

DNA polymerase III α subunit from *Mycobacterium tuberculosis* H37Rv: Homology modeling and molecular docking of its inhibitor

Gagan Chhabra¹, Aparna Dixit², Lalit C. Garg^{1*}

¹Gene Regulation Laboratory, National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi – 110067, India; ²Gene Regulation Laboratory, School of Biotechnology, Jawaharlal Nehru University, New Delhi – 110067, India; Lalit Garg - Email: lalit@nii.res.in, lalitcgarg@yahoo.com; Phone: +91 11 26703652; Fax: +91 11 26742125; *Corresponding author

Received March 03, 2011; Accepted March 16, 2011; Published March 26, 2011

Abstract:

The alpha subunit of Mycobacterial DNA polymerase III holo enzyme catalyzes the polymerization of both DNA strands. The present investigation reports three dimensional (3-D) structure model of DNA polymerase III α subunit of *Mycobacterium tuberculosis* H37Rv (MtbDnaE1) generated using homology modeling with the backbone structure of DNA polymerase III α of *Thermus aquaticus* as a template. The model was evaluated at various structure verification servers, which assess the stereo chemical parameters of the residues in the model, as well as structural and functional domains. Comparative analysis of MtbDnaE1 structure reveals the structure of its catalytic domain to be unrelated to that of the human. Successful docking of known inhibitor of bacterial DNA polymerases, 251D onto the modeled MtbDnaE1 was also performed. Therefore, the structure model of MtbDnaE1, a potential anti-mycobacterial target, opens a new avenue for structure-based drug designing against the pathogen.

Keywords: *Mycobacterium tuberculosis*; DNA polymerase III α subunit; homology modeling; drug designing

Abbreviations: aa, amino acid(s); PolIII α , DNA polymerase III alpha subunit; Taq Pol III α , Pol III α of *Thermus aquaticus*; MtbDnaE1, PolIII α of *Mycobacterium tuberculosis*.

Background:

Mycobacterium tuberculosis is the prime killer among infectious agents, accounting for 7% of all deaths and 26% of all preventable deaths [1]. The cure of tuberculosis is a special problem in the field of chemotherapy. Many of the drugs employed to treat the disease are used only for treating infections caused by Mycobacteria. Treatment of the active TB cases always includes simultaneous therapy with two or more of the frontier drugs such as isoniazid, ethambutol, rifampicin, and streptomycin, which are used to decrease the rate of emergence of resistant strains as well as to increase the antibacterial effect [2]. Recent outbreaks of tuberculosis caused by multidrug-resistant strains, mainly in individuals infected with HIV have created a worldwide interest in expanding current therapeutic programs. Analysis of complete genome sequences of the pathogen *M. tuberculosis* [3] and the host *Homo sapiens* (The Genome International Consortium, 2001) allows one to identify the functions unique to the host and the pathogen, thus facilitating the development of drugs specifically targeting the pathogen. Even among the pathways shared by the host and the pathogen, there are several proteins from pathways involved in lipid metabolism, carbohydrate metabolism, amino acid metabolism, energy metabolism, vitamin and cofactor biosynthesis, and nucleotide metabolism, which do not bear similarity to host proteins [4]. The enzymes in the pathways of *M. tuberculosis*, which do not exhibit similarity to any protein from the host,

represent attractive potential drug targets. Amino acid sequence of the probable DNA polymerase III alpha subunit (Rv1547) of *M. tuberculosis* does not exhibit significant identity (below BLASTp e-value threshold of 0.005) with its counterpart in human host, and therefore can be a potential drug target against the pathogen.

