Structural insights into the interaction of helicase and primase in *Mycobacterium tuberculosis*

Dhakaram Pangeni Sharma¹, Ramachandran Vijayan¹, Syed Arif Abdul Rehman², Samudrala Gourinath¹*¹

¹School of Life Sciences, Jawaharlal Nehru University, Delhi-110067, India.
²MRC, Protein Phosphorylation & Ubiquitylation Unit, University of Dundee, Dundee, UK

*To whom correspondence should be addressed: Dr. Samudrala Gourinath, School of Life Sciences, Jawaharlal Nehru University, Delhi-110067, India.
Email: sgourinath@mail.jnu.ac.in, Tel. No: +91-11-26704513; Fax: +91 -11-26742916/2558

**Keywords:** DnaG primase, DnaB helicase, *Mycobacterium tuberculosis*, X-ray Crystallography, Molecular Dynamics, Protein-Protein Interactions.

**Abstract:**

The helicase-primase interaction is an essential event in DNA replication and is mediated by the highly variable C-terminal domain of primase (DnaG) and N-terminal domain of helicase (DnaB). To understand the functional conservation despite the low sequence homology of the DnaB-binding domains of DnaGs of eubacteria, we determined the crystal structure of the helicase-binding domain of DnaG from *M. tuberculosis* (*Mt*DnaG-CTD), and did so to a resolution of 1.58 Å. We observed the overall structure of *Mt*DnaG-CTD to consist of two subdomains, the N-terminal globular region (GR) and the C-terminal helical hairpin region (HHR), connected by a small loop. Despite differences in some of its helices, the globular region was found to have broadly similar arrangements across the species, whereas the helical hairpins showed different orientations. To gain insight into the crucial helicase-primase interaction in *M. tuberculosis*, a complex was modeled using the *Mt*DnaG-CTD and *Mt*DnaB-NTD crystal structures. Two nonconserved hydrophobic residues (Ile605 and Phe615) of *Mt*DnaG were identified as potential key residues interacting with *Mt*DnaB. Biosensor binding studies showed a significant decrease in the binding affinity of *Mt*DnaB-NTD with the Ile605Ala mutant of *Mt*DnaG-CTD compared to native *Mt*DnaG-CTD. The loop, connecting the two helices of the HHR, was concluded to be largely responsible for the stability of the DnaB-DnaG complex. Also, *Mt*DnaB-NTD showed micromolar affinity with DnaG-CTDs from *E. coli* and *H. pylori* and unstable binding with DnaG-CTD from *V.*
The interacting domains of both DnaG and DnaB demonstrate the species-specific evolution of the replication initiation system.

**Introduction:**

DNA replication is a fundamental process in all domains of life and requires a large molecular machine coordinating the action of dozens of discrete factors to ensure accurate genome inheritance (1). The interaction between the bacterial replicative ring helicase (DnaB) and the primase (DnaG) is essential for the formation of the functional primosome at oriC to initiate the DNA replication (2, 3). DnaB helicase unwinds duplex DNA into single-stranded DNA (ssDNA), with this unwinding fueled by the hydrolysis of nucleoside triphosphate at the replication fork (4-6); and DnaG primase uses the newly formed ssDNA as a template for the de novo synthesis of RNA primers to initiate the DNA replication (7). The interaction between DnaB and DnaG mutually stimulates both of their activities (8). DnaG increases both the NTPase and helicase activities of DnaB (9) while DnaB increases and modulates the synthesis of RNA primers by DnaG (7, 10, 11). The DnaG primase consists of three domains with distinct activities; an N-terminal zinc-binding domain, a central RNA polymerase domain and a C-terminal helicase-binding domain. The C-terminal domain of DnaG (DnaG-CTD) interacts with the N-terminal domain of DnaB and stimulates the DnaB helicase activity (12). The stability of the interaction between DnaG primase and DnaB helicase varies among eubacteria. The helicase-primase interaction has been reported to be weak in *Escherichia coli* (13), moderate in *Helicobacter pylori* (14, 15) and strong in *Bacillus stearothermophilus* (9). Despite its relatively low resolution, the only available structure of the complex of DnaG-CTD with DnaB helicase in *Bacillus stearothermophilus* did provide insight into the helicase-primase interaction pattern (16). Since the helicase-binding domain of primases is poorly conserved, we expect the structures from different organisms to provide insight into the mode of interaction between DnaG and DnaB.

DNA replication in *Mycobacterium tuberculosis* (*Mt*) is not well understood. This intracellular pathogen is the main cause of human tuberculosis (17), and according to the recent World Health Organisation report (WHO report 2017), tuberculosis is the leading cause of death from a single infectious agent. This pathogen can, under favorable conditions, hide in a dormant state in the host for long periods of time without producing conspicuous disease symptoms (18). Specifically, *M. tuberculosis* progresses to a state of non-replicating persistence to lie dormant in the host for such long periods of time (19). Mutations generated
by error-prone DNA replication also enhance *M. tuberculosis* virulence and lead to antibiotic resistance (20). Part of our lack of understanding of DNA replication in *M. tuberculosis* arises from the homologs of some essential proteins for efficient DNA replication in *E. coli* having not been identified in mycobacteria. In addition, the regulators of DnaA activity (Hda in *E. coli*, YabA in *B. subtilis*), replicative helicase loader protein (DnaC in *E. coli*, DnaI in *B. subtilis*) and replication terminator proteins (Tus in *E. coli* and RTP in *B. subtilis*), have not been characterized in *M. tuberculosis* (21-24). Recently, Rv0004 was found to have a role in *M. tuberculosis* DNA replication; it interacts with DNA and the DnaB (25). The mechanism by which the replisome of *M. tuberculosis* is formed from the DnaG-DnaB complex and various accessory proteins is still not clear. *M. tuberculosis* DnaB (*Mt*DnaB) has recently been functionally characterized (26). The *Mt*dnaB gene encodes 874 amino acid residues, 416 of which (from Leu400 to Asn815) form an intein (27). An intein is an internal segment of amino acids that are excised from a protein precursor to generate a functional protein that splits off from the main protein during posttranslational maturation (27). The lack of structural data for the DnaB-DnaG complex and the incomplete structural information for the individual proteins are obstructing our efforts to understand the molecular details that reinforce this essential interaction in *M. tuberculosis*. However, the crystal structure of the NTD of *Mt*DnaB has been reported (28). To further study the mode of interaction between *Mt*DnaB and *Mt*DnaG, we determined the structure of the helicase-binding domain of *Mt*DnaG.

The *Mt*DnaG-CTD structure was observed to differ from the other available DnaG-CTD structures. We used crystal structures of *Mt*DnaB-NTD reported earlier (28) and the currently determined structure of *Mt*DnaG-CTD to derive a model of helicase-primase interactions in *M. tuberculosis*, and provided further evidence for this model by performing in vitro experiments. We identified single amino acid residues at the helicase-binding domain of *Mt*DnaG that, when mutated to alanine, resulted in proteins with altered interactions with DnaB helicase. Further analysis revealed the possibility of cross-species interaction of *Mt*DnaB with other DnaG in eubacteria. Overall, these results substantially helped us to understand the mechanism of helicase-primase interactions specifically in *M. tuberculosis* and expand our knowledge on the diversity of bacterial DNA replication strategies, some of which may be organism specific.

**Materials and methods:**
Cloning of expression constructs for heterologous expression:

*Mycobacterium tuberculosis* (H37rv) genomic DNA and specific primers (Table S1) were used for the polymerase chain reaction (PCR) to amplify the coding regions of *Mt*DNA-G-CTD (477 base pairs) and *Mt*DNA-B-NTD (591 base pairs). The amplified products of *Mt*DNA-G-CTD and *Mt*DNA-B-NTD were cloned into the Nhe1/Xho1 site of the expression vector pET21c (Novagen, Madison, WI, USA) and the Nde1/Xho1 site of the expression vector pET28b (Novagen, Madison, WI, USA), respectively, with each containing the C-terminal His6 tag. Similarly, dnaG-CTD from *Vibrio cholerae* (O395 strain) was cloned into the Nco1/Xho1 site of pET28b (Rehman thesis). DNA sequencing confirmed the sequences of all constructs. The cloning of *Hp*DNA-G-CTD and *Ec*DNA-G-CTD was performed previously in our laboratory (15, 29).

