Analysis of \textit{rpoB} and \textit{pncA} mutations in the published literature: an insight into the role of oxidative stress in \textit{Mycobacterium tuberculosis} evolution?

Denise M. O’Sullivan, Timothy D. McHugh and Stephen H. Gillespie*

Centre for Medical Microbiology, Department of Infection, Royal Free & University College Medical School, Royal Free Campus, University College London, Rowland Hill Street, London NW3 2PF, UK

Received 14 October 2004; returned 15 December 2004; revised 19 January 2005; accepted 22 January 2004

\textbf{Introduction:} It is perceived wisdom that within the host macrophage, \textit{Mycobacterium tuberculosis} frequently encounters oxidative stress. Exposure of bacteria to reactive oxygen intermediates can have a mutagenic effect on the DNA. Various mutations are thought to arise as a consequence, including the oxidation of guanine residues, leading to G\textsuperscript{\textsubscript{7,8}}-dihydro-8-oxoguanine\textsuperscript{2} (8-oxodG), resulting in a ‘GO lesion’.\textsuperscript{2} The strong preference of guanine

\textbf{Methods:} We measured the relative contribution of oxidative stress by recording the percentage of single nucleotide substitutions reported in the genes \textit{rpoB} and \textit{pncA} that confer resistance to the antimicrobials rifampicin and pyrazinamide, respectively, and determined whether there is an excess of G\textsuperscript{\textsubscript{7,8}}-dihydro-8-oxoguanine\textsuperscript{2} or G\textsuperscript{\textsubscript{7,8}}-dihydro-8-oxoguanine\textsuperscript{2}-T\textsubscript{\textsuperscript{7}}A substitutions.

\textbf{Results:} Out of 840 clinical isolates reported with single nucleotide mutations in the \textit{rpoB} gene, 67\% were G\textsuperscript{\textsubscript{7,8}}-dihydro-8-oxoguanine\textsuperscript{2} → T\textsubscript{\textsuperscript{7}}A changes, and 3\% were G\textsuperscript{\textsubscript{7,8}}-dihydro-8-oxoguanine\textsuperscript{2} → T\textsubscript{\textsuperscript{7}}A substitutions. These figures were compared to the \textit{pncA} gene, where out of 114 isolates, 30\% of the single nucleotide mutations were G\textsuperscript{\textsubscript{7,8}}-dihydro-8-oxoguanine\textsuperscript{2} → A\textsubscript{\textsuperscript{7}}T transitions and 9\% were G\textsuperscript{\textsubscript{7,8}}-dihydro-8-oxoguanine\textsuperscript{2} → T\textsubscript{\textsuperscript{7}}A changes.

\textbf{Conclusions:} While there is an excess of G\textsuperscript{\textsubscript{7,8}}-dihydro-8-oxoguanine\textsuperscript{2} → A\textsubscript{\textsuperscript{7}}T changes in the \textit{rpoB} gene, this was not the case in the \textit{pncA} gene. Fifty-three percent of mutations within the \textit{rpoB} gene were C → T mutations of the type S531L. Although this mutation gives a fitness disadvantage, it is less than other common mutations, so it is more likely that that fitness is the determinant of surviving mutation rather than oxidative stress because of the small numbers of other C → T and G → A mutations at other sites (12\%). There was no evidence of oxygen free radicals damaging the guanine bases in either gene.

