

Title: Structural Implications of Mutations Conferring Rifampin Resistance in *Mycobacterium leprae*.

Running Title: Vedithi S.C et al., RpoB Mutations –M. leprae 2017

Authors: Sundeep Chaitanya Vedithi^{1,2,*}, Sony Malhotra¹, Madhusmita Das², Sheela Daniel², Nanda Kishore³, Anuja George⁴, Shantha Arumugam², Lakshmi Rajan², Mannam Ebenezer², David B Ascher⁵, Eddy Arnold⁶ & Tom L Blundell^{1,*}

Affiliations:

1. Department of Biochemistry, University of Cambridge, Tennis Court Rd., CB2 1GA, UK
2. Molecular Biology and Immunology Division, Schieffelin Institute of Health Research & Leprosy Center (SIH R & LC), Karigiri, Vellore, Tamil Nadu – 632106, India.
3. Department of Dermatology, Father Muller Medical College & Hospital, Mangalore, Karnataka - 575 002, India
4. Department of Dermatology, Trivandrum Medical College, Trivandrum, Kerala – 695011, India
5. Department of Biochemistry and Molecular Biology, Bio21 Institute, University of Melbourne, Parkville, VIC 3052, Australia.
6. Center for Advanced Biotechnology and Medicine (CABM), and Rutgers University Department of Chemistry and Chemical Biology, 679 Hoes Lane, Piscataway, NJ 08854, USA

*** Corresponding Authors:**

Professor Sir Tom Blundell FRS
Director of Research and
Professor Emeritus of Biochemistry

Dr Sundeep Chaitanya Vedithi
Department of Biochemistry
Sanger Building
University of Cambridge
80 Tennis Ct. Rd., CB2 1GA, UK

Email: tom@cryst.bioc.cam.ac.uk; tlb20@cam.ac.uk;
scv26@cam.ac.uk
Ph: +44 1223 766033

***Mycobacterium leprae* rpoB gene Sequence:**

>NC_002677.1:c2276805-2273269 *Mycobacterium leprae* TN chromosome, complete genome

```
GTGCTGGAAGGATGCATCTTGCAGATTTTCGGCCAGAGCAAGACAGACGTTAGTCCCTAGCCAGAGTCCGCC
CCCAAAGTTTCGCCCAACAACCTCCGTGCCCGGCGGCCAACCGAATTTCAATTTGCCAAGCTCCGCGAACC
GCTTGAGGTTCCGGGGCTACTTGATGTGCAGACTGATTCATTTGAGTGGTTGATCGGATCGCCGTGCTGG
CGTGCAGCGGCCGCAAGCCGCGGCATCTCAAGCCGGTGGTGGTCTCGAAGAGGTGCTCTACGAGCTGT
CGCCGATCGAGGATTTCTCCGGCTCAATGTCATTTGCTTTCTCCGATCCCCGTTTTGACGAAGTCAAGGC
GCCCCGTCGAAGAGTGC AAAAGACAAGGACATGACGTACGCGCCCCGCTGTTTCGTACGCGCCGAGTTCATC
AACAACAACACCGGGGAGATCAAGAGCCAGACGGTGTATATGGGCGACTTCCCTATGATGACTGAGAAGG
GAACCTTCATCATCAACGGGACCGAGCGTGTGCTGTTAGCCAGCTGGTGCCTCCCTGGAGTATACTT
CGACGAGACGATCGACAAGTCCACAGAAAAGACGCTGCATAGTGTCAAGGTGATTCACGCGCGGTGCC
TGGTTGGAATTCGATGTGATAAACGCGACACCGTGGTGTCCGCAATTGACCGGAAGCGCCGCAACCCG
TCACGGTGTCTCTCAAAGCGCTAGGTTGGACCAGTGCAGCAGATCACCGAGCGTTTTCGTTTTCTCCGAGAT
CATGCGCTCGACGCTGGAGAAGGACAACACAGTTGGCACCGACGAGGCGCTGCTAGACATCTATCGTAAG
TTGCGCCAGGTGAGCCGCCGACTAAGGAGTCCGCGCAGACGCTGTTGGAGAACCCTGTTCTTCAAGGAGA
AACGCTACGACCTGGCCAGGGTTGGTCTGTTACAAGGTCAACAAGAAGCTCGGGTTGCACGCCGTGAGTT
GATCACGCTCGTCCACGCTGACCGAAGAGGATGTGCTGCCACCATAGAGTACCTGGTTGCTCTGCATGAG
GGTCAGTCGACAATGACTGTCCAGGTGGGGTAGAAGTGCCAGTGGAACTGACGATATCGACCACCTTCG
GCAACCGCCGGCTGCGCACGGTCCGGCAATTGATCCAGAACCAGATCCGGTCCGTTATGTCGCGGATGGA
GCGGGTGGTCCGGGAGCGGATGACCACCCAGGACGTCGAGGCGATCACGCCGCGACGCTGATCAATATC
CGTCCGGTGGTCCCGCTATCAAGGAATTTCTTCGGCACCAGCCAGCTGTGCGAGTTCATGGATCAGAACA
ACCCTCTGTCCGGCCTGACCCACAAGCGCCGGCTGTGCGCGCTGGGCCCGGGTGGTTTTGTCCGCTGAGCG
TGCCCCGGCTAGAGGTTCCGTGACGTGCACCTTTCGCACTACGGCCGGATGTGCCCGATCGAGACTCCGGAG
GGCCCGAACATAGGTTGATCGGTTCAATTGTCGTTGACGCGCGGGTCAACCCCTTCGGGTTTCATCGAAA
CACCGTACCGCAAAGTGGTTGACGGTGTGGTCCAGCGACGAGATCGAATACTTGACCGCTGACGAGGAAGA
CCGCCATGTGCTGGCGCAGGCCAACTCGCCGATCGACGAGGCCGGCCGCTTCCCTCGAGCCGCGCGTGTG
GTGCGCCGCAAGGCGGGCAGGTGGAGTACGTGGCTCGTCCGAGGTGGATTACATGGATGCTCGCCAC
GCCAGATGGTGTCCGTGGCCACAGCGATGATTCCTTCTTGGAGCACGACGACGCCAACCCTGCCCTGAT
GGGCGCTAACATGCAGCGCCAAGCGGTTCCGTGGTGCAGCGAAGCACCGTTGGTGGGTACCGGTATG
GAGTTGCGCGCGGCCATCGACGCTGGCCACGCTCGTGTGCGGAGAAGTCCGGGGTGTGATCGAGGAGTTT
CCGCCGACTACATCACCGTGTGATGGCCGATGACGGCACCCGGCGGACTTATCGGATGCGTAAGTTTCGCGCG
CTCCAACCACGGCACCTGCGCCAACCAGTCCCCGATCGTGGATGCGGGGGATCGGGTTCGAGGCCGGCCAA
GTGATTGCTGACGGTCCGTGCACTGAGAACGGCGAGATGGCGTTGGCAAGAAGTTGCTGGTGGCGATCA
TGCCGTGGGAGGGTCACAACTACGAGGATGCGATCATCTGTCTAACCGACTGGTTCGAAGAGGACGTGCT
TACTTCGATTCACATTTGAGGAGCATGAGATCGACGCCCGTGCACCAAGCTGGGTGCTGAGGAGATCACC
CGGGCATTCCCAACCTCCGATGAGGTGCTAGCGACTTGGACGAGCGGGGCATCGTCCGGATTGGCG
CGGAGGTTTCGTGACGGTGTATCTCTGGTTGGCAAGGTCAACCCGAAAGGGGGAAACTGAGCTGACACCGGA
AGAGCGGTTGCTGCGGGCGATCTTCGGCGAAAAAGGCCCGCGAGGTCCGTGACACGCTCGCTGAAGGTGCCA
CACGGCGAATCCGGCAAGGTGATCGGCATTCGGGTGTTCTCCCATGAGGATGACGACGAGCTGCCCGCCG
GCGTCAACGAGCTGGTCCGTGTCTACGTAGCCAGAAGCGCAAGATCTCTGACGGTGACAAGCTGGCTGG
GCGGCACGGCAACAAGGGCGTGTGCGCAAGATCTTGCCTGCCGAGGATATGCCGTTTCTGCCAGACGGC
ACCCCGGTGGACATCATCTCAACACTCACGGGTGCCCGCGCGATGAACGTCGGTTCAGATCTTGAAAA
CCCACCTTGGGTGGGTAGCCAAGTCCGGCTGGAAGATCGACGTGGCCGGCGGTATACCGGATTTGGGCGGT
CAACTTGCCTGAGGAGTTGTTGACGCTGCGCCCAACCAGATCGTGTGACCCCCGTTGTTGACGGCGCC
AAGGAAGAGGAACTACAGGGCTGTTGCTCTCCACGTTGCCCAACCAGCGACGGCGATGTGATGGTGGGCG
GCGACGGCAAGGCGGTGCTCTTCGATGGGCGCAGCGGTGAGCCGTTCCCTTATCCGGTGACGGTTGGCTA
CATGTACATCATGAAGCTGCACCACTTGGTGGACGACAAGATCCACGCCCGCTCCACCGGCCGTACTCG
ATGATTACCCAGCAGCCGTTGGGTGGTAAGGCACAGTTCGGTGGCCAGCGATTCGGTGTGATGGAGTGTCT
GGCCATGCAGGCCTACGGTGCGGCCTACACGCTGCAGGAGCTGTTGACCATCAAGTCCGACGACACCCGT
CGGTCCGGTCAAGGTTTACGAGGCTATCGTTAAGGGTGAGAACATCCCCGAGCCGGGCATCCCCGAGTCCG
TTCAAGGTGCTGCTCAAGGAGTTACAGTCTGCTGTCTCAACGTCGAGGTGCTGCTCGTCCGACGGTGGCG
CGATCGAGTTGCGCAAGGTGAGGATGAGGACCTCGAGCGGGTGC GGCCAACCTCGGTATCAACTTGTCT
CCGCAACGAATCGGCGTCCATAGAAGATCTGGCTTAG
```