Site-directed mutagenesis

*Mt*DnaG-CTD mutants (*Mt*DnaG-CTD I605A and *Mt*DnaG-CTD F615A) were prepared by following a PCR-based method using the pET21c expression vector encoding for the wild-type (*Mt*DnaG-CTD) protein as a template. The amplification of the expression vector was performed using the KOD hot-start DNA polymerase (EMD Millipore), following the manufacturer guidelines, and specific pairs of primers (Table S1) were used for the introduction of the desired mutation. PCR products were treated with DpnI enzyme (New England Biolabs) and transformed into DH5α *E. coli* cells. The mutations were confirmed by performing DNA sequencing.

Overexpression and purification of recombinant proteins:

pET21c-*Mt*DnaG-CTD and point mutant (*Mt*DnaG-CTD I605A and *Mt*DnaG-CTD F615A) constructs were transformed into *E. coli* BL21 (DE3). For expression, 1% of the overnight-grown culture of a single colony was added to two liters of Luria-Bertani (LB) broth containing 100 μg/ml ampicillin, 1 mM glucose, and 1 mM MgCl₂. Cells were grown to an optical density of 0.6 (in 600 nm wavelength) at 37°C. Overexpression of protein was induced overnight at 16°C with 0.5 mM isopropyl-beta-D thiogalactopyranoside (IPTG). The cells were harvested at 8,000 rpm for 5 min and mixed with lysis buffer A (30 mM Tris-HCl [pH 7.5], 150 mM NaCl, 5 mM imidazole, 5 mM MgCl₂, 5 mM beta-mercaptoethanol [betaME], 100 μM phenylmethylsulfonyl fluoride [PMSF]). After the addition of lysozyme (0.1 mg/ml), the cell suspension was incubated at 4°C for 30 min. The cell suspension was sonicated (Branson Ultrasonic Systems) in an ice-water mixture at 25% of the amplitude with...
a pulse of 20 seconds each interspersed with an interval of 30 seconds, and cell lysate was

treated with 0.1% Triton X-100, followed by incubation for 30 minutes on a rotating rocker at

4°C. The cell lysate was cleared by being subjected to centrifugation at 13,000 rpm for 45

minutes at 4°C and then purified using Ni-NTA Sepharose resin (GE Healthcare, Sweden)

pre-equilibrated in buffer A. Proteins were eluted with buffer B (30 mM Tris-Cl [pH 7.5],

150 mM NaCl, 250 mM imidazole, 5 mM βME and 100 μM PMSF). The protein was then

concentrated using AmiconUltra-4 centrifugal filter ultracel-10K (Millipore, Billerica,

MA, USA) and purified further by using a gel filtration chromatography Hi-Load Superdex-

75 16/60 column (GE Healthcare) pre-equilibrated with buffer C (30 mM Tris-Cl [pH 7.5],

150 mM NaCl, 5% glycerol and 5 mM βME). Peak fractions were checked using SDS–12% PAGE and pooled together.

HpDnaG-CTD and EcDnaG-CTD were purified as described earlier (15, 29). MtDnaB-NTD

and VcDnaG-CTD constructs were transformed into E. coli BL21 (DE3) cells. Cells

containing plasmids were grown in LB media in the presence of either 50 μg/ml kanamycin

(MtDnaB-NTD) or 100 μg/ml ampicillin (VcDnaG-CTD) up to an OD₆₀₀ of 0.6-0.8 at 37°C.

Overexpression of protein was induced overnight at 20°C with 0.2 mM IPTG. Bacterial cell

pellets were re-suspended in lysis buffer A and cells were disrupted by subjecting them to

sonication. The lysate was cleared by subjecting them to centrifugation at 13,000 g for 45

minutes at 4°C; and the proteins here were purified by being using Ni-NTA Sepharose beads

pre-equilibrated in buffer A, and eluted with buffer B. Final purification was performed using

gel filtration on a Hi-Load Superdex-75 16/60 column pre-equilibrated with buffer G (20 mM

HEPES pH 7.4 and 150 mM NaCl). The homogeneity of protein was checked with (12%) SDS-PAGE.

Overexpression and purification of the selenomethionine-labeled MtDnaG-CTD:

Selenomethionine (Se-Met)-labeled MtDnaG-CTD was prepared to provide phases necessary

for structure determination. Se-Met-labeled protein was purified under reducing conditions to

prevent the formation of a mixture of reduced and oxidized Se-Met-labeled proteins. Protein

labeling was carried out using media supplied by Molecular Dimensions. The concentration

of selenomethionine was maintained at about 25 mg/liter. Initially, the primary culture of

MtDnaG-CTD-transformed E.coli BL21 (DE3) was grown in LB medium overnight. The

next morning, cells were harvested by centrifuging the overnight grown culture at 4,000 rpm

for 5 minutes. The pellet obtained was resuspended in a purely selenomethionine medium and
the process was repeated once again to remove any traces of LB medium. After inoculation of
the secondary culture in complete selenomethionine medium, the culture was allowed to
grow for 4 hours until the OD$_{600}$ reached 0.5 at 37°C. After this procedure, the cells were
induced with 1 mM IPTG, and then shaken for another 8 hours at 30°C, followed by being
harvested at 8,000 rpm for 3 minutes and stored at −80°C for downstream processing. The
remaining steps of the purification were similar to those followed for the native MtDnaG-
CTD, as described above.

Crystallization of MtDnaG-CTD:
Native MtDnaG-CTD (6 mg/mL) and selenomethionine-labeled MtDnaG-CTD (5 mg/mL)
were subjected to extensive crystallization trials using the hanging drop method. Initially, the
automated machine MOSQUITO™ (AIRF, JNU) was used to set hanging crystallization
drops containing 500 nL protein solution plus 500 nL reservoir solution in 96-well Greiner-
CELLSTAR plates for screening with Hampton and Molecular Dimensions crystallization
screens. Small protein needles were grown using the reservoir solution of Hampton Peg Ion
screen 25 (20% w/v PEG 3350, 0.2 M magnesium acetate tetrahydrate). The conditions were
replicated and improved in 24-well Linbro plates by mixing 2 μl protein solution with 2 μl
reservoir solution and equilibrating the resulting mixture against a 500 ml reservoir solution
at 4°C. The best selenomethionine-labeled protein crystals were grown with a reservoir
solution of 18% w/v PEG 3350, 0.15 M magnesium acetate and MOPS buffer pH 7.6.
Crystals appeared after one week and grew to full size within two weeks.

Data collection, processing and structure determination:
The crystals were flash frozen in liquid nitrogen with a cryoprotectant solution containing
20% PEG 400 mixed with mother liquor (18% w/v PEG 3350, 0.15 M magnesium acetate
and MOPS buffer with pH 7.6). The X-ray data for Se-Met-labeled crystals were collected at
the BM14 synchrotron beamline (ESRF, Grenoble, France) at a selenium peak wavelength of
0.9786 Å. The data sets were indexed, integrated and scaled using HKL-2000 data processing
software (30). The input diffraction data were prepared using programs of the CCP4 suite
(31, 32). A partial structure was solved using the single wavelength anomalous scattering
protocol of Auto-Rickshaw of the EMBL-Hamburg automated crystal structure determination
platform ((33). Anomalous data were used to calculate FA values using the program
SHELXC (34). Based on an initial analysis of the data, the maximum resolution for
substructure determination and initial phase calculation was set to 2.0 Å. All of the three
heavy atoms expected (three selenomethionine atoms for each molecule of the asymmetric unit) were found using the program SHELXD (35). The correct hand for the substructure was determined using the programs ABS (36) and SHELXE (37). Initial phases were calculated after density modification using SHELXE. The initial phases were improved by carrying out density modification and phase extension to a resolution of 1.57 Å using the program RESOLVE (38). The initial model was built using the auto-build program ARP/wARP (39). The regions missing from this model were built manually using the COOT graphics package (40), and refinement of the resulting model was carried out with REFMAC5 (41). The structure was further improved after the iterative model building and refinement using COOT (42) and REFMAC5 (43), respectively. The final crystal structure was well refined with excellent electron density and validated by using PROCHECK (44) of the CCP4 suite. The figures presented here were generated using PyMOL (45). The data collection and refinement statistics are given in Table 1. Coordinate and structure factor files for the crystals of MtDnaG-CTD have been deposited with the Protein Data Bank under accession code PDB ID 5Z51.