DNA polymerases play fundamental roles in DNA replication and repair. Among the five known eubacterial DNA polymerases (I-V), polymerase III (PolIII) is accountable for catalyzing DNA replication, and whereas others (PolI, II, IV and V) are involved in supplementary roles in replication and repair. The eubacterial PolIII holoenzyme, is a complex made up of ten subunits namely the replicase (α, ϵ, θ), the clamp loader ($\gamma, \delta, \delta', \zeta, \chi, \psi$), and the sliding clamp (β) [5, 6]. The α subunit works as a replicative DNA polymerase at the replication fork and plays a central role in the complex [7]. Based on amino acid sequence comparisons, DNA-dependent DNA polymerases are divided into six families denoted by A, B, C, D, X, and Y [8], of which A, B and C family polymerases play critical roles in replication. Family A includes eubacterial DNA polymerase I whereas family B represents all archaeal and eukaryotic replicative polymerases. The α subunit of PolIII is a member of family C [8]. The structures of the α subunit from family C have been determined recently from *Thermus aquaticus* (TaqPolIII α) [9] and

Escherichia coli (EcoPolIII α) [10]. Although the crystal structure of MtbDnaE1 has not yet been resolved, the amino acid sequence of MtbDnaE1 shows high sequence identity with the TaqPolIII α and EcoPolIII α . The present study has been conducted to understand and elucidate the 3-D structure of DNA polymerase III alpha subunit of *M. tuberculosis* (MtbDnaE1) by homology modeling. This work describes, for the first time a structural model of MtbDnaE1. Knowledge of these structural features of MtbDnaE1 is essential for establishing its catalytic mechanisms of action at the molecular level, as well as to target the protein for designing effective and selective drug. We have also attempted to predict the interaction between DNA polymerase III alpha and its potential inhibitor namely 251D *in silico* via the process of docking.

Methodology:

Phylogenetic analysis and sequences alignment:

For phylogenetic analysis, the amino acid sequence of the MtbDnaE1 [11] was aligned against homologues sequences collected from the SwissProt (<http://www.expasy.org/sprot/>) database at PSI-BLAST E-value: 0.001 by using ConSurf server [12]. Multiple sequence alignment was performed using ClustalW [13]. Putative phylogenetic trees prediction for multiple sequence alignment by cluster algorithms was done by TreeTop- Phylogenetic Tree Prediction server [14]. In the bootstrap, multiple alignment was resampled 100 times, i.e. 100 trees were generated.

Structure modeling and validation, and identification of functional domains:

A blast search for MtbDnaE1 sequence (accession number NP_216063) was performed to search for the most suitable templates. The structures of DNA polymerase III alpha subunit from *Thermus aquaticus* (PDB entry 2HPI_A) and *Escherichia coli* (PDB entry 2HNNH_A) were taken from Protein Database (PDB) [15]. Multiple sequence alignment of the target and templates was performed using ClustalW. The final structural models were built by using the modv2 version of the MODELLER program [16]. The generated 3-D structural models were visualized by Swiss PDB Viewer [17]. The models were evaluated at various structure verification servers as PROCHECK and WHAT_CHECK. The best model was chosen on the basis of its stereo chemical quality account using PROCHECK [18] and WHAT_CHECK [19]. The functional domains of MtbDnaE1 were obtained by submitting amino acid sequence to Pfam and InterPro servers [20, 21].

Docking of the inhibitor 251D onto the active site of MtbDnaE1:

The ligand molecule 251D is the hybrid compound of anilinoauracil, 3-(4-hydroxybutyl)-6-(3-ethyl-4-methylanilino) uracil (HB-EMAU), and fluoroquinolone, {1-cyclopropyl-6,8-difluoro-7-[1-(3-methyl)piperazinyl]-4-oxo-1,4-dihydroquinoline}-3-carboxylic acid (340E) [22]. The ligand molecule was retrieved from CORINA [23]. The Modeller-generated three dimensional structure of MtbDnaE1 was taken as receptor molecule. The AutoDock 4.0 suite was used as molecular-docking tool [24]. The Graphical User Interface program "Auto-Dock Tools" was used to prepare, run, and analyze the docking simulations. Kollman united atom charges, salvation parameters and polar

