Nucleotide Metabolism and DNA Replication

DIGBY F. WARNER, JOANNA C. EVANS, and VALERIE MIZRAHI
MRC/NHLS/UCT Molecular Mycobacteriology Research Unit, DST/NRF Center of Excellence for Biomedical TB Research, Institute of Infectious Disease and Molecular Medicine, and Division of Medical Microbiology, University of Cape Town, P/Bag X3 Rondebosch 7700, South Africa

ABSTRACT The development and application of a highly versatile suite of tools for mycobacterial genetics, coupled with widespread use of “omics” approaches to elucidate the structure, function, and regulation of mycobacterial proteins, has led to spectacular advances in our understanding of the metabolism and physiology of mycobacteria. In this article, we provide an update on nucleotide metabolism and DNA replication in mycobacteria, highlighting key findings from the past 10 to 15 years. In the first section, we focus on nucleotide metabolism, ranging from the biosynthesis, salvage, and interconversion of purine and pyrimidine ribonucleotides to the formation of deoxyribonucleotides. The second part of the article is devoted to DNA replication, with a focus on replication initiation and elongation, as well as DNA unwinding. We provide an overview of replication fidelity and mutation rates in mycobacteria and summarize evidence suggesting that DNA replication occurs during states of low metabolic activity, and conclude by suggesting directions for future research to address key outstanding questions. Although this article focuses primarily on observations from Mycobacterium tuberculosis, it is interspersed, where appropriate, with insights from, and comparisons with, other mycobacterial species as well as better characterized bacterial models such as Escherichia coli. Finally, a common theme underlying almost all studies of mycobacterial metabolism is the potential to identify and validate functions or pathways that can be exploited for tuberculosis drug discovery. In this context, we have specifically highlighted those processes in mycobacterial DNA replication that might satisfy this critical requirement.

The first edition of Molecular Genetics of Mycobacteria (1) was published very shortly after the release of the complete genome sequence of Mycobacterium tuberculosis H37Rv (2). Armed with that resource, we searched for genes that might be involved in mycobacterial nucleotide metabolism and DNA replication (3). Our analysis at the time relied entirely on the homology-based identification of genes that had been discovered and characterized in other organisms; however, by confirming the presence, or suggesting the absence, in M. tuberculosis of homologs of genes of known function, it provided a useful framework for subsequent studies of the reactions and pathways underlying nucleotide metabolism and DNA replication in this major human pathogen. At that stage, the field of mycobacterial genetics was in its infancy, and little was known about the function of individual mycobacterial genes and their encoded proteins in these or other metabolic pathways. However, over the past 13 years, spectacular technical advances have been made that have had a massive impact on the broader field of general bacteriology and, more importantly in the context of this book, have driven the postgenomic revolution in our understanding of fundamental mycobacterial physiology and metabolism. The development, in particular, of a powerful toolkit for random, targeted, and conditional mutagenesis of mycobacterial genomes has allowed gene function to be probed under a variety of conditions. This has provided crucial insights into the reactions and pathways that govern nucleotide metabolism and DNA replication and has paved the way for the identification of candidate drug targets.
of conditions. In turn, this has enabled the compilation of catalogs of genes (conditionally) essential for mycobacterial growth and/or survival, while providing new insights into the biology of mycobacteria.

More recently, genetic advances have been matched by parallel developments in the use of “omics” approaches to elucidate the structure, function, and regulation of mycobacterial proteins. Therefore, in this article, we provide an update of the genetics of nucleotide metabolism and DNA replication in mycobacteria, highlighting key findings from the past decade or so. As far as possible, section headings from the first edition have been retained in order to enable direct assessment of progress in each area—and to highlight those aspects that have received little attention in the intervening period. As with the previous version, this article is focused primarily on observations in M. tuberculosis; however, comparisons with, and insights from, other mycobacterial species as well as better characterized bacterial models such as Escherichia coli are included where appropriate. Finally, a common theme underlying almost all investigations of mycobacterial metabolic function is the potential to identify, or validate, functions or pathways that can be exploited for tuberculosis (TB) drug development; for this reason, we have attempted to highlight processes in mycobacterial DNA replication that might satisfy this criterion.

**BIOSYNTHESIS OF dNTPs**

**Purine Ribonucleotide Synthesis**

The first half of the de novo purine biosynthetic pathway in mycobacteria involves the generation of an aminoimidazole moiety attached to a ribose. In the second half of the pathway, the 4-C and 5-C atoms of the imidazole are modified with the appropriate substituents, leading to the final cyclization reaction that generates inosine 5’-monophosphate (IMP) (4). The first step in the synthesis of purines involves the transfer of the β,γ-diphosphoryl moiety of ATP to the C1-hydroxyl group of α-d-ribose-5-phosphate by a class I ribose-phosphate pyrophosphokinase (PRPPase) (prsA; Rv1017c) to yield 5-phospho-α-d-ribose 1-diphosphate (PRPP) (5, 6). In addition to functioning as a central metabolite in both de novo and salvage nucleoside metabolism, PRPP is also required for the biosynthesis of NAD, NADP, histidine, and tryptophan (7, 8). Furthermore, members of the Corynebacteriaceae family, which includes mycobacteria, have evolved a mechanism for utilizing PRPP in the biosynthesis of cell wall arabinogalactan (9). As the sole PRPPase in M. tuberculosis, PrsA is responsible for the intracellular provision of PRPP and displays a significantly higher specific activity than other bacterial PRPPases (6). While most class I PRPPases are allosterically inhibited by purine nucleosides, the M. tuberculosis enzyme is only inhibited by ADP, with GDP displaying weak inhibition (10). It is likely that the less stringent repression of PRPPase activity in M. tuberculosis reflects the crucial role that this enzyme plays in the provision of PRPP for use in multiple metabolic pathways.

Following the generation of PRPP, the amidophosphoribosyltransferase, PurF (purF; Rv0808), catalyzes the first committed step in purine biosynthesis by replacing the pyrophosphate group of PRPP with the amide nitrogen of L-glutamine, yielding 5-phospho-β-d-ribosyl-amine and L-glutamate (11). Interestingly, loss of purF function results in a transient loss in culturability of M. tuberculosis smegmatis in stationary phase—a phenotype that has yet to be fully explained (11). In an ATP-dependent reaction, phosphoribosylamine-glycine ligase (purD; Rv0772) then ligates glycine to the amino group of 5-phospho-β-d-ribosyl-amine, producing 5-phosphoribosylglycinamide (GAR), and thereby providing atoms C-4, C-5, and N-7 of the purine base. The formyl group from N10-formyltetrahydrofolate (fTHF) is transferred to GAR by 5-phosphoribosylglycinamide formyltransferase (purN; Rv0956) to produce N-formyl glycinamide ribonucleotide (fGAR) and tetrahydrofolate, thereby providing the C-8 atom (12). The structural modification of M. tuberculosis PurN enzyme revealed distinct features that could potentially be exploited for drug discovery in other organisms (12). In M. tuberculosis, this step can also be performed by a second ATP-dependent phosphoribosylglycinamide formyltransferase II (purT; Rv0389), which uses formate and Mg2+ to form a formyl phosphate intermediate before transferring the formyl group to GAR. In another ATP-driven reaction, phosphoribosylformylglycinamide synthase II (purL; Rv0803) catalyzes the addition of an amide group derived from glycine to the C-5 ring position to form fGAR and glutamate. The presence of a phosphoribosylformylglycinamide synthase I (purQ; Rv0788) homolog in M. tuberculosis suggests that, as in other Gram-positive bacteria, including M. leprae (13), PurL probably requires two additional products, PurQ and PurS, for activity (14). Under such circumstances, the PurL domain possesses the phosphoribosylformylglycinamide synthase activity and PurQ acts as the glutaminase (15, 16). The M. tuberculosis PurS homolog is yet to be identified but may be the conserved hypothetical protein encoded by Rv0708A, located upstream of purQ in an operon (2).
The fifth step in the pathway involves the cyclization of fGAR to form the five-membered ring of the purine base, and it is carried out by the ATP-dependent 5′-phosphoribosyl-5-aminoimidazole synthase (purM; Rv0809). The addition of C-6 from bicarbonate requires the coordinated activity of two enzymes: first, the phosphoribosylaminomimidazole carboxylase ATPase subunit (purK; Rv3276c) catalyzes the ATP-dependent ligation of bicarbonate and the N-5 amino group of aminomidazole ribonucleotide, yielding N5-carboxyaminoimidazole ribonucleotide (CAIR); CAIR is then converted to carboxyaminoimidazole ribonucleotide by phosphoribosylaminomimidazole carboxylase (purE; Rv3275c), with the carboxylate carbon becoming C-6 of the purine ring and one of the carboxylate oxygen becoming atom O-6. Phosphoribosylaminomimidazole-succinocarboxamide synthase (purC; Rv0780) catalyzes the ATP-dependent conversion of 5′-phosphoribosyl-5-aminoimidazole-carboxylic acid to 5′-phosphoribosyl-4-(N-succino-carboxamid)-5-aminoimidazole by ligation of the carboxylate group of carboxyaminoimidazole ribonucleotide to the amino group of aspartate, thereby providing the N-1 atom of the final purine base. Fumarate is then released from N-1 through the activity of adenylosuccinate lyase (purB; Rv0777). The bifunctional enzyme, 5-aminoimidazole-4-carboxamide ribonucleotide transformylase/inosine monophosphate cyclohydrolase (purH; Rv0957) catalyzes the final steps in the pathway to IMP, which is ultimately converted to AMP or GMP (4). Structural studies of M. tuberculosis PurH revealed a bound nucleotide, 4-carboxy-5-aminoimidazole ribonucleotide, in the cyclohydrolase active site, suggesting that it may be a cyclohydrolase inhibitor (4). Interestingly, neither M. tuberculosis nor M. leprae contains a homolog of the purine repressor, PurR (2, 13). The absence of PUR box-like sequences in the region upstream of purB in M. tuberculosis suggests that PurB may be subject only to allosteric regulation by purine pools.