Construction of the helicase-primase complex model in M. tuberculosis:

The structure of MtDnaB-NTD reported earlier by the Tsodikov group (28) and our current structure of MtDnaG-CTD, as well as the structure of the BstDnaB-BstDnaG-CTD (helicase-primase) complex (16) as a template, were used to model the MtDnaB-NTD/MtDnaG-CTD (helicase-primase) complex of M. tuberculosis. The structure of MtDnaB-NTD was superimposed on that of BstDnaB-NTD, and the globular region (GR) of MtDnaG-CTD was superimposed on the GR of BstDnaG-CTD of the complex. To superimpose the helical hairpin region (HHR) of MtDnaG-CTD with the HHR of the BstDnaB-BstDnaG-CTD complex, the HHR of MtDnaG-CTD was rotated by about 32°, keeping the linker between the HHR and the GR as the center of rotation, to obtain the lowest root mean square deviation (RMSD) between the BstDnaG-CTD and MtDnaG-CTD structures. This modification maximized the interactions between primase and helicase. The structure of the modeled complex was relaxed, by carrying out molecular energy minimization using the GROMACS molecular dynamics package (46), in order to eliminate bad contacts. Similarly, we constructed models of three complexes of MtDnaB-NTD, one with EcDnaG-CTD, and the other two with VcDnaG-CTD and HpDnaG-CTD, respectively, to study the possible interactions and stability levels of the heterocomplex.
Molecular dynamics simulation for binding affinity estimation:

Molecular dynamics simulations were performed on the models of the \textit{MtDnaB-NTD} complex with DnaG-CTD, DnaG-CTD I605A, and DnaG-CTD F615A, using the GROMACS 5.1.4 package, with the GROMOS 96 force field. To create the \textit{in silico} mutants, isoleucine 605 and phenylalanine 615 were each changed to alanine using PyMOL to make the two different DnaG-CTD mutants. These proteins were each solvated in a dodecahedron solvation box by using a simple point charge (SPC) of water molecules. \textit{Na}^{+} and \textit{Cl}^{-} ions were included to neutralize the system. The energy was minimized using 500 steps of steepest descent, followed by 1,000 steps of conjugate gradient. For maintaining the system in a stable environment (300 K, 1 bar), Berendsen temperature coupling and Parrinello-Rahman pressure coupling were employed and were set to 0.1 and 2.0 ps for temperature and pressure, respectively. A particle mesh Ewald (PME) algorithm was run to measure the electrostatic and Van der Waals interactions; here, the cut-off distance for short-range VdW (rvdw) interactions was set to 1.4 nm, and Coulomb cut-off (r coulomb) and a neighbor list (rlist) was set to 0.9 nm. The LINCS algorithm was used to measure all of the constrained bond lengths and the time step was set to 0.002 ps. The complexes in a medium were equilibrated for 100 ps in NPT and NVT ensembles. Finally, a 25 ns molecular dynamics simulation was carried out for each tested structure. All trajectories were stored every two ps for further analysis. Similar molecular dynamics simulation steps, as described above, were used on the three models of the complexes of \textit{MtDnaB-NTD}, i.e., with \textit{EcDnaG-CTD}, \textit{HpDnaG-CTD}, and \textit{VcDnaG-CTD}. The binding free energies were calculated by using the molecular mechanics Poisson-Boltzmann surface area (MM/GBSA) method as implemented in GROMACS tool. In our study, the free binding energy (\(\Delta G\)) was calculated by using van der Waals (\(\Delta E_{\text{vdW}}\)), electrostatic (\(\Delta E_{\text{elec}}\)), polar solvation (\(\Delta G_{\text{polar}}\)) and nonpolar (\(\Delta G_{\text{non-polar}}\)) energy contributions.

Biosensor binding studies:

To determine the difference between the binding affinity of \textit{MtDnaB-NTD} for \textit{MtDnaG-CTD} and those of \textit{MtDnaB-NTD} for the \textit{MtDnaG-CTD} mutants and other bacterial DnaG-CTDs, biosensor-based binding experiments were performed using an Autolab surface plasmon resonance (SPR) instrument at the Advanced Instrumentation Research Facility, Jawaharlal Nehru University, New Delhi, India. The surface of SPR chip (a self-assembled monolayer of 11-mercaptopoundecanoic acid [MUA] on a gold surface; Autolab) was first activated with carbodiimide (EDC; 0.2 M) and \textit{N}-hydroxysuccinimide (NHS; 0.05 M)/\textit{N}-ethyl-\textit{N}-(diethyl
aminopropyl). MtDnaG-CTD, MtDnaG-CTD mutants, and other bacterial DnaG-CTDs were immobilized (separately on different chips) onto the activated sensor surface at a concentration of 20 μg/ml in filtered (0.22-μm pore size) and degassed 10 mM sodium acetate buffer (pH 4.5). The chip had two channels: channel one was used for immobilization of ligand; and channel two was used as a blank (the signals of the analyte with a ligand-free surface). After ligand immobilization, the surface was blocked with 100 mM ethanolamine at pH 8.5, followed by regeneration using 1 M NaCl. The running buffer constituents were the same as those recommended for HBS BIAcore running buffer (10 mM HEPES [pH 7.4], 150 mM NaCl, 3 mM EDTA, 0.05% P-20 surfactant). The association kinetics for DnaG-CTDs was monitored for 400 seconds, followed by monitoring of the dissociation kinetics for the next 300 seconds. MtDnaB-NTD samples of various concentrations were prepared in running buffer and injected at the rate of 20 μl/min across the sensor surface. The concentrations of MtDnaB-NTD used against MtDnaG-CTD were 1500, 1000, 750, 500, 250 and 125 nM, whereas, for the MtDnaG-CTD mutants and other bacterial DnaG-CTDs, the concentrations used were 2000, 1500, 1,000, 750, 500 and 250 nM. Bovine serum albumin protein samples at various concentrations (2000, 1500, 1,000, 750, 500 and 250 nM) were used as negative controls. Signal changes on the activated/blocked control panel were subtracted from the DnaG-DnaB binding interactions using an inline reference signal, and the subtracted sensorgrams were analyzed. The surface was regenerated with a buffer consisting of two manually delivered pulses of 1 M NaCl. All of the data were recorded at 25°C. The data analysis was performed using Autolab SPR kinetic evaluation software.

Results:

The overall structure of MtDnaG-CTD:

To obtain a more detailed view of the interaction between the helicase and primase of Mycobacterium tuberculosis, we set out to determine a high-resolution crystal structure of the helicase-binding domain of MtDnaG (MtDnaG-CTD) and indeed did so, to a resolution of 1.58 Å (Fig. 2b). Molecular replacement failed to give any solution, indicative of the low homology and structural similarity with all of the previously determined DnaG-CTDs. Se-Met-labeled protein crystals were used to obtain anomalous data and experimental phases. The Se-Met-labeled protein crystal diffracted X-rays to 1.58 Å at BM14 of ESRF and belonged to the space group P2₁2₁2 with cell dimensions of a = 48.2 Å, b = 137.5 Å, c = 36.5 Å (α= β= γ = 90°). The final model yielded an R factor of 18.1% and R_free of 20.8%, with good
electron density (Fig. 2a). One complete MtDnaG-CTD (residues 481 to 637) and one helical hairpin cleaved from the MtDnaG-CTD (residues 569 to 637) were identified per asymmetric unit (Fig. 2b) and these two protein chains were stabilized with a few van der Waals interactions (Fig. 2d). The overall CTD structure of MtDnaG displayed two subdomains, the N-terminal globular region (GR) subdomain (residues 482 to 562) and C-terminal helical hairpin region (HHR) subdomain (residues 568 to 637), connected by a small loop (6 residues) (Fig. 2c). Inspection of the MtDnaG-CTD structure showed it to lack any beta sheet and to consist instead primarily of alpha helices, specifically ten of them with the first eight forming the GR and the last two forming the HHR (Fig. 2c). The gel filtration chromatography profile of MtDnaG-CTD in solution on a calibrated Superdex G75 16/60 column revealed a sharp elution peak at 70.3 ml, and an observed molecular weight of ~38 kDa, indicating MtDnaG-CTD (~19 kDa) to exist in a dimeric state in solution (Fig. 1). Electron density corresponding to the GR or N-terminal residues (residues 481 to 568) of MtDnaG-CTD was not visible in one monomer due to the cleavage of MtDnaG-CTD. MtDnaG-CTD cleaved into two fragments (GR and HHR), and the fragments were also identified using MALDI-TOF (Fig. S2). A total of 232 amino acid residues, one polyethylene glycol (PEG) molecule, one acetate molecule and 320 water molecules were determined to be present per asymmetric unit. A Ramachandran analysis showed 99.1% of the residues in the favored region, 0.5% in the generously allowed region, and 0.4% in the disallowed region.