hydrogens were added into the receptor PDB file for the preparation of protein in docking simulation. Gasteiger charge was assigned to the ligand molecule which is a non-peptide structure, and subsequently non-polar hydrogens were merged. Grid points spacing in the grid box (x, y, and z:78, 78, and 92 Å) was kept at 0.375 Å. AutoGrid 4.0 Program, supplied with AutoDock 4.0 (compiled and run under Linux operating system) was used to produce grid maps. The best conformers, out of the total 100 studied during the docking process, were selected on the basis of Lamarckian Genetic Algorithm (LGA). The individuals were selected randomly with the population size of 150. Maximum number of energy evaluation was set to 250,000,00, maximum number of generations 270,000, maximum number of top individual that automatically survived set to 1, mutation rate of 0.02, crossover rate of 0.8. Step sizes were 0.2 Å for translations, 5.0° for quaternions and 5.0° for torsions. For docking simulation, cluster tolerance was set at 0.5 Å and external grid energy was kept at 1000 kcal with maximum initial energy 0.0 kcal. A total of 10 LGA runs were performed with the maximum number of retries taken as 10000. All the AutoDock docking runs were performed in Intel Xeon CPU @ 3.2 GHz of HP origin, with 2 GB DDR RAM. The molecular interactions between the ligand (251D) and the protein (MtbDnaE1) were analyzed using "Auto-Dock Tools" (Version 1.50).

Discussion:

Phylogenetic analysis:

The phylogenetic analysis of the MtbDnaE1 amino acid sequence showed the MtbDnaE1 to be closest to that of *M. leprae* (Supplementary Figure 1). However, it appears to have branched away much earlier from that of *M. paratuberculosis*. Surprisingly, it showed greater closeness to its counterparts in the bacteria of other genus such as *S. coelicolor*, *A. aeolicus* and *T. aquaticus*.

Structural model and overall architecture:

As the crystal structure of TaqPolIII α (2HPI_A) is a full-length structure and the EcoPolIII α crystal structure (2HNNH_A) is only a large fragment structure, available PDB structure of TaqPolIII α (2HPI_A) was used as a template to generate the homology model of MtbDnaE1. The structural model of MtbDnaE1 reveals that it is organized into an irregular pyramid around a central cavity (Figure 1A & 1B). The quality assessment of the predicted model by the PROCHECK program that uses the Ramachandran plot [25], shows that the modeled MtbDnaE1 has 89.9% residues in most favorable regions, 9.5% residues occurring in allowed regions and 0.6% residues were found in the disallowed regions. These values are comparable with the stereochemical data (87.5%, 12.2 and 0.3%, respectively) of the X-ray - resolved structure of the TaqPolIII α . All main chain and side chain parameters were found to be in the 'better' region. The observed G-factor score (-0.10) of the present model was much above the G-factor score (-0.50) of a reliable model. Planar groups in the modeled structure were 100% within limits. The structure verification server such as WHAT_CHECK also validated the modeled MtbDnaE1 structure. Thus, the modeled structure of MtbDnaE1 is comparable to the structurally resolved polIII α subunit of *T. aquaticus*.