RpoB Wt:(Highlighted in the gene above)

```
GTCGAGGCGATCACGCCGAGACGCTGATCAATATCCGTCCGGTGGTCCCGCTATCAAGGAATTTCTTCGGCACC
AGCCAGCTGTGCGAGTTCATGGATCAGAACAACCCCTCTGTCCGGCTGACCCACAAGCGCCGGCTGTCCGGCGTG
GGCCCGGGTGGTTTTGTCCGCTGAGCGTGC CGGGCTAGAGGTCCGTGACGTGCACCCCTTCGCACTACGGCCGGATG
TGCCCGATCGAGACTCCGGAGGGCCCCAACATAGGTTCTGATCGGTTCAATTGTC
```

RpoB Wt. SEQUENCE ALIGNMENT:

```
Query 1 CGACAATGAACCGATCAGACCTATGTTCCGGGCCCTCCGGAGTCTCGATCGGGCACATCCG 60
|||||
Sbjct 276353 CGACAATGAACCGATCAGACCTATGTTCCGGGCCCTCCGGAGTCTCGATCGGGCACATCCG 276412

Query 61 GCCGTAGTGCGAAGGGTGCACGTCACGGACCTCTAGCCCGGCACGCTCACGCGACAAACC 120
|||||
Sbjct 276413 GCCGTAGTGCGAAGGGTGCACGTCACGGACCTCTAGCCCGGCACGCTCACGCGACAAACC 276472

Query 121 ACCCGGGCCCAGCGCCGACAGCCGGCGCTTGTGGGTGAGGCCGACAGAGGGTTGTTCTG 180
|||||
Sbjct 276473 ACCCGGGCCCAGCGCCGACAGCCGGCGCTTGTGGGTGAGGCCGACAGAGGGTTGTTCTG 276532

Query 181 ATCCATGAAGTGCAGACAGCTGGCTGGTGCCGAAGAATTCCTTGATAGCGGCGACCACCGG 240
|||||
Sbjct 276533 ATCCATGAAGTGCAGACAGCTGGCTGGTGCCGAAGAATTCCTTGATAGCGGCGACCACCGG 276592

Query 241 ACGGATATTGATCAGCGTCTGCGGCGTGATCGCCTCGAC 279
|||||
Sbjct 276593 ACGGATATTGATCAGCGTCTGCGGCGTGATCGCCTCGAC 276631
```

Translated Protein:

279 nucleotides, 93 amino acids, structure: sequence

```
1 GTCGAGGCGATCACG CCGCAGACGCTGATC AATATCCGTCCGGTG GTCGCCGCTATCAAG
1 V E A I T P Q T L I N I R P V V A A I K
61 GAATTCTTCGGCACC AGCCAGCTGTGCGAG TTCATGGATCAGAAC AACCCCTCTGTCGGGC
21 E F F G T S Q L S Q F M D Q N N P L S G
121 CTGACCCACAAGCGC CGGCTGTCGGCGCTG GGCCCGGGTGGTTG TCGCGTGAGCGTGCC
41 L T H K R R L S A L G P G G L S R E R A
181 GGGCTAGAGGTCCGT GACGTGCACCCTTCG CACTACGGCCGGATG TGCCCGATCGAGACT
61 G L E V R D V H P S H Y G R M C P I E T
241 CCGGAGGGCCGAAC ATAGGTCTGATCGGT TCATTGTCG
81 P E G P N I G L I G S L S
```

RpoB D441V: (Mutated Residue highlighted in Yellow)

>RpoB_26_forward

AGCTGTCGCAGTTCATGGTTCAGAACAACCCTCTGTTCGGGCCTGACCCACAAGCGCCGGCTGTTCGGCG
 CTGGGCCCGGGTGGTTTGTTCGCGTGAGCGTGCCGGGCTAGAGGTCCGTGACGTGCACCCCTTCGCACTA
 CGGCCGGATGTGCCCGATCGAGACTCCGGAGGGCCCCGAACATAGGTCTGATCGGTTTCATTGTCTGA

RpoB D441V SEQUENCE ALIGNMENT:

```

Query   1      AGCTGTCGCAGTTCATGGTTCAGAACAACCCTCTGTTCGGGCCTGACCCACAAGCGCCGGC  60
          |||
Sbjct  276552  AGCTGTCGCAGTTCATGGATCAGAACAACCCTCTGTTCGGGCCTGACCCACAAGCGCCGGC  276493

Query   61      TGTCGGCGCTGGGCCCGGGTGGTTTGTTCGCGTGAGCGTGCCGGGCTAGAGGTCCGTGACG  120
          |||
Sbjct  276492  TGTCGGCGCTGGGCCCGGGTGGTTTGTTCGCGTGAGCGTGCCGGGCTAGAGGTCCGTGACG  276433

Query   121     TGCACCCCTTCGCACTACGGCCGGATGTGCCCGATCGAGACTCCGGAGGGCCCCGAACATAG  180
          |||
Sbjct  276432  TGCACCCCTTCGCACTACGGCCGGATGTGCCCGATCGAGACTCCGGAGGGCCCCGAACATAG  276373

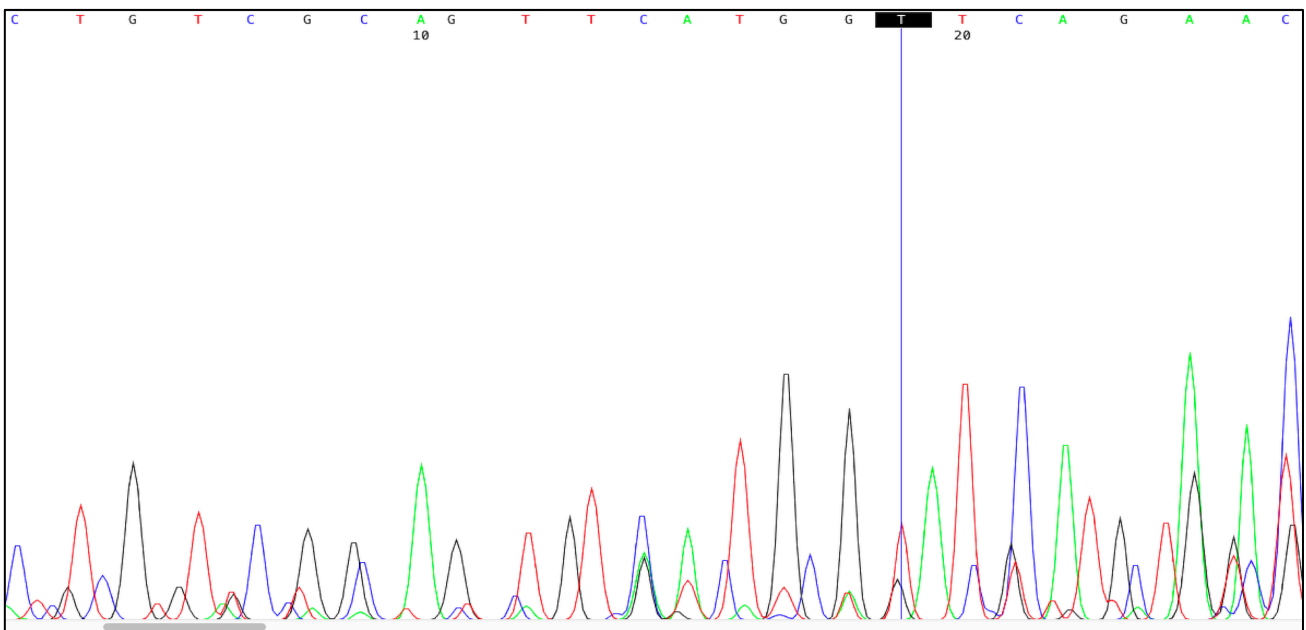
Query   181     GTCTGATCGGTTTCATTGTCTG  200
          |||
Sbjct  276372  GTCTGATCGGTTTCATTGTCTG  276353
    
```

Translated Protein:

201 nucleotides, 66 amino acids, structure: [AG](#) [sequence](#) [A](#)

```

1 CTGTCGCAGTTCATG GTTCAGAACAACCCT CTGTCGGGCCTGACC CACAAGCGCCGGCTG
1 L S Q F M V Q N N P L S G L T H K R R L
61 TCGGCGCTGGGCCCG GGTGGTTTGTTCGCGT GAGCGTGCCGGGCTA GAGGTCCGTGACGTG
21 S A L G P G G L S R E R A G L E V R D V
121 CACCCTTCGCACTAC GGCCGGATGTGCCCG ATCGAGACTCCGGAG GGCCCGAACATAGGT
41 H P S H Y G R M C P I E T P E G P N I G
181 CTGATCGGTTTCATTG TCG
61 L I G S L S
    
```