IMP can be converted into AMP by adenylosuccinate synthase (purA; Rv0357c). Fumarate is then released from adenylosuccinate through β-elimination by adenylosuccinate lyase (purB; Rv0777) (17). Adenylate kinase (adk; Rv0733) catalyzes the reversible, Mg2+-dependent transfer of a terminal phosphate from ATP/GTP to AMP to form ADP (18), which is then converted to ATP by nucleoside diphosphate kinase (ndkA; Rv2445c) (19). M. tuberculosis NdkA has both nucleoside monophosphate kinase activity, suggesting that this enzyme may potentially be involved in RNA and DNA metabolism in addition to nucleotide metabolism (18). The conversion of IMP to GMP begins with the NAD+-dependent oxidation of IMP to xanthosine 5′-monophosphate by IMP dehydrogenase (IMPDH, GuaB). Although M. tuberculosis possesses three guaB homologs, only guaB2 (Rv3411c) was found to encode a functional IMPDH enzyme when expressed in E. coli (20). This enzyme was susceptible to inhibition by diphenyl urea-based derivatives. These compounds displayed potent antimycobacterial activity that was diminished by overexpression of all three guaB genes, suggesting that IMPDH is the target (20). GMP synthase (guaA; Rv3396c) then catalyzes the addition of an amide nitrogen from L-glutamine to yield GMP, which is converted to GDP by guanylate kinase (gmk; Rv1389) (21) and to GTP by NdkA (Rv2445c) (19). The domain structure of M. tuberculosis Gmk revealed differences in domain dynamics and GMP binding of relevance to the design of specific inhibitors (21). In addition to its role in purine nucleoside biosynthesis, M. tuberculosis NdkA has been shown to localize within the nucleus of mammalian cells and cause superoxide radical-mediated DNA cleavage (22, 23). This protein also appears to play a role in inhibiting phagosome biogenesis processes (24), possibly due to its ability to sequester the extracellular ATP that accumulates at the site of inflammation as part of the host cell defense mechanism (25).

The stringent response

The stringent response is a broadly conserved mechanism that operates in bacteria in response to nutritional stress (26), and its role in M. tuberculosis pathogenesis has been the subject of intense investigation (27, 28). It is mediated by (p)ppGpp, which is produced by the stringent response regulator, designated RelMTb in M. tuberculosis. RelMTb (Rv2583c) (29, 30) is a bifunctional enzyme possessing both RelA-like ATP:GTP/GDP/ITP 3′-pyrophosphoryltransferase and SpoT-like Mn2+-dependent (p)ppGpp 3′-pyrophosphohydrolase activities on a single polypeptide (31, 32). RelMTb is critical for the establishment and maintenance of chronic M. tuberculosis infection in both mice (28) and guinea pigs (33), and functional inactivation of this protein is associated with a significant alteration in the M. tuberculosis transcriptome, with affected genes including those associated with cell wall synthesis and immunogenicity (28, 29). In M. tuberculosis, expression of relMTb is regulated by SigE, which, in turn, is regulated by the PPK1-regulated MprAB two-component system (34). Populations of actively growing M. smegmatis cells have been found to display a bimodal distribution of low and high rel expression levels, with high expression proposed to confer an enhanced ability to respond to...
stress (35). In *E. coli*, increased concentrations of (p)ppGpp mediate large-scale transcriptional changes by directly altering the stability of the RNA polymerase (RNAP) complex at regulated promoters. Together with the general transcription factor DksA, which interacts directly with RNAP in *E. coli*, (p)ppGpp redirects the RNAP to exert its effects on the expression of certain genes (36). While there are no homologs of DksA in mycobacteria, the transcriptional regulator CarD (Rv3583c) may perform an analogous function, because (p)ppGpp is ineffective at inducing stringent control in the absence of CarD, which forms a complex with the RNAP at rRNA and ribosomal protein loci (37).

Recently, a novel, (p)ppGpp synthetase was identified in *M. smegmatis* on a bifunctional protein (MSMEG_5849) that also possesses the ability to hydrolyze DNA: RNA duplexes through the activity of an RNase HIII catalytic domain (38). As discussed below (“Elongation and Termination”), this represents another example of the fusion of an RNase H domain to a separate catalytic function in mycobacteria (*M. tuberculosis* contains the Rv2228c-encoded RNase HI-CobC protein; see Table 2, below) and, in the case of *M. smegmatis*, implies a role for (p)ppGpp in coordinating DNA replication with the stringent response via inhibition of the major replisome component, DnaG primase.

**Pyrimidine Ribonucleotide Synthesis**

The *de novo* synthesis of pyrimidine nucleotides is carried out in six enzymatic steps that culminate in the formation of UMP. In *M. tuberculosis*, the genes encoding five of the enzymes involved in pyrimidine synthesis, as well as the regulatory protein PyrR (pyrR; Rv1379), are located on the *pyr* operon. Some of these genes are essential for growth of *M. tuberculosis in vitro* (39). However, much of what is known about pyrimidine biosynthesis in *M. tuberculosis* is inferred from studies in other bacteria.

Pyrimidine biosynthesis begins with the formation of carbamoyl phosphate from glutamine, bicarbonate, and two molecules of MgATP, by carbamoyl-phosphate synthase, an enzyme comprising a small subunit (*carA*; Rv1383) and a large subunit (*carB*; Rv1384). Aspartate carbamoyltransferase (ATCase) (*pyrB*; Rv1380) catalyzes the condensation of aspartate and carbamoyl phosphate, yielding N-carbamoylasparrtate, after which dihydroorotate (*pyrC*; Rv1381) catalyzes the closure of the pyrimidine ring structure to form dihydroorotate. It is likely that the mechanism of allosteric regulation of mycobacterial ATCase is similar to that of *E. coli*, in which ATP activates and CTP inhibits the activity of ATCase, thereby regulating the intracellular ratios of purine and pyrimidine nucleotides. The oxidation of dihydroorotate to orotate is carried out by dihydroorotate dehydrogenase (*pyrD*; Rv2139), resulting in the formation of the 5,6-double bond of the pyrimidine base. The fifth step in the pathway is then catalyzed by orotate phosphoribosyltransferase (OPRT) (*pyrE*; Rv0382c), a type I PR Tase that converts orotate to orotidine 5′-monophosphate (OMP) (40). This is the first committed step in *de novo* pyrimidine biosynthesis because it is the last step for which there is no chemical intermediate that can be derived from the pyrimidine salvage pathway. Submicromolar pyrimidin-2(1H)-one-based inhibitors of the *M. tuberculosis* orotate phosphoribosyltransferase enzyme have been identified but, as yet, have not been tested for antimycobacterial activity (40). OMP is further decarboxylated to UMP by OMP decarboxylase (*pyrF*; Rv1385). Subsequent reversible phosphorylation of UMP by uridylyl kinase (UMPK) (*pyrH*; Rv2883c) yields UDP, which is then used in the synthesis of all other pyrimidine nucleotides (41). *M. tuberculosis* UMPK differs from homohexameric UMPKs in other bacteria in that it is tetrameric, although its specificity for UMP as the phosphoryl group acceptor is similar to that of other UMPKs (42). UDP is then converted to UTP by NdkA, and CTP is formed from UTP via the activity of CTP synthase (*pyrG*; Rv1699).

In *M. smegmatis*, regulation of the *pyr* operon occurs via translational repression and involves the nucleotide-regulated binding of PyrR to occlude the first ribosome binding site in the *pyr* mRNA (43). This is consistent with evidence from *M. tuberculosis*, in which PyrR has been shown to bind the mRNA-binding loop in a UMP/UTP-dependent manner, and probably as a dimer (44). This regulatory mechanism differs from that in Gram-positive bacteria such as *Bacillus subtilis* (45) owing to the absence of transcription antiterminator and attenuation terminator sequences in mycobacterial 5′-*pyr* regions (43). *M. tuberculosis* pyrR expression is upregulated by hypoxia, suggesting that an attenuation of pyrimidine biosynthesis accompanies adaptation of *M. tuberculosis* to the hypoxic conditions prevailing in granulomas (46). In addition to its regulatory function, the PyrR from *M. tuberculosis* may also display weak UPRTase activity (44), as observed in other organisms (45).