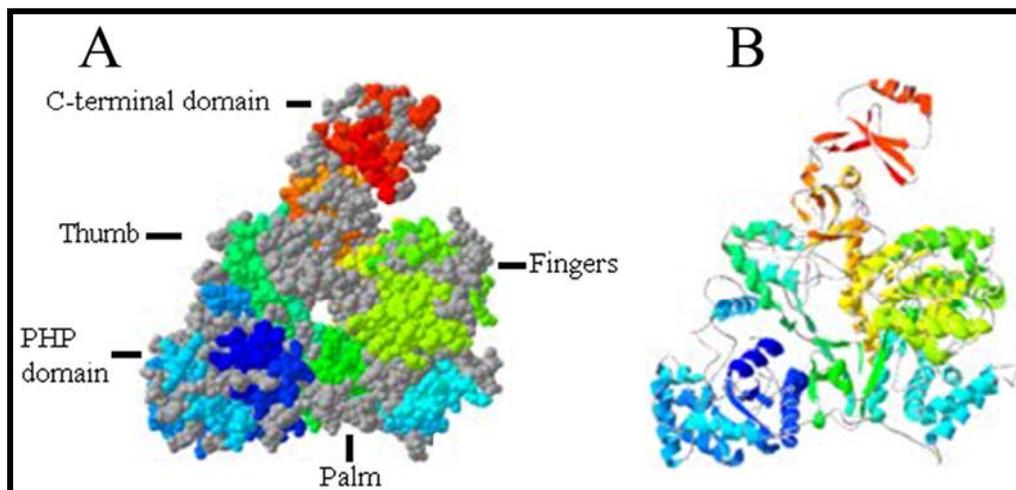


Figure 1: (A) The homology-modeled structure (spacefill model) of MtbDnaE1 with the predicted cavity and domains, using TaqPolIII α (PDB entry 2HPI_A) as a template. (B) The ribbon structure of the predicted model as visualized by Swiss PDB Viewer. All the structures are shown by secondary structure succession color scheme.

Conserved domains:

The functional domains analysis of the modeled MtbDnaE1 revealed that the MtbDnaE1 contains three conserved structural domains. The N-terminal is formed by the PHP (polymerase and histidinol phosphatases) domain [26], a putative phosphodiesterase domain (aa 11-231; TaqPolIII α 1-285). The C-terminus of MtbDnaE1 contains OB-fold (oligonucleotide/oligosaccharide-binding fold) domain that binds to nucleic acids. This domain is similar to the c-terminal domain, or CTD of TaqPolIII α (aa 1013-1220). In between the PHP domain and OB-fold domain, MtbDnaE1 contains a highly conserved region of 499 residues (aa 263-762). In TaqPolIII α , this region contains three domains- the palm (aa 286-492 and 575-622), the thumb (aa 493-574) and the fingers (aa 623-835). In MtbDnaE1, the region of 260 residues (aa 763-1023) shows no similarity with any structural domain in Pfam. However, ~45% residues in this region are identical to the β binding domain of TaqPolIII α (aa 836-1012) that contains the internal binding site for the β -clamp [27]. It is possible that the region from residues 763-1023 in MtbDnaE1 also contains a α -binding domain.

Active site residues:

In all PolIII α that have been studied, catalysis is mediated by two divalent metal ions that are anchored by three crucial aspartate or glutamate acidic residues. Two of the acidic residues must be adjacent or separated by a single amino acid [28]. Based on these motifs, the MtbDnaE1 sequence was

compared to sequences of other PolIII α sequences to identify candidates for key active-site residues. The presence of conserved active site comprising P-D-X-D-X-D could be detected at amino acids 420-425 in the MtbDnaE1 (data not shown). The three catalytic aspartates of MtbDnaE1 (D421, D423, and D587) align with the three absolutely conserved aspartate residues of TaqPolIII α (D463, D465, and D618). Thus, like all other known polymerases, the MtbDnaE1 is likely to utilize the same two-metal-ion catalytic mechanism [29]. It is likely that two of the aspartates are involved in the coordination of the two Mg²⁺ ions that are critical to the phosphotransferase activity. The third aspartate acts as a general base to activate the primer strand for nucleophilic attack on the α phosphate group of the incoming nucleotide [30]. Three other conserved residues in the MtbDnaE1 (G383, S384, and K585) correspond to the G425, S426, and K616 of the TaqPolIII α . The glycine and serine lie in a loop which forms part of the incoming nucleotide binding pocket and are conserved across all the 50 PolIII α (data not shown). The lysine forms a salt bridge with the phosphate group of the terminal 3' base of the primer [31] and is absolutely conserved as a positive residue in family C polymerases. The MtbDnaE1 consists of a cluster R410, R416, R736, K737 corresponding to the cluster of four highly conserved arginine residues (R452 and R458 from the palm and R766 and R767 from the fingers domain) in TaqPolIII α that interact with the dATP (triphosphate) and lies approximately 10Å away from the catalytic aspartates on the palm domain [14, 15], with an exception of the fourth arginine having been replaced by lysine (Figure 2A).