The MtDnaG-CTD structure differs from other DnaG-CTD structures in the HHR:

The 1.58-Å resolution of the current M. tuberculosis DnaG-CTD structure was higher than those of the other four available DnaG-CTD crystal structures, i.e., those determined for E. coli (resolution to 2.8 Å, PDB ID 1T3W), B. stearothermophilus (2.9 Å, 2R6A), H. pylori (1.7 Å, 4EHS) and V. cholerae (2.4 Å, 4IM9) (15, 16, 53). Structural alignments of MtDnaG-CTD with these structures were done, using LSQMAN (54), to evaluate the differences between them. We specifically aligned the GRs of these structures to show the diversity in the relative orientation of the HHR. (Fig. 3). The RMSDs between MtDnaG-CTD and EcDnaG-CTD, BstDnaGCTD, HpDnaG-CTD, and VcDnaG-CTD were found to be 1.89 Å (45 Cα atoms), 1.85 Å (71 Cα atoms), 1.59 Å (62 Cα atoms), and 1.77 Å (43 Cα atoms), respectively. However, DnaG-CTDs have been found to share less than 10% sequence identity (Fig. S1). The total structural alignments of MtDnaG-CTD with EcDnaG-CTD, BstDnaGCTD, HpDnaG-CTD, and VcDnaG-CTD using RAPIDO (55) yielded RMSDs of 12.58 Å (101 Cα atoms), 6.81 Å (131 Cα atoms), 5.24 Å (108 Cα atoms) and 4.28 Å (122 Cα
atoms), respectively (Table S2). Inspection of the superpositions of these structures suggested the GR of MtDnaG-CTD to be structurally more similar to that of HpDnaG-CTD than to those of the other DnaG-CTDs. Overall, the arrangement of helices in MtDnaG-CTD was observed to be similar to that in VcDnaG-CTD (Table S2). The number of alpha helices present in each CTD structure was also found to differ: ten helices in the MtDnaG-CTD structure, six each for HpDnaG-CTD and VcDnaG-CTD, and seven each for EcDnaG-CTD and BstDnaG-CTD (Fig. S4). These differences in the number of alpha helices resulted for the most part in offsets between the boundaries of the helices in the structure-based alignment (Fig. 3). Comparison of the MtDnaG-CTD structure with other DnaG-CTD structures showed a larger deviation between the entire structures than between the GRs. The HHRs from the different sources adopted different orientations relative to the GRs and showed much greater structural differences than did the GRs. Moreover, the B factors of the HHRs and linkers were found to, in general, be greater than those of the GRs (Table S3), indicating relatively high flexibility levels for the HHR and linker regions. Also, note the lower overall B-factor of the MtDnaG-CTD structure than of the other known CTD structures (Table S3). Interestingly, despite all of the DnaG-CTD structures consisting of a GR as well as HHR, multiple sequence alignment (MSA) of amino acid sequences (Fig. S1b) and structural alignment did not yield any significant homology.

MtDnaG-CTD interacts at the dimer-dimer interface of MtDnaB-NTD using mostly hydrophobic residues of the HHR:

A previous investigation showed MtDnaB-NTD (residues 21–197, including the linker) to form a stable dimer in solution, and a truncated form of the protein (residues 21–134) to be monomeric, indicating the region spanning residues 135–197 to be required for dimer stability (28). Therefore, we also generated these monomeric and dimeric MtDnaB-NTD constructs, using the same method as described earlier (28), for our studies. Specifically, biosensor binding studies of each of these two constructs with MtDnaG-CTD was carried out using surface plasmon resonance by passing the respective MtDnaB-NTD construct on immobilized MtDnaG-CTD. The results of these two experiments showed MtDnaG-CTD to not interact with the monomeric form of MtDnaB-NTD (Fig. S3b) but to do so with the dimeric form (Fig. S3c).

The lack of a high-resolution structure of the DnaB-DnaG complex and the fragmented structural information for the individual proteins have impeded our attaining a detailed understanding of this crucial binary protein interaction. The only complex structure of the
active hexameric BstDnaB (PDB: 2R6A) was determined as a trimer of dimer and three
BstDnaG-CTDs were found at the dimer-dimer interface of BstDnaB (16). Similarly, the
crystal structure of the MtDnaB-NTD (PDB: 2R5U) has been reported to form a hexameric
ring having two distinct interfaces; an extensive hydrophobic interface, stabilizing a dimer
(monomer-monomer) of MtDnaB-NTD, and other less extensive interface formed between
the dimers, connecting three of them into a hexamer (28). The dimer-dimer interface of
MtDnaB-NTD was considered as a DnaG binding site (28). Here, we produced a model of the
M. tuberculosis DnaB-DnaG complex by one molecule of MtDnaG-CTD interacted to two
molecules of MtDnaB-NTD, involved at the dimer-dimer interface similar to BstDnaB-
DnaG-CTD complex structure (16). Superimpositions of the GR of the MtDnaG-CTD crystal
structure with GRs from the crystal structures of EcDnaG-CTD, VcDnaG-CTD, BstDnaG-
CTD, and HpDnaG-CTD respectively (Fig. 3), showed the orientation of the HHR to vary
considerably between the different species and suggested the flexibility of the HHR.
Inspection of the HHR in the unbound form of BstDnaG-CTD showed it to be oriented 65°
avay from its orientation in the helicase-bound form (Fig. S5a), and this difference was
attributed to the flexibility of the HHR. Therefore, the HHR of MtDnaG-CTD was rotated by
32° to model the MtDnaG-CTD/MtDnaB-NTD complex using the BstDnaB/DnaG-CTD
(PDB: 2R6A) complex structure as a template (Fig. S5). A molecular dynamics simulation of
the model was carried out for 25 nanoseconds and revealed the stability of the model after the
first few nanoseconds of the simulation (Fig. 4c). When producing our model of the MtDnaG-
CTD/MtDnaB-NTD complex, we maximized the number of interactions between the HHR of
MtDnaG-CTD and MtDnaB-NTD (Fig. 4a). DnaG mutations in the part of the extreme C-
terminal region, containing the HHR, were shown to be disruptive to the DnaB–DnaG
complex (56), hence indicating this part of the C-terminal region to be crucial for the
helicase-primase interaction. Our modeling highlighted the amino acid residues present in the
helicase-primase complex interface that may participate in these crucial interactions and
stabilize the complex (Fig. 4a). The low homology between the DnaG-CTD sequences (Fig.
S1a) together with the presence of different yet consistently hydrophobic amino acid residues
at the M. tuberculosis (Fig. 4a, 5b) and B. stearothermophilus (16) helicase-primase interface
suggests the involvement of hydrophobic interactions in stabilizing the helicase-primase
complex. According to the model, I605 and F615 from the primase of M. tuberculosis make
extensive hydrophobic interactions with helicase, and these residues, in particular, may be
playing a crucial role in stabilizing the helicase-primase complex (Fig. 4b).
Ile605 in the HHR is crucial for helicase-primase interaction in *Mycobacterium tuberculosis*:

To elucidate the main drivers of the helicase-primase association and to identify hotspots, we turned to quantify individual contributions of potential key residues, guided by a structural analysis of our proposed *M. tuberculosis* DnaG-CTD/DnaB-NTD complex model. To accomplish this goal, we mutated a couple of crucial contact residues modeled to be present at the interface of the *M. tuberculosis* DnaG-DnaB complex. Inspection of our model (Fig. 4a) indicated a large non-conserved hydrophobic residue, Ile605, to be the most prominent surface-accessible hydrophobic residue on the HHR of *Mt*DnaG-CTD (Fig. 4b). Inspection of our model indicated Ile605, of the loop connecting αH9 and αH10 of the HHR, to be stabilizing the complex. Ile605 present at the crucial position, flexible loop within the HHR, through that the whole HHR get hooked by this residue. Our analysis also indicated Phe615, another large hydrophobic residue of *Mt*DnaG-CTD, to be present at the *M. tuberculosis* DnaG-DnaB interface (Fig. 4b). This residue, being present at helix 10 of *Mt*DnaG-CTD, was observed to be less accessible than *Mt*DnaG-CTD Ile605 present at the connecting loop of the HHR. The stability levels of the helicase (*Mt*DnaB-NTD) in complex with native *Mt*DnaG-CTD, mutant *Mt*DnaG-CTD F615A, and mutant *Mt*DnaG-CTD I605A were analyzed by carrying out molecular dynamics simulations (Fig. 4c). Furthermore, the binding energy levels of all of the complex models were calculated using the MM/GBSA method of GROMACS (Table S4). The binding energy values of helicase (*Mt*DnaB-NTD) in complex with native *Mt*DnaG-CTD, mutant *Mt*DnaG-CTD F615A, and mutant *Mt*DnaG-CTD I605A were -100 kJ/mol, -75.3 kJ/mol and -60 kJ/mol, respectively (Table S4). The binding affinities for the native and mutant *Mt*DnaG-CTDs in complex with *Mt*DnaB-NTD were also measured in vitro using surface plasmon resonance (SPR) spectroscopy (Fig. 5). Native and mutant *Mt*DnaG-CTDs were immobilized on three different chips under identical conditions, and different concentrations of *Mt*DnaB-NTD were passed as an analyte with HBS buffer over the immobilized proteins. The $K_D$ values of MtDnaB-NTD in complex with MtDnaG-CTD, MtDnaG-CTD F615A, and MtDnaG-CTD I605A were measured to be 370 nM, 444 nM, and 2.8 µM, respectively (Fig. 5). The SPR results indicated a 10-fold decrease in the binding affinity of MtDnaB-NTD with MtDnaG-CTD I605A compared to that of MtDnaB-NTD with native MtDnaG-CTD. Although the measured difference in binding affinity to the native (*Mt*DnaG-CTD) and mutated (*Mt*DnaG-CTD I605A) proteins was relatively large considering that only a single residue was mutated, there was still observed significant affinity between the helicase and the mutated (*Mt*DnaG-CTD I605A) primase. In
our current study, we noticed the important role of the loop connecting the two helices of the HHR in stabilizing the helicase-primase complex. This major role of the loop within the HHR in the stability of the helicase-primase complex was indicated despite no common amino acid residues present at the same position of different DnaG-CTDs. Both *in silico* calculations and *in vitro* experiments (SPR) suggested a significant role of Ile605 of *Mt*DnaG in the stabilization of the helicase-primase (*Mt*DnaB/*Mt*DnaG) complex in *M. tuberculosis*.

