



Studies on The Interaction Between DnaG Primase and ssDNA Template in *Mycobacterium tuberculosis**

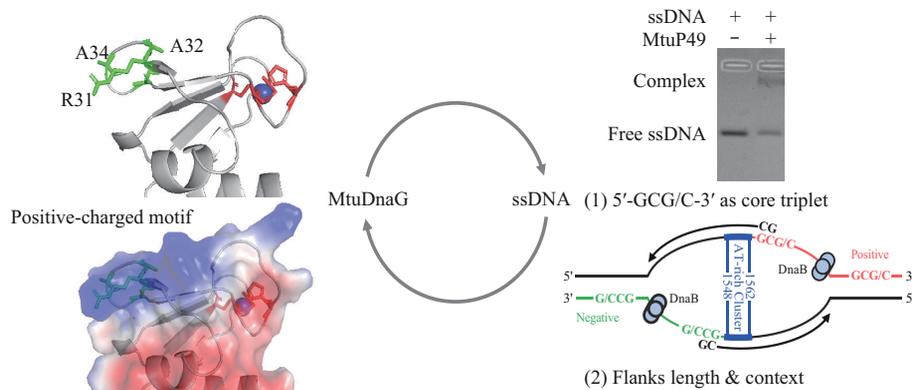
CHEN Jiang^{1,2,3}, LUO Hao^{1,3}, ZHANG Zhi-Ming^{1,3}, SONG Xu², WANG Gang-Gang¹**

¹Key Laboratory of Environmental and Applied Microbiology, Key Laboratory of Environmental Microbiology of Sichuan Province, Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu 610041, China;

²College of Life Sciences, Sichuan University, Chengdu 610064, China;

³University of Chinese Academy of Sciences, Beijing 100049, China)

Graphical abstract



Abstract Objective DnaG primase in *Mycobacterium tuberculosis* (MtuDnaG) plays a vital role in DNA replication, making it a target for novel antituberculosis drug discovery. However, the mechanism of MtuDnaG priming is not fully understood, which hinders the screening of MtuDnaG inhibitors. In this work, the specific recognition sites (SRS) in ssDNA for MtuDnaG binding was investigated and the interactions between MtuDnaG and ssDNA template was discussed. **Methods** By biochemical and biophysical methods, the binding of the didomain of MtuDnaG (MtuP49, containing the zinc-binding domain and RNA polymerase domain) to ssDNA template with various trinucleotide sites was evaluated, the affinity of MtuP49 to ssDNA template was measured. **Results** The present study suggested the 5'-GCG/C-3' as the potential SRS in ssDNA for specific binding to MtuDnaG. Besides, 5'-GCG/C-3' sites were further identified within the *oriC* region of *M. tuberculosis* genome. Importantly, the 3' sequence flanking the 5'-GCG/C-3' site markedly affected the binding affinity of ssDNA to MtuP49. Mutagenesis studies showed that substitution of residue Arg31 in the zinc-binding domain affected the binding activity of MtuP49 to template ssDNA. Combined with the predicted structure of MtuP49, an intramolecular rearrangement of zinc-binding domain relative to the RNA polymerase domain was implied to be essential in the binding of MtuP49 to template ssDNA. **Conclusion** This study firstly identified the SRS in ssDNA for MtuDnaG binding, the key factors affecting MtuDnaG binding to ssDNA was revealed. The above results provide evidence to shed light on the mechanism of MtuDnaG priming, and pave the way for development of novel DnaG-targeted antituberculosis drugs.

* This work was supported by grants from the Science and Technology Department of Sichuan Province (2022YFSY0028), The National Natural Science Foundation of China (31470742, U1432102, 31270783), and the 100 Talents Program of the Chinese Academy of Sciences.

** Corresponding author.

Tel: 86-28-82890828, E-mail: wanggg@cib.ac.cn

Received: November 13, 2023 Accepted: February 26, 2024

Key words DnaG primase, *Mycobacterium tuberculosis*, specific recognition site, domain rearrangement, DNA replication

DOI: 10.16476/j.pibb.2023.0445

Tuberculosis (TB) caused by *Mycobacterium tuberculosis* infection has been the public health concern—in 2021, there were an estimated of 1.6 million deaths from TB around the world^[1]. Despite that World Health Organization has been adopting End TB Strategy for years, the global control and management of TB is still facing challenges due to the spreading of multi-drug resistant (MDR) and extensively-drug resistant (XDR) strains^[2]. Therefore, it is urgent to develop novel anti-TB drugs^[3-4].