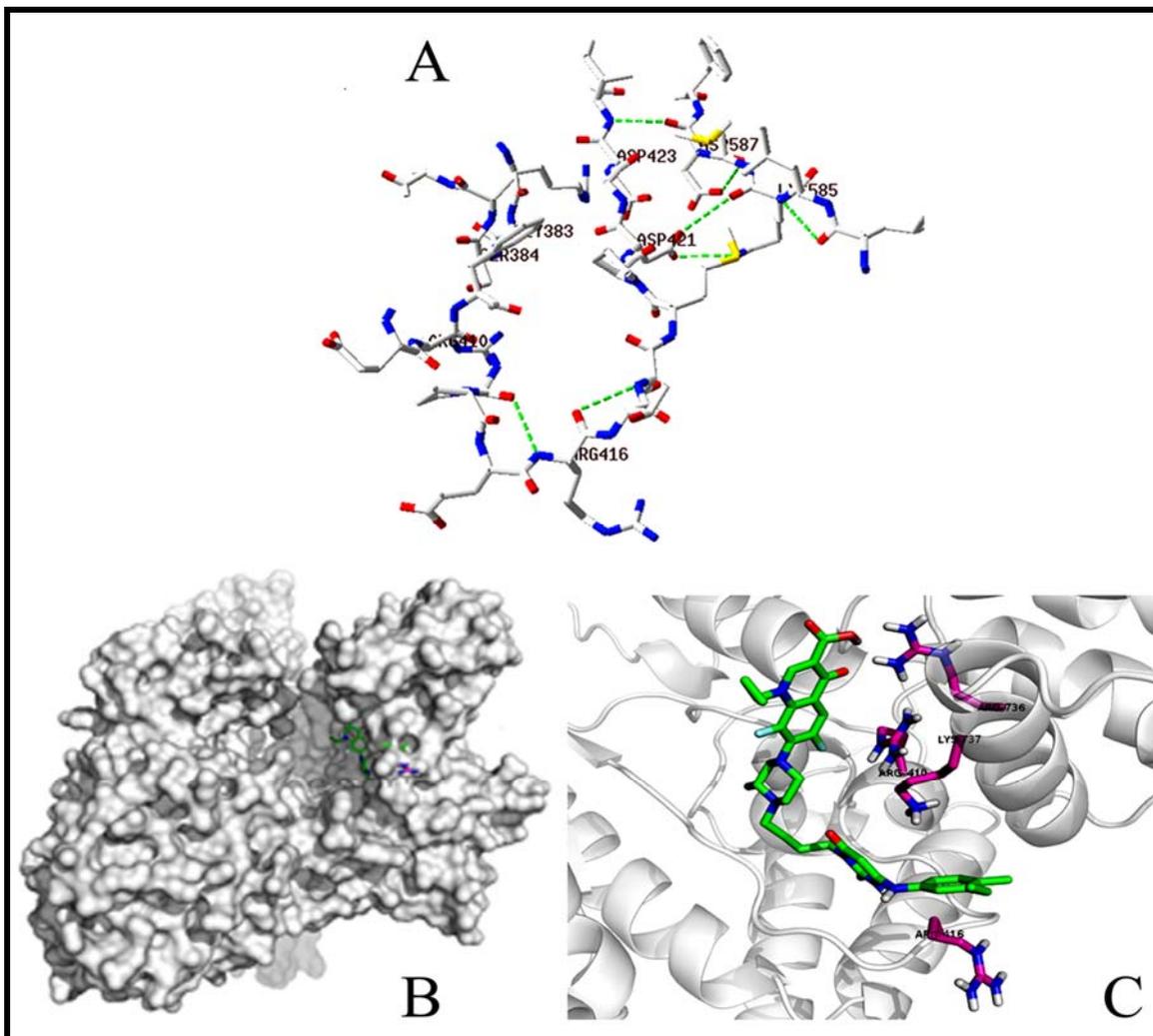


Figure 2: (A) Active site cleft of homology-modeled MtbDnaE1 showing conserved residues involved in catalytic function in stick form and colored white (carbon), blue (nitrogen), and red (oxygen). Hydrogen bonds are depicted as green lines (B) Lowest energy structure for MtbDnaE1 docked with 251D. MtbDnaE1 protein is displayed as cyan colored surface, the docked ligand is displayed as a stick model. (C) Detailed view of 251D docked into the MtbDnaE1. The ligand is shown as a green stick model. The protein backbone is displayed as cyan colored ribbon diagram, while active site residues (Arg410, Arg416, Arg736, and Lys737) are listed and shown as stick diagram.

Docking of MtbDnaE1 with 251D:

Compound 251D has been shown as a potent inhibitor of bacterial replicative DNA polymerase [22]. Further, 251D has been reported to show low in vitro cytotoxicity [22]. Therefore, molecular docking of the homology modeled MtbDnaE1 was performed with 251D (Figure 2B). In this study, semi-flexible docking was performed, where the target protein MtbDnaE1 was kept as non-flexible while the ligand (251D) was kept flexible to facilitate random degrees of freedom (including torsional and spatial degrees of freedom) spanned by the translational and rotational parameters. All the amide bonds were set rigid while all other bonds in the ligands were allowed to rotate freely. As required by the AutoDock, pre-calculated grid maps were assigned to each atom type, present in the ligand. AutoGrid, a part of ADT calculates the energy of non-covalent contacts between the protein and ligand at different grid points. In the present study, the area of interest was selected on the basis of amino acid residues, which are implicated in binding with the incoming nucleotides. Thermodynamic properties like free energy of bindings (ΔG_b) and inhibition constant (K_i) were generated by docking of the 251D onto the MtbDnaE1. The estimated free energy of bindings was -16.04 Kcal/mol and estimated inhibition constant was 1.74pM. The 251D interacts with the active