(54) Title: MUTATIONS IN THE katG GENE USEFUL FOR DETECTION OF M. TUBERCULOSIS

A method for selectively detecting M. tuberculosis is provided employing restriction fragment length polymorphism analysis of an enzymatic digest of the M. tuberculosis katG gene.
# FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AL</td>
<td>Albania</td>
<td>ES</td>
<td>Spain</td>
<td>LS</td>
<td>Lesotho</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AM</td>
<td>Armenia</td>
<td>FI</td>
<td>Finland</td>
<td>LT</td>
<td>Lithuania</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT</td>
<td>Austria</td>
<td>FR</td>
<td>France</td>
<td>LU</td>
<td>Luxembourg</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AU</td>
<td>Australia</td>
<td>GA</td>
<td>Gabon</td>
<td>LV</td>
<td>Latvia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AZ</td>
<td>Azerbaijan</td>
<td>GB</td>
<td>United Kingdom</td>
<td>MC</td>
<td>Monaco</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BA</td>
<td>Bosnia and Herzegovina</td>
<td>GE</td>
<td>Georgia</td>
<td>MD</td>
<td>Republic of Moldova</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BB</td>
<td>Barbados</td>
<td>GH</td>
<td>Ghana</td>
<td>MG</td>
<td>Madagascar</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BE</td>
<td>Belgium</td>
<td>GN</td>
<td>Guinea</td>
<td>MK</td>
<td>The former Yugoslavia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BF</td>
<td>Burkina Faso</td>
<td>GR</td>
<td>Greece</td>
<td>ML</td>
<td>Mali</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BG</td>
<td>Bulgaria</td>
<td>HU</td>
<td>Hungary</td>
<td>MN</td>
<td>Mongolia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BJ</td>
<td>Benin</td>
<td>IE</td>
<td>Iceland</td>
<td>MR</td>
<td>Mauritania</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BR</td>
<td>Brazil</td>
<td>IL</td>
<td>Israel</td>
<td>MW</td>
<td>Malawi</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BY</td>
<td>Belarus</td>
<td>IS</td>
<td>Iceland</td>
<td>MX</td>
<td>Mexico</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CA</td>
<td>Canada</td>
<td>IT</td>
<td>Italy</td>
<td>NE</td>
<td>Niger</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CF</td>
<td>Central African Republic</td>
<td>JP</td>
<td>Japan</td>
<td>NL</td>
<td>Netherlands</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG</td>
<td>Congo</td>
<td>KE</td>
<td>Kenya</td>
<td>NO</td>
<td>Norway</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CH</td>
<td>Switzerland</td>
<td>KG</td>
<td>Kyrgyzstan</td>
<td>NZ</td>
<td>New Zealand</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CI</td>
<td>Côte d’Ivoire</td>
<td>KP</td>
<td>Democratic People’s</td>
<td>PL</td>
<td>Poland</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CM</td>
<td>Cameroon</td>
<td>KR</td>
<td>Republic of Korea</td>
<td>PT</td>
<td>Portugal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CN</td>
<td>China</td>
<td>KZ</td>
<td>Kazakhstan</td>
<td>RO</td>
<td>Romania</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CU</td>
<td>Cuba</td>
<td>LC</td>
<td>Saint Lucia</td>
<td>RU</td>
<td>Russian Federation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CZ</td>
<td>Czech Republic</td>
<td>LI</td>
<td>Liechtenstein</td>
<td>SD</td>
<td>Sudan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DE</td>
<td>Germany</td>
<td>LK</td>
<td>Sri Lanka</td>
<td>SE</td>
<td>Sweden</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DK</td>
<td>Denmark</td>
<td>LR</td>
<td>Liberia</td>
<td>SG</td>
<td>Singapore</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EE</td>
<td>Estonia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI</td>
<td>Slovenia</td>
<td>SK</td>
<td>Slovakia</td>
<td>SN</td>
<td>Senegal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SZ</td>
<td>Swaziland</td>
<td>TD</td>
<td>Chad</td>
<td>TG</td>
<td>Togo</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TJ</td>
<td>Tajikistan</td>
<td>TM</td>
<td>Turkmenistan</td>
<td>TR</td>
<td>Turkey</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TT</td>
<td>Trinidad and Tobago</td>
<td>UA</td>
<td>Ukraine</td>
<td>UG</td>
<td>Uganda</td>
<td></td>
<td></td>
</tr>
<tr>
<td>US</td>
<td>United States of America</td>
<td>UZ</td>
<td>Uzbekistan</td>
<td>VN</td>
<td>Viet Nam</td>
<td></td>
<td></td>
</tr>
<tr>
<td>YU</td>
<td>Yugoslavia</td>
<td>ZW</td>
<td>Zimbabwe</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
MUTATIONS IN THE katG GENE USEFUL FOR DETECTION OF M. TUBERCULOSIS

This is a continuation-in-part application U.S. patent application Serial No. 08/418,782, filed April 7, 1995, which is a continuation-in-part of U.S. patent application Serial No. 08/228,662, filed April 18, 1994, both of which are incorporated herein by reference.

Background of the Invention

Despite more than a century of research since the discovery of Mycobacterium tuberculosis, the aetiological agent of tuberculosis, this disease remains one of the major causes of human morbidity and mortality. There are an estimated 3 million deaths annually attributable to tuberculosis (see, D. Snider, Rev. Inf. Dis., S335 (1989)), and although the majority of these are in developing countries, the disease is assuming renewed importance in the West due to the increasing number of homeless people and the impact the AIDS epidemic (see, R.E. Chaisson et al., Am. Res. Resp. Dis., 23, 56 (1987); D. E. Snider, Jr. et al., New Engl. J. Med., 326, 703 (1992); M.A. Fischl et al., Ann. Int. Med., 117, 177 (1992) and ibid, at 184.


Several investigators have associated the toxicity of INH for mycobacteria with endogenous catalase activity. See, for example, "Isonicotinic
acid hydrazide," in F. E. Hahn, *Mechanism of Action of Antibacterial Agents*, Springer-Verlag (1979) at pages 98-119. This relationship was strengthened by a recent report by Ying Zhang and colleagues in *Nature*, **358**, 591 (1992) which described the restoration of INH susceptibility in an INH resistant


A conclusive diagnosis of tuberculosis depends on the isolation and identification of the etiologic agent, *Mycobacterium tuberculosis*, which generally requires 3-8 weeks. Design of an appropriate therapeutic regimen depends on the results of subsequent antituberculosis susceptibility testing by the agar dilution method and produces additional delays of 3-6 weeks (Roberts et al., "Mycobacterium" in *Manual of Clinical Microbiology*, 5th Ed.; A. Balows et al., Eds.; American Society for Microbiology: Washington; pp. 304-399 (1991). Identification and drug resistance testing can now also be accomplished more quickly by using the BACTEC radiometric method. (Tenover et al., *J. Clin. Microbiol.*, **31**, 767-779 (1993) and Huebner et al., *J. Clin. Microbiol.*, **31**, 771-775 (1993). Acid fast bacilli are detected in the BACTEC bottle, and and identification is made using a nucleic acid hybridization technique on the BACTEC-derived growth. Drug susceptibility testing is then conducted using the same BACTEC growth to inoculate fresh BACTEC bottles containing various antituberculosis drugs. This procedure reduces the time needed to generate a complete analysis, but the total time required to report susceptibility results for MTB is typically in excess of 20 days. The need to minimize the transmission of newly identified drug resistant strains of MTB requires the development of much more rapid identification procedures.

The rapid detection of *M. tuberculosis* directly from clinical samples has been possible recently by virtue of the availability of polymerase
chain reaction (PCR) and the recognition of diagnostic sequences amplified by the appropriate primers. The ability to conduct PCR analyses depends on having a high enough gene or gene product concentration so that the molecular tools work efficiently even when the organism numbers are low. Thus, the most efficient molecular assays used to detect *M. tuberculosis* depend on the IS6110 insertion sequence (about 10 copies) or the 16S ribosomal RNA (thousands of copies). See, respectively, K.D. Eisenach et al., *J. Infect. Dis.*, 61, 997 (1990) and N. Miller et al., *Abstracts ASM*, Atlanta, GA (1993) at page 177. However, these methods do not provide any information regarding the drug-resistance phenotype of the *M. tuberculosis* strain.

Recently, B. Heym et al. (PCT WO 93122454) disclose the use of polymerase chain reaction to amplify portions of the *katG* gene of putative resistant strains. The PCR products were evaluated by single-strand conformation polymorphism (SSCP) analysis, wherein abnormal strand motility on a gel is associated with mutational events in the gene. For example, in five strains, a single base difference was found in a 200 bp sequence, a G to T transversion at position 3360. This difference would result in the substitution of Arg-461 by Leu. However, carrying out SSCP on a given clinical sample can be a laborious procedure that requires sequencing to confirm whether mutations or deletions predictive of drug resistance are in fact present in the target gene.

There is a continuing need in the art to develop a simple test permitting the rapid identification of *M. tuberculosis* and its drug-resistance phenotype.
Summary of the Invention

The present invention provides a method to rapidly identify strains of *M. tuberculosis* which are resistant to isoniazid (INH). The method is based on our discovery that certain mutations in the katG gene of *M. tuberculosis* which confer INH resistance coincidentally result in the addition or deletion of restriction sites, which are recognized by various restriction enzymes. For example, the wild-type (WT) katG gene of *M. tuberculosis* contains an *NciI-MspI* restriction site spanning codon 463, which site is absent in the corresponding codon 463 in a number of INH resistant strains due to a single base mutation in that codon. An *NciI-MspI* restriction site is a site cleaved by both *NciI* and *MspI*. This site is represented by the nucleotide sequence CCGGG (see Table 1).

Alternatively, or in addition, some INH resistant strains have a single base mutation in codon 315 in the katG gene that produces a new *MspI* restriction site associated with the corresponding codon 315 in the INH resistant strain. Some INH resistant strains have a single base mutation at codon 337 that results in the deletion of a *RsaI* restriction site otherwise present at the corresponding position in the WT gene; and some have a single base mutation at codon 264 that eliminates a *CfoI* restriction site. These mutations may be present singly or in combination in INH resistant *M. tuberculosis* strains.

When used in reference to nucleotide position, codon position or restriction site position, the term "corresponding" is defined to mean the same absolute location on two different *M. tuberculosis* katG genes, wherein absolute location is defined by the numbering system used in Figure 7 (SEQ ID NO:20). For example, a wild-type codon 463 represented by CGG at nucleotide positions 1456-1458 on a wild-type katG gene of *M. tuberculosis* and a mutant codon 463 represented by CTG at the same nucleotide positions 1456-1458 on a katG gene of an INH resistant strain of *M. tuberculosis* are considered to be corresponding codons.

The determination of whether one or more of these identifying mutations in the katG gene are present in a strain of *M. tuberculosis* can be made by employing the techniques of restriction fragment length polymorphism (RFLP)
analysis. Therefore, in an embodiment directed to the identification of a mutation in codon 463 that is associated with INH resistance, the present assay comprises the steps of:

(a) amplifying a portion of the katG gene of an *M. tuberculosis* isolate to yield a detectable amount of DNA comprising the nucleotide position occupied by base 1457 of the *M. tuberculosis* katG gene consensus sequence depicted in Figure 7 (SEQ ID NO:20); and

(b) determining whether an *N*<sup>c</sup>-I-*M*<sup>sp</sup>I restriction site is absent in codon 463 of said katG gene, wherein said absence is indicative of an INH resistant strain of *M. tuberculosis*.

The RFLP technique involves cleaving the DNA with a restriction endonuclease which cleaves at an *N*<sup>c</sup>-I-*M*<sup>sp</sup>I restriction site to yield at least one DNA fragment and determining whether the number and location of the fragments is indicative of the absence of an *N*<sup>c</sup>-I-*M*<sup>sp</sup>I restriction site in codon 463 of said katG gene, wherein said absence is indicative of an INH resistant strain of *M. tuberculosis*, preferably by employing the techniques of gel electrophoresis.

If the amplified DNA of step (a) contains no *N*<sup>c</sup>-I-*M*<sup>sp</sup>I restriction sites, then the DNA fragment yielded in step (b) will be identical to the amplified DNA of step (a). This can occur where the portion of the katG gene amplified in step (a) is from an INH resistant strain of *M. tuberculosis* having a mutation in codon 463 that removes the *N*<sup>c</sup>-I-*M*<sup>sp</sup>I restriction site spanning that codon in the wild-type katG gene, and having no other additional *N*<sup>c</sup>-I-*M*<sup>sp</sup>I restriction sites.

In order for the amplified DNA to yield a meaningful RFLP pattern, the portion of the katG gene amplified in step (a) will be of sufficient length to produce fragments of sufficient length to visualize using gel electrophoresis. In the above-described embodiment, for example, the portion amplified will contain a sufficient number of bases to either side (5′ or 3′) of codon 463 such that cleavage at a site spanning that codon will yield fragments that can be visualized using gel electrophoresis.

In another embodiment of the invention directed to the additional identification of a mutation in codon 315 associated with INH resistance, the
amplified DNA of step (a) further comprises at least one MspI restriction site and
the nucleotide position occupied by base 1013 (Figure 7, SEQ ID NO:20), and
the determination made in step (b) further includes whether an MspI restriction
site associated with codon 315 is present, wherein said presence is indicative of
an INH resistant strain. For example, RFLP can also be employed to determine
whether the number and location of the fragments is indicative of the codon 315
MspI restriction site. Preferably, the portion of the katG locus which is amplified
is a minor portion of the entire katG gene, i.e., less than 1500 base pair, more
preferably less than 1000 base pair, and is isolated and amplified by polymerase
chain reaction, as described hereinbelow. The term "location" refers to the Rf
(relative electrophoretic mobility) of a given fragment on the gel.