**Salvage and Interconversion Pathways**

In contrast to most intracellular pathogens, which utilize either one pathway or the other, pathogenic mycobacteria express enzymes of *de novo* biosynthesis as well as salvage pathways for purines and pyrimidines (2, 47, 48).
Since the \textit{de novo} synthesis of nucleosides is energy-intensive, it is tempting to speculate that organisms such as \textit{M. tuberculosis} preferentially engage salvage pathways during chronic, persistent infection when energy stores are likely to be low; however, experimental data in support of this notion are lacking. The majority of genes involved in the salvage pathways of \textit{M. tuberculosis} are individually dispensable for growth \textit{in vitro} \cite{39, 49}, in macrophages \cite{50} and in animal tissue \cite{51}. There are, however, notable exceptions, as outlined below.

**Purine salvage**

The \textit{de novo} synthesis of AMP and GMP from IMP is irreversible, but the purine salvage pathway facilitates the reconversion of preformed purine bases and nucleosides from the external environment to their corresponding purine nucleotides \cite{52}. The \textit{M. tuberculosis} genome encodes a number of enzymes capable of interconverting purine bases, nucleosides, and nucleotides (Table 1). Adenosine deaminase (\textit{add}; Rv3133c) catalyzes the irreversible hydrolytic deamination of (deoxy)adenosine to (deoxy)inosine. The reversible phosphorylation of the N-glycosidic bond of \(\beta\)-purine (deoxy)ribonucleotides, generating \(\alpha\)-(deoxy)ribose 1-phosphate and the corresponding purine base, is carried out by purine nucleoside phosphorylase (PNP) (\textit{deoD}; Rv3307) \cite{53}. In contrast to the human PNP, the \textit{M. tuberculosis} enzyme has greater specificity for guanosine N7 methyl analogs than natural nucleosides \cite{53}. However, in keeping with the trimeric structure of both the human and \textit{M. tuberculosis} enzymes, as opposed to the hexameric structure of most prokaryotic PNP\textquotesingle s, the \textit{M. tuberculosis} PNP is unable to phosphorylate adenosine \cite{53}. The identification of unique hydrogen-bonding patterns upon inhibition of \textit{M. tuberculosis} PNP by immucillin-H suggests that the rational design of mycobacterial PNP-specific inhibitors may be feasible \cite{54}.

Adenosine kinase (AK) (\textit{adoK}; Rv2202c) catalyzes the Mg\textsuperscript{2+}-dependent phosphorylation of adenosine to AMP. Although AKs are found in most eukaryotes, fungi, plants, and parasites, they are seldom found in bacteria \cite{47}. The \textit{M. tuberculosis} enzyme was the first bacterial AK to be characterized \cite{47} and functions as a homodimer, unlike most AKs, which are monomeric.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Rv no.</th>
<th>Protein</th>
<th>Reaction catalyzed</th>
<th>Essentiality\textsuperscript{a}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purine salvage</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>\textit{add}</td>
<td>Rv3133c</td>
<td>Adenosine deaminase</td>
<td>(d)Ado + H\textsubscript{2}O \rightleftharpoons (d)ino + NH\textsubscript{3}</td>
<td>Nonessential\textsuperscript{b}</td>
</tr>
<tr>
<td>\textit{deoD}</td>
<td>Rv3307</td>
<td>Purine nucleoside phosphorylase</td>
<td>Purine nucleoside + P\textsubscript{i} \rightleftharpoons purine base + (\alpha)-ribose-1-phosphate</td>
<td>Nonessential</td>
</tr>
<tr>
<td>\textit{adoK}</td>
<td>Rv2202c</td>
<td>Adenosine kinase</td>
<td>Ado + ATP \rightleftharpoons AMP + ADP</td>
<td>Nonessential</td>
</tr>
<tr>
<td>\textit{hpt}</td>
<td>Rv3624c</td>
<td>Hypoxanthine-guanine phosphoribosyl transferase</td>
<td>Hypoxanthine (guanine) + PRPP \rightleftharpoons IMP (GMP) + P\textsubscript{i}</td>
<td>Nonessential</td>
</tr>
<tr>
<td>\textit{apt}</td>
<td>Rv2584c</td>
<td>Adenine PRTase</td>
<td>Adenine + PRPP \rightleftharpoons AMP + P\textsubscript{i}</td>
<td>Nonessential</td>
</tr>
<tr>
<td>\textit{iunH}</td>
<td>Rv3393</td>
<td>Nucleoside hydrolase</td>
<td>Inosine (uridine) + H\textsubscript{2}O \rightleftharpoons D-ribose + hypoxanthine (uracil)</td>
<td>Nonessential</td>
</tr>
<tr>
<td>Pyrimidine salvage</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>\textit{dut}</td>
<td>Rv2697c</td>
<td>dUTPase</td>
<td>dUTP + H\textsubscript{2}O \rightleftharpoons dUMP + P\textsubscript{i} + H\textsuperscript{+}</td>
<td>Essential \textit{in vitro}</td>
</tr>
<tr>
<td>\textit{dcd}</td>
<td>Rv0321</td>
<td>dCTP deaminase:dUTPase</td>
<td>dCTP + H\textsubscript{2}O \rightleftharpoons dUTP + NH\textsubscript{3} and dUTP + H\textsubscript{2}O \rightleftharpoons dUMP + P\textsubscript{i} + H\textsuperscript{+}</td>
<td>Nonessential</td>
</tr>
<tr>
<td>\textit{thxA}</td>
<td>Rv2764c</td>
<td>Thymidylate synthase</td>
<td>CH\textsubscript{2}H\textsubscript{4}folate + dUMP \rightleftharpoons H\textsubscript{2}folate + dTMP</td>
<td>Nonessential</td>
</tr>
<tr>
<td>\textit{thxY}</td>
<td>Rv2754c</td>
<td>Flavin-dependent thymidylate synthase</td>
<td>CH\textsubscript{2}H\textsubscript{4}folate + dUMP + (FADH\textsubscript{2}) + NADPH \rightleftharpoons H\textsubscript{3}folate + dTMP + NADP\textsuperscript{+}</td>
<td>Essential \textit{in vitro}</td>
</tr>
<tr>
<td>\textit{tmk}</td>
<td>Rv3247c</td>
<td>Thymidylate kinase</td>
<td>dTMP + ATP \rightleftharpoons ADP + dTDP</td>
<td>Nonessential</td>
</tr>
<tr>
<td>\textit{ndkA}</td>
<td>Rv2445c</td>
<td>Nucleoside diphosphate kinase</td>
<td>Ribonucleoside diphosphate + ATP \rightleftharpoons ribonucleoside triphosphate + ADP</td>
<td>Nonessential</td>
</tr>
<tr>
<td>\textit{cdd}</td>
<td>Rv3315c</td>
<td>Cytidine deaminase</td>
<td>(d)Cytidine + H\textsubscript{2}O \rightleftharpoons (d)uridine + NH\textsubscript{3}</td>
<td>Nonessential</td>
</tr>
<tr>
<td>\textit{deoA}</td>
<td>Rv3314c</td>
<td>Thymidine phosphorylase</td>
<td>Thymidine + P\textsubscript{i} \rightleftharpoons 2-deoxyribose-1-phosphate + thymine</td>
<td>Nonessential</td>
</tr>
<tr>
<td>\textit{upp}</td>
<td>Rv3309c</td>
<td>Uracil PRTase</td>
<td>Uracil + PRPP \rightleftharpoons UMP + P\textsubscript{i}</td>
<td>Nonessential</td>
</tr>
<tr>
<td>\textit{cmk}</td>
<td>Rv1712</td>
<td>Cytidylate kinase</td>
<td>(d)CMP + ATP \rightleftharpoons (d)CDP + ADP</td>
<td>Essential \textit{in vitro}</td>
</tr>
</tbody>
</table>

\textsuperscript{a}In \textit{in vitro} essentiality, as determined by TraSH \cite{39, 49}.

\textsuperscript{b}All of the genes listed as nonessential for growth \textit{in vitro} are also dispensable for adaptation and survival in macrophages \cite{50} or for survival during infection in mice \cite{51}.

\textbf{Nucleotide Metabolism and DNA Replication}

\textbf{TABLE 1} Genes involved in purine and pyrimidine salvage pathways in \textit{M. tuberculosis}

\textbf{ASMscience.org/MicrobiolSpectrum}
Moreover, the activity of *M. tuberculosis* AK is stimulated by monovalent metal ions such as K⁺, but not by inorganic phosphate, as for other AK enzymes (47, 55, 56). Interestingly, despite the presence of adenosine deaminase and PNP enzymes, mycobacteria preferentially use AK to dephosphorylate adenosine (55). Since the proinflammatory response is likely to result in adenosine accumulation (elevated levels of host adenosine deaminase are a diagnostic marker of tuberculous pleurisy [57]), it is possible that *M. tuberculosis* expresses an AK that preferentially utilizes adenosine over deoxyadenosine in order to exploit this purine reservoir (58). The observation that AK catalyzes the first step in the conversion of 2′-methyladenosine into an analog that is active against *M. tuberculosis* (58) has provided a proof-of-concept for targeting the purine salvage pathway for TB drug discovery.