**Mycobacterium** DnaB-NTD interacts to DnaG-CTD of other organisms with low binding affinity: The helicase-binding domain of DnaG and primase-binding region of DnaB are the least conserved regions of their respective proteins. Despite the insignificant homology between the various DnaG-CTD sequences, the DnaG-CTD structures were found to be similar, especially in their all having globular and HHR regions. A functional complementation of *Ec*DnaB with *Hp*DnaB has been reported earlier (57), indicative of an interaction between *Hp*DnaB with *Ec*DnaG. To further test whether the cross-species helicase-primase interaction takes place even with such a low sequence similarity, we performed a study of the interaction of *Mt*DnaB-NTD with other DnaG-CTDs. We studied the specificity in the interaction of DnaB with non-cognate DnaG by generating an *in silico* model of the complex and further tested the biophysical interactions of *Mycobacterium* DnaB-NTD with other DnaG-CTDs, i.e., from *E. coli*, *H. pylori*, and *V. cholerae*. The steps used to *in silico* build and analyze the models of *Mt*DnaB-NTD in complex with DnaG-CTDs from *E. coli*, *H. pylori*, and *V. cholerae* (Fig. 6) were similar to the strategy used in producing the model of the *M. tuberculosis* helicase-primase complex. The free energy values of the binding of native *Mt*DnaB-NTD to *Ec*DnaG-CTD, *Hp*DnaG-CTD, and *Vc*DnaG-CTD were -13.7 kJ/mol, -71.5 kJ/mol and -58.4 kJ/mol respectively. The *in silico* studies suggested the *Mt*DnaB-NTD/*Hp*DnaG-CTD complex to be more stable than the *Mt*DnaB-NTD/*Ec*DnaG-CTD and *Mt*DnaB-NTD/*Vc*DnaG-CTD complexes. We also carried out biosensor binding studies using surface plasmon resonance (SPR) spectroscopy. Specifically, the binding affinities of native *Mt*DnaB-NTD for *Ec*DnaG-CTD, *Hp*DnaG-CTD, and *Vc*DnaG-CTD were measured *in vitro* using SPR spectroscopy. The *Kd* values for the complexes of *Mt*DnaB-NTD with *Ec*DnaG-CTD (Fig. 7a) and *Hp*DnaG-CTD (Fig. 7b) were determined from these studies to be 7.4 µM and 3.4 µM, respectively, while *Mt*DnaB-NTD formed an unstable association with *Vc*DnaG-CTD (Fig. 7c). However, the cross species helicase-primase binding did not show the order
as predicted by the in silico studies. But both in silico as well as in vitro cross- study on MtDnaB-NTD to DnaG-CTD of other bacteria showed the significant decrease in binding affinities compared to the binding affinity of native MtDnaG-CTD and MtDnaB-NTD interaction. Moreover, the decrease in the binding affinity of DnaG to DnaB might affect the activity of these proteins during the replication process.

**Discussion**

In this study, the crystal structure of the DnaB helicase-binding domain of DnaG primase (MtDnaG-CTD) of *Mycobacterium tuberculosis* was determined to gain insight into the helicase-primase interaction, crucial for the survival of eubacteria. The complex between the DnaG primase and the DnaB helicase unwinds duplex DNA at the eubacterial replication fork and synthesizes the RNA primers required for initiation of DNA replication (58, 59). An alignment of DnaG-CTD sequences showed high diversity. Therefore, the MtDnaG-CTD crystal structure could not be solved using molecular replacement, and we instead used SAD phasing. Analysis of the determined structure highlighted the flexibility of the DnaG-CTD scaffold, which has been observed in other DnaG-CTDs, and supported the existence of hinge bending within the protein. The hinge point located in the linker between the globular region (GR) and hairpin region (HHR) apparently allows these two regions of the protein to become close together or far apart, resulting in an alternation between elongated and contracted shapes. Also, the flexibility of the helicase-binding domain has been shown to be due in large part to the identified hinge point, and to be the main factor maximizing the interactions with DnaB helicase. A comparison of MtDnaG-CTD with other DnaG-CTD structures suggested some degree of conservation of the protein fold, in particular in the GR. Overall; the HHR appears to have the major role in the interaction with DnaB and shows a high degree of diversity in eubacteria.

We also constructed in silico a hypothetical model of the MtDnaB-NTD/MtDnaG-CTD complex, based on the *Bst*DnaB/DnaG-CTD complex crystal structure (16), and found the hypothetical model to be consistent with previous mutagenesis studies (56). In the model of the MtDnaB-NTD/MtDnaG-CTD complex, the HHR of MtDnaG-CTD covers the interface otherwise making dimer-dimer interactions in the hexameric form MtDnaB-NTD. Also, Ile605 and Phe615 are large non-conserved hydrophobic residues present at the HHR of MtDnaG and, according to our model, interact with MtDnaB (Fig. 4a). Our analysis also indicated Ile605, a hydrophobic residue present on the loop of the HHR of MtDnaG-CTD, to interact with Val82 and Ala85 of MtDnaB-NTD and hence to be crucial for the formation of
the complex. Our *in silico* study of the *MtDnaB/MtDnaG* complex also indicated a decrease in binding energy upon mutation of *MtDnaG-CTD* Ile605 to alanine as well as upon the mutation of its Phe615 to alanine (Table S4). To provide further evidence for the results of these binding affinity calculations, we also carried out *in vitro* studies involving expressing both mutants of *MtDnaG-CTD* and using SPR spectroscopy to measure their affinities for DnaB-NTD. In this biosensor binding study, the affinity of the Ile605Ala *MtDnaG-CTD* mutant for *MtDnaB-NTD* was shown to be ten times weaker than the affinity of the native *MtDnaG-CTD* for *MtDnaB-NTD*. These findings helped us validate the model and identify Ile605 as one of the key *MtDnaG* residues stabilizing the association of *MtDnaG* with *MtDnaB*.

Analysis of our current structure, biosensor results and the results of the previous studies of DnaB-DnaG complexes (15, 16) led us to conclude that the loop connecting the two helices of the HHR is apparently largely responsible for the stability of this complex. The sequence and structural analysis indicated the non-conserved residue Ile605, present at the helical hairpin loop of DnaG, to be specific to *M. tuberculosis* and to be a key residue for the interaction between DnaB and DnaG (Fig. 4). The earlier study in *H. pylori*, residue Phe537 of the helical hairpin loop of DnaG-CTD has been shown to play a key role in the stability of the DnaB-DnaG complex (15). The observation of amino acid residues present at the helical hairpin loop of DnaG-CTDs suggested the presence of non-conserved hydrophobic residues that play a significant role in the stability of the DnaB-DnaG complex. The flexibility of the loop apparently allows for maximizing the access of its hydrophobic residues for contacts with DnaB-NTD.