Keywords: \textit{M. tuberculosis}, \textit{rpoB} gene, \textit{pncA} gene

Introduction
To survive inside the macrophage, the tubercle bacillus must protect itself against intracellular bactericidal mechanisms including reactive oxygen intermediates.\textsuperscript{1,2} These may arise through the by-products of aerobic metabolism; hydrogen peroxide, hydroxyl radicals and superoxide or from environmental agents such as UV irradiation or redox-cycling agents. Reactive oxygen intermediates are bactericidal or bacteriostatic \textit{in vitro}, but their concentration in the human macrophage \textit{in vivo} is still uncertain. The genes \textit{katG}, which encodes a catalase-peroxidase protein, and \textit{aphC}, an alkyl hydroperoxide reductase protein, are believed to be involved in protecting mycobacteria against millimolar concentrations of hydrogen peroxide.\textsuperscript{3,4} Bacteria that become resistant to isoniazid have a reduced defence against oxidative damage due to a mutation in \textit{katG}.\textsuperscript{5} The effect of \textit{aphC} should be protective but Dhandayuthapani \textit{et al.} failed to demonstrate expression in \textit{M. tuberculosis} H37Rv, leading to the interpretation that expression of this gene is inversely correlated to isoniazid sensitivity in mycobacterial species that were tested.\textsuperscript{6,7} Oxygen free radicals cause the oxidation of guanine residues in DNA generating 7,8-dihydro-8-oxoguanine\textsuperscript{2} (8-oxodG), resulting in a ‘GO lesion’.\textsuperscript{7} The strong preference of guanine
bases for cytosine during replication is disrupted by 8-oxodG residues so that adenine can be misincorporated on the opposite strand. If the mismatch is not repaired, a further round of replication results in a G → T mutation. Adenine can be oxidized to form 8-oxodadenine (8-oxoDA) which is less mutagenic than 8-oxodG and correctly pairs with thymine but it has been shown to mispair with guanine in vitro.

*Mycobacterium tuberculosis* can repair the effects of oxidative damage using the GO repair system composed of at least three proteins: MutM, MutY and MutT. MutM can remove oxidized guanine from the DNA and subsequent repair restores the original G/C pair. If the GO lesion is not removed before replication, translesion synthesis can be accurate and result in C/G. However, translesion synthesis by replicative DNA polymerases is inaccurate, leading to the misincorporation of adenine opposite the GO lesion. The MutY protein removes this mismatch and repair by polymerases results in a C/G pair. The GO lesion may produce only a minor distortion of the DNA backbone so it is perhaps not the ideal substrate for nucleotide excision repair. An increase in A/T → C/G transversions has been shown in *mutT* mutant strains and in G/C → T/A transversions in *mutM* and *mutY* mutant strains.

Oxidation can cause DNA damage through oxidized cytosine-5-hydroxycytosine (5-OH-C) and/or 5-hydroxuracil (5-OH-U) leading to C → T transitions. Oxidation of cytosine to 5,6-dihydroxy-5,6-dihydrocytosine (Cg) breaks down further to 5-OH-C, 5-OH-U and 5,6-dihydroxy-5,6-dihydrouracil (Ug) which is poorly repaired leading to C → T or G → A transition unless the uracil is removed by a uracil DNA glycosylase encoded by *ung* in *M. tuberculosis*. 5-OH-U and Ug are likely sources of the observed C → T transitions as they are present in oxidatively damaged DNA. Both G/C → A/T base pairs are damaged by reactive oxygen species with about equal efficiency but recent studies suggest that G/C damage is more likely to lead to G/C → A/T transitions.

Some strains of *M. tuberculosis*, notably the Beijing family, are associated with an increased risk of resistance and some consider this may be due to an increased mutation rate. Sequencing studies show that there are miss-sense mutations in three putative *mut* genes but no experimental evidence is available that directly links these lineages with an increased mutation rate.

The aim of this study is to review the literature describing resistance mutations in phenotypically-resistant clinical isolates of *M. tuberculosis*. As rifampicin and pyrazinamide resistance is thought to be usually preceded by a *katG* mutation, review of the rifampicin resistance mutations may provide an indication of the impact of oxidative stress to the *M. tuberculosis* genome. If this hypothesis is correct, then damage to the guanine base will be observed by a high percentage of G → T mutations or on the opposite strand a C → A mutation. Alternatively, it may result in the deamination of cytosine which is the source of C → T transitions.

**Methods and results**

Relevant published studies were identified through computerized searches of the PubMed database (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi). The search terms that were entered into the NCBI database were: rifampicin AND tuberculosis AND sequencing AND resistance. The term rifampicin was also used and in subsequent searches, it was substituted with pyrazinamide. Single base pair changes conferring antimicrobial resistance that occurred on the sense strand of DNA were collated. Deletions and inversions were excluded from the analysis and representative isolates from outbreaks/clusters were counted once. For each isolate, the following data were recorded: base pair change, amino acid change and codon affected. A total of 33 papers were collected for analysis of *rpoB*, and seven papers were collected for analysis of *pncA*. Countries of origin were collected as a control factor for geographic bias. The papers describing the point mutations in the *rpoB* gene in clinical isolates originated from India, Taiwan, Japan, Shanghai, Brazil, Peru, Lebanon, Latvia, USA (New York, Pennsylvania), Mexico, Australia, South Africa, Turkey, Greece, Spain, Germany, Italy, Yemen, Thailand, Malaysia, Myanmar, Philippines, Canada, Korea, Singapore, Japan, Peru, Turkey, Canada, Bangladesh, Azerbaijan and Scotland.