DNA replication is a critical process for the viability of all organisms, which requires the coordination of several macromolecular machines^[5-8]. Since DNA polymerase is unable to produce the daughter strands *de novo*, the synthesis of a short RNA primer by the enzyme primase is essential^[9]. In bacteria, the primase DnaG is composed of an N-terminal zinc-binding domain (ZBD), an RNA polymerase domain (RPD), and a C-terminal helicase-binding domain (HBD)^[10]. Due to the marked differences from its counterpart in humans, in terms of sequence homology, structure, and the primer synthesis mechanism^[11-12], DnaG primase is proposed to be a promising target for the development of novel antibacterial drugs^[10].

Many trials have been carried out on antituberculosis drugs targeting the DnaG primase from *M. tuberculosis* (MtuDnaG)^[13-14]. However, a lack of understanding of mechanism by which MtuDnaG interacts with the template DNA hinders drug discovery efforts. If the interface between MtuDnaG and ssDNA template is known, new inhibitors could be designed to disrupt the interactions between enzyme and ssDNA template. It is known that, prior to initiating the synthesis of primers, bacterial DnaG primase would firstly recognize a specific recognition site (SRS) for achieving the binding with single-stranded DNA (ssDNA) template, which are typically consists of trinucleotide^[15]. Besides, SRSs varies among different bacteria. For example, in *Escherichia coli*, DnaG catalyzes primer synthesis on 5'-CTG-3'. DnaG in *Aquifex aeolicus* preferentially primes 5'-CCC-3', and DnaG of *Bacillus subtilis* dominantly initiates on 5'-CTA-3'^[11]. Yet, the SRS for MtuDnaG is still unknown, and the mechanism by which MtuDnaG recognizes it to

initiate primer synthesis remains unclear^[2,16].

In the present study, the SRS for MtuDnaG was investigated using the MtuDnaG ZBD-RPD didomain (referred to hereafter as MtuP49). Under the optimized condition, we found that MtuP49 showed highly binding affinity to the trinucleotide 5'-GCG/C-3', suggesting that these two sites might be the potential SRS for MtuDnaG. Besides, 5'-GCG/C-3' sites were also identified in the *oriC* region of *M. tuberculosis* genome, whose binding affinity to MtuP49 was significantly affected by the length of 3' sequence flanking the SRS. Mutagenesis studies suggested that ZBD was involved in the binding of template ssDNA. These results shed light on the mechanism by which MtuDnaG recognizes and binds to template ssDNA, paving the way for development of novel DnaG-targeted antituberculosis drugs.

1 Materials and methods

1.1 Cloning and purification of MtuP49

DNA region encoding MtuP49 (residues 1–476 of MtuDnaG) was amplified by PCR from a plasmid encoding full-length DnaG from *M. tuberculosis* H37Rv^[17], with primers 5'-AATGGATCCATGGG-AAATCGCATACCAG-3' and 5'-CCGCTGGAGTTA-GGAATGGACTGTTATATC-3'. The PCR product was cloned into vector pGEX-6p-1 and confirmed by DNA sequencing. Protein expression was carried out in *E. coli* BL21 (DE3) cells. In brief, transformed cells were cultured in Luria-Bertani broth containing 100 mg/L ampicillin at 37°C. When the bacterial culture reached an A_{600} value of about 0.6, 0.5 mmol/L isopropyl β -D-1-thiogalactopyranoside (IPTG) was added, and the cells were further incubated overnight at 18°C. After centrifugation at 4°C, the cell pellet was resuspended in lysis buffer (100 mmol/L phosphate buffer (PB, $\text{NaH}_2\text{PO}_4/\text{Na}_2\text{HPO}_4$), pH 6.0, 500 mmol/L NaCl, 5% (v/v) glycerol, 10 mmol/L MgCl_2 , 1 mmol/L ATP, and 1 mmol/L dithiothreitol (DTT)) and lysed by sonication (200 W for 30 min, 2 s pulse, 2 s pause). After that, the sample was centrifugated at 12 000 r/min for 30 min and the supernatant was applied to Glutathione Sepharose 4B beads (GE Healthcare) equilibrated with lysis buffer. After washing with lysis