The hypoxanthine-guanine and adenine PRTases are encoded by *hpt* (Rv3624c) and *apt* (Rv2584c), respectively. Both are type I enzymes that catalyze the Mg²⁺-dependent, reversible transfer of the 5′-phosphoribosyl moiety from PRPP to the N-9 position of the purine. Guanine and hypoxanthine are substrates for Hpt, giving rise to GMP and IMP, respectively (52), whereas Apt utilizes adenine as a substrate to produce AMP (59). *M. tuberculosis* Hpt has a preference for hypoxanthine over guanine as a substrate and, unlike the Hpt from most other organisms, is unable to utilize xanthine as a substrate (52). Finally, *M. tuberculosis* also possesses a homolog of nucleoside hydrolyase (*iumH;* Rv3393), an enzyme that preferentially utilizes inosine and uridine as substrates but can likely catalyze irreversible hydrolysis of all of the common purine and pyrimidine nucleosides. These enzymes are of particular importance in parasitic protozoa, such as *Crithidia fasciculate*, which lack a *de novo* purine synthetic pathway and therefore rely solely on salvage mechanisms for purine supply (60).

### Pyrimidine salvage

As in the case of purines, mycobacteria are able to reutilize pyrimidine bases and nucleosides derived from preformed nucleotides via the pyrimidine salvage pathway (61). The majority of enzymes involved in salvage pathways are dispensable *in vitro* and during growth in macrophages and in animal tissue (Table 1); in contrast, the mono-functional dUTPase (*dut;* Rv2697c) is distinguished by virtue of its essentiality in *M. tuberculosis* (39, 49) and *M. smegmatis* (62). This homotrimeric enzyme (63) converts dUTP to dUMP, thus providing the immediate precursor of thymidine nucleotide. Since *M. tuberculosis* lacks both a dCMP deaminase and a thymidine kinase (2), dUTPase activity is of particular importance because it provides the only source of d(U) TMP for DNA biosynthesis (64) (Fig. 1). Interestingly, although Dut is essential for growth of *M. tuberculosis* *in vitro*, it is not the only enzyme with dUTPase activity: dCTP deaminase (*dcd, Rv0321*), a homotrimeric enzyme that converts dCTP to dUTP and is dispensable *in vitro* (39, 49), also displays dUTPase activity—a feature only observed previously in archaea (65). Biochemical studies on recombinant Dcd from *M. tuberculosis* showed that, although its dUTPase activity is lower than that of Dut (Rv2697c), the bifunctionality of this enzyme allows the formation of dUMP from both dCTP and dUTP, with dUMP being formed at the same rate that dCTP is deaminated. In light of these findings, the essentiality of Dut will need to be established *in vivo* in order to validate this enzyme as a novel TB drug target.

Thymidylate synthase (TS) converts dUMP to dTMP. *M. tuberculosis* contains both the classical TS enzyme, ThyA (Rv2764c), and the evolutionarily unrelated flavin-dependent TS, ThyX (Rv2754c) (66). Both enzymes catalyze the reductive methylolation of dUMP by N⁵, N¹⁰-methylenetetrahydrofolate to produce dTMP and dihydrofolate, but ThyX additionally requires FADH₂ and NAD(P)H as reductant (67). ThyA is dispensable for growth *in vitro* and *in vivo*, whereas ThyX is essential *in vitro* (68). Although the biological significance of distinct TS enzymes in a single organism remains obscure (69), the *in vitro* essentiality of *M. tuberculosis* ThyX in the presence of a functional ThyA enzyme suggests that ThyX may serve other cellular functions in addition to providing dTMP (68, 70). This uncertainty notwithstanding, ThyX has been the subject of intense investigation and has emerged as a popular target for structure-guided TB drug discovery efforts (67, 71, 72). However, as in the case of Dut, the essentiality of ThyX *in vivo* must still be established in order to validate this target. Following the conversion of dUMP to dTMP by TS, thymidylate kinase (*tmk;* Rv3247c) catalyzes the phosphorylation of dTMP to form dTDP, and nucleoside diphosphate kinase (*ndkA;* Rv2445c) phosphorylates dTDP to produce dTTP.

Cytidine deaminase (*cdd;* Rv3315c) catalyzes the *Zn²⁺*-dependent hydrolytic deamination of cytidine or 2′-deoxyctydine to uridine or 2′-deoxyuridine, respectively (73, 74). Located downstream of *cdd*, on an operon, is the gene encoding thymidine phosphorylase (*deoA;* Rv3314c), another pyrimidine salvage enzyme, which catalyzes the reversible hydrolysis of thymidine to β-2-deoxyribose-1-phosphate and thymine. Mycobacteria also express a cytidylate kinase (*cmk;* Rv1712), which catalyzes the reversible γ-phosphoryl transfer from ATP.
to CMP or dCMP, generating CDP or dCDP, respectively (75, 76). The *M. tuberculosis* enzyme, which is essential for growth *in vitro* (39, 49), preferentially phosphorylates dCMP and is thought to play a role in the recycling of nucleotides derived from RNA degradation (75).

As described previously, PyrR displays weak UPRTase activity (44). However, the major UPRTase in *M. tuberculosis* is probably provided by Upp (Rv3309c), which catalyzes the conversion of uracil and PRPP to UMP, the common precursor of all pyrimidine nucleotides. This enzyme may play a particularly important role in pyrimidine salvage in *M. tuberculosis* because the organism lacks both uridine nucleosidase and uridine phosphorylase enzymes, which catalyze conversion of uracil to uridine, as well as uridine kinase and uridine monophosphatase, which convert uridine to UMP (2).

Formation of Deoxyribonucleotides from Ribonucleotides

The reduction of ribonucleotides to their corresponding deoxyribonucleotides is carried out by ribonucleotide reductase (RNR). *M. tuberculosis* and other mycobacteria contain *nrdE* (Rv3051c) and *nrdF2* (Rv3048c) genes, which encode an essential class Ib RNR that is regulated by the regulator NrdR (Rv2718c) (77, 78).

Class Ib RNRs are composed of two homodimeric proteins (the large R1 or α subunit, NrdE, and the small R2 or β subunit, NrdF) in an α₂β₂ configuration (79). Association of the subunits is essential for catalytic activity, with both the substrate and allosteric effector binding sites located on R1, and the stable free radical on a tyrosine residue on R2 that is involved in the reduction of the ribonucleotide. In addition, the presence of an Fe(III)-Fe(III) radical diiron site on R2 is necessary for activity of the *M. tuberculosis* NrdEF2 enzyme (79). The C-terminus of the small subunit is required for holoenzyme association, which has facilitated the design of small peptides that competitively inhibit binding of R2 to R1, thereby abrogating enzymatic activity (80–82). Interestingly, *M. tuberculosis* contains another hypothetical class Ib small subunit encoded by *nrdF1* (Rv1981c). Although this protein contains the critical tyrosine residue, biochemical and genetic studies have confirmed that it is unable to substitute for NrdF2 to form a functional class Ib enzyme (77, 78). The essentiality of NrdEF2, together with its vulnerability to chemical inhibition, has validated this enzyme as a TB drug target (78). In contrast, the role—if any—of the vitamin B₁₂-dependent class II RNR, NrdZ, in deoxyribonucleoside triphosphate (dNTP) provision in *M. tuberculosis* remains obscure. As part of the DosR regulon

**FIGURE 1** Pathways for the formation of dTTP in *M. tuberculosis*. The steps for which homologs are absent are shown in gray, and enzymes that are essential for growth *in vitro* are shown in shaded boxes. Adapted from references 64 and 62. doi:10.1128/microbiolspec.MGM2-0001-2013.f1
(83, 84), nrdZ is induced in response to hypoxia; however, loss of nrdZ function was found to confer no discernible phenotype in vitro or in vivo (77).

NrdH, which is encoded by the first gene in the nrdHIE operon (Rv3053c), is a member of the thioredoxin family of proteins that are involved in the maintenance of redox homeostasis. NrdH redoxins are widespread among prokaryotes that are not capable of producing glutathione; in mycobacteria, these proteins function as hydrogen donors, providing the electrons that are required for reconstitution of the RNR active site radicals (85). NrdH sequences are conserved throughout mycobacteria, but the genomic arrangement of the nrdHIE operon varies between pathogenic and nonpathogenic mycobacteria. In nonpathogenic mycobacteria, NADPH-dependent FMN reductase is located upstream of the nrdHIE operon, while hypothetical proteins are found in this location in pathogenic mycobacteria. The significance of this is as yet unknown but may become clearer once the function of these hypothetical proteins is elucidated. The fourth gene in the nrd operon encodes Nrdl (Rv3052c), a stimulatory protein that affects the activity of class Ib RNRs in E. coli by reduction of ribonucleotides (86). Nrdl, which represents a novel class of flavodoxin (87), is essential for class Ib RNR activity in Streptococcus pyogenes (88) but is not required for growth of M. tuberculosis in vitro.