Most papers examined for this review reported a sequence of a segment of the *rpoB* gene containing the 81 bp rifampicin resistance defining region (RRDR) to identify mutations or used a commercially available rapid test, the PCR based reverse hybridization line probe assay (Inno-LIPA Rif TB test, Innogenetics, Belgium) and real-time PCR using probes. IS6110 typing was used to differentiate between isolates and identify clusters. Garcia et al. evaluated PCR-ELISA to detect mutations in *M. tuberculosis* cultures and smear-positive specimens which agreed overall with DNA sequencing. Direct sequencing was used to analyse the total length (561 bp) of the *pncA* gene in all papers analysed.

A total of eight papers analysed for mutations in *rpoB* were excluded from the study as they contained only amino acid data and it was not possible to deduce the mutational events that led to the amino acid substitution. In some cases, it was possible to deduce the base pair change that had occurred just from the amino acid information. Another paper was excluded as the authors had induced spontaneous mutations in the laboratory strain *M. tuberculosis* H37Rv. One study was eliminated due to errors in the reported mutations.

The base pair change of C → T at codon S531L was the most commonly reported single base pair mutation in the *rpoB* gene of clinical isolates of *M. tuberculosis* occurring in 456 out of 840 isolates (Table 1). Possibly even more significant, this mutation occurs in 54% of isolates including deletions, insertions, single and double base pair changes when looking at 1050 clinical isolates. The next most common mutation in the *rpoB* gene is C → G (11%). C → A mutations occurred in 840 isolates (9%), G → T mutations occurred in 21 out of 840 isolates (2.5%). The majority of mutations reported occurred within the 81 bp RRDR, which is between codons 507 and 533 (Figure 1). Mutations occurring outside the RRDR were reported at codons 146, 381, 490, and 572. Mutations conferring resistance to pyrazinamide were evenly distributed along the *pncA*. There was no hotspot of mutation (Figure 2). Most frequent mutation in this gene is A → C occurring in 24 out of 114 isolates (21%). C → A and G → A mutations occurred equally in five out of 120 isolates (4.4%) (Table 2).
Discussion

The hypothesis tested in this study was that oxidative damage is the major mutational mechanism in *M. tuberculosis*. An excess of G·C to T·A mutations would indicate oxidative damage due to the production of 8-oxodG and an excess of C·G to T·A would be indicative of oxidative damage due to the production of 5-OH-C.

Most mutagenic processes will result in a transition (purine to purine/pyrimidine to pyrimidine). The C→T change is the most common base substitution arising from oxidative damage of DNA. The high percentage (66%) of C→T mutations in the *rpoB* gene could support the idea that oxidative stress drives mutations in this gene as a result of damage to dC residues. However, one should also see an equal percentage of G→A transitions as a result of the same lesion occurring on the opposite strand but in this review only 0.3% are G→A.

Another explanation of the high percentage of C→T mutations is that existing DNA repair mechanisms may be inefficient in recognizing this base pair change so this mutation is not corrected and becomes fixed in the population. However, this possibility can be eliminated when looking at the *pncA* gene as we found only 13.2% C→T mutations and on the opposite strand 16.7% G→A. A G·C→A·T transition occurs through the deamination of cytosine unless the uracil is removed by a uracil DNA glycosylase encoded by the *ung* gene in *M. tuberculosis*. The high percentage of C→T transitions in the *rpoB* gene may be due to the methylation of guanine, which is a well known inducer of G·C→A·T transitions, but this effect was only seen in *rpoB* not *pncA*.

The isolates in many of the studies were isoniazid-resistant so have a reduced capacity to deal with oxidative stress. This would result in a higher than predicted proportion of C·G to T·A and G·C to T·A, but this was not observed.