buffer, the fusion protein-bound beads were rinsed with hydrolysis buffer (100 mmol/L PB, pH 6.5, 100 mmol/L NaCl, 5% (v/v) glycerol, 10 mmol/L $MgCl_2$, 1 mmol/L ATP, and 1 mmol/L DTT) and incubated with PreScission Protease overnight at 4°C. The eluate was then sequentially purified using SP Sepharose and Superdex 75 gel filtration columns (GE Healthcare) into storage buffer (20 mmol/L PB, pH 6.5, 400 mmol/L NaCl, 10% (v/v) glycerol, 10 mmol/L $MgCl_2$, and 2 mmol/L DTT). The resulting MtuP49 was found to be >95% pure by SDS-PAGE and concentrated to >10 g/L for storage at -80°C.

1.2 Trinucleotide screening by agarose-gel electrophoretic mobility shift assay (EMSA)

DnaG primase recognizes and binds to the specific trinucleotide SRS to initiate primer synthesis, but the SRS for MtuDnaG is not known. To screen it, 64 (*i. e.*, 4³) ssDNA templates characterized as 5'-CAGA(CA)₅-XYZ-(CA)₃-3' were prepared as described previously^[18-19], where 5'-XYZ-3' represents the test trinucleotide. All ssDNA were obtained from Sangon Biotech (Shanghai) Co., Ltd., which was dissolved in ddH₂O before quantification by Nanodrop™ Lite UV spectrophotometer (Thermo, USA).

In order to establish a fast label-free EMSA^[20-22], based the Ultra GelRed staining dye (Nanjing Vazyme Biotech. Co., China) to screen the potential SRS of MtuDnaG, staining method for 23-nt ssDNA detection was firstly determined. As shown in Figure 1a, the ssDNA was more effectively stained by the pre-staining method when using the Ultra GelRed staining dye. Besides, the ability to quantify ssDNA using 1×Ultra GelRed dye was assessed. Band intensities of gradient concentrations of 5'-CAGA(CA)₅-CTG-(CA)₃-3' (CTG) were measured with ImageJ software, and an excellent linearity between the band intensities and concentrations of ssDNA was found when the ssDNA concentration was <30 μmol/L (the relative integrated density (RID) was approximately 15 000 for ssDNA at 30 μmol/L) (Figure 1b). Accordingly, 10 μmol/L of ssDNA was applied for the subsequent screening.

Studies indicated that MtuDnG could maintain its solubility and activity in the presence of DNA and nucleotides^[23]. Therefore, for binding, ssDNA was diluted into the optimized binding buffer (40 mmol/L CAPS, pH 9.0, 400 mmol/L NaCl, 10 mmol/L $MgCl_2$,

1 mmol/L ATP) before adding MtuP49, making a 50 μl reaction system including 10 μmol/L MtuP49 and 5 μmol/L ssDNA. The mixtures were then incubated at 20°C for 30 min. After that, they were analyzed by the established EMSA assay. Briefly, samples above were loaded onto 2% agarose gels containing 1×Ultra GelRed stain, and were resolved by running the gels in 0.5×TB buffer at 20 V/cm for 10 min, with the one without MtuP49 being the blanks. Gels were then imaged using a Gel Doc™ XR⁺ system (Bio-Rad, USA), and the RID for free ssDNA bands was then determined by ImageJ software with a light background with 300 pixels of rolling ball radius. Proportion of free ssDNA to the total ssDNA from the blanks was calculated: the less the free ssDNA detected, the higher the binding affinity of MtuP49 to the ssDNA containing the specific trinucleotide.