DNA REPLICATION

In contrast to eukaryotes—in which replication and segregation of DNA are separated temporally—the cell cycle in bacteria is not divided into discrete stages; instead, cell growth, DNA replication, chromosome segregation, and the initial assembly of the cell division machinery can overlap (89). As might be expected, this requires that the activity of the replication machinery is carefully regulated to ensure that the generation of a second chromosomal copy is coordinated with the pathways that function in chromosomal segregation and cell division. In the remaining sections of this article, we will provide some insight into the genes involved in the replication of the 4.4-Mb M. tuberculosis chromosome, with a special focus on the composition and operation of the mycobacterial replisome. In addition, we will highlight briefly the key pathways and interactions that coordinate the initiation of DNA replication with growth of the mycobacterial cell. As will become evident, in contrast to many other aspects of mycobacterial metabolism—including the biosynthesis and salvage of nucleotides described above—there is limited biochemical and mechanistic insight into many aspects of mycobacterial DNA replication, with some notable exceptions. Therefore, wherever possible, the analysis will be informed by current knowledge from other bacterial systems, in particular the E. coli model.

The Mycobacterial Replisome

A large, multiprotein complex and numerous interacting partners function in bacterial DNA replication (90). Many components of the replication machinery are conserved and have been most thoroughly characterized in E. coli (91) and B. subtilis (92). Based on evidence from these organisms, the model bacterial “replisome” comprises two main subcomplexes: (i) the primosome, which itself includes the replicative helicase that unwinds the duplex DNA, and the primase, which is activated by DnaB to produce short RNA primers for discontinuous lagging-strand synthesis (93); and (ii) the DNA polymerase III holoenzyme (Pol III HE), which is made up of the Pol III core, the β2-sliding clamp, and the clamp-loader complex (93).

Until very recently, it was thought that the Pol III HE consisted of two core polymerase subassemblies (each comprising an α polymerase subunit, 3′→5′ proofreading exonuclease, ε, and the ε stabilizer, θ), which were required for concurrent leading and lagging strand synthesis at a single replication fork. However, new evidence has established that a third (αθ) core is tethered to the clamp loader to form a tripolymerase replisome (91, 94): a single Pol III core is involved in leading strand synthesis, while the other two Pol III cores extend multiple Okazaki fragments on the lagging strand (95). Current models of the E. coli replisome are therefore based on a Pol III HE that contains three separate subcomplexes: a seven-subunit DnaX clamp loader complex (τδθγδθγ) binds three Pol III cores through the τ-α interaction, while the β2 sliding-clamp processivity factor tethers the Pol III core to the DNA through a separate interaction with α. Of the remaining DnaX subunits, δ and δ′ serve to load and unload the β2 sliding-clamp, while χ and ψ stabilize the interactions between DnaX, δ, and δ′ (93).

The genes known to be involved in DNA replication in M. tuberculosis are summarized in Table 2. On inspection, it is apparent that the mycobacterial replisome lacks several components that perform key functions in the E. coli system. However, comparative genomic analyses have established that this gene complement is typical of
many bacteria and has contributed to the definition of a basic bacterial replication module that contains the replication initiator protein, DnaA, the DnaB helicase, DnaG primase, Pol III\(\alpha\), the \(\beta\) sliding clamp, the \(\epsilon\) proofreading subunit, \(\tau\), \(\delta\) and \(\delta'\), single-stranded DNA (ssDNA)-binding protein, DNA ligase, and Pol I (93, 96). In the sections below, we highlight some of the major advances that have been made since the previous edition of this article was published in identifying and understanding the components and functions of the conserved mycobacterial replisome, as well as the genes required for the initiation of DNA replication.

### TABLE 2 Genes known to be involved in DNA replication in M. tuberculosis

<table>
<thead>
<tr>
<th>Gene</th>
<th>Rv Number</th>
<th>Function</th>
<th>Catalytic activity</th>
<th>Essential for growth in vitro*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Initiation complex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dnaA</td>
<td>Rv0001</td>
<td>Replication initiator</td>
<td>ATPase</td>
<td>Essential</td>
</tr>
<tr>
<td>dnaB</td>
<td>Rv0058</td>
<td>Replicative helicase</td>
<td>ATPase</td>
<td>Essential</td>
</tr>
<tr>
<td><strong>Primosome</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dnaG</td>
<td>Rv2343c</td>
<td>DNA primase</td>
<td>RNA primase</td>
<td>Essential</td>
</tr>
<tr>
<td>priA</td>
<td>Rv1402</td>
<td>Primosome replication factor; replication restart</td>
<td>ATPase</td>
<td>Essential</td>
</tr>
<tr>
<td><strong>Elongation – the DNA Pol III core</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dnaE1</td>
<td>Rv1547</td>
<td>(\alpha) Subunit, polymerase activity</td>
<td>DNA polymerase</td>
<td>Essential</td>
</tr>
<tr>
<td>dnaQ</td>
<td>Rv3711c</td>
<td>(\epsilon) Subunit, proofreading activity</td>
<td>Exonuclease</td>
<td>Nonessential</td>
</tr>
<tr>
<td>dnaQ-uvrC</td>
<td>Rv2191</td>
<td>Chimeric protein; N-terminal 3'-5' exonuclease domain, **C-terminal UvrC-like endonuclease domain</td>
<td>Exonuclease/**Endonuclease</td>
<td>Nonessential</td>
</tr>
<tr>
<td><strong>Elongation – the DNA Pol III clamp loader complex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dnaZXY</td>
<td>Rv3721c</td>
<td>(\tau) and (\gamma) subunits</td>
<td>ATPase</td>
<td>Essential</td>
</tr>
<tr>
<td>holA</td>
<td>Rv2413c</td>
<td>(\delta) subunit</td>
<td>Nonessential</td>
<td></td>
</tr>
<tr>
<td>holB</td>
<td>Rv3644c</td>
<td>(\delta') subunit</td>
<td>ATPase</td>
<td>Essential</td>
</tr>
<tr>
<td><strong>Elongation – the mycobacterial replisome</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dnaN</td>
<td>Rv0002</td>
<td>(\beta) sliding clamp</td>
<td></td>
<td>Essential</td>
</tr>
<tr>
<td>ssb</td>
<td>Rv0054</td>
<td>Single-stranded DNA binding protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>polA</td>
<td>Rv1629</td>
<td>DNA Pol I</td>
<td>DNA polymerase</td>
<td>(Essential)*</td>
</tr>
<tr>
<td>lgiA</td>
<td>Rv3014c</td>
<td>DNA ligase, DNA replication</td>
<td>NAD-dependent DNA ligase</td>
<td>Essential</td>
</tr>
<tr>
<td>lgiB</td>
<td>Rv3062</td>
<td>DNA ligase</td>
<td>ATP-dependent DNA ligase</td>
<td>Nonessential</td>
</tr>
<tr>
<td>lgiC</td>
<td>Rv3731</td>
<td>DNA ligase</td>
<td>ATP-dependent DNA ligase</td>
<td>Nonessential</td>
</tr>
<tr>
<td>lgiD</td>
<td>Rv0938</td>
<td>DNA ligase, DSB repair</td>
<td>ATP-dependent DNA ligase</td>
<td>Essential</td>
</tr>
<tr>
<td>gyrA</td>
<td>Rv0006</td>
<td>DNA gyrase, subunit A (DNA topoisomerase II)</td>
<td>Topoisomerase, ATPase</td>
<td>Essential</td>
</tr>
<tr>
<td>gyrB</td>
<td>Rv0005</td>
<td>DNA gyrase, subunit B (DNA topoisomerase II)</td>
<td>Topoisomerase, ATPase</td>
<td>Essential</td>
</tr>
<tr>
<td>topA</td>
<td>Rv3646c</td>
<td>DNA topoisomerase I</td>
<td>Topoisomerase, ATPase</td>
<td>Essential</td>
</tr>
<tr>
<td>mha-cobC</td>
<td>Rv2228c</td>
<td>Chimeric protein; N-terminal RNase HI domain, **C-terminal CobC-like (\alpha)-ribazole phosphatase domain</td>
<td>RNase/**(\alpha)-ribazole phosphatase</td>
<td>Essential</td>
</tr>
<tr>
<td>mhb</td>
<td>Rv2902c</td>
<td></td>
<td>Rnase</td>
<td>Nonessential</td>
</tr>
<tr>
<td><strong>Other DNA polymerases</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dnaE2</td>
<td>Rv3370c</td>
<td>(\alpha) Subunit, TLS polymerase activity</td>
<td>DNA polymerase</td>
<td>Nonessential</td>
</tr>
<tr>
<td>dinB1</td>
<td>Rv1537</td>
<td>DNA Pol IV.I</td>
<td>DNA polymerase</td>
<td>Nonessential</td>
</tr>
<tr>
<td>dinB2</td>
<td>Rv3056</td>
<td>DNA Pol IV.II</td>
<td>DNA polymerase</td>
<td>Nonessential</td>
</tr>
<tr>
<td>polX</td>
<td>Rv3856c</td>
<td>DNA Pol X</td>
<td>Unknown(i)</td>
<td>Nonessential</td>
</tr>
</tbody>
</table>

*In vitro essentiality, as determined by TraSH (39, 49).
*b Rv0054 did not satisfy the strict criterion for essentiality in the study by Griffin et al. (49) owing to a paucity of TA sites within the open reading frame; no transposon (Tn) insertions were identified in any of the five possible TA dinucleotides, suggesting that it is likely to be essential.
*c A truncation in the polymerase domain of Rv3856c is predicted to preclude catalytic activity (98).