D441Y: (Mutated Residue highlighted in Yellow)

>RpoB_25

GATTCATG**T**ATCAGAACAACCCCTCTGTCGGGCCTGACCCACAAGCGCCGGCTGTCGGCGCTGGGCCCCG
GGTGGTTTGTTCGCGTGAGCGTGCCGGGCTAGAGGTCCGTGACGTGCACCCTTCGCACTACGGCCGGAT
GTGCCCGATCGAGACTCCGGAGGGCCCCGAACATAGGTCTGATCGGTTTCATTGTCTGA

RpoB: D441Y: Sequence Alignment:

```
Query 3      TTCATGTATCAGAACAACCCCTCTGTCGGGCCTGACCCACAAGCGCCGGCTGTCGGCGCTG 62
          ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 276541 TTCATGGATCAGAACAACCCCTCTGTCGGGCCTGACCCACAAGCGCCGGCTGTCGGCGCTG 276482

Query 63     GGCCCGGGTGGTTTGTTCGCGTGAGCGTGCCGGGCTAGAGGTCCGTGACGTGCACCCTTCG 122
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 276481 GGCCCGGGTGGTTTGTTCGCGTGAGCGTGCCGGGCTAGAGGTCCGTGACGTGCACCCTTCG 276422

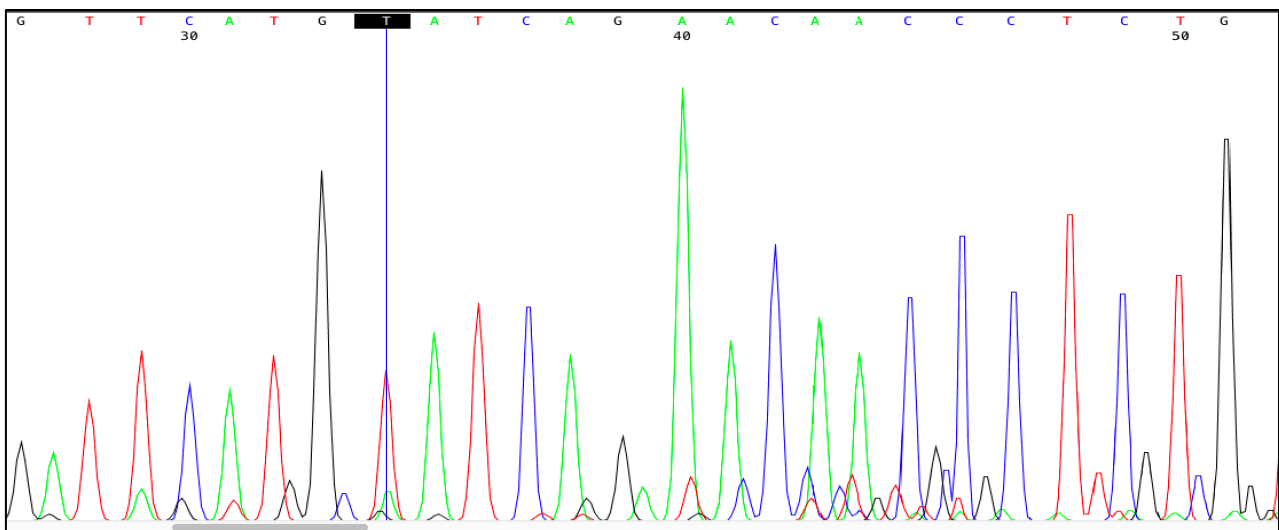
Query 123    CACTACGGCCGGATGTGCCCGATCGAGACTCCGGAGGGCCCCGAACATAGGTCTGATCGGT 182
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 276421 CACTACGGCCGGATGTGCCCGATCGAGACTCCGGAGGGCCCCGAACATAGGTCTGATCGGT 276362

Query 183    TCATTGTCTG 191
          ||||||||
Sbjct 276361 TCATTGTCTG 276353
```

Translated Protein:

192 nucleotides, 63 amino acids, structure: [GA](#) [sequence](#) [A](#)

```
1 TTCATGTATCAGAAC AACCCCTCTGTCGGGC CTGACCCACAAGCGC CGGCTGTCGGCGCTG
1 F M Y Q N N P L S G L T H K R R L S A L
61 GGCCCGGGTGGTTTGTTCGCGTGAGCGTGCC GGGCTAGAGGTCCGT GACGTGCACCCTTCG
21 G P G G L S R E R A G L E V R D V H P S
121 CACTACGGCCGGATG TGCCCGATCGAGACT CCGGAGGGCCCCGAAC ATAGGTCTGATCGGT
41 H Y G R M C P I E T P E G P N I G L I G
181 TCATTGTCTG
61 S L S
```



H476R: (Mutated Residue highlighted in Yellow)