1.3 Acquisition of apparent dissociation constant by gradient EMSA

The ssDNA with various binding affinities to MtuP49 in fast EMSA were selected to acquire their apparent dissociation constant (K_D , indicating the protein concentration at 50% free ssDNA). 5 μmol/L of ssDNA were incubated with increasing concentration of MtuP49 (0–27 μmol/L) to form MtuP49/ssDNA complex, then the samples were resolved on gel and the proportions of free ssDNA were measured according to the procedure as described above. Values for K_D were calculated by non-linear fitting of the MtuP49 concentrations and the intensities of the free ssDNA band^[23-25].

1.4 Isothermal titration calorimetry (ITC)

The ssDNA with various binding affinities to MtuP49 above were selected for ITC analysis, using a PEAQ-ITC instrument (Malvern Instruments) at 20°C. MtuP49 was diluted to 20 μmol/L using the binding buffer (40 mmol/L CAPS, pH 9.0, 400 mmol/L NaCl, 10 mmol/L $MgCl_2$, 2 mmol/L ATP) and filtrated through a 0.22 μm filter. The selected ssDNA was diluted to 100 μmol/L. Approximately 40 μl of ssDNA was titrated into MtuP49 within 13 injections, with 3 μl/120 s for each except for the first titration of 0.6 μl/0.8 s. The binding constant (K_d) was calculated by fitting the data using a one-site model.

1.5 Identification of SRS within *oriC* of *M. tuberculosis* genome

In the genome of *M. tuberculosis* H37Rv

(Accession ID: CCP45131), the *oriC* region (nucleotides 1 349 to 2 162) contains an AT-rich cluster (nucleotides 1 548 to 1 562)^[26]. The specific binding sites for MtuDnaG are expected to be adjacent to the AT-rich cluster^[9]. Thus, ssDNA containing the potential SRS around this AT-rich cluster were extracted, which consists of the central SRS and 10-nt flanking sequence on each side. The binding affinity of these ssDNA to MtuP49 was analyzed by EMSA above.

1.6 MtuP49 mutants and binding affinity evaluation

Structure of MtuP49 was predicted by using AlphaFold2^[27] and was superimposed on the structure of the ZBD-RPD didomain of DnaG from *A. aeolicus* (PDB: 2au3), then Zn^{2+} was modelled into the MtuP49 structure. After that, residues on ZBD associated with ssDNA binding were predicted by Graphbind, iDRNA-ITF, and NCBRPred^[28-30]. As results, 4 residues (R31, A32, A34, and S36) were selected for mutagenesis, which was performed by the QuikChange mutagenesis method^[31]. Mutations were

confirmed by DNA sequencing, and mutants were expressed and purified according to the procedures described above. The binding affinity of mutants to ssDNA containing the potential SRS (5'-GGATG-CGAGTGGTGAGTTC-3') was evaluated by the established EMSA.

1.7 Data analysis

Graphing and statistical analysis were performed in GraphPad Prism v5.01 (San Diego, CA, USA), Origin Pro 9 (Origin Lab Corporation, Wellesley Hills, MA, USA), and protein structures were drawn using PyMOL2. Significant differences were evaluated by one-way analysis of variance and the Tukey test, with at least three replicates. Data are expressed as the mean±standard error of the mean (SEM).

2 Results

2.1 SRS screening by EMSA

To screen the potential SRS of MtuDnaG, the staining and quantification of ssDNA by Ultra GelRed dye was optimized. (Figure 1a, b). Moreover, the pH