### Table 1. Distributions of mutations within the *rpoB* gene of 840 clinical isolates of *M. tuberculosis*

<table>
<thead>
<tr>
<th>Nucleotide change</th>
<th>Codon (no. of isolates)</th>
<th>Total no. of isolates (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C→T</td>
<td>381 (1), 521 (1), 522 (9), 526 (89), 531 (456)</td>
<td>556 (66.2)</td>
</tr>
<tr>
<td>C→G</td>
<td>511 (1), 522 (2), 526 (58), 531 (34)</td>
<td>95 (11.3)</td>
</tr>
<tr>
<td>C→A</td>
<td>509 (2), 513 (3), 522 (2), 526 (1)</td>
<td>8 (0.9)</td>
</tr>
<tr>
<td>G→C</td>
<td>528 (2)</td>
<td>2 (0.2)</td>
</tr>
<tr>
<td>G→T</td>
<td>146 (1), 490 (1), 516 (18), 527 (1)</td>
<td>21 (2.5)</td>
</tr>
<tr>
<td>A→G</td>
<td>521 (1), 532 (2)</td>
<td>3 (0.3)</td>
</tr>
<tr>
<td>T→C</td>
<td>511 (5), 533 (20)</td>
<td>25 (3)</td>
</tr>
<tr>
<td>A→A</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>G→A</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>A→C</td>
<td>513 (7), 518 (3), 526 (1)</td>
<td>11 (1.3)</td>
</tr>
<tr>
<td>A→G</td>
<td>513 (1), 515 (1), 516 (4), 526 (24)</td>
<td>30 (3.6)</td>
</tr>
<tr>
<td>A→T</td>
<td>513 (6), 516 (69), 526 (13), 572 (1)</td>
<td>89 (10.6)</td>
</tr>
</tbody>
</table>

*a*Indicates transitions while all other nucleotide changes are transversions.

### Figure 1. Distribution of mutations within the *rpoB* gene in *M. tuberculosis*. The arrows and numbers indicate the number of codons that were omitted from the diagram as they had no mutation.

### Figure 2. Distribution of mutations within the *pncA* gene in *M. tuberculosis*. All mutations with no number were reported once. The arrows and numbers indicate the number of codons that were omitted from the diagram as they had no mutation.
Analysis of rpoB and pncA mutations in the published literature

Table 2. Distributions of mutations within the pncA gene of 114 clinical isolates of M. tuberculosis

<table>
<thead>
<tr>
<th>Nucleotide change</th>
<th>Codon (no. of isolates)</th>
<th>Total no. of isolates (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C → T&lt;sup&gt;a&lt;/sup&gt;</td>
<td>10 (1), 46 (6), 51 (1), 76 (2), 134 (2), 141 (1), 142 (1), 171 (1)</td>
<td>15 (13.2)</td>
</tr>
<tr>
<td>C → G</td>
<td>57 (1), 103 (1)</td>
<td>2 (1.8)</td>
</tr>
<tr>
<td>C → A</td>
<td>46 (2), 62 (1), 148 (1), 168 (1)</td>
<td>5 (4.4)</td>
</tr>
<tr>
<td>G → C</td>
<td>9 (1), 68 (1), 121 (1), 136 (1), 138 (1)</td>
<td>5 (4.4)</td>
</tr>
<tr>
<td>G → T</td>
<td>1 (1), 7 (2), 136 (1), 180 (1)</td>
<td>5 (4.4)</td>
</tr>
<tr>
<td>G → A&lt;sup&gt;a&lt;/sup&gt;</td>
<td>7 (1), 14 (1), 17 (1), 68 (3), 78 (1), 97 (3), 103 (1), 136 (2), 139 (4), 146 (1), 171 (1)</td>
<td>19 (16.7)</td>
</tr>
<tr>
<td>T → C&lt;sup&gt;c&lt;/sup&gt;</td>
<td>4 (1), 9 (1), 14 (1), 58 (2), 59 (1), 67 (1), 68 (2), 85 (1), 103 (3), 139 (1), 155 (1), 172 (1)</td>
<td>16 (14.0)</td>
</tr>
<tr>
<td>T → A</td>
<td>7 (1), 9 (1), 45 (1), 64 (1), 68 (1), 71 (1), 85 (3), 139 (1), 155 (2), 159 (1)</td>
<td>1 (0.9)</td>
</tr>
<tr>
<td>T → G</td>
<td>4 (1), 9 (1), 49 (1), 53 (1), 57 (1), 63 (1), 76 (2), 96 (1), 103 (2), 135 (2), 137 (2), 141 (3), 142 (1)</td>
<td>13 (11.4)</td>
</tr>
<tr>
<td>A → C</td>
<td>10 (5), 12 (2), 49 (1), 53 (1), 57 (1), 63 (1), 76 (2), 96 (1), 103 (2), 135 (2), 137 (2), 141 (3), 142 (1)</td>
<td>24 (21.1)</td>
</tr>
<tr>
<td>A → G&lt;sup&gt;c&lt;/sup&gt;</td>
<td>8 (1), 10 (2), 47 (1), 51 (1), 82 (1), 96 (3)</td>
<td>9 (7.9)</td>
</tr>
<tr>
<td>A → T</td>
<td>0</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Indicates transitions while all other nucleotide changes are transversions.