The pattern of fragments produced on a gel by electrophoresis of a
restriction digest of an amplified portion of the katG gene of an M. tuberculosis
strain of interest, such as an INH resistant strain, is preferably compared to the
pattern produced in a digest of an equivalent portion of the katG gene of a
wild-type (WT) control strain of M. tuberculosis, which strain is INH sensitive.
The term "equivalent" is defined herein to mean that any two portions of the katG
gene would comprise the same number and location of restriction sites being
analyzed (e.g., sites recognized by CfoI, RsaI, MspI, and/or NciI) if the portions
both were selected from a portion of the DNA of SEQ ID NO: 20 (i.e., if there
were no mutations altering the number of restriction sites of the type being
analyzed), and that the portions do not differ in size before cleavage to the extent
that the number of fragments obtained cannot be compared following side-by-side
gel electrophoresis and visualization of the resultant fragments, as described
hereinbelow. For example, the control katG DNA can correspond to an
equivalent portion of SEQ ID NO:20 (Figure 7, upper sequence) comprising one
or more of the codons of interest (e.g., codons 315 or 463) and their associated
restriction sites. As discussed below, such a portion of DNA can be derived from
strain H37Rv MC. A positive control corresponding to DNA fragments derived
from a known INH resistant strain may also be used.
In the embodiment of the assay of the invention directed to the determination of the presence or absence of a NciI-MspI restriction site associated with codon 463, gel electrophoresis is employed to compare the number and location of the DNA fragments to the number and location of DNA fragments derived from cleavage of DNA derived from an equivalent portion of the katG gene wherein the NciI-MspI restriction site at codon 463 is present, wherein a determination of the absence of the restriction site at codon 463 in the katG gene is indicative of an INH resistant strain of M. tuberculosis. Preferably, the control DNA sequence of the portion of the katG gene wherein the codon 463 restriction site is present corresponds to a portion of SEQ ID NO.20 (Figure 7, upper sequence). For example, the control DNA may contain five NciI-MspI restriction sites in each DNA molecule prior to cleavage, and the DNA of step (a), which is derived from an INH resistant strain, may contain four NciI-MspI restriction sites in each DNA molecule prior to cleavage. The assay also preferably includes positive control DNA fragments derived from an INH resistant strain which does not include the codon 463 NciI-MspI restriction site in the katG gene.

The present invention also provides method for selectively detecting M. tuberculosis in a DNA sample, wherein the DNA is amplified to generate a detectable amount of amplified DNA comprising a katG DNA fragment which consists of base 904 through base 1523 of the M. tuberculosis katG gene. The generation of this katG DNA fragment is indicative of the presence of M. tuberculosis in the sample. Preferably, the DNA sample is a human biological tissue or fluid, more preferably a human biological fluid. The DNA sample is most preferably human sputum. Advantageously, the method of the invention can be performed on clinical human sputum samples with minimal pretreatment of the clinical sample.

In a preferred embodiment of the M. tuberculosis detection method, the katG DNA fragment has a restriction site that comprises either a G or a C at the nucleotide position occupied by base 1013 in codon 315 of the M. tuberculosis katG gene as depicted in Figure 7 (SEQ ID NO: 20), and the method
further comprises contacting the katG DNA fragment with a restriction
endonuclease, preferably MspI, that cleaves either at the restriction site
comprising a G at the nucleotide position occupied by base 1013 of codon 315,
or at the restriction site comprising a C at the nucleotide position occupied by
base 1013 of codon 315, but not both, yielding at least one cleaved fragment.
The at least one cleaved fragment is electrophoresed to yield an electrophoretic
mobility pattern comprising the at least one cleaved fragment, and the mobility
pattern is analyzed to selectively detect the presence of M. tuberculosis in the
sample. Preferably, restriction fragment length polymorphism (RFLP) analysis is
used to analyze the electrophoretic mobility pattern generated by gel
electrophoresis of the cleaved fragments to detect M. tuberculosis.

In another embodiment of the M. tuberculosis detection method,
M. tuberculosis is selectively detected in a DNA sample by:

(a) amplifying the DNA to generate a detectable amount of
amplified DNA comprising a katG DNA fragment comprising base 904 through
base 1523 of the M. tuberculosis katG gene, wherein the katG DNA fragment
further comprises a restriction site comprising either a G or a C at the nucleotide
position occupied by base 1013 in codon 315 of the M. tuberculosis katG gene as
depicted in Figure 7 (SEQ ID NO: 20);

(b) contacting the katG DNA fragment with a restriction
endonuclease, preferably MspI, that cleaves either at said restriction site
comprising a G at the nucleotide position occupied by base 1013 of codon 315,
or at said restriction site comprising a C at the nucleotide position occupied by
base 1013 of codon 315, but not at both of said restriction sites, to yield at least
one cleaved fragment;

c) electrophoresing, preferably using gel electrophoresis, the at
least one cleaved fragment to yield an electrophoretic mobility pattern comprising
the at least one cleaved fragment; and

d) analyzing the mobility pattern, preferably using RFLP, to
selectively detect the presence of M. tuberculosis in the sample.
In a further embodiment of the *M. tuberculosis* detection method, *M. tuberculosis* is selectively detected a sample containing DNA by:

(a) amplifying the DNA to generate a detectable amount of amplified DNA comprising a *katG* DNA fragment comprising base 904 through base 1523 of the *M. tuberculosis* *katG* gene as depicted in Figure 7 (SEQ ID NO: 20);

(b) contacting the *katG* DNA fragment with a restriction endonuclease, preferably *MspI*, that cleaves at C/CGG to yield at least one cleaved fragment;

(c) electrophoresing, preferably using gel electrophoresis, the at least one cleaved fragment to yield a mobility pattern comprising the at least one cleaved fragment; and

(d) analyzing the mobility pattern, preferably using RFLP, to selectively detect the presence of *M. tuberculosis* in the sample.

In a particularly preferred embodiment, the *M. tuberculosis* detection method further comprises determining whether or not the *katG* DNA fragment has a S315T mutation, preferably using a restriction digest followed by gel electrophoresis of the digested DNA and RFLP analysis of the electrophoretic mobility patterns, wherein the presence of a S315T mutation is indicative of an INH-resistant strain of *M. tuberculosis*.

The present invention also provides oligonucleotides and subunits thereof useful in pairs as primers to initiate the polymerase chain reaction (PCR). Subunits of at least seven bases in length are preferred. PCR is useful both to amplify *katG* DNA so as to prepare both the target DNA of step (a) of the present process, as well as the DNA which is used to prepare the control digest.

The present invention also provides isolated, purified DNA represented by the consensus sequence derived for the *M. tuberculosis* *katG* gene. This DNA was found to occur in nature as the *katG* gene of *M. tuberculosis* strain H37Rv MC, as maintained at the Mayo Clinic, and is also referred to as the wild-type (WT) DNA. The present invention also includes isolated, purified DNA encoding the consensus amino acid sequence encoded by
the consensus wild-type \textit{katG} DNA, as well as DNA sequences that differ in
sequence but which also encode this amino acid sequence (a consensus catalase
peroxidase polypeptide) and can be employed to provide the isolated, purified
polypeptide represented by the consensus amino acid sequence, which

polypeptide is also provided by the invention.

The polypeptide of the invention can be prepared by expression in
transformed host cells, such as bacteria, yeast, plant, or insect cells transformed
with the DNA sequences of the present invention, operatively linked to regulatory
regions functional in the transformed host cells. The polypeptide can be used as a

standard \textit{M. tuberculosis} catalase peroxidase, to correlate enzymatic activity
(relative level, loss and restoration), with INH modification and degradation and
drug resistance in \textit{M. tuberculosis}.

The present invention also provides a kit comprising, separately
packaged in association:

(a) a pair of oligonucleotide primers selected so as to amplify
a portion of the DNA of the \textit{M. tuberculosis} \textit{katG} gene
comprising base 1457 in codon 463 or base 1013 in codon
315, as depicted in Figure 7 (SEQ ID NO:20); and

(b) an amount of a restriction endonuclease such as \textit{MspI},
effective to cleave the amplified portion of said DNA at a
restriction site comprising said base 1457 or said base
1013.

The present kits will also preferably comprise instruction means
for carrying out the present assay, i.e., a printed package insert, tag or label, or an
audio or video tape. The present kits will also preferably comprise a control
DNA digest prepared by amplifying a portion of the consensus DNA of SEQ ID
NO: 20 (Figure 7), that is equivalent to the portion defined and amplified by the

pair of primers, followed by digestion of the DNA with a suitable restriction
endonuclease such as \textit{MspI}.

The present invention is exemplified by the use of \textit{NcoI}, \textit{MspI},
\textit{CfoI}, and \textit{RsaI} digestions, with the use of \textit{MspI} digestion being preferred;
however, any restriction endonuclease having a restriction site spanning all or a
portion of codon 463, codon 315, codon 337, or codon 264, which portion
contains the site of the single base mutation associated with INH resistance as identified in Table 2, may be used, as desired. For example, the restriction endonucleases listed in Table 1 can be employed. Particularly preferred are restriction endonucleases having a restriction site that contains the position occupied by base 1457 in codon 463, or base 1013 in codon 315, as depicted in Figure 7 (SEQ ID NO:20).
<table>
<thead>
<tr>
<th>Specificity</th>
<th>Restriction Site</th>
<th>Restriction Enzyme</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>C/CGC</td>
<td>AcI&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>GC/NGC&lt;sup&gt;b&lt;/sup&gt;</td>
<td>BsoFI, Fmu4HI, Bsp61,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BssFI, BssXI, Cac824I,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CcoP2151, CcoP2161, FbrI,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ital, Utr960I</td>
</tr>
<tr>
<td>10</td>
<td>R/GCGCY&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Bsp143II, HaeII, Bme14ZI,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BsmHI, Bst1473II, Bsr161,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Btr34II, HinHI, LpnI, Ngol</td>
</tr>
<tr>
<td>15</td>
<td>G/CGC</td>
<td>CfoI, HhaI, BcaI, CcoP951,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Csp1470I, FnuDIII, Hin61,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hin7I, HinGUI, HinPIL,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IllnSII, IllnS2I, MnnIV, SciNI</td>
</tr>
<tr>
<td>20</td>
<td>GACGCNNNNNN/NNNNNN</td>
<td>Hgal</td>
</tr>
<tr>
<td></td>
<td>(SEQ ID NO.22)</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>GT/AC</td>
<td>Rsal, AflI,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Asp16HI, Asp17HI,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Asp18HI, Asp29HI,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CcoP73I, Csp6I, CviQI,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CviRII</td>
</tr>
<tr>
<td>30</td>
<td>GC/NGC</td>
<td>BsoFI, Fmu1HI, Bsp61,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BssFI, BssXI, Cac824I,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CcoP2151, CcoP2161, FbrI,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ital, Utr960I</td>
</tr>
<tr>
<td>35</td>
<td>C/CGC</td>
<td>AcI&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>40</td>
<td>C/CGC</td>
<td>AcI&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>GC/NGC</td>
<td>BsoFI, Fmu4HI, Bsp61,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BssFI, BssXI, Cac824I,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CcoP2151, CcoP2161, FbrI,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ital, Utr960I</td>
</tr>
<tr>
<td>45</td>
<td>CMG/CKG&lt;sup&gt;d&lt;/sup&gt;</td>
<td>MspAlII, NspBII</td>
</tr>
<tr>
<td></td>
<td>R/CCGGY&lt;sup&gt;e&lt;/sup&gt;</td>
<td>BsrFI, Cfr10I, Bco118I,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bse118I, Bsp21I, BssAI</td>
</tr>
</tbody>
</table>
C/CGG

\( MspI, Bsu192I, BsuFl, \)
\( FinII, HapII, Hin2I, Hin5I, \)
\( HpaII, MnlII, MnoI, MspI, \)
\( Pde137I, Pme35I, SecII, \)
\( SfaGI, SfiI34I, \)
\( Uba1128I, Uba1141I, \)
\( Uba1267I, Uba1338I, \)
\( Uba1355I, Uba1439I \)

5

Cuts 463-R
(sensitive)

CC/SGG\(^f\)

\( NciI, BcnI, AhaI \)

C/CGG

\( MspI, Bsu192I, BsuFl, \)
\( FinII, HapII, Hin2I, Hin5I, \)
\( HpaII, MnlII, MnoI, MspI, \)
\( Pde137I, Pme35I, SecII, \)
\( SfaGI, SfiI34I, \)
\( Uba1128I, Uba1141I, \)
\( Uba1267I, Uba1338I, \)
\( Uba1355I, Uba1439I \)

10

Cuts 463-L
(resistant)

CAG/NNN/CTG

\( AlwNI \)

CC/WGG\(^g\)

\( BstNI, BstOI, MvaI \)

20

/CCWGG

\( EcoRII \)

---

\(^a\)\( AcII \) cleaves the complementary strand of the \( katG \) gene; \(^b\)\( N = C \) or \( G \) or \( A \) or \( T \);
\(^c\)\( R = A \) or \( G \); \(^d\)\( M = A \) or \( C \), \( K = G \) or \( T \); \(^e\)\( Y = C \) or \( T \); \(^f\)\( S = C \) or \( G \); \(^g\)\( W = A \) or \( T \).
<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>264-A (sensitive)</td>
<td>(SEQ ID NO:8)</td>
<td>847</td>
<td>GTC GAA ACA GCG <strong>GCG</strong> CTG ATC GTC GGC</td>
<td>873</td>
<td></td>
</tr>
<tr>
<td>5 264-T (resistant)</td>
<td>(SEQ ID NO:9)</td>
<td></td>
<td>GTC GAA ACA GCG <strong>ACG</strong> CTG ATC GTC GGC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>337-Y (sensitive)</td>
<td>(SEQ ID NO:10)</td>
<td>1066</td>
<td>CTC GAG ATC CTG <strong>TAC</strong> GGC TAC GAG TGG</td>
<td>1092</td>
<td></td>
</tr>
<tr>
<td>337-C (resistant)</td>
<td>(SEQ ID NO:11)</td>
<td></td>
<td>CTC GAG ATC CTG <strong>TGC</strong> GGC TAC GAG TGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 315-S (sensitive)</td>
<td>(SEQ ID NO:12)</td>
<td>1000</td>
<td>GAC GCG ATC ACC <strong>AGC</strong> GGC ATC GAG GTC</td>
<td>1026</td>
<td></td>
</tr>
<tr>
<td>315-T (resistant)</td>
<td>(SEQ ID NO:13)</td>
<td></td>
<td>GAC GCG ATC ACC <strong>ACC</strong> GGC ATC GAG GTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 463-R (sensitive)</td>
<td>(SEQ ID NO:14)</td>
<td>1444</td>
<td>AAG AGC CAG ATC <strong>CGG</strong> GCA TCG GGA TTG</td>
<td>1470</td>
<td></td>
</tr>
<tr>
<td>463-L (resistant)</td>
<td>(SEQ ID NO:15)</td>
<td></td>
<td>AAG AGC CAG ATC <strong>CTG</strong> GCA TCG GGA TTG</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*The underlined codons represent the sites where the indicated single base mutations confer INH resistance. The bold bases indicate restriction sites as follows: G/CGC for *CfoI* in 264-A (sensitive); GT/AC for *RsaI* in 337-Y (sensitive); C/CGG for *MspI* in 315-T (resistant) and 463-R (sensitive). For ease of reference, the partial sequences shown in this table include the 12 bases to either side of the affected codon; the numbering system is the same as used for the wild-type consensus sequence in Figures 1 and 7. The full sequence of bases to either side of the affected codon is shown in Figure 7. In each of the sensitive/resistant pairs shown in this table, the upper sequence is the consensus, wild-type sequence (INH-sensitive) and the lower sequence is the mutant (INH-resistant) sequence.*

b Codon 264 GCG = Ala (A)
ACG = Thr (T)
c Codon 337 TAC = Tyr (Y)
TGC = Cys (C)
d Codon 315 AGC = Ser (S)
ACC = Thr (T)
ecod-on 463 CGG = Arg (R)
CTG = Leu (L)
**Brief Description of the Drawings**

Figure 1, panels A-D, depicts the consensus, wild-type DNA sequence of the *M. tuberculosis* *katG* gene as the upper of the pair of sequences (61-2295) (SEQ ID NO: 1). This DNA sequence data has been submitted to Gen Bank and has been assigned accession number U06262. The lower of the pair of sequences depicts nucleotide sequence 1970-4190 of the *KpnI* fragment bearing the *katG* gene as depicted in Figure 6 of Institute Pasteur et al. (published PCT application WO 93/22454). This sequence (SEQ ID NO: 2) has been deposited in the EMBL data library under accession number X68081 (Gen Bank X68081. gb_ba). Dots (.) above the sequence mark every tenth base. The upper sequence is in lower case in areas where variation in the sequence among isolates described hereinbelow and the consensus sequence was found. The arrow before position 70 and after position 2291 of the upper sequence indicate the coding sequence of the *katG* gene.