GGGGCCCCGGTTAGTGCGAAGGG**C**GCACGTCACGGACCTCTAGCCCGGCACGCTCACGCGAC
 AAACCACCCGGGCCAGCGCCGACAGCCGGCGCTTGTGGGTGAGGCCCGACAGAGGGTTGTT
 CTGATCCATGAACTGCGACAGCTGGCTGGTGCCGAAGAATTCCTTGATAGCGGCGACCACCG
 GACGGATATTGATCAGCGTCTGCGGCGTGATCGCCTCGAC

```

Query 12 TAGTGCGAAGGGCGCACGTCACGGACCTCTAGCCCGGCACGCTCACGCGACAAACCACCC 71
          |||
Sbjct 276417 TAGTGCGAAGGGTGCACGTCACGGACCTCTAGCCCGGCACGCTCACGCGACAAACCACCC 276476

Query 72 GGGCCCAGCGCCGACAGCCGGCGCTTGTGGGTGAGGCCCGACAGAGGGTTGTTCTGATCC 131
          |||
Sbjct 276477 GGGCCCAGCGCCGACAGCCGGCGCTTGTGGGTGAGGCCCGACAGAGGGTTGTTCTGATCC 276536

Query 132 ATGAACTGCGACAGCTGGCTGGTGCCGAAGAATTCCTTGATAGCGGCGACCACCGGACGG 191
          |||
Sbjct 276537 ATGAACTGCGACAGCTGGCTGGTGCCGAAGAATTCCTTGATAGCGGCGACCACCGGACGG 276596

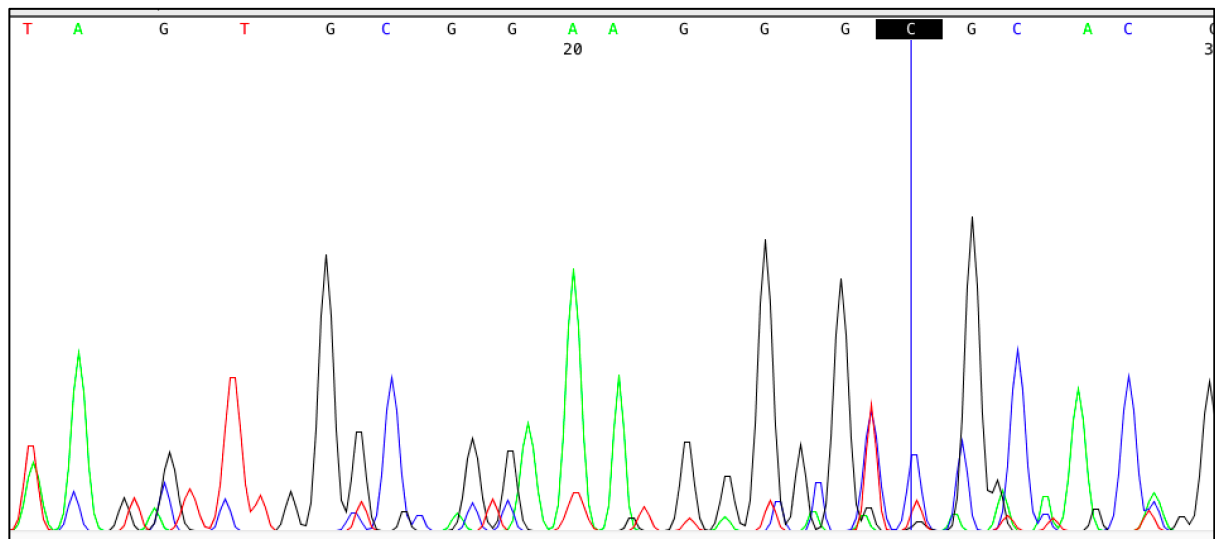
Query 192 ATATTGATCAGCGTCTGCGGCGTGATCGCCTCGAC 226
          |||
Sbjct 276597 ATATTGATCAGCGTCTGCGGCGTGATCGCCTCGAC 276631
  
```

Translated Reverse frame 1

226 nucleotides, 75 amino acids, structure: sequence C

```

1 GTCGAGGCGATCACG CCGCAGACGCTGATC AATATCCGTCCGGTG GTCGCCGCTATCAAG
1 V E A I T P Q T L I N I R P V V A A I K
61 GAATTCTTCGGCACC AGCCAGCTGTGCGAG TTCATGGATCAGAAC AACCCCTCTGTCGGGC
21 E F F G T S Q L S Q F M D Q N N P L S G
121 CTGACCCACAAGCGC CGGCTGTGCGGCGCTG GGCCCGGTGGTTTG TCGCGTGAGCGTGCC
41 L T H K R R L S A L G P G G L S R E R A
181 GGGCTAGAGGTCCGT GACGTGCGCCCTTCG CACTAACCGGGGCC
61 G L E V R D V R P S H * P G P
  
```



S437L: (Mutated Residue highlighted in Yellow)

CCGGCGCTTGTGGGTCAGGCCCGACAGAGGGTTGTTCTGATCCATGAACTGCAACAGCTGGCTGGTGC
CGAAGAATTCCTTGATAGCGGCGACCACCGGACGGATATTGATCAGCGTCTGCGGCGTGATCGCCTCG
AC

Query 1 CCGGCGCTTGTGGGTCAGGCCCGACAGAGGGTTGTTCTGATCCATGAACTGCAACAGCTG 60
|||||
Sbjct 276494 CCGGCGCTTGTGGGTCAGGCCCGACAGAGGGTTGTTCTGATCCATGAACTGCGACAGCTG
276553

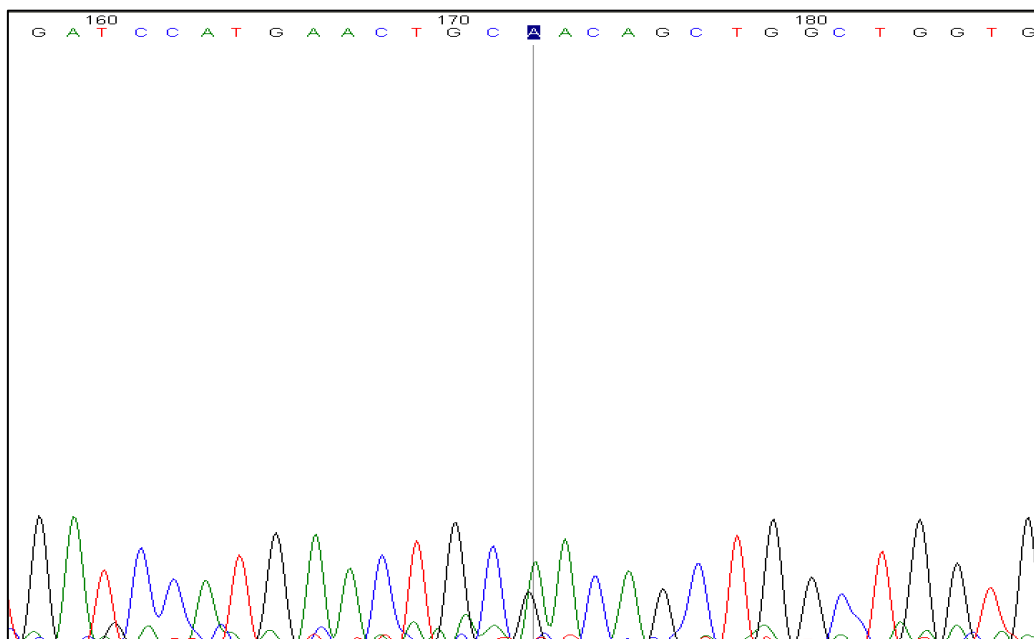
Query 61 GCTGGTGCCGAAGAATTCCTTGATAGCGGCGACCACCGGACGGATATTGATCAGCGTCTG 120
|||||
Sbjct 276554 GCTGGTGCCGAAGAATTCCTTGATAGCGGCGACCACCGGACGGATATTGATCAGCGTCTG
276613