The binding affinity of *MtDnaB* for *MtDnaG* was measured to be in the nanomolar range while the binding affinities of *MtDnaB* with other DnaGs from different species were measured in the micromolar range (Fig. 5, Fig. 7). Such a relatively low level of heterologous binding affinity might be expected to result in relatively poor helicase and primase activities in the case of complementarity with other organisms. A suboptimal activity level of any DNA replication initiation protein would, in turn, be expected to slow down replication fork progression (2). Though the helicase of *H. pylori* showed functional complementarity with *E. coli* (57), such complementarity of this helicase seems to be specific to just a few organisms. Therefore, *MtDnaB* may show complementarity with *E. coli* and *H. pylori* but not in *V. cholerae*. The pattern of interactions between DnaG and DnaB seems to vary from one organism to the next, and therefore any change in the affinity between these proteins may affect not only the initiation of replication but also overall DNA replication.
Based on the crystal structure and our model, we can suggest a possible mechanism to explain the regulation of DnaB-DnaG activity at the replication fork. Bacterial primases primarily adopt a compact state but can transition to an extended conformation (60). In the unbound form or free form of DnaG, the HHR may fold back to interfere with the RNA polymerase domain or the Zn-binding domain. According to the proposed mechanism, the moment when DnaG-CTD comes into contact to helicase the GR side of DnaG-CTD might interact to the one molecule of MtDnaB-NTD and this binding might enforce for the changing the orientation of HHR from the unbound state of DnaG. The HHR would bend in such a way as to properly bind to the second molecule of DnaB-NTD with comparatively stronger than first one, and the bending of the HHR away from the RNA polymerase domain or Zn-binding domain may lead to the activation of primase by coming into contact with DnaB. Because DnaG cannot effectively function as an independent molecule, i.e., unbound from DnaB, this mechanism would also provide a means to prevent unwanted priming at non-replicative sites in the cell. Future studies to further address this model in the context of the replisome will highlight how interactions between DnaG and the rest of the replisome synergize to regulate replication of the lagging strand.

In conclusion, the structural and biophysical data provided by our work have shed new light on the structural assembly and mode of binding of DnaG-CTD with DnaB-NTD in M. tuberculosis and unveiled mechanistic details and selectivity determinants for other eubacteria. The non-conserved interacting domains of DnaG-DnaB complex expand the study on the aspects of the species-specific evolution of the replication system.

Acknowledgments:
S.G. acknowledges DBT, DST-PURSE, and UGC RNW, University Grant Commission (SAP), D.P.S. acknowledges UGC for the fellowship, and R.V. acknowledges DST-SERB for the fellowship. We thank the Advanced Instrumentation Research Facility (AIRF) and JNU for providing the SPR facility. We also thank the staffs of BM14, ESRF for helping us with the collection of high-resolution X-ray data.

Author contributions:
SG and DPS conceived the idea and designed the experiments. DPS solved the MtDnaG-CTD structure. SG, DPS and RV analyzed the structures and constructed the models. DPS
generated SDM and performed all of the biosensor studies. RV performed the MD simulations of all of the models and SAAR performed cloning. DPS wrote the manuscript taking input from SG, and SG reviewed the manuscript.

Abbreviations:

Mt, Mycobacterium tuberculosis; Ec, Escherichia coli; Hp, Helicobacter pylori; Bst, Bacillus stearothermophilus; Vc, Vibrio cholerae; DnaG-CTD, DnaG C-terminal domain; DnaB-NTD, DnaB N-terminal domain; RMSD, root mean square deviation;

References:


Table 1: Crystallographic data and refinement statistics:

<table>
<thead>
<tr>
<th>Crystallographic data</th>
<th>734</th>
</tr>
</thead>
<tbody>
<tr>
<td>X-ray source</td>
<td>BM14, ESRF, France</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.97</td>
</tr>
<tr>
<td>Space group</td>
<td>P2₁2₁2₂</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Unit cell parameters</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a, b, c (Å)</td>
<td>48.28, 137.55, 36.51</td>
</tr>
<tr>
<td>α, β, γ (°)</td>
<td>90, 90, 90</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>68.78 -1.58</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>92.8 (99.8)</td>
</tr>
<tr>
<td>Total no. of observations</td>
<td>150161</td>
</tr>
<tr>
<td>No. of unique observations</td>
<td>40035</td>
</tr>
<tr>
<td>Redundancy</td>
<td>3.8 (3.9)</td>
</tr>
<tr>
<td>Average I/σ (I)</td>
<td>19.66 (1.94)</td>
</tr>
<tr>
<td>CC (1/2)</td>
<td>0.92 (0.72)</td>
</tr>
<tr>
<td>$R_{r.i.m.}$</td>
<td>0.088</td>
</tr>
<tr>
<td>Crystal mosaicity (°)</td>
<td>0.76</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Refinement</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>R-factor (%)</td>
<td>18.14</td>
</tr>
<tr>
<td>Free R-factor (%)</td>
<td>20.88</td>
</tr>
<tr>
<td>Mean B factor (Å²)</td>
<td>16.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of atoms</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein/PEG/acetate/water</td>
<td>1812/17/7/320</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RMS deviation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Bond length (Å)</td>
<td>0.017</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.64</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ramachandran plot (%)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>In preferred region</td>
<td>99.1</td>
</tr>
<tr>
<td>In allowed region</td>
<td>0.5</td>
</tr>
<tr>
<td>Outliers</td>
<td>0.4</td>
</tr>
<tr>
<td>PDB ID</td>
<td>5Z51</td>
</tr>
</tbody>
</table>
Figure legends:

**Fig. 1:** *Mt*DnaG-CTD is a dimer in solution. (a) Size-exclusion chromatograph of the C-terminal domain of primase (*Mt*DnaG-CTD) passed through a HiLoad 16/60 Superdex 75 column. Ten-milliliter fractions were collected. The elution volume (70.3 ml) and the elution pattern of the protein are displayed. (b) The molecular mass of the eluted DnaG-CTD was deduced from the standard plot to be about ~35kDa, corresponding to the dimeric state of the protein. (c) SDS-PAGE showing fractions purified by gel filtration. The proteins were separated on SDS–12% PAGE and stained with Coomassie brilliant blue. Lane M shows the molecular markers; lane P shows the gel filtration fraction.

**Fig. 2:** Crystal structure of *Mt*DnaG-CTD. (a) A portion of the molecule shown with a 2Fo-Fc electron density map contoured at 1σ within 1.5 Å of selected atoms. (b) Cartoon diagram showing the crystal structure of the *Mt*DnaG primase CTD. The overall structure was observed to have one globular region connected to the helical hairpin region by a flexible loop. Chain A is colored purple, and chain B is cyan. One complete *Mt*DnaG-CTD (chain A) and one helical hairpin cleaved from this CTD (chain B) were identified per asymmetric unit. (c) Cylindrical representation of the MtDnaG-CTD crystal structure showing the arrangement of the ten helices connected by loops. Helices αH1 - αH8 formed the globular region and helices αH9 – αH10 formed the helical hairpin region. (d) Amino acid residues showed the hydrophobic interaction between the chain A and chain B in the crystal structure of *Mt*DnaG-CTD.

**Fig. 3:** Superpositions of the globular region of *Mt*DnaG-CTD with the globular regions of other DnaG-CTDs provided insight into the differences in the shapes and orientations of HHRs. Superposition of the globular region of the *Mt*DnaG-CTD (purple) crystal structure with the globular region of each of (a) *Ec*DnaG-CTD (green), (b) *Vc*DnaG-CTD (cyan), (c) *Bst*DnaG-CTD (violet) and (d)*H*pr*DnaG-CTD (wheat). The helical hairpin regions (HHRs) of the different structures showed different orientations.

**Fig. 4:** Model of the *M. tuberculosis* helicase-primase complex. (a) Model of the *M. tuberculosis* helicase-primase complex. Experimentally determined structures of *Mt*DnaB-NTD (PDB: 2R5U) and *Mt*DnaG-CTD (PDB: 5Z51), as well as that of the *Bst*DnaB-*Bst*DnaG-CTD complex (PDB: 2R6A), were used to generate the model of the complex. In the complex model, two molecules of DnaB-NTD interacts with one molecule of DnaG-CTD.
The model was further relaxed to eliminate bad contacts using molecular energy minimization with the help of the GROMACS molecular dynamics package. (b) The crucial residues of the helical hairpin showing hydrophobic interactions stabilizing the helicase-primase complex. (c) RMSD analysis of the trajectories of native and mutants primase models. Time-dependent Cα RMSDs for all residues of DnaB-NTD complexed with the native DnaG-CTD (black), I605A mutant DnaG-CTD (red), and F615A mutant DnaG-CTD (green) showed structural convergence from the initial time point. No effect of the mutations on the DnaG-CTD structure was observed, but a significant effect on the binding affinity of the helicase for the primase was found.