Figure 2 depicts the *katG* amino acid consensus sequence derived from 15 strains of *M. tuberculosis* (SEQ ID NO: 7).

Figure 3 schematically depicts the *NciI* restriction sites for the part B amplicon of *katG*. The (*) depicts the site of the Arg→Leu mutation which is found in some INH resistant *M. tuberculosis* strains.

Figure 4 depicts the results of a gel electrophoresis of the *NciI* digest of the part B amplicon of 14 strains of *M. tuberculosis* (1-14).

Figure 5 schematically depicts *MspI* and *RsaI* restriction sites and resulting RFLP fragments for a portion of the *M. tuberculosis* *katG* gene. For *MspI*, restriction maps for wild-type (W+), single (315 Ser→Thr or 463 Arg→Leu) mutants and the double (315 Ser→Thr and 463 Arg→Leu) mutant are shown. For *RsaI*, restriction maps for wild-type (W+) and the 337 W→C mutant are shown.
Figure 6 schematically depicts CfoI restriction sites and resulting RFLP fragments for a portion of the *M. tuberculosis* katG gene. Restriction maps for wild-type (W+) and the 264 A→T mutant are shown.

Figure 7, panels A-C, depict as the upper of the pair of sequences the consensus, wild-type DNA sequence of the *M. tuberculosis* katG gene (SEQ ID NO:20), and as the lower of the pair of sequences the amino acid consensus sequence encoded thereby (SEQ ID NO:21). This information is updated from that presented in Figures 1 (SEQ ID NO:1) and 2 (SEQ ID NO:7) and the numbering system is as used therein. The amino acid and nucleotide sequences are arranged in this figure so as to facilitate convenient determination of which codons encode which amino acid in the polypeptide sequence.

Figure 8 depicts the RFLP patterns produced by an *Msp*I restriction digest of an amplified portion of the DNA of the katG genes of wild-type and mutant strains of *M. tuberculosis*, wherein the mutant DNA contains mutations at either codon 315 or codon 463, or both.

Figure 9 depicts the RFLP patterns produced by an *Rsa*I restriction digest of an amplified portion of the DNA of the katG genes of wild-type and mutant strains of *M. tuberculosis*, wherein the mutant DNA contain a mutation at codon 337.

Figure 10 depicts the RFLP patterns produced by a CfoI restriction digest of an amplified portion of the DNA of the katG genes of wild-type and mutant strains of *M. tuberculosis*, wherein the mutant DNA contains a mutation codon 264.

**Detailed Description of the Invention**

Wild type strains of *M. tuberculosis* are highly susceptible to isoniazid (INH) with minimum inhibitory INH concentration (MIC or ICₘᵢₙ) ≤0.02 μg/ml, and a susceptible strain is considered to be one with an ICₘᵢₙ < 1.0 μg/ml. At the Mayo Clinic, Rochester, Minnesota, clinical strains of *M. tuberculosis* (including MDR-TB strains) were identified which exhibit intermediate to high level resistance to INH (ICₘᵢₙ range 1.0 to >32 μg/ml).
Many of these strains, especially those highly resistant to INH (>4.0 μg/ml), exhibited diminished catalase activity as assessed by a semiquantitative technique. The mean semiquantitative catalase was 16.5 mm for 6/15 strains with INH IC\textsubscript{min} < 1.0 μg/ml and 13.3 mm for 9/15 strains with IC\textsubscript{min} ≥ 1.0 μg/ml.

To develop the present assay, it was first necessary to determine whether some *M. tuberculosis* strains have decreased INH sensitivity as a result of *katG* gene mutations. Therefore, the nucleic acid sequences of the *katG* genes for both INH sensitive (IC\textsubscript{min} < 1.0 μg/ml) and INH resistant (IC\textsubscript{min} ≥ 1.0 μg/ml) *M. tuberculosis* strains were determined. From the DNA sequencing data generated, a *katG* consensus sequence was derived, and *katG* sequences from all 15 *M. tuberculosis* strains (INH sensitive and INH resistant) were compared to the consensus sequence to determine *katG* deviations.

Five of nine INH resistant strains (INH IC\textsubscript{min} ≥ 1.0 μg/ml) had one or more missense mutations; one had a nonsense mutation; one had an 8 base pair deletion; and two had no mutations in the coding sequences. All of the five strains with missense mutations had a common G to T transversion at base 1457 in codon 463 (bases 1456-1458) causing replacement of arginine with leucine and loss of an *NciI-MspI* restriction site. Two of those having mutations at codon 463 also showed a G to C transversion at base 1013 in codon 315 (bases 1012-1014) causing replacement of serine with threonine. A third contained a G to A transversion at base 859 in codon 264 (bases 859-861) resulting in the replacement of alanine by threonine, and a fourth contained an A to G transversion at base 1079 in codon 337 (bases 1078-1080), causing tyrosine to be replaced by cysteine. The numbering system is shown in Figure 1 (SEQ ID NO:1). The affected codons and portions of the DNA sequences on either side are shown for both INH sensitive and INH resistant strains in Table 2.

Six INH sensitive strains (INH IC\textsubscript{min} < 1.0 μg/ml) were also sequenced and found to have from none to 5 amino acid differences with the consensus sequence of all 15 strains, but none of the mutations affected codons 463, 315, 264, or 337 or their overlapping restriction sites. Restriction analysis of a total of 32 sensitive and 43 resistant strains revealed a common restriction fragment length polymorphism (RFLP) in nearly half (19) of the 43 of INH
resistant strains, but only one of the INH sensitive strains. Specifically, 44% of the INH resistant had lost the NciI-MspI restriction site at the locus of codon 463 while only 1 of 32 sensitive strains had this restriction polymorphism.

Subsequently, the frequency of codon 463 (R→L) and codon 315 (S→T) mutations in 97 M. tuberculosis clinical isolates was determined. These isolates were obtained from patients treated at Mayo Clinic and samples referred from other health care institutions. Restriction fragment length polymorphism (RFLP) analysis using the MspI restriction enzyme, which cleaves at a site spanning the consensus codon 463 site and at a site comprising a portion of the mutant codon 315 site on the katG gene of M. tuberculosis, was performed on amplified DNA from 97 clinical isolates. Comparison of the resulting RFLP patterns and IC_{min} for isoniazid revealed that of the 90 INH-resistant strains, approximately 10% had both mutations, 20% had the 315 S→T mutation only, and 26% had the 463 R→L mutation only. Thus, 51 of the 90 resistant strains were identified by RFLP as having mutations at codons 463, 315 or both, resulting in a detection of over 50% of the resistant strains by this molecular method in a single experiment. Only one of the seven INH sensitive strains was found to have the 463 R→L mutation, and none of the INH sensitive strains had the 315 S→T mutation. Greater INH resistance (> 4.0 μg, INH/ml) is associated with the 315 S→T mutation, but not if the 463 R→L mutation is also present.

These results indicate that two mutations, arginine→leucine in codon 463 and serine→threonine in codon 315 of the M. tuberculosis catalase-peroxidase (katG) gene occur in a significant fraction of INH resistant M. tuberculosis strains (INH IC_{min} ≥ 1.0 μg/ml). Furthermore, these single base mutations can be determined using a rapid relatively simple method, i.e., PCR amplification, digestion and monitoring for a loss of an NciI and/or an MspI restriction site at codon 463, and the addition of an MspI restriction site at codon 315, by RFLP, as described in detail hereinbelow. Other restriction endonucleases can be used to determine whether or not these single base
mutations exist in a katG gene of interest, as long as the restriction site cleaved by the restriction endonucleases contains the affected base, such that the endonuclease cleaves the wild-type sequence but not the corresponding mutant sequence, or vice versa. Although in a preferred embodiment of the invention, the number and location of the fragments is determined by gel electrophoresis, the presence or absence in the digest of a fragment comprising the indicated restriction sites can be determined by other methods known to the art, including immunoassays (dot blots and reverse dot blots), DNA probes, microtiter well capture and the like.

It was further found that the RFLP patterns produced by a MspI digest of the M. tuberculosis DNA fragment katG 904-1523 (Figure 8, lanes B-S) are specific for M. tuberculosis and thus allow selective detection of M. tuberculosis. MspI digestion of any M. tuberculosis katG gene fragment containing the MspI restriction sites depicted in Figure 5 is expected to yield a similar M. tuberculosis-specific RFLP pattern. In contrast, MspI digestions of DNA samples containing other microorganisms, such as mycobacteria other than M. tuberculosis (MOTT), yield either different, distinguishable RFLP patterns, or no detectable restriction fragments at all.

When oligonucleotide primers, preferably katG904 and katG1523, are used in a PCR to generate the 620 base pair M. tuberculosis DNA amplicon katG 904-1523 in a sample that contains M. tuberculosis, typically the PCR yields no amplicon at all for non-M. tuberculosis samples (i.e., the primers do not function to create an amplified product). Thus, the generation of the katG 904-1523 amplicon itself is indicative of the presence of M. tuberculosis in the sample.

Subsequent enzymatic digestion of the amplicon, preferably using restriction endonuclease MspI, can be used to confirm the determination of M. tuberculosis.

The present invention will be further described by reference to the following detailed examples. The 58 clinical strains of Mycobacterium tuberculosis used in Examples 1 and 2 were obtained from the Mycobacteriology Laboratory at the Mayo Clinic, Rochester, Minnesota, and the 17 M. tuberculosis DNA preparations were obtained from the GWL Hansen's Disease Center, Louisiana State University, Baton Rouge, Louisiana. The strain designated
H37Rv MC has been maintained at the Mayo Clinic for over 50 years, and therefore was isolated before INH became available as a treatment modality for tuberculosis (circa 1952). H37Rv was deposited in the American Type Tissue Collection, Rockville, Maryland in 1937 by A. Karlson of the Mayo Clinic under the accession number ATCC 25618, and has been freely available to the scientific community since. An apparent variant of this strain is disclosed in PCT WO 93/22454 (SEQ ID NO: 2, herein). The ATCC strains 27294 and 25618 were recovered from the same patient in 1905 and 1934, respectively. All clinical *M. tuberculosis* strains were confirmed as *M. tuberculosis* using routine identification techniques described by J.A. Washington, "Mycobacteria and *Nocardia,*" in: *Laboratory Procedures in Clinical Microbiology*, 2d ed., Springer-Verlag, NY (1985) at pages 379-417.

For the 15 *M. tuberculosis* strains for which complete *katG* DNA sequencing was performed, susceptibility testing was done at the Mayo Clinic using Middlebrook 7H11 agar (DiMed, Inc., St. Paul, Minnesota 55113) and the 1% proportion method described in *Manual of Clinical Microbiology*, 5th ed., A. Balows et al., eds., *Amer. Soc. Microbiol.* (1991) at pages 1138-52. The same method was used at the Mayo Clinic to determine susceptibility for an additional 43 *M. tuberculosis* strains for which restriction fragment length polymorphisms (RFLP) were determined. Isoniazid concentrations tested using this method included: 0.12, 0.25, 1.0, 2, 4, 8, 16, 32 μg/ml for the 15 strains sequenced and 1.0 and 4.0 μg/ml for the remaining 45 strains. Isoniazid resistance was defined as a maximum inhibitory concentration (*IC*<sub>min</sub>) ≥1.0 μg/ml. Susceptibility testing was performed elsewhere for an additional 17 *M. tuberculosis* strains for which DNA lysates were provided by Diana L. Williams, Baton Rouge, Louisiana. These strains were of diverse geographical origin. 10 of these 17 strains, originated from Japan. The remaining 7 *M. tuberculosis* INH resistant strains included multiple drug resistant strains from recent multiple drug resistant tuberculosis (MDR-TB) nosocomial epidemics in New York, New York and Newark, New Jersey. All were INH resistant (*IC*<sub>min</sub> ≥1.0 μg/ml), and had resistance to at least one other drug. For all strains provided by Williams, the 1%
direct proportion method was used, but the concentration of INH tested, and the
media used varied as to site.

To conduct a semiquantitative test of catalase activity, *M. tuberculosis* strains were propagated on Lowenstein-Jensen media deeps

contained in 20 x 150 mm screw-capped tubes. One ml of a 30% hydrogen
peroxide (EM, Science, Gibbstown, New York 08027) and 10% Tween 80
(Aldrich Chemical Co., Milwaukee, Wisconsin 53233) solution mixed in a 1:1
ratio was applied to the surface of growth. After 5 minutes, the highest (mm) of
the column of bubbles (O₂) generated was recorded.

**Example 1. DNA Isolation and Polymerase Chain Reaction.**

A. **DNA Isolation.** For *M. tuberculosis* strains obtained from
Mayo Clinic samples, DNA was extracted from cells using phenol (Boehringer
Mannheim, Indianapolis, Indiana 46250-0414) and TE (1.0 M Tris HCl pH 8.0,

0.1M EDTA, Sigma, St. Louis, Missouri 63778) in a ratio of 600 μl:400 μl and

0.1 mm zirconium beads (Biospec Products, Bartlesville, Oklahoma 74005). The
mixture was processed in a mini-bead beater for 30 seconds and allowed to stand
for an additional 15 minutes. Following a brief centrifugation to sediment the
zirconium beads, DNA in the supernatant was extracted using the IsoQuick kit
(MicroProbe Corp., Garden Grove, California 92641).