The replicative Pol III\(\alpha\) subunit in M. tuberculosis

A feature of the M. tuberculosis genome is that it encodes two dnaE-type Pol III\(\alpha\) subunits, DnaE1 and DnaE2. In a study that redefined notions of the strict role of C-family DNA polymerases in high-fidelity DNA replication, Boshoff et al. (97) established that the DnaE1 and DnaE2 subunits fulfill separate roles: DnaE1 is the essential replicative polymerase in M. tuberculosis, whereas DnaE2 functions in nonessential, error-prone translesion synthesis (TLS). This work identified DnaE2 as a central...
player in the DNA damage response in *M. tuberculosis* and, importantly, demonstrated a critical role for the TLS polymerase in virulence and the emergence of drug resistance *in vivo*. Subsequently, DnaE2 was shown to operate with two accessory factors as part of a split *imuA*-*imuB*DnaE2 mutagenic cassette, with ImuB acting as a potential hub protein that interacts functionally with the other two cassette proteins, as well as the β₂ sliding clamp (98). The elucidation of a functional analog of *E. coli* Pol V suggested that the components of the tripartite mycobacterial mutasome might offer compelling targets for novel drugs designed to limit mutagenesis (99). In addition, it reinforced the idea that, together with potential differences in the architecture of the respective DnaE active sites (100), differential protein interactions might contribute to the specialist function of the dnaE1- and dnaE2-encoded subunits, which remains poorly understood (93).

All major DNA Pol IIIα structural features are readily identifiable in both DnaE1 and DnaE2, except for the C-terminal τ-interacting domain, which is absent in DnaE2 (98). Since the α-τ interaction enables simultaneous leading and lagging strand synthesis by Pol III HE in *E. coli*, the absence of this region might account for the inability of DnaE2 to function as a chromosomal replicase in *M. tuberculosis*. Another critical interaction in the *E. coli* replisome is that which occurs between Pol IIIα and the dnaQ-encoded 5′→3′ exonuclease proofreading subunit. The *M. tuberculosis* genome encodes two DnaQ-like proteins, Rv3711c and Rv2191; although antigenic evidence suggests that Rv3711c is actively expressed during host infection (101), the contribution of either DnaQ homolog to replicative fidelity, or differential α subunit function, is unknown. Considering the inferred role of DnaE2 in TLS across DNA damage lesions, it seems very likely that structural determinants such as active site architecture contribute significantly to the polymerase function (100). Again, however, this possibility requires further investigation. Recent evidence suggests that the related DnaE3-type α subunit in *B. subtilis* is essential for the extension of RNA primers to enable PolC-mediated DNA synthesis (92); therefore, while the nonessentiality of dnaE2 precludes a similar role for the second α subunit in *M. tuberculosis*, it remains to be determined whether DnaE2 has a function in mycobacterial DNA replication separate from its key role as a damage-responsive translesion polymerase.

### A role for other DNA polymerases at the replication fork?

Although the α subunit functions as dominant replicase in bacterial DNA synthesis, there is increasing evidence to suggest that chromosomal replication is characterized by dynamic DNA polymerase exchange (102). For example, of the four accessory DNA polymerases in *E. coli*, Pol I (*polA*) and Pol II (*polB*) contribute directly to replication fidelity during normal chromosomal replication (102) through their respective roles in high-fidelity maturation of Okazaki fragments during lagging strand synthesis (Pol I), and as a back-up replicative polymerase during transient dissociation of the Pol III HE from either leading or lagging strand (Pol II). The remaining accessory polymerases, DNA Pol IV (*dinB*) and Pol V (*imuDC*), are both members of the Y family of specialist translesion polymerases that function primarily in the DNA damage response but can also access the lagging strand under conditions of elevated expression (102).

*M. tuberculosis* does not possess a Pol II enzyme but does encode two homologs of Pol IV, DinB1 and DinB2. However, unlike the *E. coli* model in which both Pol IV and Pol V are induced as part of the SOS regulon, the mycobacterial DNA damage response does not include either *dinB1* or *dinB2* and is instead limited to the two α subunits, DnaE1 and DnaE2 (98). Moreover, only DinB1 possesses a consensus β clamp binding motif, suggesting that DinB2 must interact with another protein(s) in order to access the replication fork. Therefore, while recent work has established that the *M. smegmatis* homolog of *M. tuberculosis* DinB2 is a functional DNA polymerase with a tendency to promote G:T and T:G mismatches (103), the activity and function of the DinB-type Y-family polymerases in *M. tuberculosis* remain poorly understood: deletion of *dinB1* and/or *dinB2* does not result in any discernible phenotype in *M. tuberculosis* in various assays *in vitro* or during mouse infection *in vivo* (104). Further research will, therefore, be required to elucidate the contribution, if any, of either mycobacterial Pol IV homolog to replication fidelity.

### Initiation of Replication

The replication of the chromosome constitutes a major event in the bacterial cell cycle and must be carefully coordinated with cell growth and division. For this reason, bacteria have evolved rigorous control mechanisms to regulate the initiation of DNA replication and ensure that it does not occur at random sequences throughout the chromosome (105). Instead, replication is initiated at
a single site—termed the origin of replication, oriC—and proceeds bidirectionally around the chromosome until the two replication forks meet in the replication terminus (ter), a region located approximately opposite oriC. Moreover, since it must occur only once during the cell cycle, a diverse array of regulatory mechanisms—many of which are centered on the activity of the initiator protein, DnaA—ensure that the assembly of the replication machinery is triggered at the appropriate stage (106).

The prereplication complex

Oligomerization of DnaA on oriC is critical for replication initiation since it results in the opening of the DNA duplex to allow loading of DnaB. In E. coli, DnaA proteins bind throughout oriC at specific 9-bp binding sites (DnaA boxes) that are closely spaced to facilitate the formation of a higher-order nucleoprotein scaffold comprising over 20 DnaA monomers (106). The formation of this “prereplication complex” (pre-RC) causes the melting of a short AT-rich region to create an open DNA-DnaA platform for loading the replicative helicase and the other replisome components. For this reason, mechanisms regulating replication initiation in E. coli are focused on the pre-RC complex, in particular the activity of DnaA, and include (i) the prevention of de novo DnaA synthesis via sequestration of the dnaA promoter, (ii) the depletion of available DnaA levels through the location of alternative DnaA binding sites elsewhere in the chromosome, (iii) the inactivation of DnaA by DnaN- and/or Hda-dependent stimulation of intrinsic DnaA ATPase activity, and (iv) SeqA-dependent sequestration of oriC (106).

The key features of the E. coli model are retained in M. tuberculosis: the DNA-ATP interaction is critical for replication initiation and, as in E. coli, the dnaA promoter remains active during replication to ensure progression through the cell cycle (107). M. tuberculosis oriC is located in the 527-bp intergenic region between dnaA and dnaN and contains multiple predicted and confirmed DnaA binding sites. Interestingly, this region also serves as a common locus for the insertion of IS6110 transposable elements. To date, however, there is no evidence to suggest the insertions have any effect on the replication process, including the timing of initiation. Instead, these sites have proved practically useful as markers for RFLP fingerprinting of clinical M. tuberculosis isolates (108).

There are 13 closely spaced DnaA boxes in the oriC region of M. tuberculosis, of which 11 have been shown to interact with DnaA (109). Since M. tuberculosis DnaA appears to associate weakly with isolated DnaA box elements (110), it is thought that the close spacing of the DNA boxes favors formation of the nucleoprotein complex through cooperative binding of DnaA monomers (111). Moreover, in an important departure from the E. coli model (where binding of DnaA to ATP is sufficient to promote DnaA binding, oligomerization, and DNA strand separation), both ATP- and ADP-bound DnaA exhibit similar affinities for DnaA boxes and, in M. tuberculosis, the hydrolysis of ATP is required for rapid oligomerization of DnaA on oriC (112). This observation is consistent with the absence in M. tuberculosis of an Hda homolog as well as an equivalent of the datA locus, which in E. coli, acts as principle negative regulators of ATP-DnaA (113). In combination, these observations further support the idea that the intrinsic ATPase activity of M. tuberculosis DnaA is critical in regulating replication, although a mycobacterial IciA (inhibitor of chromosomal initiation) protein has been shown to inhibit DnaA-mediated open complex formation by binding to a specific AT-rich region of oriC (114).

It has recently been shown that MtrA, the response regulator in the essential MtrAB two-component regulatory system in M. tuberculosis, binds a 7-bp sequence motif within the dnaA promoter region (115). This suggests a direct link between the initiation of replication and the processes involved in septation and cell division (116, 117) and is further supported by the observation that DnaA colocalizes with cardiolipin during cell cycle progression (118). In addition to MtrA, there is some evidence to suggest that multiple transcriptional regulators might control expression of dnaA in response to a variety of environmental conditions, including metal availability, oxygen limitation, and low pH (119). These analyses have also identified Mce1R as a putative dnaA regulator, thereby elucidating a possible functional association between cell growth and the activity of the mammalian cell entry module (120).