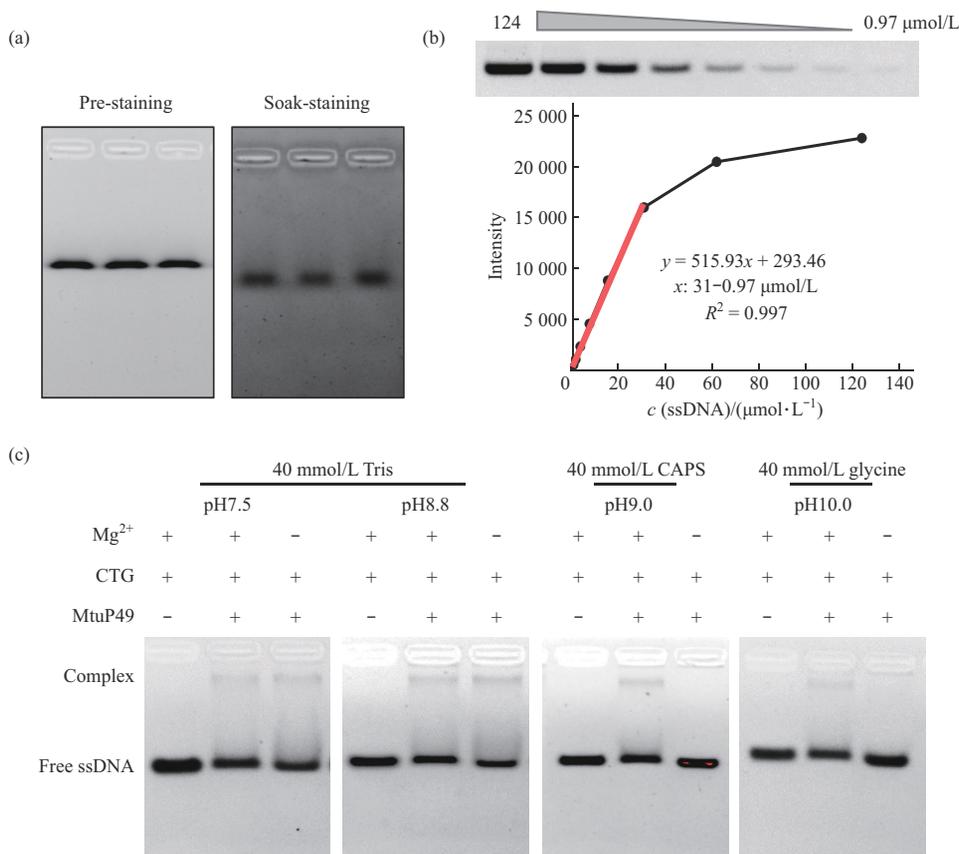


Fig. 1 Establishment of agarose-gel EMSA

(a) Comparison of ssDNA staining by pre-staining and soak-staining methods. (b) Linear fitting between the concentration of ssDNA and the corresponding relative integrated density (RID). (c) Binding behavior between MtuP49 and CTG at different pH.

2.2 Validation by gradient EMSA and ITC assays

To validate the above results, ssDNA containing 5'-GCC-3', 5'-GCG-3', 5'-AGT-3', 5'-GAC-3', 5'-AAA-3',

and 5'-CCC-3' sites were selected for analysis by gradient EMSA and ITC assays. For gradient EMSA, 5 $\mu\text{mol/L}$ of ssDNA was titrated with 0–27 $\mu\text{mol/L}$ MtuP49 (Figure 3). The K_D values for ssDNA