B. **PCR Using Primer Pairs A1-A4 and B1-B2.** The DNA
sequence for *katG* (EMBL no. X6808124) employed to design primers is
depicted in Figure 1(A-D), lower strand. The PCR method of R.K. Saiki et al.,

*Science*, **239**, 487 (1988) was used to amplify the *katG* gene (ca. 2220 base pairs)
in two segments which were designated A and B. Genomic DNA preparations (2
μl) were used with primers A1 (5'

TCGGACCATAACGCGCTTCCTGTTGGACGAG 3') (SEQ ID NO:3) and A4

(5' AATCTGCTTCGCGACGAGGTCGTGCTGAC 3') (SEQ ID NO:4) or B1

(5' CACCCCCGACGAATGGGACAAACAGTTTCTC 3') (SEQ ID NO:5) and

B2 (5' GGGTCTGACAATAATCCGGCCGGGCAACACC 3') (SEQ ID NO:6).

The PCR mixture (50 μl) contained 10 mM TRIS, pH 8.3, 50 mM
KCl, 1.5 mM MgCl₂, 0.2 mM each of dATP, dTTP, dGTP, dCTP, 1 μM of each
primer pair, 10% glycerol, 1.25 units/50 µl AmpliTaq DNA polymerase (Perkin
Elmer Cetus). The mixture was overlaid with mineral oil and subjected to 4 min
at 95°C followed by 50 cycles of 1 min at 94°C and 2 min at 74°C. A 1495 base
pair product from the first half of katG was generated from the A1-A4 primers
and 1435 base pair product was generated with the B1-B2 primer pair.

Example 2. DNA Sequencing and Homology Analysis.

The polymerase chain reaction (PCR) products were prepared for
sequencing using the Magic™ PCR Preps DNA Purification System (Promega
Corp., Madison, Wisconsin 53711). The DNA sequences were determined in
both directions using the Taq dye-deoxy terminator cycle sequencing kit and
373A DNA sequencer (Applied Biosystems, Foster City, California 94404) using
a series of internal sequencing primers which provided appropriate coverage of
katG.

The sequence data were analyzed using version 7 of the Genetics
Computer Group sequence analysis software, as disclosed by J. Devereux et al.,
a consensus sequence was derived to which all M. tuberculosis strains were
compared. This consensus sequence is depicted in Figure 1 (A-D) (SEQ ID
NO:1) as the upper strand, and is compared to the sequence for katG (EMBL no.
X6808124), depicted as the lower strand. The two sequences have 98.6%
identity, as determined by the GCG program BESTFIT. The DNA sequence data
has been submitted to Gen Bank and can be referenced by the accession numbers
UO6262 (H37Rv MC), UO6258 (ATCC 25618), UO6259 (ATCC 27294),
UO6260 (G6108), UO6261 (H35827), UO6270 (L6627-92), UO6271 (L68372),
UO6264 (L11150), UO6268 (L24204), UO6269 (L33308), UO6265 (L16980),
UO6266 (L1781), UO6272 (TMC306), UO6263 (L10373), and UO6267
(L23261). An updated, more complete and accurate M. tuberculosis katG gene
sequence is presented in Figure 7 (A-C) (SEQ ID NO:20).

The DNA data was then translated, aligned for comparison and a
consensus amino acid sequence was generated (Figure 2) (SEQ ID NO: 7). The
consensus amino acid sequence (SEQ ID NO:21) generated from the DNA of
SEQ ID NO:20 is also presented in Figure 7.

In general, the overall sequence agreement between INH sensitive
and resistant strains was very high; the only deviations are those shown in Table

3.
TABLE 3.

Analysis of Catalase-Peroxidase (*katG*) Gene in *M. tuberculosis* Strains

<table>
<thead>
<tr>
<th>Strain</th>
<th>INH MIC(^a) (µg/ml)</th>
<th>Amino Acid Codon(^b)</th>
<th>2</th>
<th>10</th>
<th>17</th>
<th>90</th>
<th>224</th>
<th>243</th>
</tr>
</thead>
<tbody>
<tr>
<td>H37Rv MC</td>
<td>&lt;0.12</td>
<td>P-S</td>
<td>S-N</td>
<td>Q-E</td>
<td>A-S</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATCC 25618</td>
<td>&lt;0.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATCC 27294</td>
<td>0.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G6108</td>
<td>&lt;0.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H35827</td>
<td>0.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L6627-92</td>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L68372</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L11150</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L24204</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L33308</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L16980</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1781</td>
<td>32</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>W(^*)</td>
<td></td>
</tr>
<tr>
<td>TMC 306</td>
<td>&gt;32</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L10373</td>
<td>&gt;32</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L23261</td>
<td>&gt;32</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Consensus P S W Q A
<table>
<thead>
<tr>
<th>Strain</th>
<th>INH (μg/ml)</th>
<th>Catalase</th>
<th>Amino Acid (Codon)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H37Rv MC</td>
<td>&lt;0.12</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>ATCC 25618</td>
<td>&lt;0.12</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>ATCC 27294</td>
<td>0.12</td>
<td>28</td>
<td>A-D</td>
</tr>
<tr>
<td>G6108</td>
<td>&lt;0.12</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>M-I</td>
<td>0.25</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>L6627-92</td>
<td>0.5</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>L68372</td>
<td>1</td>
<td>8</td>
<td>Y-C</td>
</tr>
<tr>
<td>L11150</td>
<td>8</td>
<td>28</td>
<td>R-L</td>
</tr>
<tr>
<td>L24204</td>
<td>8</td>
<td>36</td>
<td>S-T</td>
</tr>
<tr>
<td>L33308</td>
<td>8</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>L16980</td>
<td>16</td>
<td>15</td>
<td>S-T</td>
</tr>
<tr>
<td>L1781</td>
<td>32</td>
<td>5</td>
<td>A-T</td>
</tr>
<tr>
<td>TMC 306</td>
<td>&gt;32</td>
<td>5</td>
<td>A-V</td>
</tr>
<tr>
<td>L10373</td>
<td>&gt;32</td>
<td>5</td>
<td>A-D</td>
</tr>
<tr>
<td>L23261</td>
<td>&gt;32</td>
<td>5</td>
<td>R-L</td>
</tr>
<tr>
<td>M-I</td>
<td></td>
<td></td>
<td>W-R</td>
</tr>
</tbody>
</table>

**Consensus:**

A S Y A R W A M

---

aMIC denotes Maximum Inhibitory Concentration, INH denotes isoniazid

aA denotes alanine, C cysteine, D aspartic acid, E glutamic acid, F phenylalanine, G glycine, I isoleucine, K lysine, L leucine, M methionine, N asparagine, P proline, Q glutamine, R arginine

S serine, T threonine, V valine, W tryptophan, Y tyrosine, B bpd B base pair deletion

cTGG→TGA (W→stop codon)

8 base pair deletion corresponding to wild type coordinates 98-105 creates a new TAG stop codon beginning 11 bp from coordinate 97.
The data in Table 3 show that six strains, H37Rv MC, ATCC 25618, H35827, L6627-92, L11150, and L33308, are completely homologous to the consensus at the indicated sites. Four are INH sensitive (INH IC_{min} < 1.0 \mu g/ml) and two are INH resistant (IC_{min} \geq 1.0 \mu g/ml). All other strains listed in Table 3 had 1 to 5 differences with the consensus and there was no strong correlation between the number of differences and INH sensitivity.

In the group of INH resistant strains, the most frequent change observed was the conversion of arginine at codon 463 to leucine. This was detected in five of nine isolates examined. There was not a consistent correlation between the loss of catalase activity and INH resistance since strains L11150 and L24204 had high levels of enzymatic activity, yet were INH resistant. Moreover, several other INH resistant strains showed catalase activity near the mean activity (16.5 mm) of the sensitive strains. Two other isolates had lost the ability to make normal katG gene product due either to an eight bp deletion (L10373, semiquantitative catalase, 3mm) or a nonsense mutation (TMC 306, semiquantitative catalase 5 mm). It was not possible to determine if, or how, any of the deviations from the consensus reported in Table 3 affect catalase activity or cause INH resistance. However, the change at codon 463 is frequent enough that is indicative of resistance.

The DNA sequence analysis indicated that the codon 463 occurs in the context of an NciI-MspI restriction site (both enzymes recognize the same site). Thus, when in the wild type sequence depicted in Fig. 1 at bases 1455-1458, CCGGG, is changed to CCTGG, it is no longer recognized (or cleaved) by either of these enzymes. The 1435 bp amplicon produced from the half of katG gene containing codon 463 normally has five NciI-MspI restriction sites whereas the codon altered strains have only four sites, as shown in Figure 3. The loss of the site in question causes a unique restriction fragment length polymorphism (RFLP), which can be readily adapted to assay for resistant strains, as described in Example 3, below.
Example 3. RFLP Analysis: *MspI* - *NcoI* site in Codon 463

For restriction fragment length polymorphism (RFLP) analysis, a 1435 base pair amplimer (produced using the B1-B2 primers) representing the 3' half of the *katG* gene was generated using PCR and then digested with *NcoI* or *MspI* (Sigma Chemical Co., St. Louis, Missouri 63178). The gene fragments were analyzed with agarose gel electrophoresis using 2% Metaphor agarose (FMC BioProducts, Richland, Maine 04811). The gel was stained with ethidium bromide and photographed. The investigator who performed all restriction digests and electrophoresis was blinded as to the INH IC₅₀ results.

The results of this experiment are depicted in Figure 4, wherein Lane 1 denotes strain H37Rv MC, IC₅₀ = < 0.12 µg/ml; (2) L6627-92, 0.5 µg/ml; (3) L68372, 1.0 µg/ml; (4) L16980, 16 µg/ml; (5) L39791, 16 µg/ml; (6) L1781, 32 µg/ml; (7) L9118, 4 µg/ml; (8) L11150, 8 µg/ml; (9) L24204, 8 µg/ml; (10) L68858, < 0.12 µg/ml; (11) 1115A < 0.12 µg/ml; (12) L23261, > 32 µg/ml; (13) 1341, > 32 µg/ml; (14) M10838, > 32 µg/ml; (15) molecular weight standard: PCR markers (United States Biochemical Corp., Cleveland, Ohio 44122). The digests obtained from resistant strains can be readily visually detected and differentiated from digests from susceptible strains.

Subsequently, a total of 75 *M. tuberculosis* strains (including the 15 strains sequenced) were analyzed for their loss of the appropriate restriction site. Of these strains, 32 were INH sensitive and 43 were INH resistant. The data showed that 19 (44%) of the 43 resistant strains had lost the expected restriction site in codon 463. One of the 33 (2.9%) sensitive strains had lost this restriction sites as well. None of the six sensitive strains listed in Table 3 lost this site.
Example 4. Determination of the Presence or
Absence of Mutations at Codons 264, 315, 337 or 463
in the M. tuberculosis katG Gene

A. Materials.

Primer pairs used for polymerase chain reaction were
katG904/katG1523 (nucleotide sequences 5' AGC TCG TAT GGC ACC GGA
AC 3' (SEQ ID NO:16) and 5' TTG ACC TCC CAC CCG ACT TG 3' (SEQ ID
NO:17)) and katG633/katG983 (nucleotide sequences 5' CGG TAA GCG GGA
TCT GGA GA 3' (SEQ ID NO:18) and 5' CAT TTC GTC GGG GTG TTC GT
3' (SEQ ID NO:19)). Subunits thereof that hybridize to the amplified DNA under
the conditions described hereinbelow may also be used.

Polyacrylamide was obtained from National Diagnostics, Tris
Borate EDTA solution (6X, cat. no. T6400), magnesium chloride, and
dithiothreitol (DTT) from Sigma Chemical Company (St. Louis, MO), TEMED
(cat. no. 161-0800) and ethidium bromide (EtBr) from Biorad, ammonium
persulfate from Intermountain Sci., and nucleotides (dATP, dGTP, dCTP and
dTTP, 100mM solutions) from Boehringer Mannheim Biochemicals. dUTP was
obtained from Pharmacia. AmpErase™ uracil-N-glycosylase (UNG) and
AmpliTaq™ were obtained from Perkin Elmer.

Restiction endonucleases were obtained as follows: MspI from
Sigma (cat. no. R-4506) 10 u/μl with blue palette buffer; Rsal from New England
Biochemical (cat. no. 167S) 10 u/μl with NEB buffer 1; and CfoI from Promega
(cat. no. R624) 10 u/μl with buffer B.

100 mM nucleotide concentrates obtained from Boehringer
Mannheim Biochemicals were used to make the dNTP stock solution, which was
1.25 mM in each nucleotide. Specifically, 10 μl each of dATP, dGTP, dCTP, and
dTTP concentrates were added to 760 μl water. dNTP(U) stock solution, also
1.25 mM in each nucleotide, was made from the same 100 mM dATP, dGTP and
dCTP concentrates, and 100 mM dUTP concentrate from Pharmacia. Ten μl of
each of the four concentrates was added to 760 μl water to make the stock
solution.
10X PCR buffer consisted of 100 mM Tris, pH 8.3, 500 mM KCl, and 15 mM MgCl2. PCR mix "A" consisted of 1X PCR buffer, 200 µM each dATP, dGTP, dCTP, and dUTP, 1 µM each katG904 (SEQ ID NO:16) and katG1523 (SEQ ID NO:17) primers, 10% glycerol, 10 units/ml AmpErase™UNG, and 0.025 units/µl AmpliTaq. PCR mix B consisted of 1X PCR buffer, 200 µM each dATP, dGTP, dCTP, and dUTP, 1 µM each katG904 (SEQ ID NO:16) and katG1523 (SEQ ID NO:17) primers, 10% glycerol, and 0.025 units/µl AmpliTaq. PCR mix "C" consisted of 1X PCR buffer, 200 µM each dATP, dGTP, dCTP, and dUTP, 1 µM each katG633 (SEQ ID NO:18) and katG983 (SEQ ID NO:19) primers, 10% glycerol, 10 units/ml AmpErase™UNG, and 0.025 units/µl AmpliTaq.

Gel loading solution (Blue Juice) was obtained from Sigma (cat. no. G-2526). Gels were photographed on a UV transilluminator (UVP) with Polaroid 667 black and white film (31/4 X 41/4 inch) through an orange filter. DNA extracts (target DNA) were prepared as described in Example I(A).