The mycobacterial primosome

In E. coli, DnaA recruits the hexameric DnaB replicative helicase to the origin in complex with the loader protein, DnaC, which suppresses the ATPase activity of DnaB and so remains bound until loading is complete. Following loading, the DnaG primase triggers dissociation of DnaC from DnaB, unlocking the helicase activity of DnaB for the initiation of strand separation. Recent work has confirmed the physical interaction of M. tuberculosis DnaA and DnaB and, further, has implicated DnaB in controlling DnaA complex formation
and the interaction with oriC (121). However, like many other organisms, M. tuberculosis does not possess a homolog of the DnaC helicase loader (Table 2), suggesting that this activity is provided by a non-orthologous protein, or that a PriA/PriB/DnaT-type replication restart complex functions as a helicase loader at the origin (96). Alternatively, DnaA alone might be sufficient for DnaB loading, an idea that is supported by recent insights into the structure of DnaB from H. pylori—another organism lacking a DnaC helicase loader (122).

Elongation and Termination

Following strand separation and early unwinding of the parental chromosomal DNA by the DnaB replicative helicase, Pol III HE is loaded onto each of the two replication forks. As detailed above, the mycobacterial replisome consists of three core polymerases connected to the central clamp loader via a τ subunit, which also acts as a link to the DnaB helicase. The previous edition of this article (3) noted that M. tuberculosis dnaZX does not contain a programmed ribosomal frameshift signal for the production of both τ and the shorter γ subunit, a feature common to several other bacteria. When viewed in the light of recent evidence that three Pol III replicases are tethered to the τ subunit of the clamp loader complex to form a tripolymerase replisome (91, 94, 95), this observation appears consistent with the idea that all bacterial replisomes comprise trimeric Pol III replicases (95, 96).

Leading strand synthesis is highly processive and involves the continuous extension of DNA; in contrast, lagging-strand replication is rate-limiting and requires discontinuous replication via the extension and ligation of Okazaki fragments. Here, DNA primase fulfills an essential role in producing short RNA primers for extension by Pol III. M. tuberculosis DnaG was recently expressed and applied in a novel high-throughput enzyme assay to identify inhibitors of the priming reaction (123). Following completion of Okazaki fragment synthesis, a second switch occurs so that Pol III is replaced by Pol I, which catalyzes high-fidelity DNA synthesis across the resulting gap. Consistent with previous predictions, M. tuberculosis Rv2228c has been shown to be essential for growth of M. tuberculosis in vitro (49) and encodes a bifunctional protein that fuses RNase H and CobC-like α-ribazole phosphatase activities in a single polypeptide. The protein is unusual in that, in addition to functioning as a classic RNase HI in cleaving RNA/DNA hybrids, it also has double-stranded RNase activity (124).

DNA Unwinding: Helicases and Topoisomerases

In order to replicate the chromosome, the replication complex must gain access to separate leading and lagging strands for template-directed DNA synthesis. This requires the concerted action of helicases and topoisomerases: helicases are motor enzymes that unwind the DNA duplex into the separate single-stranded templates, whereas topoisomerases control supercoiling and resolve topological stress to enable translocation of the replisome (125).

In the E. coli replisome, the τ complex connects leading- and lagging-strand polymerases to the replicative helicase, DnaB, which acts on the lagging strand to unwind the DNA and, through its interaction with DnaG, stimulates primase activity. The M. tuberculosis genome encodes more than 15 putative helicase or helicase-related proteins, including an essential DnaB helicase that, like its E. coli counterpart, forms a hexameric helicase complex (126). Single-molecule analyses of replisome components have shown that E. coli DnaB is highly processive (127). This processivity is not maintained in the context of a full replisome, however, where lagging-strand synthesis is rate-limiting during replication of a double-stranded template (127). While similar analyses have not been performed for the corresponding M. tuberculosis proteins, accurate measurements of replisome processivity could provide critical insight into the inferred relationship between the mycobacterial replication rate and intrinsic fidelity in the absence of postreplicative mismatch repair, as discussed below.

The first edition of this article noted that, unlike E. coli, which encodes four topoisomerases, the M. tuberculosis genome contains only a type I topoisomerase, topA, and the type II topoisomerase, gyrAB, which introduce transient single-strand (type I) or double-strand (type II) breaks in DNA to relax supercoiling. Owing to their essentiality for the growth of M. tuberculosis, both enzymes represent attractive drug targets. However, whereas DNA gyrase has been heavily exploited as a target for the various fluoroquinolone antibiotics, inhibitors of bacterial type I topoisomerases are rare (128).

Separate physical interactions have been elucidated between TopA and ribokinase (129) and between TopA and MazF (130) in M. tuberculosis. Ribokinase is a close homolog of the AK enzyme described above (purine salvage) and catalyzes the phosphorylation of β-ribose to yield ribose-5-phosphate, a key intermediate in multiple cellular processes including the synthesis of
DNA, RNA, and ATP. Notably, the interaction between the *M. tuberculosis* TopA and ribokinase proteins was reported to inhibit the TopA-catalyzed relaxation of supercoiled DNA. In contrast, the same TopA-ribokinase interaction stimulated ribokinase activity. The significance of this interaction for mycobacterial physiology and cell function remains to be determined; however, it has been suggested that cross-regulatory mechanisms control d-ribose utilization in *M. tuberculosis*, particularly during growth under nutrient-limited conditions (129). The second reported interaction was with MazF, the stable toxin protein of the mazEF-encoded toxin-antitoxin pair (131). Recently, Huang and He (130) showed that the interaction between TopA and MazF inhibits MazF ribonuclease activity. Moreover, as with ribokinase, this interaction in turn eliminated the ability of TopA to relax supercoiled DNA, suggesting MazF-mediated inhibition of TopA as a possible mechanism to regulate mycobacterial growth.

### Replication Fidelity and the *M. tuberculosis* Mutation Rate

In most bacteria, the fidelity of DNA replication depends on the selectivity of the replicative polymerase for the correct nucleotide during template-directed synthesis, the removal of any misincorporated nucleotides by the 3′→5′ exonuclease activity of the replicative polymerase itself or its interacting proofreading subunit, and the operation of a postreplicative mismatch repair (MMR) system that detects polymerase errors that have escaped proofreading (102). In *E. coli*, these three highly conserved mechanisms are estimated to contribute $10^{-5}$ (polymerase selectivity), $10^{-2}$ (proofreading), and $10^{-3}$ (MMR), respectively, to the observed overall error rate of $\sim 10^{-10}$ per base pair per round of replication (3). The absence of MMR would be expected, therefore, to lead to an elevated mutation rate, particularly in an organism such as *M. tuberculosis* whose genome is characterized by a high GC content. For this reason, the discovery that *M. tuberculosis* lacks an identifiable MMR system provoked considerable speculation and prompted several experimental investigations since it was first made by Mizrahi and Andersen (132) 15 years ago through comparative genomic analyses.

Despite the prediction that loss of MMR should compromise replication fidelity, *M. tuberculosis* is not a natural “mutator”: consistent with previous estimates (97), the mycobacterial mutation rate was recently calculated at $\sim 2 \times 10^{-10}$ per base pair per round of replication (133), a value that corresponds well with other bacterial systems including *E. coli*. Although there is a possibility that mycobacteria encode a nonorthologous alternative to MMR (132), it seems more likely that the activity of other repair components has enabled *M. tuberculosis* to mitigate the lack of this function (134–136). For example, it has recently been shown that the key nucleotide excision repair helicase, UvrD1, fulfills a critical role in limiting recombination-associated mismatches (137). In addition, intrinsic features of the genome itself might limit the risk of replication errors: short sequences of nucleotide repeats pose a significant problem to replicative DNA polymerases and can result in frameshift mutations, but the GC-rich *M. tuberculosis* genome appears to have been under strong selection to restrict the number of repeat regions through context-dependent codon choice (135).

As noted above, in addition to MMR, mechanisms ensuring the insertion of the correct nucleotide by the replicative polymerase represent another critical factor in replication fidelity. The first part of this article highlighted advances in our understanding of the functions and regulation of the mycobacterial RNRs in supplying dNTPs for various DNA metabolic pathways. In this context, it is interesting to consider very recent evidence that a reduction of the intracellular dNTP pool as a consequence of ATP depletion can enhance replication fidelity in *E. coli* (138): it seems possible, for example, that the closely regulated supply of dNTPs might limit the risk of insertion errors—an effect that would be further enhanced in combination with an inherently slower replicative polymerase (139). Again, more research will be required to establish accurate estimates of the intracellular dNTP concentration in the different phases of the mycobacterial cell cycle, as well as to determine the speed and processivity of the mycobacterial replicase. Certainly, the current lack of structural data for most of the replisome components—with a few notable exceptions (126, 140)—implies that a major effort will be needed to produce sufficient soluble protein for the requisite assays.