site residues (R410, R416, R736, and K737) of the MtbDnaE1 via hydrophilic-hydrophilic contact (hydrogen bond), hydrophobic-hydrophobic contact and hydrophobic-hydrophilic contacts (Figure 2C and Table 1 see Supplementary material). The present study thus predicts the possible interaction of the inhibitor 251D with the MtbDnaE1, its position with respect to the active site and binding energies to understand the nature of binding.

Conclusion:

The crystal structure of full-length *Thermus aquaticus* PolIII α represents the first crystal structure of a cellular replicative polymerase which has approximately 43% sequence identity with the MtbDnaE1. This high degree of sequence identity led us to design the structural model of MtbDnaE1. In fact, a protein sequence with over 30% sequence identity to a known crystal structure can often be modeled with an accuracy equivalent to a low resolution X-ray structure [32]. The structural model of MtbDnaE1 generated on the template TaqPolIII α in this study represents the first structure model of MtbDnaE1 in the absence of its crystal structure. An evaluation of the stereochemical quality of the modeled structure by using PROCHECK and WHAT-CHECK programs shows the reliability of the modeled protein. The presence of conserved catalytically active site residues D421, D423, and D587, provides insight into the catalytic mechanism of this enzyme and indicates that it may be similar to that of *Thermus aquaticus* and *E. coli*. Docking with the known inhibitor 251D demonstrates the role played by the conserved active site residues in the MtbDnaE1. As the catalytic domains of the replicative polymerases in humans and *Mycobacterium* are not homologous, the active site residues of the MtbDnaE1 can be targeted to develop novel drugs to specifically block the Mycobacterial polymerase and thus inhibit tuberculosis progression.

Acknowledgements:

Financial support from the Department of Biotechnology, New Delhi is acknowledged. We extend our thanks to Mr Sandeep Kaushik and Dr. Debasisa Mohanty for providing help in docking studies.