B. MspI RFLP analysis.

PCR was performed by adding 2 µl of DNA extract to 48 µl PCR mix "A." Each reaction was covered with 2 drops of mineral oil. Temperature was cycled (Perkin Elmer DNA Thermo Cycler model 480) for 1 cycle of (5'-37°C; 5'-95°C) and 40 cycles of (1'-94°C; 0.5'-60°C; 0.75'-72°C) and a 72°C soak. MspI (10 u/µl) was diluted 1:10 in 100 mM MgCl2. The amplified DNA (base pairs 904 through 1523) of a wild-type katG gene contains 7 MspI restriction sites (Figure 5); of the 8 fragments produced in an MspI restriction digest, 4 are of sufficient length to be visualized using gel electrophoresis (see Figure 5 for a restriction map). Diluted MspI (1 µl) was mixed with 9 µl of the PCR reaction mixture containing the amplified DNA. The digest was incubated at 37°C for 2 hours, then heated to 65°C for 10 minutes. Subsequently, 10 µl of the digest plus 4 µl blue juice was electrophoresed on 6% polyacrylamide for 0.4 hour at 200 V. The gel was stained in EtBr (0.5 mg/ml 1XTBE) for 5 minutes and photographed.
Results are shown in Figure 8. Lanes C, D, F, G, H, K, L, N, and Q show the wild-type genotype at codons 315 (AGC) and 463 (CGG) evidenced by 4 restriction of sufficient length to be visualized using gel electrophoresis (228, 153, 137, and 65 base pairs, respectively, see Figure 5). Lanes M and O show an RFLP indicating a mutation at codon 315 that adds a new MspI restriction site, causing the 153 base pair fragment to be shortened to 132 base pair and become difficult to resolve from the 137 base pair fragment. The resulting 3 fragment pattern (65, 132/137 and 228 base pair) is indicative of an INH resistant strain. Lanes E, I and P show an RFLP indicating a mutation at codon 463 that eliminates an MspI restriction site, evidenced by the 3 visible fragments produced by cleavage versus the 4 produced by the wild-type genotype. The resulting 3 fragment pattern (153, 202, and 228 base pair) is indicative of an INH resistant strain. Lanes B and J show an RFLP indicating mutations at both codon 315 and codon 463. The resulting gain and loss of MspI restriction sites produces a distinctive 3 fragment RFLP pattern (132, 202 and 228 base pair) indicative of an INH resistant strain (see Figure 5 for a restriction map).

C. Rsal RFLP analysis.

PCR was performed by adding 2 µl of DNA extract to 48 µl PCR mix "B." Each reaction was covered with 2 drops of mineral oil. Temperature was cycled (Perkin Elmer DNA Thermo Cycler model 480) for 1 cycle of 2 minutes at 94° and 40 cycles of (1°-94°; 0.5°-60°; 0.75°-72°) and a 4° soak. Rsal (10 u/µl) was diluted 1:20 in 100 mM MgCl₂/100 mM dithiothreitol. The amplified DNA (bases 904 through 1523) of a wild-type katG gene contains 2 Rsal restriction sites. Diluted Rsal (2 µl) was placed on top of the PCR reaction mixture (on the oil) and centrifuged at about 12,000 x g for 10 seconds to drop the Rsal enzyme into the mixture containing the amplified DNA. The resulting mixture was incubated overnight (15-20 hours) at 37°, after which 10 µl of the digest plus 1 µl blue juice was electrophoresed on 6% polyacrylamide for 0.4 hour at 200 V. The gel was stained in EtBr (0.5 mg/ml 1XTBE) for 5 minutes and photographed.
Results are shown in Figure 9. Lanes A, B, D, E, and F show the wild-type genotype at codon 337 (TAC), evidenced by three restriction fragments produced by cleavage at two sites. Lane C shows an RFLP indicating a mutation at codon 337 that eliminates one of the Rsal restriction sites. The resulting two fragment pattern has been observed in an INH resistant strain.

D. CfoI RFLP analysis.

PCR was performed by adding 2 µl of DNA extract to 48 µl PCR mix "C." Each reaction was covered with 2 drops of mineral oil. Temperature was cycled (Perkin Elmer DNA Thermo Cycler model 480) for 1 cycle of (5'-37°C; 5'-95°C) and 40 cycles of (1'-94°C; 0.5'-60°C; 0.75'-72°C) and a 72°C soak. The amplified DNA (bases 633 through 983) of a wild-type katG gene contains 3 CfoI restriction sites. CfoI (10 u/µl) was diluted 1:5 in 100 mM MgCl₂. Diluted CfoI (1 µl) was mixed with 9 µl of the PCR reaction mixture containing the amplified DNA. The digest was incubated at 37°C for 2 hours, then heated to 65°C for 10 minutes. Subsequently, 10 µl of the digest plus 4 µl blue juice was electrophoresed on 6% polyacrylamide for 0.4 hour at 200 V. The gel was stained in EtBr (0.5 mg/ml 1XTBE) for 5 minutes and photographed.

Results are shown in Figure 10. Lanes A-C show the wild-type genotype at codon 264 (GCG), evidenced by 4 restriction fragments produced by cleavage at three sites. Lane E shows an RFLP indicating a mutation at codon 264 that eliminates one of the CfoI restriction sites. The resulting three fragment pattern has been observed in an INH resistant strain.

Example 5. Rapid Simultaneous Detection of M. tuberculosis (MTB) and Determination of Isoniazid (INH) Susceptibility Directly from Sputum

Five microliter aliquots of 785 ethanol-fixed sputum samples from 365 patients were screened for MTB and INH resistance using the MspI RFLP analysis disclosed in Example 4(B). Primers katG904 (SEQ ID NO:16) and katG1523 (SEQ ID NO:17) were used in a polymerase chain reaction as
described in Example 4 to produce a 620 base pair katG gene fragment or "amplicon" (base pairs 904 through 1523), which was digested with MspI. The resulting restriction fragment pattern was visualized using gel electrophoresis. A result considered "positive for MTB" was defined as production of a katG amplicon that generated an RFLP pattern (see Figure 8) indicative of wild-type MTB or mutant MTB (MTB containing a S315T or R463L mutation, or both, in the 620 base pair katG amplicon). A negative result was defined as the failure to produce a katG amplicon (i.e., failure to produce the 620 base pair segment in the PCR), or, in rare cases, the production of a katG amplicon followed by generation of an RFLP pattern that differed from the RFLP pattern know to be associated with wild-type or mutant (S315T, R463L or S315T/R463L) MTB.

The results of this PCR-RFLP assay were compared to results for acid-fast bacilli (AFB) staining by the Ziehl-Neelsen method, and to results of culture and INH susceptibility testing using the BACTEC radiometric method. Technologists performed PCR-RFLP after AFB bacilli stains and cultures were done, and were blinded to the results for AFB stains and cultures. Patient charts were also reviewed for clinical correlation.

Seventy of 785 (8.9%) sputa were AFB stain-positive. MTB was cultured from 48 of these 70 samples. For 9 other AFB stain-positive samples, MTB was not isolated, however MTB was isolated from these same patients from a recent prior sputum. Eight of these 9 patients were receiving antituberculous therapy at the time the sputum was collected for the study.

Mycobacteria other than MTB (MOTT) were exclusively cultured from 9 other AFB stain-positive specimens [M. avium intracellulare (6), M. fortuitum (2), M. kansasii (1)]. No mycobacteria (MTB or MOTT) were cultured from 2 AFB stain-positive sputa obtained from two patients whose recent prior sputa did not grow mycobacteria. No clinical laboratory information (including whether prior cultures for mycobacteria were done) was available for the 2 remaining AFB stain-positive but culture-negative sputa.

The results for the PCR-RFLP were positive for MTB for 45 of the 48 AFB stain-positive samples that grew MTB. Two of these 3 discordant samples had few AFB on stain and 1 had rare AFB on stain. Additional 5
microliter aliquots from these 3 discordant samples were tested and produced positive results for PCR-RFLP. For the 9 samples that were AFB stain-positive, MTB culture-negative, and where recent prior samples from the same patient were MTB culture-positive, PCR-RFLP results were positive for MTB. In no case was a \textit{katG} amplicon generated for the 9 samples from which MOTT were recovered on culture. PCR-RFLP results were also negative for MTB for the 4 remaining AFB stain-positive, culture-negative samples.

For AFB stain-positive samples from which MTB was isolated and for which susceptibility testing was performed (n=45), 7 isolates (15.6\%) were INH resistant. All of these INH resistant isolates were detected by the PCR-RFLP and analysis of the RFLP pattern showed them to carry the S315T mutation. These 7 isolates were from 7 different patients and 4 had different susceptibility patterns for other drugs.

The PCR-RFLP method produced a \textit{katG} amplicon for 39 of the 715 AFB stain-negative samples. MTB was cultured from only 3 of these samples. However, for 21 of the 39 samples, although MTB was not cultured, MTB was recovered from recent prior samples from the same patient and/or the patient was receiving antituberculous therapy at the time the study sample was collected. Comprehensive clinical and laboratory chart reviews were not available for the remaining 15 patients.

This PCR-RFLP MTB \textit{katG} assay, which can be performed in one working day, is thus a reliable, rapid method for detecting MTB and determining INH susceptibility directly from AFB stain-positive sputa.

All publications, patents and patent documents are incorporated by reference herein, as though individually incorporated by reference. The invention has been described with reference to various specific and preferred embodiments and techniques. However, it should be understood that many variations and modifications may be made while remaining within the spirit and scope of the invention.
SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Cockerill, Franklin R.
    Kline, Bruce C.
    Uhl, James R.

(ii) TITLE OF INVENTION: Detection of Isoniazid Resistant Strains
    of M. Tuberculosis

(iii) NUMBER OF SEQUENCES: 22

(iv) CORRESPONDENCE ADDRESS:
    (A) ADDRESSEE: Schwegman, Lundberg & Woessner
    (B) STREET: 3500 IDS Center
    (C) CITY: Minneapolis
    (D) STATE: MN
    (E) COUNTRY: USA
    (F) ZIP: 55402

(v) COMPUTER READABLE FORM:
    (A) MEDIUM TYPE: Floppy disk
    (B) COMPUTER: IBM PC compatible
    (C) OPERATING SYSTEM: PC-DOS/MS-DOS
    (D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
    (A) APPLICATION NUMBER: US
    (B) FILING DATE:
    (C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
    (A) NAME: Woessner, Warren D.
    (B) REGISTRATION NUMBER: 30,440
    (C) REFERENCE/DOCKET NUMBER: 150.185US1

(ix) TELECOMMUNICATION INFORMATION:
    (A) TELEPHONE: 612-339-0331
    (B) TELEX: 612-339-3061
(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 2235 base pairs
   (B) TYPE: nucleic acid
   (C) STRANDEDNESS: single
   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

AGGAATGCTG TGCCCGAGCA ACACCCACCC ATTACAGAAA CCACCACC CGG AGCCGCTAGC 60
AACGGCTGTC CCGTCCGTGG TCATAATGAA TAACCCCGTCG AGGGCGGCGG AAACCAGGAC 120
TGCTGGCCCA ACCCGCTTCA TAATGGATA CAGCTCACAAA ACCCGGCGGT CTGGTGGCCG 180
ATGGGGTGGG CGTTCGACTA TGCCGGCGGG GTGGGCGACCA TCCCGTGTGA CGCCTGACG 240
CGGGACATCG AGGAAGGTAT GAGCACCTCG CAGCCGTGTT GCCCCCGCGA CTACGSCCA 300
TACGGGCCC GCCTTATCGG GACGGCGGTG GCGCTGGCAG GGCCCAGGAC CACCTACCGC 360
GGCCGGCGCC GGCGGGGGGG CGGCTCGCA CGGCTGGCGC CGTTAACAG CTGGCCGACG 420
AACGGCCAGCT TGGACAGGCG CGGCGGCGCTG CTGGGGCCCG TCGAGAAGAA GTACGCGCAAG 480
AAGCTCTCAGT GGGCGACACT GTGTGTTTTT GCCGGCAACT GCCGCGCGTA ATCGATGGGC 540
TTCAAGACGT TCGGGGGCTG GCGGCTCAGT GGGACCCCGA TGAGGCTAT 600
TGCCGCAAGG AGGCCACTTG GCGCGGCGAT GAGGGTTACG GCGGAAAAG GCATCTGGAG 660
AACCCGCTGG CGCGGCTGCA GATGGGGCTG ATCTACGTGA ACCCGGAGGG GCCGAACGGC 720
AACCCGGACC CCAAGGCGGC GCGTGGCTAC ATCGGGCAAG CGTTGCGCCG CATGGCCATG 780
AAGCACTGCG AAACAGCGGC GCTGATCGTC GCGCGTCACA CTGGGTGAA GACCCATGGC 840
GCCGGCCCGG CCAGCCTGGT CCCCGCGGCGA CACCGAGGCTG CTCGCGCTGGA GCAGATGGGC
900
TTGGGCTCGGA AGAGCTCGTA TGGGCAAGGA ACCGGTAAGG GCCGACTAC CAGCGCCATC
960
GAGGTGTGAT GTAGCAACAC CCAGGCAGGAA TGGGCAACAAT GTTCTGCTGGA GATCCTGTC
1020
GGTACGAGTG GGAGATCGAT GAAAGCCTCT GCTGGCCCTT GGAATACAC CGCCAAGGAC
1080
GCCGCCGTTG CCGCACCATT CCGGAGCCGG TTCGCGCGGC CAGGCGCCTC CCGGACGATG
1140
CCTGCCACTG ACCTCGCTTG GGGGTGGAAT CCAGATCTATG AGCGAGTACAC CCGTCGCTGG
1200
CTGGAAACATT CCGGCGAGAG TTCGCCAAGG CCTGGTACAA CACGTCGATCCAC
1260
CGAGACATGG TCCTCCGTGC GAGATACCTT CGGCCGCTGG TCCGCAAGGA GACCCCGCTG
1320
TGCGAGGGTC CGGTCTCCGT GCCGAGCCAC GCCCTCGTGG GCCGAGCAGGA GAATGCGACG
1380
CTTAAAGACC GATCGCGGGC CTGGGATTG TACGTCACAGC AGTCTGTTTC GACCCCGATG
1440
GGGGCGGCGT CTGCTTCTCG TGCCAGCGAG AAGCGCGGCCG CGGCAACGGG TGTCGGGATC
1500
CGCCTGCAGC CACAGTGGG GTGGGAGTTC ACGAACCGCG ACAGGGATCT CGGCGAAGGT
1560
ATTGCGCACCC TGGAGAGGAT CCCAGAGTCA TCTCAAACCG CGCGCCCCGGG GAACATCAA
1620
GTGTCCTTCGG CCGACCTCGT CGTGCTCCGT GGCCTGTCGG CCATAGAGAA AGCAGCGAAG
1680
GGCGCTGGCC ACAACACTAC GGCAGGCCCTAC CCCCCGGGCC GCAGGGAATGC GTGCGAGGAA
1740
CAAGCCGAGG TGGAATCCTT TGCGCTGCTG GACGCCAAAGG CAGATGCGTT CCGAAAACAT
1800
CTCGGAAGGG CCAACCGGTG GCGGCGGAG TACATGCTCG TCGACAAGGG GAACCAGTCT
1860
ACGCTCGATG CCCCCAGATT GACCGTGCTG GTAGGTTGCC TGGCGCTCCT CGGGCAAACCT
1920
ACAAGCCTGGT ACCCTGGGGT ATGGTACCG AGGCGTCCGA GTCACTGACC AAGCACTTTT
1980
TCGTGAACCT GCTCGACATG GGTATCACCT GGGAGCCCTC GCCAGCAGAT GACGGGACCT
2040
ACCAGGGCAA GGATGGCAGT GGCAAGGTGA AGTGGGACGG CAGCCGCGTG GACCCTGGCT
2100
TCGGGGTCCAA CTCGGAGTGG CGGGCCGCTTG TCGAGGCTTA TGGCCGCCGAT GACGCCAGC
2160
CGAAGTTCTCG GTCCGACTTC GTGCCGCTGCT GGGACAAGGT GATGAACCTC GACAGTTCCG
2220
ACGTGGCGCTG ATTCG
2235