The final fidelity mechanism relates to proofreading function. Here, it is tempting to speculate that a putative intrinsic 3′→5′ exonuclease activity in the PHP domain of DnaE-type polymerases provides an additional fidelity measure (93). However, while the PHP domains of both *M. tuberculosis* DnaE1 and DnaE2 contain conserved amino acids predicted to be required for exonuclease activity, it is uncertain if this function is preserved in either of the mycobacterial dnaE-type subunits and, further, whether it contributes to relative α subunit fidelity.
Replication During States of Low Metabolic Activity

The replicative status of *M. tuberculosis* during the various stages of infection is of critical importance in understanding the immunobiology of the host-pathogen interaction and is a key determinant of the susceptibility of the bacillus to antibiotics and host immune effectors. Despite the absence of definitive evidence, slow growth is often cited as a fundamental determinant of mycobacterial virulence. In contrast, the demonstrated association of reduced growth and metabolic activity with a state(s) in which the organism shows reduced susceptibility to antimycobacterial agents is an established component of mycobacterial pathogenesis (141). This is an important concept invoked in attempts to elucidate the biological mechanisms subverting antibiotic efficacy as well as in understanding of the observed utility of isoniazid prevention therapy for the prevention of relapse: as a frontline anti-TB agent and inhibitor of cell wall biosynthesis, isoniazid requires active replication for maximal activity (142).

The well-defined Wayne model of nonreplicating persistence is based on the notion that hypoxia recapitulates a major characteristic of some TB lesions (143). When cultured under oxygen-limiting conditions, *M. tuberculosis* ceases replication through gradual and programmed shutdown of macromolecular synthesis, including DNA, which ceases as the oxygen tension falls below 1%. However, the dependence on continued ATP production for viability during this process suggests that cellular metabolism remains active (144, 145). Moreover, mycobacteria in hypoxia-induced dormancy appear to be diploid, and the resumption of growth on re-exposure to oxygen is characterized by the synchronous initiation of RNA synthesis, followed by cell division and then finally DNA replication (146).

In the mouse model, several studies have identified a correlation between total and viable bacterial counts that is suggestive of the existence of nonreplicating (or very slowly replicating) bacilli in “static equilibrium” with the host immune system during chronic infection (147). More recently, however, the application of an unstable “clock” plasmid by Gill et al. (148) has instead revealed that infecting bacilli maintain a slow but constant replication, with total bacterial population numbers under the control of a continuously active immune response. This is a key observation and is supported by evidence from whole-genome sequencing of latent *M. tuberculosis* infection in the nonhuman primate model (133). It remains unclear, however, whether replication is limited to a specific subpopulation that is able to generate sufficient ATP (and biomass) to drive chromosomal replication and cell division.

Targeting the Replication Machinery for Drug Discovery

As illustrated in Table 2, most genes involved in DNA replication are essential for the growth of *M. tuberculosis*. However, while this is a general feature of all bacterial pathogens, the replication machinery remains relatively under-exploited as a source of potential targets for novel antibacterial agents, with inhibitors of DNA gyrase representing the major exception (96, 149). Recent reports indicate that this gap is being addressed. For example, Dallmann et al. (150) developed an assay that enabled identification of both species-specific and broad-spectrum inhibitors of bacterial replicases. In addition, compounds have been identified that target RctB, the protein essential for replication of the second chromosome, *chrII*, in *Vibrio cholerae* (151), as well as *E. coli* RecA (152). For *M. tuberculosis*, small-molecule inhibitors of DnaG have been identified from biochemical screens *in vitro* (123) but have not been tested in whole-cell assays. In contrast, naphthoquinones derived from natural product extracts with demonstrated activity against *M. tuberculosis* have been shown to inhibit DNA gyrase B by a novel mechanism (153). However, it is not known whether the biochemical and whole-cell inhibitory activities are related.

There is also increasing evidence to suggest that disruption of essential protein-protein interactions might offer a viable alternative to targeting enzyme function. For example, multiple studies have described the identification of small molecules and peptides that target the β-clamp processivity factor, suggesting the potential to block access of essential DNA replication and repair components to the replication fork, including the Pol IIIα subunit (154–156). Similarly, others have explored the use of small-molecule inhibitors of single-stranded DNA-binding protein interactions in various organisms including *M. tuberculosis* (157). As noted below, however, the design of DNA replication-specific protein-protein interaction inhibitors for *M. tuberculosis* will require the construction of comprehensive maps describing the critical interactions involving replisome components (158, 159), as has been done for organisms such as *E. coli* and *B. subtilis* (160, 161).

DIRECTIONS FOR FUTURE RESEARCH

The past decade or so has witnessed exciting progress in our understanding of the pathways and functions involved in mycobacterial DNA replication. However, as with all areas of mycobacterial metabolism, the central question that continues to motivate fundamental research...
of this nature is whether key pathways and functions have adapted specifically to pathogenesis and, in turn, the extent to which these components might be targeted for novel interventions. The increasing availability of whole-genome sequencing data from clinical *M. tuberculosis* strains isolated within populations, and from individual *M. tuberculosis*-infected patients through the course of infection and/or relapse, raises the prospect of identifying specific genotype-phenotype correlations that have contributed to lineage divergence. Comparative genomics have already been used to infer the operation of strong diversifying selection in the adaptive evolution of specific DNA replication and repair components (162, 163); however, as noted previously for similar analyses of DNA repair genes (99), the demonstration that an observed DNA replication polymorphism results in a measurable phenotypic consequence remains elusive. Instead, a significant component of future research should involve the generation of a complete and multilayered picture of the different functional and regulatory interactions that govern DNA replication in *M. tuberculosis* and ensure that chromosomal replication is coordinated with mycobacterial cell division—in *vitro* as well as within relevant disease models and, ultimately, within the obligate human host.

This article has detailed the significant progress made toward understanding the operation and regulation of nucleotide biosynthetic and salvage pathways in maintaining intracellular pools of dNTPs and precursors for DNA synthesis and repair, as well as in identifying potential vulnerabilities for drug targeting. The number of genes involved, as well as the presence in some cases of alternative mechanisms for key steps, nevertheless suggests the potential capacity of *M. tuberculosis* to respond to dynamic and hostile environments. An important question therefore relates to the differential regulation of specific components, and the potential interaction between different pathways, as well as their roles in different disease stages. For example, do dNTP levels fluctuate as a consequence of growth phase in *M. tuberculosis*, as reported for *E. coli* (164)? Does the interplay between replisome function and dNTP provision determine overall replication fidelity? And is this modulated under specific conditions (e.g., hypoxia, stress) or in response to exogenous (host-derived) stimuli (e.g., nitric oxide, low pH)?

As highlighted above, advances in optical imaging have provided critical data to complement genetic analyses in refining models of DNA replication (91, 95). It is expected, therefore, that the application of similar technology to *M. tuberculosis* will elucidate the components and subcellular localization of the mycobacterial replisome and its interacting partners during cell division, including chromosomal replication and segregation. Questions here include the following: What is the composition of the mycobacterial replisome? What is the inherent processivity of the fully constituted replisome, and how does it compare with other known bacterial systems? How does the observed rate of cell division correlate with the replication rate, and is there a link between replication rate and error rate? Alternatively, are there other factors that mediate replication fidelity? Similarly, do the accessory polymerases function in chromosomal replication and, if so, what is their role in determining the overall fidelity of replication? This last question is of special relevance given the inferred dependence of *M. tuberculosis* on chromosomal rearrangements and point mutations for its microevolution within the host, including the development of drug resistance. Moreover, given the absence of predicted β clamp-binding motifs in some replication and repair proteins, what are the additional factors that might regulate access to the replication fork?

The recent application of single-cell time-lapse imaging has yielded conflicting evidence of the timing and nature of mycobacterial cell division (165, 166), which will need to be resolved. Moreover, the cryptic role of a diploid state in nonreplicating persistence requires further attention, since it remains unclear whether this is functionally significant and, if so, how it is regulated: which components are critical for the completion of DNA synthesis during metabolic shutdown, and what are the factors involved in organizing the separate chromosomes? In turn, this highlights the need for further research into the factors that coordinate replication and cell division in *M. tuberculosis*. Where does DNA replication occur within the mycobacterial cell, particularly considering that the size of the replisome almost certainly precludes the possibility that replication occurs within the mycobacterial nucleoid (105)? Is the *M. tuberculosis* ori tethered to the cell pole? Also, given evidence of transcriptional repression of replication components during stationary phase (167), is there a direct link between the stringent response and replication shutdown?

The previous edition of this article noted the presence in the *M. tuberculosis* genome of putative components of DNA restriction modification systems. However, despite increasing evidence implicating differential DNA methylation status in bacterial physiology and virulence (168), the contribution of epigenetic modifications to *M. tuberculosis* pathogenesis remains under-explored.
It is expected, therefore, that the application of novel sequencing technologies (168) to bacilli isolated from clinical and experimental TB infections will be a significant research area in the near future. Finally, there is emerging (and resurgent) interest in the potential utility of DNA replication as a source of new antibacterial drug targets (96, 149). Realizing this potential will require the development of novel screens for potential enzyme and protein-protein interaction inhibitors (123); in addition, it will necessitate the generation of structural data for the mycobacterial DNA replication proteins. It is impossible, therefore, to predict what the next version of this article will contain; however, we look forward to another exciting decade in the molecular genetics of DNA replication in M. tuberculosis.

ACKNOWLEDGMENT

This work was funded by grants from the South African Medical Research Council (to V. M.), the National Research Foundation (to V. M. and D. F. W.), and the Howard Hughes Medical Institute (Senior International Research Scholar’s grant to V. M.).

REFERENCES

Nucleotide Metabolism and DNA Replication


Nucleotide Metabolism and DNA Replication