Query 121 CGGCGTGATCGCCTCGAC 138
|||||
Sbjct 276614 CGGCGTGATCGCCTCGAC 276631

Translated Reverse frame 1

138 nucleotides, 46 amino acids, structure: sequence

1 GTCGAGGCGATCAG CCGCAGACGCTGATC AATATCCGTCCGGTG GTCGCCGCTATCAAG GAATTCCTCGGCACC
1 V E A I T P Q T L I N I R P V V A A I K E F F G T
76 AGCCAGCTGTTG CAG TTCATGGATCAGAAC AACCCCTCTGTCGGGC CTGACCCACAAGCGC CGG
26 S Q L L Q F M D Q N N P L S G L T H K R R



II: Multiple sequence alignment of the β -subunit of RNA Polymerase in Mycobacterial Species *M. abscessus*, *M. ulcerans*, *M. kansasii*, *M. leprae* and *M. tuberculosis*.

(The residue positions for mutations were highlighted in yellow)

CLUSTAL multiple sequence alignment by MUSCLE (3.8)

```

abscessus      -MAGQVVQYGRHRK-----RRNYARISEVLELPLNLEIQTKS
ulcerans      MLEGRILADFRQTD--ASLHQGRPQSSSNSSVPGAPNRVVFSAKLREPLEVPGLLDVQTD
kansasii      MLEGRILADFRQSKTDASPAQSRPQSSSNSSVPGAPNRVVFSAKLREPLEVPGLLDVQTD
leprae       MLEGCILPDFGQSKTDVSPQSRPQSSPNNSVPGAPNRISFAKLREPLEVPGLLDVQTD
tuberculosis  MLEGCILADSRQSKTAASPSRSPQSSSNSSVPGAPNRVVFSAKLREPLEVPGLLDVQTD
               : * ::      : .                               * .:.*: * **:*.*::**.*

abscessus      YDWFLKE-----GLEMFRDISPIEDFTGNLSLEFVDYRLGEPK
ulcerans      FEWLIGSQRWRESAAQRGD---ATPVGGLEEVLYELSPIEDFSGSMSLSFSDFPRFDEVK
kansasii      FEWLIGSQRWRSIAIARGEAAAGGINPIGGLEEVLDELSPIDFSGSMSLSFSDFPRFDEVK
leprae       FEWLIGSPCWRAAAASRGD---LKPVGGLEEVLYELSPIEDFSGSMSLSFSDFPRFDEVK
tuberculosis  FEWLIGSPRWRESAAERGD---VNPVGGLEEVLYELSPIEDFSGSMSLSFSDFPRFDDVK
               ::*:: .                               ** *:: :*****.*.:**.* * *.: : *

abscessus      YDLEESKNRDATYAAPLRVKVRLI IKETGEVKEQEVFMGDFPLMTDTGTFFVINGAERVIV
ulcerans      APVDECKDKDMTYAAPLFVTAEFINNTGEIKSQTVFMGGFPMTEKGTFIINGTERVVV
kansasii      APVDECKDKDMTYAAPLFVTAEFINNTGEIKSQTVFMGDFPMTEKGTFIINGTERVVV
leprae       APVEECKDKDMTYAAPLFVTAEFINNTGEIKSQTVFMGDFPMTEKGTFIINGTERVVV
tuberculosis  APVDECKDKDMTYAAPLFVTAEFINNTGEIKSQTVFMGDFPMTEKGTFIINGTERVVV
               :*:.*.* ***** *.. : * :*:**.*.* *****.*.:**.***.***.*

abscessus      SQLVRSPPVYFNEKIDKN-GRENYDATIIPNRGAWLEYETDAKDIVYVRIDRTRKLPVTV
ulcerans      SQLVRSPPVYFDETIDKSTDKLHLSVKVIPSARGAWLEFDVVKRDTVGVRIDRKRQPVTV
kansasii      SQLVRSPPVYFDETIDKSTDKLHLSVKVIPSARGAWLEFDVVKRDTVGVRIDRKRQPVTV
leprae       SQLVRSPPVYFDETIDKSTDKLHLSVKVIPSARGAWLEFDVVKRDTVGVRIDRKRQPVTV
tuberculosis  SQLVRSPPVYFDETIDKSTDKLHLSVKVIPSARGAWLEFDVVKRDTVGVRIDRKRQPVTV
               *****.***:*.***.* . . :...:*.*****:.* * * *****.* *:*

abscessus      LLRALGFSSDQEIIDLLGDSEYLRNTLEKDSSTENTEQALLEIYERLRPGEPPPTVENAKSL
ulcerans      LLKALGWSNEQ-IHERFGFSEIMMGTLEKDNTAGTDEALLDIYRKLRPGEPPPTKESAQTL
kansasii      LLKALGWTSEI-ITERFGFSEIMMSTLEKDNTAGTDEALLDIYRKLRPGEPPPTKESAQTL
leprae       LLKALGWTSEI-ITERFGFSEIMRSTLEKDNTVGTDEALLDIYRKLRPGEPPPTKESAQTL
tuberculosis  LLKALGWTSEI-IVERFGFSEIMRSTLEKDNTVGTDEALLDIYRKLRPGEPPPTKESAQTL
               **.***.:... * : :* * * : .*****.* *.:**.***.* ***** *.*:*

abscessus      LYSRFFDPKRYDLASVGRYKTNKKLHLKHRLFNQKLAEPVDSSETGEIVAAEETVLDRRK
ulcerans      LENLFFKEKRYDLARVGRYKVNKKLGL-----NAGQPITSS-----
kansasii      LENLFFKEKRYDLARVGRYKVNKKLGL-----NTNHPITTT-----
leprae       LENLFFKEKRYDLARVGRYKVNKKLGL-----HAGELITSS-----
tuberculosis  LENLFFKEKRYDLARVGRYKVNKKLGL-----HVGEPIITSS-----
               * . ** . ***** *****.***.* * : * . :

abscessus      LDEIMNVLESNANSEVFELEGSVIDEPVEIQSIKVYVPNDEEERTTTVIGNALPDSEVKC
ulcerans      -----TLTEEDV-VATIEYLVRLHEGQTAMTAPGGVEVPVE---