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2221 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:
AGGAAATGCTG TGCCCCAGCA ACACCCACCC ATTACAGAAA CCACCACCGG AGCCGCTAGC
60
AACGGCTGTC CCGTCGGGGG TCATATGAAA TACCCCCGTG AGGGGCGGGG AAACCAGGAC
120
TGGTGGCCAAC ACCGGCTCAA TCTGAAGGTA CTGCACACAA ACCGGGACGT CGCTGACCCG
180
ATGGTGGCGG CTTTGACTTA TGCCGCGGAG GTGCCGACCA GTGCGACTTGA CGCCCTGAGC
240
CGGGACATCG AGGAAGTGAT GACCACCTCG CAGCCGTGGT GGCCGCCGGA CTACGGCCAC
300
TACGGGCCGC TGGTTATCCG GATGGGCGTG CAGCTGCGCG GCACCTACCG CATCCACGAC
360
GGCCCGGCAG GGCCGGGGGG CGGCCATGCA CGGTTGCCGC CGCTTAACAG CTGGCCGCGAC
420
AACGCCAGCT TGGACAAGGC GCACGGCGCTG CTGGCGCGCG TCAAGAAGAA GTACGGCAAG
480
AGCTCCTCAT GGCGGACCT GTATGTTTTTC GCCGCAACC GCTGCCCTCG GAATCGATGG 540
GCTTCAAGAC GTTCCGGTTTC GGCTTCCGGC GTGACCCAGTGGGAGACCCGATCAGCTATAT 600
TGGGGCAAGG AAGCACCCTGTCCGGCATGACCGTTACA CGGTAAGCGATCCGGAGAAC 660
CCGGCTGGCAG CCGTGACAGTC GGGCTGATTC AAGGCTGAACC CGGAGGGCGCCAACG 720
CCGGACCCCA TGCCCGCGGCC GGTCCGACATT CGCGAGACGTT TCCGGCCGATGACCCATGAC 780
GACGTGCAA AAGATCGCGCTGATGAGGCTCCGACATT TGCTGTAAGAC ACATGGCCGCAACG 840
GGGCGGCGGTG ATCTCGGCTGGG CGCGAACCCT GCTTTTGTCTCGGTGAGACGCTGCTTGGT 900
GCTGTGAAAGA GCTGCTATGGG CACCGGAAACC GTAAAGGACGCGATCACCAG CGGCATCGAG 960
GTGGTATGGA CGAAGACCCCG AAGAAATTGGG GAAACAGCTT TCCCTCGAGATCTCTGATCGG 1020
TACGTGATGGG AGCTGACGAA GACCCCTGCT GGGGTGTGGCA CATACACCGGCCAGGCACGGC 1080
GGGGCTGGCG CACCCATCCC GAGCCGCTTCC GGGGCGCAGGGCGCTCCCCC AGCGATGCTG 1140
GCACTGACC TCTCGCTGCGT GTGGATTCG ACTATGAGGCGATCAGCCCGTGCTCGCTGCTG 1200
GAACACCCCG AGGAATAATGCG CAGCGAGTTC CGCAAGACGGT TCTAAAAGCT GATCCACCGA 1260
GACATGGGTC CGGTGCGGAG ATACCTTGCGC GGGCTGTTGCC CAACAGCAGAC CCTGCTGTGG 1320
CGGATACCGG TCCCTGGCTGAG CAGCACGACC GTGTCGGCGA AGCGAATTGC CAGCCTTAAG 1380
AGCCAGATCC GGGCATCGGG ATGGACTGTC TCAAGCTAGTTGGAGACCCGATGGGCGCGG 1440
GGTGTGGTGT TCATGACCTAG CGCAAGACGGCGGTGGCGAGCCAAGCTGGGCTGGC 1500
CGGCACAGG TCGGGTTGGGA GGTCAACGAC CCCGACGGGAT CGGCGCAAGG TCATTCGAC 1560
CCTGAGAAGA TCCAGGAGTC ATTCACTGCG CGCGGAACA TCAAAGTGCT TCTCGCCGAC 1620
CTGGTGGTGC TCGGGTGCTG TGGCCCCACTA GAGAAAGCAG CAAAGGCGGC TGGCCACAAC
1680
ATCAGGGTGC CCTCCACCCCGGGCGGCACCAGTGCGTGCG AGGGAACAAAC CGACGGGGA
1740
TGCTTGGGCCG TGCTTGGAGCC CAAGGCAGAT GGCTCTCGGAA ACTACCTCGG AAAGGGCAAC
1800
CGTGGGCCGC CGAGTACATC GCTGGCTGAC AAGGGCGACC TGCTTACGCT CAGTGCCCG
1860
GAGATGACGG TGCTGTGAGG TGGCCTGGC GCCTCGGCG CAAACTACAA GCGTTACCG
1920
CTGGGGCGTGTC ACACCGAGGC CTCCGAGTCA CTGACCAACCG ACTTCTGCCT GAACTGTCT
1980
GACATGGGTA TCACCTGGGA GCCCTCGGCA GCAGATGACG GGACCTACCA GGGCAAGGAT
2040
GCCAGTGCCA AGGTGAAGTG GACCGCGACG CGCGTGGACCC TGGCTCGGGA GTCCAACTCG
2100
GAGTGTGCAG CGCTTGTGCGA GGTTGAGTCGG CCGATGACGC GGCAGGGGAA GTCCGTGACA
2160
GGATTCTGGG CGCGCTGGGA CAAGGTGAGT AACCTCGACA GGGTCAAGGT CGCGTGGAT
2220
G
2221

(2) INFORMATION FOR SEQ ID NO:3:
   (i) SEQUENCE CHARACTERISTICS:
      (A) LENGTH: 30 base pairs
      (B) TYPE: nucleic acid
      (C) STRANDEDNESS: single
      (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

TOGGACCATA ACGGCTTCTC GTGGACGAG
30
(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

AATCTGCTTC GCCGACGAGG TCGTGTGAC
30

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

CACCCGACG AAATGGGACA ACAGTTTCCT
30
(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

GGGTCTGACA AATCGCGCCG GGCAAAACCC

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 740 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

Val Pro Glu Gly His Pro Pro Ile Thr Glu Thr Thr Thr Gly Ala Ala
1  5  10  15
Ser Asn Gly Cys Pro Val Val Gly His Met Lys Tyr Pro Val Glu Gly
20  25  30
Gly Gly Asn Gln Asp Trp Trp Pro Asn Arg Leu Asn Leu Lys Val Leu
35  40  45
His Gln Asn Pro Ala Val Ala Asp Pro Met Gly Ala Ala Phe Asp Tyr
50  55  60
Ala Ala Glu Val Ala Thr Ile Asp Val Asp Ala Leu Thr Arg Asp Ile
65  70  75  80
Glu Glu Val Met Thr Thr Ser Gln Pro Trp Trp Pro Ala Asp Tyr Gly
85  90  95
His Tyr Gly Pro Leu Phe Ile Arg Met Ala Trp His Ala Ala Gly Thr
100 105 110
Tyr Arg Ile His Asp Gly Arg Gly Gly Ala Gly Gly Gly Gly Met Gln Arg
115 120 125
Phe Ala Pro Leu Asn Ser Trp Pro Asp Asn Ala Ser Leu Asp Lys Ala
130 135 140
Arg Arg Leu Leu Trp Pro Val Lys Lys Tyr Gly Lys Lys Leu Ser
145 150 155 160
Trp Ala Asp Leu Ile Val Phe Ala Gly Asn Cys Ala Leu Glu Ser Met
165 170 175
Gly Phe Lys Thr Phe Gly Phe Gly Phe Gly Arg Val Asp Gln Trp Glu
180 185 190
Pro Asp Glu Val Tyr Trp Gly Lys Ala Thr Trp Leu Gly Asp Glu
195 200 205
Arg Tyr Ser Gly Lys Arg Asp Leu Glu Asn Pro Leu Ala Ala Val Gln
210 215 220
Met Gly Leu Ile Tyr Val Asn Pro Glu Gly Pro Asn Gly Asn Pro Asp
225 230 235 240
Pro Met Ala Ala Ala Val Asp Ile Arg Glu Thr Phe Arg Arg Met Ala
245 250 255
Met Asn Asp Val Glu Thr Ala Ala Leu Ile Val Gly Gly His Thr Phe
260 265 270
Gly Lys Thr His Gly Ala Gly Pro Ala Asp Leu Val Gly Pro Glu Pro
275 280 285
Glu Ala Ala Pro Leu Glu Gln Met Gly Leu Gly Trp Lys Ser Ser Tyr
290 295 300
Gly Thr Gly Thr Gly Lys Asp Ala Ile Thr Ser Gly Ile Glu Val Val
305 310 315 320
Trp Thr Asn Thr Pro Thr Lys Trp Asn Ser Phe Leu Glu Ile Leu
325 330 335
Tyr Gly Tyr Glu Trp Glu Leu Thr Lys Ser Pro Ala Gly Ala Trp Gln
340 345 350
Tyr Thr Ala Lys Asp Gly Ala Gly Ala Gly Thr Ile Pro Asp Pro Phe
355 360 365
Gly Gly Pro Gly Arg Ser Pro Thr Met Leu Ala Thr Asp Leu Ser Leu
370 375 380
Arg Val Asp Pro Ile Tyr Glu Arg Ile Thr Arg Arg Trp Leu Glu His
385 390 395 400
Pro Glu Glu Leu Ala Asp Glu Phe Ala Lys Ala Trp Tyr Lys Leu Ile
His Arg Asp Met Gly Pro Val Ala Arg Tyr Leu Gly Pro Leu Val Pro 
 420   425   430
Lys Gln Thr Leu Leu Trp Gln Asp Pro Val Pro Ala Val Ser His Asp 
 435   440   445
Leu Val Gly Glu Ala Glu Ile Ala Ser Leu Lys Ser Gln Ile Arg Ala 
 450   455   460
Ser Gly Leu Thr Val Ser Gln Leu Val Ser Thr Ala Trp Ala Ala Ala 
 465   470   475   480
Ser Ser Phe Arg Gly Ser Asp Lys Arg Gly Gly Ala Asn Gly Gly Arg 
 485   490   495
Ile Arg Leu Gln Pro Gln Val Gly Trp Glu Val Asn Asp Pro Asp Gly 
 500   505   510
Asp Leu Arg Lys Val Ile Arg Thr Leu Glu Glu Ile Gln Glu Ser Phe 
 515   520   525
Asn Ser Ala Ala Pro Gly Asn Ile Lys Val Ser Phe Ala Asp Leu Val 
 530   535   540
Val Leu Gly Gly Cys Ala Ala Ile Glu Lys Ala Ala Lys Ala Ala Gly 
 545   550   555   560
His Asn Ile Thr Val Pro Phe Thr Pro Gly Arg Thr Asp Ala Ser Gln 
 565   570   575
Glu Gln Thr Asp Val Glu Ser Phe Ala Val Leu Glu Pro Lys Ala Asp 
 580   585   590
Gly Phe Arg Asn Tyr Leu Gly Lys Gly Asn Pro Leu Pro Ala Glu Tyr 
 595   600   605
Met Leu Leu Asp Lys Ala Asn Leu Leu Thr Leu Ser Ala Pro Glu Met 
 610   615   620
Thr Val Leu Val Gly Gly Leu Arg Val Leu Gly Ala Asn Tyr Lys Arg 
 625   630   635   640
Leu Pro Leu Gly Val Phe Thr Glu Ala Ser Glu Ser Leu Thr Asn Asp 
 645   650   655
Phe Phe Val Asn Leu Leu Asp Met Gly Ile Thr Trp Glu Pro Ser Pro 
 660   665   670
Ala Asp Gly Thr Tyr Gln Gly Lys Asp Gly Ser Gly Lys Val Lys 
 675   680   685
Trp Thr Gly Ser Arg Val Asp Leu Val Phe Gly Ser Asn Ser Glu Leu 
 690   695   700
Arg Ala Leu Val Glu Val Tyr Gly Ala Asp Ala Gln Pro Lys Phe 
 705   710   715   720
Val Gln Asp Phe Val Ala Ala Trp Asp Lys Val Met Asn Leu Asp Arg
725 730 735
Phe Asp Val Arg
740

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GTCGAAACAG CCGCGCTGAT CGTCGGC
27
(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GTCGAAACAG CGACGCTGAT CGTCGGC
27

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CTCGAGATCC TGTAAGCGTA CGAGTGG
27

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

CTCGAGATCC TGTCGGCTA CGAGTGG
27
(2) INFORMATION FOR SEQ ID NO:12:
   (i) SEQUENCE CHARACTERISTICS:
       (A) LENGTH: 27 base pairs
       (B) TYPE: nucleic acid
       (C) STRANDEDNESS: single
       (D) TOPOLOGY: linear
   (ii) MOLECULE TYPE: DNA
   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GACGCGATCA CCAGCGGAT CGAGGTC
27

(2) INFORMATION FOR SEQ ID NO:13:
   (i) SEQUENCE CHARACTERISTICS:
       (A) LENGTH: 27 base pairs
       (B) TYPE: nucleic acid
       (C) STRANDEDNESS: single
       (D) TOPOLOGY: linear
   (ii) MOLECULE TYPE: DNA
   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GACGCGATCA CCACCGGCAT CGAGGTC
27

(2) INFORMATION FOR SEQ ID NO:14:
   (i) SEQUENCE CHARACTERISTICS:
       (A) LENGTH: 27 base pairs
       (B) TYPE: nucleic acid
       (C) STRANDEDNESS: single
       (D) TOPOLOGY: linear
   (ii) MOLECULE TYPE: DNA
   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

AAGGACCCAAG ATTCGCGATG CCGATTG
27
(2) INFORMATION FOR SEQ ID NO:15:

   (i) SEQUENCE CHARACTERISTICS:
       (A) LENGTH: 27 base pairs
       (B) TYPE: nucleic acid
       (C) STRANDEDNESS: single
       (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: DNA

(xii) SEQUENCE DESCRIPTION: SEQ ID NO:15:

    ANGAGCCAGA TCCTGGCATC GGGATTG
    27

(2) INFORMATION FOR SEQ ID NO:16:

   (i) SEQUENCE CHARACTERISTICS:
       (A) LENGTH: 20 base pairs
       (B) TYPE: nucleic acid
       (C) STRANDEDNESS: single
       (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: DNA

(xii) SEQUENCE DESCRIPTION: SEQ ID NO:16:

    AGCTCGTATG GCACCGGAAC
    20

(2) INFORMATION FOR SEQ ID NO:17:

   (i) SEQUENCE CHARACTERISTICS:
       (A) LENGTH: 20 base pairs
       (B) TYPE: nucleic acid
       (C) STRANDEDNESS: single
       (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: DNA

(xii) SEQUENCE DESCRIPTION: SEQ ID NO:17:

    TTGACCTCCC ACCGACTTG
    20
(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CGGTAAGCGG GATCTGGAGA
20

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CATTTCGTCG GGGTGTTCGT
20

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2331 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA
(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 70..2289

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

CGATATCCGA CACTTCGCGA TCACATCGT GATCAACGCC CGATAACACC AACCTCTGGA
60
AGGAATGCT GTG CCC GAG CAA CAC CCA CCC ATT ACA GAA ACC ACC ACC
108  Val Pro Glu Gln His Pro Ile Thr Glu Thr Thr Thr
1  5  10
GGA GCC GCT AGC AAC GGC TGT CCC GTC GTG GGT CAT ATG AAA TAC CCC
156  Gly Ala Ala Ser Asn Gly Cys Pro Val Val Gly His Met Lys Tyr Pro
15  20  25
GTC GAG GCC GGC GGA AAC CAG GAC TGG TGG CCC AAC CGG CTC AAT CTG
204  Val Glu Gly Gly Asn Gln Asp Trp Trp Pro Asn Arg Leu Asn Leu
30  35  40  45
AAG GTA CTG CAC CAA AAC CCG GCC GTC GCT GAC CCG ATG GGT GCG GGC
252  Lys Val Leu His Gln Asn Pro Ala Val Ala Asp Pro Met Gly Ala Ala
50  55  60
TTC GAC TAT GCC GCG GAG GTC GCC ACC ATC GAC GTT GAC GCC CTG ACG
300  Phe Asp Tyr Ala Ala Glu Val Ala Thr Ile Asp Val Asp Ala Leu Thr
65  70  75
CGG GAC ATC GAG GAA GTG ATG ACC ACC TCG CAG CCG TGG TGG CCC GCC
348  Arg Asp Ile Glu Glu Val Met Thr Thr Ser Gln Pro Trp Trp Pro Ala
80  85  90
GAC TAC GCC CAC TAC GGG CCG CTG TTT ATC CCG ATG GCG TGG CAC GCT
396  Asp Tyr Gly His Tyr Gly Pro Leu Phe Ile Arg Met Ala Trp His Ala
95 100 105
GCC GGC ACC TAC CGC ATC CAC GAC GCC CGC GCC GCC GCC GGG GGC GCC
444  Ala Gly Thr Tyr Arg Ile His Asp Gly Arg Gly Ala Gly Gly Gly
110 115 120 125
ATG CAG CGG TTC GGG CGG CTT AAC AGC TGG CCC GAC AAC GCC AGC TGG
492  Met Gln Arg Phe Ala Pro Leu Asn Ser Trp Pro Asp Asn Ala Ser Leu
130 135 140
GAC AAG GCG CGC CGG CTG CTG TGG CCG GTC AAG AAG AAG TAC GCC AAG
540
Asp Lys Ala Arg Arg Leu Leu Trp Pro Val Lys Lys Lys Tyr Gly Lys
145     150     155
AAG CTC TCA TGG GCG GAC CTG ATT GTT TTC GCC GCC AAC TGC GCG CTG
Lys Leu Ser Trp Ala Asp Leu Ile Val Phe Ala Gly Asn Cys Ala Leu
160 165 170
GAA TCG AGT GCC TTC AAG ACG TTC GGG TTC GCC TTC GGC CGG GTC GAC
Glu Ser Met Gly Phe Lys Thr Phe Gly Phe Gly Phe Gly Arg Val Asp
175 180 185
CAG TGG GAG CCC GAT GAG GTC TAT TGG GCC AAG GAA GCC ACC TGG CTC
Gln Trp Glu Pro Asp Glu Val Tyr Trp Gly Lys Glu Ala Thr Trp Leu
190 195 200 205
GGC GAT GAG CGT TAC AGC GGT AAG CGG GAT CTG GAG AAC CCG CTG GCC
Gly Asp Glu Arg Tyr Ser Gly Lys Arg Asp Leu Glu Asn Pro Leu Ala
210 215 220
GCC GTG CAG ATG GGG CTG ATC TAC GTG AAC CCG GAG GGG CCG AAC GCC
Ala Val Glu Met Gly Leu Ile Tyr Val Asn Pro Glu Gly Pro Asn Gly
225 230 235
AAC CCG GAC CCC ATG GCC GCG GTC GAC ATT CGC GAG ACG TTT CGG
Asn Pro Asp Pro Met Ala Ala Ala Val Asp Ile Arg Glu Thr Phe Arg
240 245 250
CGC ATG GCC ATG AAC GAC GTC GAA ACA GCG GCG CTG ATC GTC GGC GGT
Arg Met Ala Met Asn Asp Val Glu Thr Ala Ala Leu Ile Val Gly Gly
255 260 265
CAC ACT TTC GGT AAG ACC CAT GCC GCC GCC CCG GCC GAT CTG GTC GCC
His Thr Phe Gly Lys Thr His Gly Ala Gly Pro Ala Asp Leu Val Gly
270 275 280 285
CCC GAA CCC GAG GCT GCT CCG CTG GAG CAG ATG GCC TTG GCC TGG AAG
Pro Glu Pro Glu Ala Ala Pro Leu Glu Gln Met Gly Leu Gly Trp Lys
290 295 300
AGC TCG TAT GGC ACC GGA ACC GGT AAG GAC GCG ATC ACC AGC GGC ATC
Ser Ser Tyr Gly Thr Gly Thr Gly Lys Asp Ala Ile Thr Ser Gly Ile
305 310 315
GAG GTC GTA TGG ACG ACC CCG ACG AAA TGG GAC AAC AGT TTC CTC
Glu Val Val Trp Thr Asn Thr Pro Thr Lys Trp Asp Asn Ser Phe Leu
320 325 330
GAG ATC CTG TAC GGC TAC GAG TGG GAG CTG ACG AAG AGC CCT GCT GGC
1116
Glu Ile Leu Tyr Gly Tyr Glu Trp Glu Leu Thr Lys Ser Pro Ala Gly
335 340 345
GCT TGG CAA TAC ACC GCC AAG GAC GCC GGC GGT GCC GGC ACC ATC CCG
1164
Ala Trp Gln Tyr Thr Ala Lys Asp Gly Ala Gly Ala Gly Thr Ile Pro
350 355 360 365
GAC CCG TTC GCC GGG CCA GGG CGC TCC CCG ACG ATG CTG GCC ACT GAC
1212
Asp Pro Phe Gly Gly Pro Gly Arg Ser Pro Thr Met Leu Ala Thr Asp
370 375 380
CTC TCG CTG CGG GTG GAT CGG ATC TAT GAG CGG ATC ACG CGT CGC TGG
1260
Leu Ser Leu Arg Val Asp Pro Ile Tyr Glu Arg Ile Thr Arg Arg Trp
385 390 395
CTG GAA CAC CCC GAG GAA TTG GCC GAC GAG TTC GCC AAG GCC TGG TAC
1308
Leu Glu His Pro Glu Glu Leu Ala Asp Glu Phe Ala Lys Ala Trp Tyr
400 405 410
AAG CTG ATC CAC CGA GAC ATG GGT CCC GTC AGA TAC CTT GGG CCG
1356
Lys Leu Ile His Arg Asp Met Gly Pro Val Ala Arg Tyr Leu Gly Pro
415 420 425
CTG GTC CCC AAG CAG ACC CTG TCG TGG CAG GAT CCG GTC CCT GCG GTC
1404
Leu Val Pro Lys Gln Thr Leu Leu Trp Gln Asp Pro Val Pro Ala Val
430 435 440 445
AGC CAC GAC CTC GTC GCC GAA GCC GAG ATT GCC AGC CCT AAG AGC CAG
1452
Ser His Asp Leu Val Gly Glu Ala Glu Ile Ala Ser Leu Lys Ser Gln
450 455 460
ATC CGG GCA TCG GGA TTG ACT GTC TCA CAG CTA GTT TCG ACC GCA TGG
1500
Ile Arg Ala Ser Gly Leu Thr Val Ser Gln Leu Val Ser Thr Ala Trp
465 470 475
GCC GCG GCG TCG TCC GTG GGT AGC GAC AAG CGC GGC GCC GCC AAC
1548
Ala Ala Ala Ser Ser Phe Arg Gly Ser Asp Lys Arg Gly Gly Ala Asn
480 485 490
GGT GGT CGC ATC CGC CTG CAG CCA CAA GTC GGG TGG GAG GTC AAC GAC
1596
Gly Gly Arg Ile Arg Leu Gln Pro Gln Val Gly Trp Glu Val Asn Asp
495 500 505
CCC GAC GGG GAT CTG CGC AAC GTC ATT CGC ACC CTG GAA GAG ATC CAG
1644
Pro Asp Gly Asp Leu Arg Lys Val Ile Arg Thr Leu Glu Glu Ile Glu
510 515 520 525
GAG TCA TTC AAC TCC GGC CCG CGG AAG ATC AAA GTG TCC TTC GCC
1692
Glu Ser Phe Asn Ser Ala Ala Pro Gly Asn Ile Lys Val Ser Phe Ala
530 535 540
GAC CTC GTC GTG CTC GGT GCC TGT GCC GCC ATA GAG AAA GCA GCA AAG
1740
Asp Leu Val Val Leu Gly Gly Cys Ala Ala Ile Glu Lys Ala Ala Lys
545 550 555
GGC GCT GCC CAC AAC ATC ACG GTG CCC TTC ACC CCG GCC CGC ACG GAT
1788
Ala Ala Gly His Asn Ile Thr Val Pro Phe Thr Pro Gly Arg Thr Asp
560 565 570
GCG TCG CAG GAA CAA ACC GAC GTG GAA TTC GCC GTG CTG GAG CCC
1836
Ala Ser Gln Glu Gln Thr Asp Val Glu Ser Phe Ala Val Leu Glu Pro
575 580 585
AAG GCA GAT GGC TTC CGA AAC TAC CTC GGA AAG GGC AAC CCG TTG CGG
1884
Lys Ala Asp Gly Phe Arg Asn Tyr Leu Gly Asn Pro Leu Pro
590 595 600 605
GCC GAG TAC ATG CTG CTC GAC AAC GCG AAC CTG CTT ACG CTC AGT GCC
1932
Ala Glu Tyr Met Leu Leu Asp Lys Ala Asn Leu Leu Thr Leu Ser Ala
610 615 620
CCT GAG ATG ACG GTG GTA GGT GCC CTG CGC GTC CTC GCC GCA AAC
1980
Pro Glu Met Thr Val Leu Val Gly Gly Leu Arg Val Leu Gly Ala Asn
625 630 635
TAC AAG CGC TTA CCG CTG GCC GTG TTC ACC GAG GCC TCC GAG TCA CTG
2028
Tyr Lys Arg Leu Pro Leu Gly Val Phe Thr Glu Ala Ser Glu Ser Leu
640 645 650
ACC AAC GAC TTC TTC GTG AAC CTG CTC GAC ATG GGT ATC ACC TGG GAG
2076
Thr Asn Asp Phe Phe Val Asn Leu Leu Asp Met Gly Ile Thr Trp Glu
655 660 665
CCC TCG CCA GCA GAT GAC GGG ACC TAC CAG GGC AAG GAT GGC AGT GCC
2124
Pro Ser Pro Ala Asp Gly Thr Tyr Glu Gln Gyl Lys Asp Gly Ser Gly
670 675 680 685
AAG GTG AAG TGG ACC GCC AGC CGC GTG GAC CTG GTC TTC GGG TCC AAC

Lys Val Lys Trp Thr Gly Ser Arg Val Asp Leu Val Phe Gly Ser Asn
690

TCG GAG TTG CGG GCC CTT GTC GAG GTC TAT GCC GAT GAC GCG CAG

Ser Glu Leu Arg Ala Leu Val Glu Val Tyr Gly Ala Asp Ala Gln
705

CCG AAG TTC CAG GAC TTC GTC GCT GCC TGG GAC AAG GTG ATG AAC

Pro Lys Phe Val Gln Asp Phe Val Ala Ala Trp Asp Lys Val Met Asn
720

CTC GAC AGG TTC GAC GTG CGC TGATCGGCT GTATCGGCCC TGCCGCGGA

Leu Asp Arg Phe Asp Val Arg
735

TCAACCACAA CC

2331
(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 740 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xii) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Val Pro Glu Gln His Pro Pro Ile Thr Glu Thr Thr Thr Gly Ala Ala
      1      5     10
Ser Asn Gly Cys Pro Val Val Gly His Met Lys Tyr Pro Val Glu Gly
      20     25     30
Gly Gly Asn Gln Asp Trp Trp Pro Asn Arg Leu Asn Leu Lys Val Leu
      35     40     45
His Gln Asn Pro Ala Val Ala Asp Pro Met Gly Ala Ala Phe Asp Tyr
      50     55     60
Ala Ala Glu Val Ala Thr Ile Asp Val Asp Ala Leu Thr Arg Asp Ile
      65     70     75     80
Glu Glu Val Met Thr Thr Ser Gln Pro Trp Trp Pro Ala Asp Tyr Gly
      85     90     95
His Tyr Gly Pro Leu Phe Ile Arg Met Ala Trp His Ala Ala Gly Thr
     100    105    110
Tyr Arg Ile His Asp Gly Arg Gly Gly Ala Gly Gly Gly Met Gln Arg
     115    120    125
Phe Ala Pro Leu Asn Ser Trp Pro Asp Asn Ala Ser Leu Asp Lys Ala
     130    135    140
Arg Arg Leu Leu Trp Pro Val Lys Lys Tyr Gly Lys Lys Leu Ser
     145    150    155    160
Trp Ala Asp Leu Ile Val Phe Ala Gly Asn Cys Ala Leu Glu Ser Met
     165    170    175
Gly Phe Lys Thr Phe Gly Phe Gly Arg Val Asp Gln Trp Glu
     180    185    190
Pro Asp Glu Val Tyr Trp Gly Lys Glu Ala Thr Trp Leu Gly Asp Glu
     195    200    205
Arg Tyr Ser Gly Lys Arg Asp Leu Glu Asn Pro Leu Ala Ala Val Gln
     210    215    220
Met Gly Leu Ile Tyr Val Asn Pro Glu Gly Pro Asn Gly Asn Pro Asp
     225    230    235    240
Pro Met Ala Ala Ala Val Asp Ile Arg Glu Thr Phe Arg Arg Met Ala
245 250

