacting partners?</td>
</tr>
<tr>
<td></td>
<td>Can UvrD1 serve as a target for therapy during the persistence stage?</td>
</tr>
<tr>
<td></td>
<td>What is the difference in the mechanisms of monomeric and hexameric forms of Mfd?</td>
</tr>
<tr>
<td>Polymerases</td>
<td>Can ImuA’-ImuB/DnaE2 system be targeted to disarm the induced mutagenesis machinery and thus serve as a therapeutic target?</td>
</tr>
<tr>
<td></td>
<td>What are the cellular roles of DinB polymerases?</td>
</tr>
<tr>
<td></td>
<td>What is the role of Prim-Pol/PolD2? To establish a clear role of Prim-PolC/PolD1 play in BER and NHEJ. What other polymerases are present?</td>
</tr>
<tr>
<td>MMR</td>
<td>To elaborate the role of NucS in MMR and find potential downstream partners of this pathway</td>
</tr>
<tr>
<td></td>
<td>What role do UvrD1 and UvrB play, if any?</td>
</tr>
<tr>
<td></td>
<td>Does any correlation exist between the activity of NucS and hyper-mutable strains?</td>
</tr>
<tr>
<td>RER</td>
<td>Characterization of <em>MtR</em>Nase2</td>
</tr>
<tr>
<td></td>
<td>Probing the essentiality of RNase enzymes</td>
</tr>
<tr>
<td>DSB</td>
<td>How are the three pathways regulated inside the cell?</td>
</tr>
<tr>
<td>HR</td>
<td>Which mechanism of RecA loading happens in the cell? Whether end resection is a necessity or a choice?</td>
</tr>
<tr>
<td></td>
<td>Which other helicases or nucleases are involved?</td>
</tr>
<tr>
<td></td>
<td>How is the labour distributed between YggF/RuvX fit in?</td>
</tr>
<tr>
<td></td>
<td>Where does the role of YggF/RuvX fit in?</td>
</tr>
</tbody>
</table>
|              | Can Rv2119, SSBb, RadA and RecG 
mod play a role in recombination?                                                                                                                                                                      |
| NHEJ         | Can NHEJ be targeted for therapy against the latent mycobacterial population?                                                                                                                                                         |
|              | How is the activity of LigD-PE domain modulated?                                                                                                                                                                                      |
|              | What is the role of LigD in oxidation damage?                                                                                                                                                                                         |
|              | What happens to the ribonucleotides incorporated into the genome by Prim-Pols? Are they excised by PolA/RNaseH enzymes or left as a flag to mark repaired site?                                                                             |
|              | What are the other unknown polymerases involved in the pathway?                                                                                                                                                                       |
|              | What is the exact role of LigB *in vivo*?                                                                                                                                                                                             |
| SSA          | How is the nuclease activity of RecBCD regulated? What causes termination of nuclease activity upon reaching the ‘repeat’ sequence?                                                                                                     |
|              | What is the role of RecO?                                                                                                                                                                                                             |

---

**Singh, Microbiology 2017;163:1740–1758**
knowledge of the mycobacterial DNA repair systems may allow us to find new targets for therapy [196]. This concise review presents valuable and essential details of DNA repair machinery in mycobacteria. It will perhaps help in providing interesting areas for future research in the characterization of this notorious pathogen.

Funding information
The author received no specific grant from any funding agency.

Acknowledgements
The author would like to thank Professor Aidan Doherty for providing early access to his manuscript. The author would also like to thank Professor M. Vijayan, Dr Sunanda M. Williams and Anju Paul for their inputs on the manuscript. A.S. is a senior CSIR research fellow in Professor M. Vijayan’s laboratory.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References
17. Arif SM, Patil AG, Varshney U, Vijayan M. Biochemical and structural studies of Mycobacterium smegmatis MutT1, a santi-

21. Prabu JR, Manjunath GP, Chandra NR, Muniyappa K, Vijayan M. Functionally important mutations in RecA molecules and fila-
26. Sang PB, Srinath T, Patil AG, Woo EJ, Varshney U. A unique uracil-DNA binding protein of the uracil DNA glycosylase super-
27. Hassim F, Papadopoulos AO, Kana BD, Gordhan BG. A combina-
torial role for MutY and Fpg DNA glycosylases in mutation avoidance in Mycobacterium smegmatis. Mutat Res 2015;779:24–32.
30. Patil AG, Sang PB, Govindan A, Varshney U. Mycobacterium tuberculosis MutT1 (Rv2985) and ADPRase (Rv1700) proteins constitute a two-stage mechanism of 8-oxo-dGTP and 8-oxo-
31. Bhattacharji H, Gupta R, Glickman MS. DNA ligase C1 mediates the LigD-independent nonhomologous end-joining pathway of Myco-
32. Gupta R, Barkan D, Redelman-Sidi G, Shuman S, Glickman MS. Mycobacteria exploit three genetically distinct DNA double-


68. Puri RV, Singh N, Gupta RK, Tyagi AK. Endonuclease IV Is the major apurinic/apyrimidinic endonuclease in Mycobacterium tuberculosis and is important for protection against oxidative damage. PLoS One 2013;8:e71535.


Five reasons to publish your next article with a Microbiology Society journal

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as ‘excellent’ or ‘very good’.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.