kansasii      -----TLTEEDV-VATIEYLVRLHEGQATMTVPGGVEVPVE---
leprae       -----TLTEEDV-VATIEYLVRLHEGQSTMTVPGGVEVPVE---
tuberculosis  -----TLTEEDV-VATIEYLVRLHEGQTTMTVPGGVEVPVE---
               : : * * : :*: * * : : * . *.. *

abscessus      ITPADI IASMSYFFNLLNGIGYTDIDHFGNRRLRSV GELLQNFRIGLSRMERVVRRM
ulcerans      -----TDDIDHFGNRRLRTV GELIQNQIRVGM SRMERVVRRM
kansasii      -----TDDIDHFGNRRLRTV GELIQNQIRVGM SRMERVVRRM
leprae       -----TDDIDHFGNRRLRTV GELIQNQIRVGM SRMERVVRRM
tuberculosis  -----TDDIDHFGNRRLRTV GELIQNQIRVGM SRMERVVRRM
               *****.******.***.***.*.:**.***.***.*

abscessus      SIQDTSITPQQLINIRPVIAIKEFFFGSSQLS 437 QFMDQANPLAELTHKRRLSALGPGLT
ulcerans      TTQDVEAITPQTLINIRPVVAAIKEFFFGTSQLS 441 QFMDQNNPLSLGTHKRRLSALGPGLS
kansasii      TTQDVEAITPQTLINIRPVVAAIKEFFFGTSQLS 437 QFMDQNNPLSLGTHKRRLSALGPGLS
leprae       TTQDVEAITPQTLINIRPVVAAIKEFFFGTSQLS 437 QFMDQNNPLSLGTHKRRLSALGPGLS
tuberculosis  TTQDVEAITPQTLINIRPVVAAIKEFFFGTSQLS 437 QFMDQNNPLSLGTHKRRLSALGPGLS
               : **.:**.* *****.*.:**.***.***.***.***.***.***.***.*

abscessus      RERAGMEVRDV 476 YSHYGRMCPIETPEGPNIGLINSLSYARVNEFGFIETPYRKVDIDTN
ulcerans      RERAGLEVRDV 476 PSHYGRMCPIETPEGPNIGLIGLSVYARVNPFGFIETPYRKV-VD-G
kansasii      RERAGLEVRDV 476 PSHYGRMCPIETPEGPNIGLIGLSVYARVNPFGFIETPYRKV-ID-G
leprae       RERAGLEVRDV 476 PSHYGRMCPIETPEGPNIGLIGLSVYARVNPFGFIETPYRKV-VD-G
tuberculosis  RERAGLEVRDV 476 PSHYGRMCPIETPEGPNIGLIGLSVYARVNPFGFIETPYRKV-VD-G
               **** * :***** *****.***.***.***.***.***.***.***.* : * .
    
```



```

abscessus          SITDQIDYLTADEEDSYVVAQANSRLDENGRFLDDEVVCRFR-GNNTVMAKEKMDYMDVS
ulcerans           VVSDEIHYLTADEEDRHVVAQANSPIDAQGRFVEPRVLRVKAGEVEYVPSSEVDYMDVS
kansasii           LVTDEIHYLTADEEDRHVVAQANSPIDAEGRFVEPRVLRVKAGEVEYVPSSEVDYMDVS
leprae             VVSDEIEYLTADEEDRHVVAQANSPIDEAGRFLEPRVLRVKAGEVEYVPSSEVDYMDVS
tuberculosis       VVSDEIVYLTADEEDRHVVAQANSPIADAGRFEPRVLRVKAGEVEYVPSSEVDYMDVS
                  :.*:* ***** :***** :* **:: *:* * .*: :...:*****

abscessus          PKQVVAATACIPFLENDDSNRALMGANMQRQAVPLMNPEAPFVGTGMEHVAARDSGAAI
ulcerans           PRQMVSVATAMIPFLEHDDANRALMGANMQRQAVPLVLRSEAPLVGTGMEELRAAIDAGDVV
kansasii           PRQMVSVATAMIPFLEHDDANRALMGANMQRQAVPLVLRSEAPLVGTGMEELRAAIDAGDVV
leprae             PRQMVSVATAMIPFLEHDDANRALMGANMQRQAVPLVLRSEAPLVGTGMEELRAAIDAGHVV
tuberculosis       PRQMVSVATAMIPFLEHDDANRALMGANMQRQAVPLVLRSEAPLVGTGMEELRAAIDAGDVV
                  *.*.*.* *****.*:*.....*.***.***** ** *:* ..

abscessus          TAKHRGRVEHVESNEILVRRLEENGTEHEGELDRYPLAKFKRSNSGTCYNQRPVSVGD
ulcerans           VADKAGVIEEVSADYITV---MADDGTRHT-----YRMRKFARSNHGTGANQSPVVDAGE
kansasii           VAEKSGVIEEVSADYITV---MADDGTRHT-----YRMRKFARSNHGTGANQSPVVDAGE
leprae             VAEKSGVIEEVSADYITV---MADDGTRHT-----YRMRKFARSNHGTGANQSPVVDAGD
tuberculosis       VAEESGVIEEVSADYITV---MHDNGTRHT-----YRMRKFARSNHGTGANQCPVVDAGD
                  *. * :* * :.* * : :.* * . * : ** ** ** **** * * * * .*.

abscessus          VVEYNEILADGPSMELGEMALGRNVVVGFMWDGYNVEDAVIMSERLVKDDVYTSIHIEE
ulcerans           RVEAGQVIADGPCTQNGEMALGKNLLVAIMPWEHNYEDAIILSNRLEEDVLTSIHIEE
kansasii           RVEAGQVIADGPCTQNGEMALGKNLLVAIMPWEHNYEDAIILSNRLEEDILTSIHIEE
leprae             RVEAGQVIADGPCTENGEMALGKNLLVAIMPWEHNYEDAIILSNRLEEDVLTSIHIEE
tuberculosis       RVEAGQVIADGPCTDDGEMALGKNLLVAIMPWEHNYEDAIILSNRLEEDVLTSIHIEE
                  ** ..:.*.*. : *****.*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