Fig. 5: SPR sensorgram for measuring the binding affinity of the helicase for the primase. Sensorgrams showing the binding pattern of MtDnaB-NTD at various concentrations (legends) with immobilized (a) native MtDnaG-CTD, (b) F615A mutant MtDnaG-CTD, and (c) I605A mutant MtDnaG-CTD. The mutation of a single residue, Ile605 to Ala, on the helicase-binding surface resulted in an almost 10-fold decrease in the binding affinity.

Fig. 6: In silico investigations of the cross-species helicase-primase complexes. (a-c) Experimentally determined structures of MtDnaB-NTD (PDB: 2R5U) and (a) EcDnaG-CTD (PDB: 1T3W), (b) HpDnaG-CTD (PDB: 4EHS) and (c) VcDnaG-CTD (PDB: 4IM9) were used to generate models of the helicase-primase complex. In each case, the crystal structure of the BstDnaB-BstDnaG-CTD (PDB: 2R6A) complex was used as a template. The model was further relaxed to eliminate bad contacts using molecular energy minimization with the help of the GROMACS molecular dynamics package. (d) RMSD analyses of the trajectories of the three models. Time-dependent Cα-RMSDs for all residues of DnaB-NTD complexed with EcDnaG-CTD (red), HpDnaG-CTD (black) and VcDnaG-CTD (green) showed structural convergence from the initial time point. There were no structural changes during the dynamics simulation, but the DnaB-DnaG binding affinities for the cross-species complexes differed significantly from that for the native M. tuberculosis DnaB-DnaG complex.

Fig. 7: SPR sensorgrams for measuring the binding affinities of helicase with non-cognate primases. Sensorgrams showing the binding patterns of MtDnaB-NTD at various concentrations (legends) with immobilized (a) E. coli DnaG-CTD, (b) H. pylori DnaG-CTD and (c) V. cholerea DnaG-CTD. The results indicated MtDnaB-NTD to weakly bind EcDnaG-CTD and HpDnaG-CTD, and unstably interact with VcDnaG-CTD.
Figures:

Fig. 1
Fig. 2

(a) 

(b) 

(c) 

(d) 

Helical hairpin region 

Globular region 

Helical hairpin region
Fig. 4
Fig. 5

(a) $K_D = 370 \text{ nM}$

(b) $K_D = 444 \text{ nM}$

(c) $K_D = 2.82 \text{ uM}$

- Time (s)
- Response [m$^2$]
Fig. 6

(a) Ec DnaG

(b) Hp DnaG

(c) Vc DnaG

(d) 

- MtDnaB-HpDnaG
- MtDnaB-EcDnaG
- MtDnaB-VcDnaG

RMSD (nm)

Time (ns)
Fig. 7

(a) 

(b) 

(c) 

Unstable binding

Response [mM]

Response [mM]

Response [mM]

Time (s)

Time (s)

Time (s)
Supporting Data

Supplementary Fig. 1: DnaG primase. (a) Domain organization of DnaG primase. DnaG has three domains: zinc-binding domain for DNA binding, polymerase domain for primer synthesis, and C-terminal domain for helicase binding. The sequences of the zinc-binding domain and polymerase domain are conserved in DnaG primases, but the helicase-binding domain shows high variation. (b) Multiple sequence alignment of DnaG-CTD or helicase-binding domain shows insignificant homology.

Supplementary Fig. 2: Cleavage of MtDnaG-CTD into two fragments. (a) About a week old purified MtDnaG-CTD protein was loaded into 15% SDS-PAGE, yielded three bands, indicating part of the protein sample was cleaved into two fragments. This experiment confirms the crystal structure, where cleaved HHR fragment and DnaG-CTDs assembled into one asymmetric unit. The three bands indicate full-length CTD, lower twin bands represent globe region and HHR. (b) MALDI analysis of all three bands confirmed the cleavage of MtDnaG-CTD into two fragments.

Supplementary Fig. 3: SPR sensorgram: (a) SPR sensorgram for measuring the binding pattern of BSA with different concentrations of MtDnaB-NTD as a negative control. SPR sensorgrams showing the binding pattern of MtDnaG-CTD with different concentrations of (b) monomeric MtDnaB-NTD and (c) dimeric MtDnaB-NTD. Only dimeric DnaG-CTD showed a pattern indicative of binding DnaB-NTD.

Supplementary Fig. 4: Cartoons showing the helices in crystal structures of the DnaG primase C-terminal domain. Cylindrical representation of (a) EcDnaG-CTD, (b) BstDnaG-CTD, (c) HpDnaG-CTD and (d) VcDnaG-CTD showing the total number of helices present in each of the different DnaG-CTDs.

*X is very small (three residues) and it was not considered as a complete helix.

Supplementary Fig. 5: The helical hairpin region adopts different orientations in the free and bound states. Alignments of the globular regions (GRs) of the DnaG-CTDs gave insight into the differences in the shapes and orientations of the HHRs. (a) Superposition of the BstDnaG-CTD portion of the crystal structure of the BstDnaB-BstDnaG-CTD complex (blue) with the NMR structure of (uncomplexed) BstDnaG-CTD (green), showing a difference of about 65° between the orientations of their hairpins. (b) Superposition of the (uncomplexed) MtDnaG-CTD crystal structure (light blue) with the MtDnaG-CTD portion of the model of the complex (orange), showing a difference of about 32° between their hairpin orientations. (c) Structure of the BstDnaB/BstDnaG-CTD complex (PDB: 2R6A) used as a template for the MtDnaB/MtDnaG model. (d) Superposition of the structure of MtDnaB-NTD on that of the BstDnaB-NTD domain, and a superposition of the MtDnaG-CTD GR structure on the BstDnaG-CTD GR portion of the structure of the BstDnaB-BstDnaG-CTD complex. For superimposing the CTD helical hairpin structures, the helical hairpin of MtDnaG-CTD was rotated by about 32°, keeping the linker between the HHR and the GR as the center of rotation, in order to obtain the lowest root mean square deviation (RMSD) between the BstDnaG-CTD and MtDnaG-CTD structures.
Supplementary Fig. 1

(a) DnaG-primase domain organization

![Diagram showing the organization of DnaG-primase domains: Zinc binding, RNA polymerase, Helicase binding.]

(b) Alignment of DnaG-primase sequences from different species:

- M. tuberculosis
- B. stearothermophilus
- H. pylori
- E. coli
- V. cholerae

<table>
<thead>
<tr>
<th>Species</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>M. tuberculosis</td>
<td>A/L/V/E/G/F/T/S/etc.</td>
</tr>
<tr>
<td>B. stearothermophilus</td>
<td>I/E/H/E/A/L/A/Y/etc.</td>
</tr>
<tr>
<td>H. pylori</td>
<td>G/V/L/K/H/E/etc.</td>
</tr>
<tr>
<td>E. coli</td>
<td>D/E/+K/L/E/L/etc.</td>
</tr>
<tr>
<td>V. cholerae</td>
<td>R/H/L/M/E/G/L/etc.</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>Species</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>M. tuberculosis</td>
<td>L/F/G/D/L/V/A/S/R/Y/etc.</td>
</tr>
<tr>
<td>B. stearothermophilus</td>
<td>A/A/R/I/A/K/E/etc.</td>
</tr>
<tr>
<td>H. pylori</td>
<td>K/Q/G/E/L/A/etc.</td>
</tr>
<tr>
<td>E. coli</td>
<td>N/Q/E/L/A/K/etc.</td>
</tr>
<tr>
<td>V. cholerae</td>
<td>I/L/K/G/L/etc.</td>
</tr>
</tbody>
</table>
### Supplementary Fig. 2

#### (a)

<table>
<thead>
<tr>
<th>Accession</th>
<th>Mass</th>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>gi</td>
<td>289762519</td>
<td>16940</td>
<td>101 DNA primase dnaG [Mycobacterium tuberculosis GM 1503]</td>
</tr>
<tr>
<td>gi</td>
<td>115609480</td>
<td>70006</td>
<td>60 DNA primase [Mycobacterium tuberculosis H37Rv]</td>
</tr>
<tr>
<td>gi</td>
<td>3266093959</td>
<td>69991</td>
<td>60 DNA primase dnaG [Mycobacterium tuberculosis W-148]</td>
</tr>
<tr>
<td>gi</td>
<td>2761358231</td>
<td>70106</td>
<td>48 DNA primase [Mycobacterium bovis BCG str. Pasteur 1173P2]</td>
</tr>
<tr>
<td>gi</td>
<td>1165891436</td>
<td>5537</td>
<td>42 RNA polymerase beta subunit [Mycobacterium tuberculosis]</td>
</tr>
<tr>
<td>gi</td>
<td>2897644663</td>
<td>6545</td>
<td>38 RNA polymerase beta subunit [Mycobacterium tuberculosis]</td>
</tr>
<tr>
<td>gi</td>
<td>1297644649</td>
<td>6503</td>
<td>38 RNA polymerase beta subunit [Mycobacterium tuberculosis]</td>
</tr>
</tbody>
</table>