Met Asn Asp Val Glu Thr Ala Ala Leu Ile Val Gly Gly His Thr Phe
260 265 270

Gly Lys Thr His Gly Ala Gly Pro Ala Asp Leu Val Gly Pro Glu Pro
275 280 285

Glu Ala Ala Pro Leu Glu Gln Met Gly Leu Gly Trp Lys Ser Ser Tyr
290 295 300

Gly Thr Gly Thr Gly Lys Asp Ala Ile Thr Ser Gly Ile Glu Val
305 310 315 320

Trp Thr Asn Thr Pro Thr Lys Trp Asp Asn Ser Phe Leu Glu Ile Leu
325 330 335

Tyr Gly Tyr Glu Trp Glu Leu Thr Lys Ser Pro Ala Gly Ala Trp Gln
340 345 350

Tyr Thr Ala Lys Asp Gly Ala Gly Ala Gly Thr Ile Pro Asp Pro Phe
355 360 365

Gly Gly Pro Gly Arg Ser Pro Thr Met Leu Ala Thr Asp Leu Ser Leu
370 375 380

Arg Val Asp Pro Ile Tyr Glu Arg Ile Thr Arg Arg Trp Leu Glu His
385 390 395 400

Pro Glu Glu Leu Ala Asp Glu Phe Ala Lys Ala Trp Tyr Lys Leu Ile
405 410 415

His Arg Asp Met Gly Pro Val Ala Arg Tyr Leu Gly Pro Leu Val Pro
420 425 430

Lys Gln Thr Leu Leu Trp Gln Asp Pro Val Pro Ala Val Ser His Asp
435 440 445

Leu Val Gly Glu Ala Glu Ile Ala Ser Leu Lys Ser Glu Ile Arg Ala
450 455 460

Ser Gly Leu Thr Val Ser Gln Leu Val Ser Thr Ala Trp Ala Ala Ala
465 470 475 480
Ser Ser Phe Arg Gly Ser Asp Lys Arg Gly Gly Ala Asn Gly Gly Arg
485
Ile Arg Leu Gln Pro Gln Val Gly Trp Glu Val Asn Asp Pro Asp Gly
500
Asp Leu Arg Lys Val Ile Arg Thr Leu Glu Glu Ile Gln Glu Ser Phe
515
Asn Ser Ala Ala Pro Gly Asn Ile Lys Val Ser Phe Ala Asp Leu Val
530
Val Leu Gly Gly Cys Ala Ala Ile Glu Lys Ala Ala Lys Ala Gly
545
His Asn Ile Thr Val Pro Phe Thr Pro Gly Arg Thr Asp Ala Ser Gln
565
Glu Gln Thr Asp Val Glu Ser Phe Ala Val Leu Glu Pro Lys Ala Asp
580
Gly Phe Arg Asn Tyr Leu Gly Lys Gly Asn Pro Leu Pro Ala Glu Tyr
595
Met Leu Leu Asp Lys Ala Asn Leu Leu Thr Leu Ser Ala Pro Glu Met
610
Thr Val Leu Val Gly Gly Leu Arg Val Leu Gly Ala Asn Tyr Lys Arg
625
Leu Pro Leu Gly Val Phe Thr Glu Ala Ser Glu Ser Leu Thr Asn Asp
645
Phe Phe Val Asn Leu Leu Asp Met Gly Ile Thr Trp Glu Pro Ser Pro
660
Ala Asp Asp Gly Thr Tyr Gln Gly Lys Asp Gly Ser Gly Lys Val Lys
675
Trp Thr Gly Ser Arg Val Asp Leu Val Phe Gly Ser Asn Ser Glu Leu
690
Arg Ala Leu Val Gly Val Gly Ala Asp Asp Ala Glu Pro Lys Phe
705
Val Gln Asp Phe Val Ala Ala Trp Asp Lys Val Met Asn Leu Asp Arg
725
Phe Asp Val Arg
740

(2) INFORMATION FOR SEQ ID NO:22:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xii) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GACGCNNNNN NNNNNN
15
WHAT IS CLAIMED IS:

1. A method for selectively detecting *M. tuberculosis* in a sample containing DNA, said method comprising amplifying the DNA to generate a detectable amount of amplified DNA comprising a *katG* DNA fragment consisting of base 904 through 1523 of the *M. tuberculosis* *katG* gene as depicted in Figure 7 (SEQ ID NO: 20), wherein the generation of the *katG* DNA fragment is indicative of the presence of *M. tuberculosis* in the sample.

2. The method of claim 1 wherein the DNA is amplified in a polymerase chain reaction using oligonucleotide primer katG904 (SEQ ID NO:16) and oligonucleotide primer katG1523 (SEQ ID NO:17).

3. The method of claim 1 wherein the *katG* DNA fragment comprises a restriction site comprising either a G or a C at the nucleotide position occupied by base 1013 in codon 315 of the *M. tuberculosis* *katG* gene as depicted in Figure 7 (SEQ ID NO: 20), said method further comprising:

   contacting the *katG* DNA fragment with a restriction endonuclease that cleaves either at said restriction site comprising a G at the nucleotide position occupied by base 1013 of codon 315, or at said restriction site comprising a C at the nucleotide position occupied by base 1013 of codon 315, but not at both of said restriction sites, to yield at least one cleaved fragment;

   electrophoresing the at least one cleaved fragment to yield an electrophoretic mobility pattern comprising the at least one cleaved fragment; and

   analyzing the mobility pattern to selectively detect the presence of *M. tuberculosis* in the sample.

4. The method of claim 3 wherein the restriction endonuclease is *MspI*. 
5. The method of claim 3 wherein the electrophoresis comprises gel electrophoresis, and wherein the presence of *M. tuberculosis* in the sample is selectively detected using restriction fragment length polymorphism (RFLP) analysis of said electrophoretic mobility pattern.

6. The method of claim 1 wherein the sample is a biological fluid.

7. The method of claim 6 wherein the biological fluid is human sputum.

8. The method of claim 1 further comprising determining whether or not the *katG* DNA fragment has a S315T mutation, wherein the presence of a S315T mutation is indicative of an INH-resistant strain of *M. tuberculosis*.

9. The method of claim 8 wherein the *katG* DNA fragment comprises a restriction site comprising either a G or a C at the nucleotide position occupied by base 1013 in codon 315 of the *M. tuberculosis katG* gene as depicted in Figure 7 (SEQ ID NO: 20), and wherein the step of determining whether or not the *katG* DNA fragment has a S315T mutation comprises contacting the *katG* DNA fragment with a restriction endonuclease that cleaves either at said restriction site comprising a G at the nucleotide position occupied by base 1013 of codon 315, or at said restriction site comprising a C at the nucleotide position occupied by base 1013 of codon 315, but not at both of said restriction sites, to yield at least one cleaved fragment, and wherein cleavage at said restriction site is indicative of either the presence or the absence, but not both, of a S315T mutation in the *katG* DNA fragment.

10. The method of claim 9 wherein the restriction endonuclease is *MspI*, and wherein cleavage at said restriction site is indicative of the presence of a S315T mutation in the *katG* DNA fragment.
11. The method of claim 9 further comprising 
   electrophoresing the at least one cleaved fragment to yield an 
   electrophoretic mobility pattern comprising the at least one cleaved 
   fragment; and 
   analyzing the mobility pattern to determine the presence or 
   absence of a S315T mutation in the \textit{katG} DNA fragment.

12. The method of claim 11 wherein the electrophoresis comprises gel 
   electrophoresis, and wherein the presence of \textit{M. tuberculosis} in the 
   sample is selectively detected using restriction fragment length 
   polymorphism (RFLP) analysis of said electrophoretic mobility pattern.

13. A method for selectively detecting \textit{M. tuberculosis} in a sample containing 
   DNA, said method comprising: 
   (a) amplifying the DNA to generate a detectable amount of amplified 
       DNA comprising a \textit{katG} DNA fragment comprising base 904 
       through 1523 of the \textit{M. tuberculosis katG} gene, wherein the \textit{katG} 
       DNA fragment further comprises a restriction site comprising 
       either a G or a C at the nucleotide position occupied by base 1013 
       in codon 315 of the \textit{M. tuberculosis katG} gene as depicted in 
       Figure 7 (SEQ ID NO: 20); 
   (b) contacting the \textit{katG} DNA fragment with a restriction 
       endonuclease that cleaves either at said restriction site comprising 
       a G at the nucleotide position occupied by base 1013 of codon 
       315, or at said restriction site comprising a C at the nucleotide 
       position occupied by base 1013 of codon 315, but not at both of 
       said restriction sites, to yield at least one cleaved fragment; 
   (c) electrophoresing the at least one cleaved fragment to yield an 
       electrophoretic mobility pattern comprising the at least one 
       cleaved fragment; and 
   (d) analyzing the mobility pattern to selectively detect the presence of 
       \textit{M. tuberculosis} in the sample.
14. The method of claim 13 wherein the restriction endonuclease is \textit{MspI}.

15. The method of claim 13 wherein the sample is a biological fluid.

16. The method of claim 15 wherein the biological fluid is human sputum.

17. The method of claim 13 wherein the electrophoresis comprises gel electrophoresis, and wherein the presence of \textit{M. tuberculosis} in the sample is selectively detected using restriction fragment length polymorphism (RFLP) analysis of said electrophoretic mobility pattern.

18. A method for selectively detecting \textit{M. tuberculosis} in a sample containing DNA, said method comprising:

(a) amplifying the DNA to generate a detectable amount of amplified DNA comprising a \textit{katG} DNA fragment comprising base 904 through 1523 of the \textit{M. tuberculosis katG} gene as depicted in Figure 7 (SEQ ID NO: 20);

(b) contacting the \textit{katG} DNA fragment with a restriction endonuclease that cleaves at C/CGG to yield at least one cleaved fragment;

(c) electrophoresing the at least one cleaved fragment to yield a mobility pattern comprising the at least one cleaved fragment; and

(d) analyzing the mobility pattern to selectively detect the presence of \textit{M. tuberculosis} in the sample.

19. The method of claim 18 wherein the restriction endonuclease is \textit{MspI}.

20. The method of claim 18 wherein the sample is a biological fluid.

21. The method of claim 20 wherein the biological fluid is human sputum.
22. The method of claim 18 wherein the electrophoresis comprises gel electrophoresis, and wherein the presence of *M. tuberculosis* in the sample is selectively detected using restriction fragment length polymorphism (RFLP) analysis of said electrophoretic mobility pattern.
FIG. 1B
SUBSTITUTE SHEET (RULE 26)
FIG. 1D
FIG. 2
SUBSTITUTE SHEET (RULE 26)
BASE NUMBERS

5'  1  200  400  600  800  1000  1200  1400  3'
    126  417*  727  1333

RESTRICTION SITES

FIG. 3

SUBSTITUTE SHEET (RULE 26)
FIG. 4
DETECTION OF M. TUBERCULOSIS katG MUTATIONS AT CODONS 264 USING CfoI RFLP ANALYSIS

FIG. 6

264 A-T

katG

PCR PRODUCT

CfoI RFLP

W^+

983

633

31

40

150

150

129

129

22233

SUBSTITUTE SHEET (RULE 26)
FIG. 7A
SUBSTITUTE SHEET (RULE 26)
FIG. 7C
SUBSTITUTE SHEET (RULE 26)
RFLP analysis of M. tuberculosis katG mutations in codons 315 and 463.

4% Metaphor agarose (FMC), 1X TBE, 100 V

A&T = MW markers
B-S = MspI digests of PCR products (katG 904-1523)

C,D,F,G,H,K,L,N,Q wild genotype at codons 315 and 463
M,O RFLP indicating a mutation at codon 315
E,I,P RFLP indicating a mutation at codon 463
B,J RFLP indicating mutations at codons 315 and 463

FIG. 8
RFLP (RsaI) analysis of *M. tuberculosis* katG mutations in codon 337.

A - F  RsaI digest of katG (904-1523) PCR products
G  Blank
H  MWM

A, B, D, E, F  Wild genotype at codon 337
C  RFLP indicating mutation at codon 337

FIG. 9
RFLP (CfoI) analysis of *M. tuberculosis* katG mutations in codon 264

![Image of gel electrophoresis](image)

- 200 bp
- 100 bp
- 50 bp

6% PA, 200V, 20 min., EtBr stained

A, B, C, E  CfoI digest of katG (633-983) PCR product  
MWM

A, B, C  Wild genotype at codon 264  
E  RFLP indicating mutation at codon 264

FIG. 10

SUBSTITUTE SHEET (RULE 26)
INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C1201/68

According to International Patent Classification (IPC) or to both national classification and IPC.

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>WO 96 15267 A (THIRD WAVE TECH INC ; DAHLBERG JAMES E (US); LYAMICHEV VICTOR I (US) 23 May 1996 see page 213, line 11 - page 216</td>
<td>1-22</td>
</tr>
<tr>
<td>Y</td>
<td>WO 93 22454 A (PASTEUR INSTITUT ; MEDICAL RES COUNCIL (GB); ASSISTANCE PUBLIQUE (F) 11 November 1993 cited in the application see claims 1,6</td>
<td>1-22</td>
</tr>
</tbody>
</table>

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

27 August 1998

Date of mailing of the international search report

09/09/1998

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV RI/erikswijk
Tel: (+31-70) 340-2040, Tx: 31 651 epo nl, Fax: (+31-70) 340-3016

Authorized officer

Osborne, H
<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>MARTILA H ET AL: &quot;KATG mutations in isoniazid-resistant mycobacterium tuberculosis isolates recovered from finnish patients&quot; ANTIMICROBIAL AGENTS AND CHEMOTHERAPY, vol. 40, no. 9, - September 1996 pages 2187-9, XP002075617 see the whole document</td>
<td>1-22</td>
</tr>
</tbody>
</table>
### INTERNATIONAL SEARCH REPORT

**Information on patent family members**

<table>
<thead>
<tr>
<th>Patent document cited in search report</th>
<th>Publication date</th>
<th>Patent family member(s)</th>
<th>Publication date</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>CA 2203627 A</td>
<td>23-05-1996</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EP 0788557 A</td>
<td>13-08-1997</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 5719028 A</td>
<td>17-02-1998</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FR 2704002 A</td>
<td>21-10-1994</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 5633131 A</td>
<td>27-05-1997</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CA 2134103 A</td>
<td>11-11-1993</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EP 0639229 A</td>
<td>22-02-1995</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JP 7506003 T</td>
<td>06-07-1995</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JP 7059595 A</td>
<td>07-03-1995</td>
</tr>
</tbody>
</table>