abscessus          YESEARDTKLGPEEITRDIPNVSENALKNLDDRGIIVVGAEVKDGDIILVGKVPKGVTEL
ulcerans           HEIDARDTKLGAEIITRDIPNVSDEVLADLDERGIVRIGAEVRDGDILVGKVPKGETEL
kansasii           HEIDARDTKLGAEIITRDIPNVSDEVLADLDERGIVRIGAEVRDGDILVGKVPKGETEL
leprae             HEIDARDTKLGAEIITRDIPNVSDEVLADLDERGIVRIGAEVRDGDILVGKVPKGETEL
tuberculosis       HEIDARDTKLGAEIITRDIPNISDEVLADLDERGIVRIGAEVRDGDILVGKVPKGETEL
                  :* :*****.*****.*:*:* * :*:.*:* :*****.***** * * *

abscessus          TAEERLLHAI FGEKAREVRDTSLRVPHGAGGIVLDVKVFNREEGDDTLSPGVNLVVRVYI
ulcerans           TPEERLLRAIFGEKAREVRDTSLKVPHGESGKVI G I R V F S R E D - D E L P A G V N E L V R V Y V
kansasii           TPEERLLRAIFGEKAREVRDTSLKVPHGESGKVI G I R V F S R E D - D E L P A G V N E L V R V Y V
leprae             TPEERLLRAIFGEKAREVRDTSLKVPHGESGKVI G I R V F S H E D - D E L P A G V N E L V R V Y V
tuberculosis       TPEERLLRAIFGEKAREVRDTSLKVPHGESGKVI G I R V F S R E D - E D E L P A G V N E L V R V Y V
                  *.*****.*****.*.* * :.*.*.* : * * .*.***.*****:

abscessus          VQKRKIHVGDKMCGRHGNKGVISKIVPEEDMPYLPDGRPIDIMLNPLGVPSRMNIGQVLE
ulcerans           AQKRKISDGDKLAGRHGNKGVIGKILPAEDMPFLPDGTPVDIILNTHGVPRRMNIGQVLE
kansasii           AQKRKISDGDKLAGRHGNKGVIGKILAQEDMPFLPDGTPVDIILNTHGVPRRMNIGQVLE
leprae             AQKRKISDGDKLAGRHGNKGVIGKILPAEDMPFLPDGTPVDIILNTHGVPRRMNVGQVLE
tuberculosis       AQKRKISDGDKLAGRHGNKGVIGKILPVEDMPFLADGTPVDIILNTHGVPRRMNIGQVLE
                  .***** **.*.*****.*:*.* * :*:.*.* ** ** *:*:*

abscessus          LHLGMAAKN-----LGIH-----VAVPVDGANDDDVWSTIE---
ulcerans           THLGWVAKSGWNIDVANGVPEWAGKLPENLLSAQPDSIVSTPVFDGAQEAELOGLLSATL
kansasii           THLGWCAHSGWQI---AGSPDWAANLPEALRSAEPNQIVSTPVFDGAQEAELOGLLSSTL
leprae             THLGWVAKSGWKIDVAGGIPDVAWVNLPEELHHAAPNQIVSTPVFDGAKEEELQGLLSSTL
tuberculosis       THLGWCAHSGWKVDAAGKVPDWAARLPDELLEAQPNAIVSTPVFDGAQEAELOGLLSCTL
                  *** *:* . * :*:*:*:*:*: :.*.*.*

abscessus          -----EAGMARDGKTIVLYDGRTEGPFVDNRISVGVYMLKLAHMVDDKIHARSTGPYSLVT
ulcerans           PNRDGEVLVDGDGKAVLFDGRSGEPFPYPVTVGYMYIMKLHHLVDDKIHARSTGPYSMIT
kansasii           PNRDGDVLVDGDGKAVLFDGRSGEPFPYPVTVGYMYIMKLHHLVDDKIHARSTGPYSMIT
leprae             PNRDGDVMVGDGKAVLFDGRSGEPFPYPVTVGYMYIMKLHHLVDDKIHARSTGPYSMIT
tuberculosis       PNRDGDVLVDADGKAMLFDGRSGEPFPYPVTVGYMYIMKLHHLVDDKIHARSTGPYSMIT
                  :. : *:* :*:*:*:* * :* * :.* * * :*:*:*:*:*:*:*

abscessus          QQPLGGKAQFGGQRFGEMEVALEAYGAAYTLQEILTYKSDDTVGRVKTYESIVKGENIS
ulcerans           QQPLGGKAQFGGQRFGEMEVCWAMQAYGAAYTLQELLTIKSDDTVGRVKVYEAIVKGENIP
kansasii           QQPLGGKAQFGGQRFGEMEVCWAMQAYGAAYTLQELLTIKSDDTVGRVKVYEAIVKGENIP
leprae             QQPLGGKAQFGGQRFGEMEVCWAMQAYGAAYTLQELLTIKSDDTVGRVKVYEAIVKGENIP
tuberculosis       QQPLGGKAQFGGQRFGEMEVCWAMQAYGAAYTLQELLTIKSDDTVGRVKVYEAIVKGENIP
                  *****.*****.*:*:*.*.* * * * * * * * .*.*****.

abscessus          RPSVPEFRVLMKELQSLGLDVKVMDEHDNEIEMADVDEDED----ATERKVDLQKDVPE
ulcerans           EPGIPESFKVLLKELQSLCLNVEVLSSDGAIELREGEDEDLERAAANLGINLSRNESAS

```

kansasii	EPGIPESFKVLLKELQSLCLNVEVLSSDGAAIELREGEDEDLERAANLGINLSRNESAS
leprae	EPGIPESFKVLLKELQSLCLNVEVLSSDGAAIELREGEDEDLERAANLGINLSRNESAS
tuberculosis	EPGIPESFKVLLKELQSLCLNVEVLSSDGAAIELREGEDEDLERAANLGINLSRNESAS
	*.:****.**:***** *:*:*. . . **: : :*** *: : :*. . . : .
abscessus	TQKETNE
ulcerans	VEDLA--
kansasii	VEDLA--
leprae	IEDLA--
tuberculosis	VEDLA--
	:. :