References:

- [1] http://whqlibdoc.who.int/hq/2001/WHO_CDS_TB_2001.287.pdf
- [2] Bloom BR & Murray CJ. *Science* 1992 **257**: 1055 [PMID: 1509256]
- [3] Cole ST *et al. Nature* 1998 **393**: 537 [PMID: 9634230]
- [4] Anishetty S *et al. Comput Biol Chem.* 2005 **29**: 368 [PMID: 16213791]
- [5] Kelman Z & O'Donnell M. *Annu Rev Biochem.* 1995 **64**: 171 [PMID: 7574479]
- [6] Onrust R *et al. J Biol Chem.* 1995 **270**: 13366 [PMID: 7768938]
- [7] Kong XP *et al. Cell* 1992 **69**: 425 [PMID: 1349852]
- [8] Filée J *et al. J Mol Evol.* 2002 **54**: 763 [PMID: 12029358]
- [9] Bailey S *et al. Cell* 2006 **126**: 893 [PMID: 16959569]
- [10] Lamers MH *et al. Cell* 2006 **126**: 881 [PMID: 16959568]
- [11] <http://www.doe-mbi.ucla.edu/TB/>
- [12] Landau M *et al. Nucleic Acids Res.* 2005 **33**: W299 [PMID: 15980475]
- [13] Thompson JD *et al. Nucleic Acids Res.* 1994 **22**: 4673 [PMID: 7984417]
- [14] Brodsky LI *et al. Biochemistry* 1995 **60**: 923
- [15] <http://www.rcsb.org/pdb/home/home.do>
- [16] Fiser A & Sali A. *Methods Enzymol.* 2003 **374**: 461 [PMID: 14696385]
- [17] <http://www.expasy.org/spdbv>
- [18] Laskowski RA *et al. J Appl Cryst.* 1993 **26**: 283
- [19] Vriend G. *J Mol Graph.* 1990 **8**: 52 [PMID: 2268628]
- [20] Bateman A *et al. Nucleic Acids Res.* 2004 **32**: 138 [PMID: 14681378]
- [21] www.sanger.ac.uk/Software/Pfam
- [22] Butler MM *et al. Antimicrob Agents Chemother.* 2007 **51**: 119 [PMID: 17074800]
- [23] Gasteiger J *et al. Tetrahedron Comput Method.* 1990 **3**: 537
- [24] Morris GM *et al. J Comput Chem.* 1998 **19**: 1639
- [25] Ramachandran GN & Sasisekharan V. *Adv Protein Chem.* 1968 **23**: 283 [PMID: 4882249]
- [26] Aravind L & Koonin EV. *Nucleic Acids Res.* 1998 **26**: 3746 [PMID: 9685491]
- [27] Dohrmann PR & McHenry CS. *J Mol Biol.* 2005 **350**: 228 [PMID: 15923012]
- [28] Joyce CM & Steitz TA. *Annu Rev Biochem.* 1994 **63**: 777 [PMID: 7526780]
- [29] Steitz TA. *Nature* 1998 **391**: 231 [PMID: 9440683]
- [30] Brautigam CA & Steitz TA. *Curr Opin Struct Biol.* 1998 **8**: 54 [PMID: 9519297]
- [31] Sawaya MR *et al. Science* 1994 **264**: 1930 [PMID: 7516581]
- [32] Xiang Z. *Curr Protein Pept Sci.* 2006 **7**: 217 [PMID: 16787261]