In the pncA gene, single nucleotide mutations are the major form of genetic change<sup>37</sup> making it a useful control for rpoB. Although only seven research articles were available, a diverse mutation profile is found for pncA gene. There are 1235 possible single base pair mutations at 87 positions that have been reported.<sup>38</sup> It should be noted that only mutations that confer a resistant phenotype were analysed, single nucleotide mutations caused by oxidative damage that do not alter changes in the resistance phenotype are missed. An exception was the mutation found in the rpoB gene at amino acid Leu-521.<sup>23,27</sup>

The C → T mutation occurring at the S531L position was most common<sup>64</sup> but other C → T and G → A mutations were rare (12.3%). We have previously shown that the frequency of clinical isolation of different mutants correlates with their relative fitness indicating that strains with this mutation are better able to survive.<sup>64,65</sup> However, some resistance mutations appear to incur no fitness cost when measured in vitro, for example, katG mutations in isoniazid-resistant M. tuberculosis<sup>5</sup> so this argument may not be generalized. The fitness deficit resulting from resistance mutations in rpoB gene may be a unique feature of this gene and similar findings have been reported in the rpoB gene in Staphylococcus aureus mutants that had reduced fitness.<sup>56</sup> It must be recognized, therefore, that only a very limited set of mutations are likely to be consistent with retaining gene function whilst providing resistance to rifampicin suggesting the possibility that the observed mutation pattern of a high C → T percentage is the result of selective pressure.

In this paper, the hypothesis that oxidative damage is the major driving force behind mutations in M. tuberculosis is tested. There is evidence in the rpoB gene that oxidative damage leads to C → T base substitutions (induced by 5-OH-C) but this is not mirrored in the pncA gene. Also the opposite mutation G → A occurs at <0.4% in the rpoB gene. Our analysis shows that G-C→T-A mutations occur at low frequencies in both genes, which may indicate that M. tuberculosis is sufficiently competent in the repair of oxidized guanine mediated by the MutM/MutY proteins. This leads us to conclude that oxidative damage is not the driving force behind mutations in the M. tuberculosis genome and instead it is the relative fitness of the mutant strain coupled with the resistance phenotype that permits it to survive and be detected.

Acknowledgements

This work was part funded by a grant from the British Society for Antimicrobial Chemotherapy.

References

2. Russell DG, Dant J and Sturgill-Koszycki S (1996) *Mycobacterium avium* and *Mycobacterium tuberculosis*-containing vacuoles are dynamic, fusion-competent vesicles that are accessible to glycosphingolipids from the host cell plasmalemma. *Journal of Immunology* 156,4764–4773.
7. Michaels ML, Cruz C, Grollman AP et al. (1992) Evidence that MutY and MutM combine to prevent mutations by an oxidatively...
damaged form of guanine in DNA. *Proceedings of the National Academy of Sciences, USA* **89**,7022–7025.


44. Heep M, Brandstatter B, Rieger U et al. (2001) Frequency of rpoB mutations inside and outside the cluster I region in rifampin-