#### (b)

<table>
<thead>
<tr>
<th>Accession</th>
<th>Mass</th>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>gi</td>
<td>289762519</td>
<td>16940</td>
<td>187 DNA primase dnaG [Mycobacterium tuberculosis GM 1503]</td>
</tr>
<tr>
<td>gi</td>
<td>115609480</td>
<td>70006</td>
<td>160 DNA primase [Mycobacterium tuberculosis H37Rv]</td>
</tr>
<tr>
<td>gi</td>
<td>3266093959</td>
<td>69991</td>
<td>160 DNA primase dnaG [Mycobacterium tuberculosis W-148]</td>
</tr>
<tr>
<td>gi</td>
<td>2761358231</td>
<td>70106</td>
<td>92 DNA primase [Mycobacterium bovis BCG str. Pasteur 1173P2]</td>
</tr>
<tr>
<td>gi</td>
<td>2165556968</td>
<td>38451</td>
<td>49 MCE-family protein mce1B [Mycobacterium tuberculosis Ti]</td>
</tr>
<tr>
<td>gi</td>
<td>115607311</td>
<td>37806</td>
<td>49 MCE-family protein MCE1B [Mycobacterium bovis AF2122/97]</td>
</tr>
</tbody>
</table>

#### (c)

<table>
<thead>
<tr>
<th>Accession</th>
<th>Mass</th>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>gi</td>
<td>289762519</td>
<td>16940</td>
<td>159 DNA primase dnaG [Mycobacterium tuberculosis GM 1503]</td>
</tr>
<tr>
<td>gi</td>
<td>115609480</td>
<td>70006</td>
<td>146 DNA primase [Mycobacterium tuberculosis H37Rv]</td>
</tr>
<tr>
<td>gi</td>
<td>3266093959</td>
<td>69991</td>
<td>146 DNA primase dnaG [Mycobacterium tuberculosis W-148]</td>
</tr>
<tr>
<td>gi</td>
<td>2761358231</td>
<td>70106</td>
<td>82 DNA primase [Mycobacterium bovis BCG str. Pasteur 1173P2]</td>
</tr>
<tr>
<td>gi</td>
<td>2165556968</td>
<td>38451</td>
<td>53 signal recognition particle protein ffr [Mycobacterium tuberculosis Ti]</td>
</tr>
<tr>
<td>gi</td>
<td>115607311</td>
<td>37806</td>
<td>53 signal recognition particle protein ffr [Mycobacterium tuberculosis Ti]</td>
</tr>
<tr>
<td>gi</td>
<td>289762519</td>
<td>16940</td>
<td>44 signal recognition particle protein ffr [Mycobacterium tuberculosis H37Rv]</td>
</tr>
<tr>
<td>gi</td>
<td>2165556968</td>
<td>38451</td>
<td>44 signal recognition particle protein ffr [Mycobacterium tuberculosis Ti]</td>
</tr>
</tbody>
</table>

Supplementary Fig. 4

(a) PDB: 1J3W

(b) PDB: 2R6A

(c) PDB: 4EHS

(d) PDB: 4M9
Supplementary Fig. 5

(a)  

(b)  

(c)  

(d)
**Supplementary Table 1:** List of primers used for the experiments.

<table>
<thead>
<tr>
<th></th>
<th>Primer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fw-MtDnaG-CTD</strong></td>
<td>CTAGCTAGCACCCCTGTGCCCACAGCGTGAGGCAGCT</td>
</tr>
<tr>
<td><strong>Rv-MtDnaG-CTD</strong></td>
<td>CGGCTCGAGGTTCGGCCGCTGCGCTGCTCTAACAGGCT</td>
</tr>
<tr>
<td><strong>Fw-MtDnaGI605A</strong></td>
<td>CATGTCGCCGGCGAGCAAGCGGCG</td>
</tr>
<tr>
<td><strong>Rv-MtDnaGI605A</strong></td>
<td>CGCGCTCTGAGGTCGCCGCTGGCGACCTGG</td>
</tr>
<tr>
<td><strong>Fw-MtDnaGF615A</strong></td>
<td>CCACCGCGCTGGGCAGGCGACCTGG</td>
</tr>
<tr>
<td><strong>Rv-MtDnaGF615A</strong></td>
<td>TATTCGTCGCCCTGGCTCG</td>
</tr>
<tr>
<td><strong>Fw-MtDnaB-NTD</strong></td>
<td>GGAATTCCATATG ATGGC GTCAA CGACAGGAGG</td>
</tr>
<tr>
<td><strong>Rv-MtDnaB-NTD</strong></td>
<td>CCGCTCGAGCTATAC GCG CGACAGGAGG</td>
</tr>
</tbody>
</table>

*Primers for DnaG-CTD of *H. pylori*, *E. coli*, and *V. cholerae* were listed previously.

**Supplementary Table 2:** Structure-based alignment of *M. tuberculosis* DnaG-CTD with other structures using Rapido

<table>
<thead>
<tr>
<th>Name</th>
<th># aligned residues</th>
<th>RMSD rigid</th>
<th># aligned residues (rigid body)</th>
<th>RMSD flex</th>
<th># rigid bodies</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>E. coli</em></td>
<td>101</td>
<td>12.58</td>
<td>69</td>
<td>0.52</td>
<td>5</td>
</tr>
<tr>
<td><em>B. stearothermophilus</em></td>
<td>131</td>
<td>6.82</td>
<td>87</td>
<td>1.01</td>
<td>4</td>
</tr>
<tr>
<td><em>H. pylori</em></td>
<td>108</td>
<td>5.24</td>
<td>85</td>
<td>0.90</td>
<td>4</td>
</tr>
<tr>
<td><em>V. cholerae</em></td>
<td>122</td>
<td>4.28</td>
<td>77</td>
<td>0.81</td>
<td>3</td>
</tr>
</tbody>
</table>

**Supplementary Table 3:** Average B factors of different regions of the DnaG C-terminal domain

<table>
<thead>
<tr>
<th>Protein</th>
<th>Globular region (Å²)</th>
<th>Linker (Å²)</th>
<th>Hairpin helix (Å²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MtDnaG (PDB)</td>
<td>17.89</td>
<td>25.18</td>
<td>19.28</td>
</tr>
<tr>
<td>EcDnaG (IT3W)</td>
<td>61.70</td>
<td>89.60</td>
<td>66.10</td>
</tr>
<tr>
<td>BstDnaG (2R6A)</td>
<td>98.92</td>
<td>103.06</td>
<td>92.64</td>
</tr>
<tr>
<td>HpDnaG (4EHS)</td>
<td>18.4</td>
<td>24.30</td>
<td>22.70</td>
</tr>
<tr>
<td>VcDnaG (4IM9)</td>
<td>33.87</td>
<td>51.67</td>
<td>39.74</td>
</tr>
</tbody>
</table>
Supplementary Table 4: Free binding energy values of the binding of native and mutant *M. tuberculosis* primases to native *M. tuberculosis* helicase.

<table>
<thead>
<tr>
<th>No.</th>
<th>Complex</th>
<th>ΔE elec (kcal mol$^{-1}$)</th>
<th>ΔE vdW (kcal mol$^{-1}$)</th>
<th>ΔG polar (kcal mol$^{-1}$)</th>
<th>ΔG nonpolar (kcal mol$^{-1}$)</th>
<th>ΔG binding (kcal mol$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Native <em>MtDnaG</em>-CTD</td>
<td>-624.8±1.8</td>
<td>-92.3±1.3</td>
<td>960.2±4.0</td>
<td>-20.9±5.6</td>
<td>-100.3±3.7</td>
</tr>
<tr>
<td>2.</td>
<td>Mutant <em>MtDnaG</em>-CTD (I605A)</td>
<td>-200.0±3.6</td>
<td>-87.8±1.7</td>
<td>906.9±3.8</td>
<td>-3.7±6.6</td>
<td>-60.9±2.4</td>
</tr>
<tr>
<td>3.</td>
<td>Mutant <em>MtDnaG</em>-CTD (F615A)</td>
<td>-293.2±5.9</td>
<td>-91.9±1.5</td>
<td>913.8±6.82</td>
<td>-14.1±9.0</td>
<td>-75.3±3.2</td>
</tr>
</tbody>
</table>