Edited by P Kanguane

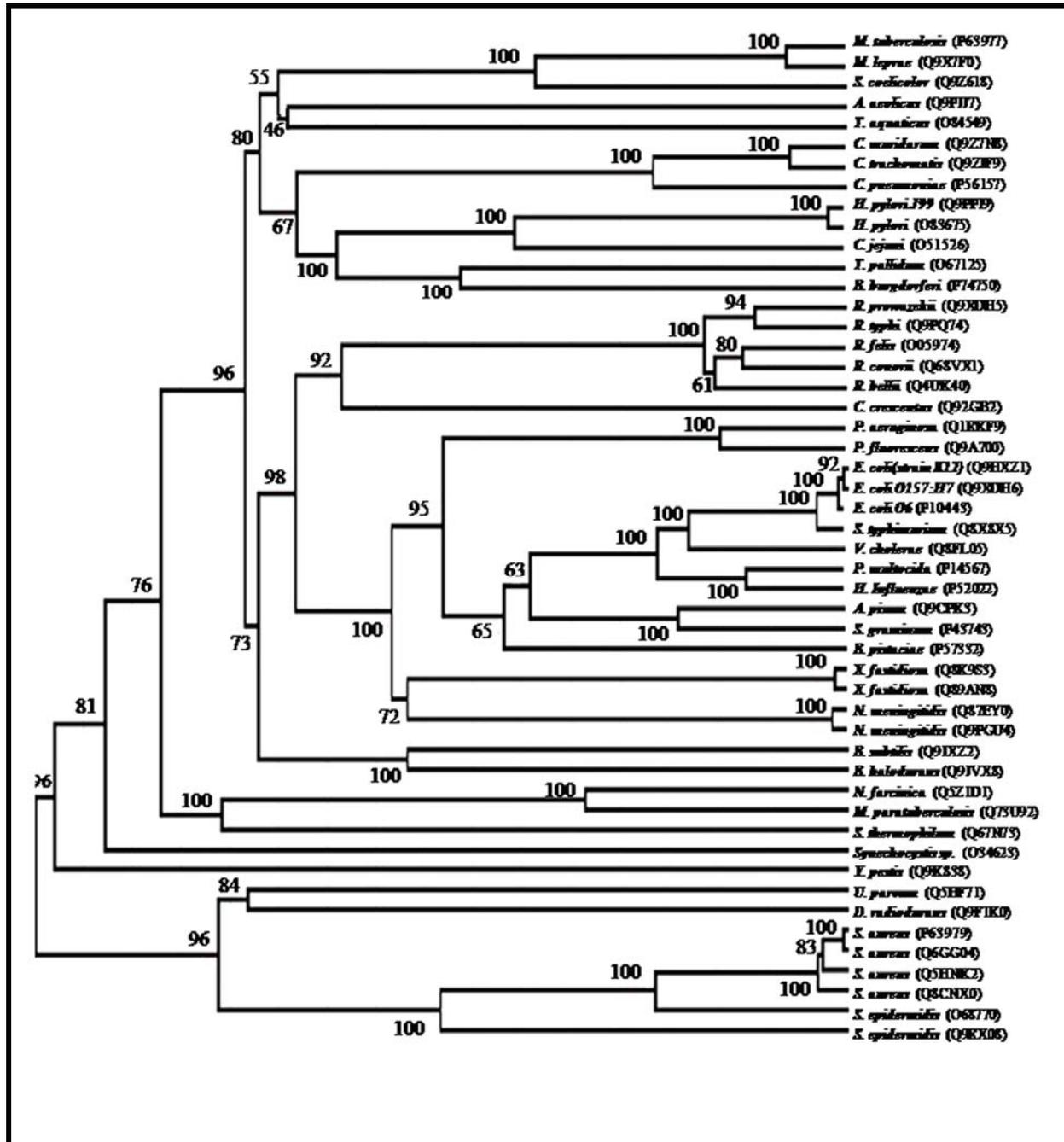
Citation: Chhabra *et al.* Bioinformation 6(2): 69-73 (2011)
provided the original author and source are credited.

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes,

Supplementary material:

Table 1: Hydrogen bond interactions between 251D and MtbDnaE1 with their bond distances

251D	MtbDnaE1	Distance (Å)
O43	R736 NH1	3.0
O43	R736 NH2	2.86



Supplementary Figure 1: TreeTop - Phylogenetic Tree for multiple sequence alignment of MtbDnaE1 in Phylip format by cluster algorithms. The numbers on the nodes are bootstrap values in percentages. The Swiss-Prot ID numbers of the PolIIIα from different bacteria are given in brackets.