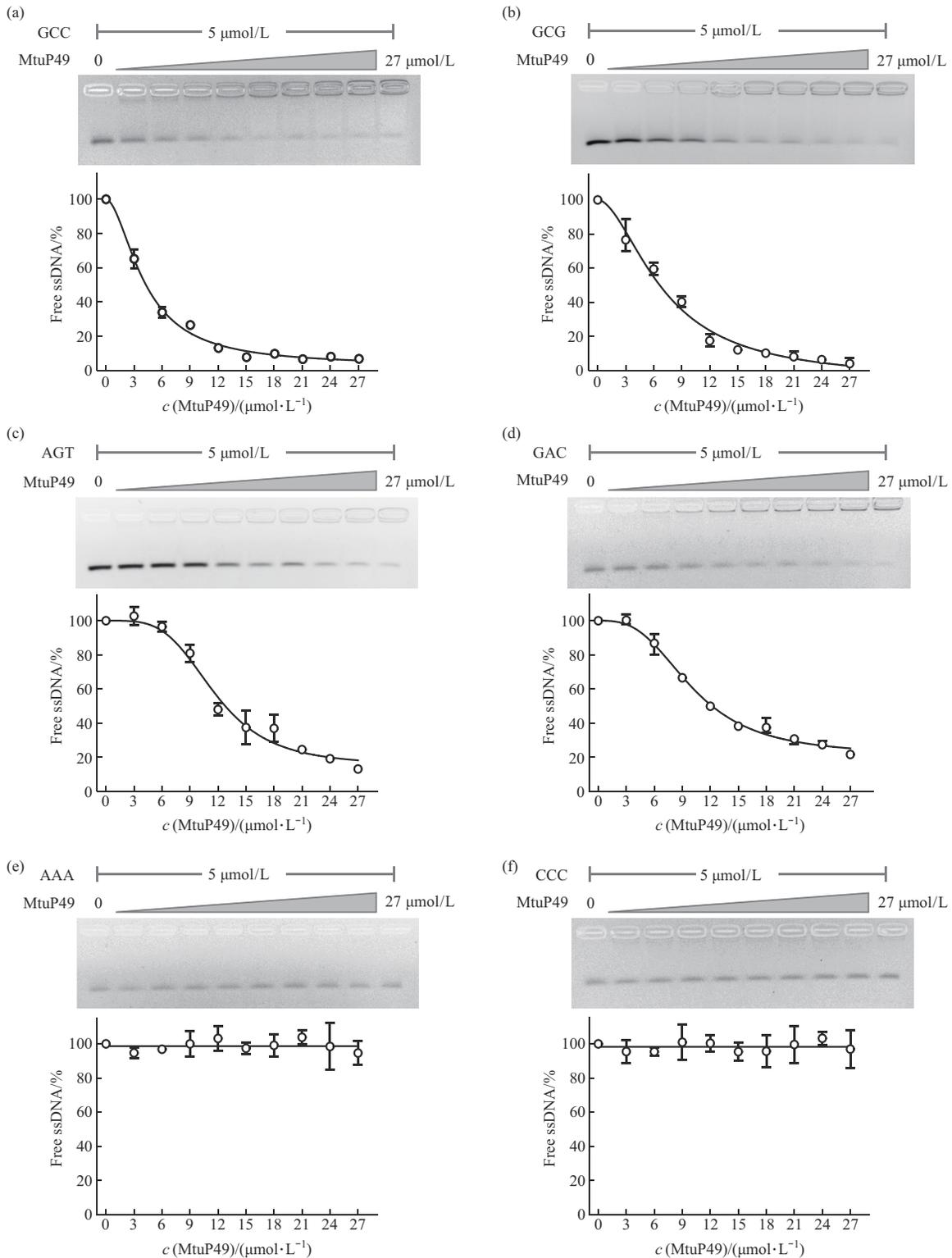


Fig. 3 Gradient EMSA analysis of MtuP49 binding with 6 ssDNAs containing various trinucleotide

(a) 5'-GCC-3' site; (b) 5'-GCG-3' site; (c) 5'-AGT-3' site; (d) 5'-GAC-3' site; (e) 5'-AAA-3' site; (f) 5'-CCC-3' site. The upper panels show representative migration of the ssDNAs in EMSA; the lower panels show nonlinear curve fits to the data, with the proportion of free ssDNA as a function of the concentration of MtuP49.

containing 5'-GCC-3' and 5'-GCG-3' were (4.19 ± 0.18) and (6.68 ± 0.20) $\mu\text{mol/L}$, respectively, lower than those with 5'-AGT-3' ((12.61 ± 0.31) $\mu\text{mol/L}$), and 5'-GAC-3' sites ((11.59 ± 0.25) $\mu\text{mol/L}$) (Table 1). The K_D values for ssDNA containing 5'-AAA-3' and 5'-CCC-3' sites were outside of the measurable experimental range. These data suggested that the affinity of MtuP49 for ssDNA with 5'-GCC-3' and 5'-GCG-3' was higher than that for 5'-AGT-3',

5'-GAC-3', 5'-AAA-3', and 5'-CCC-3'.

For ITC analysis, 20 $\mu\text{mol/L}$ of MtuP49 were titrated with 100 $\mu\text{mol/L}$ of the above ssDNA templates. The ITC curves were shown in Figure 4. The results showed that MtuP49 had higher affinity for ssDNA containing 5'-GCC-3' and 5'-GCG-3' sites than those for the others, as evidenced by significantly lower K_d values ((2.5 ± 0.05) and (2.18 ± 0.11) $\mu\text{mol/L}$, respectively) (Table 1). Taken together, the results of

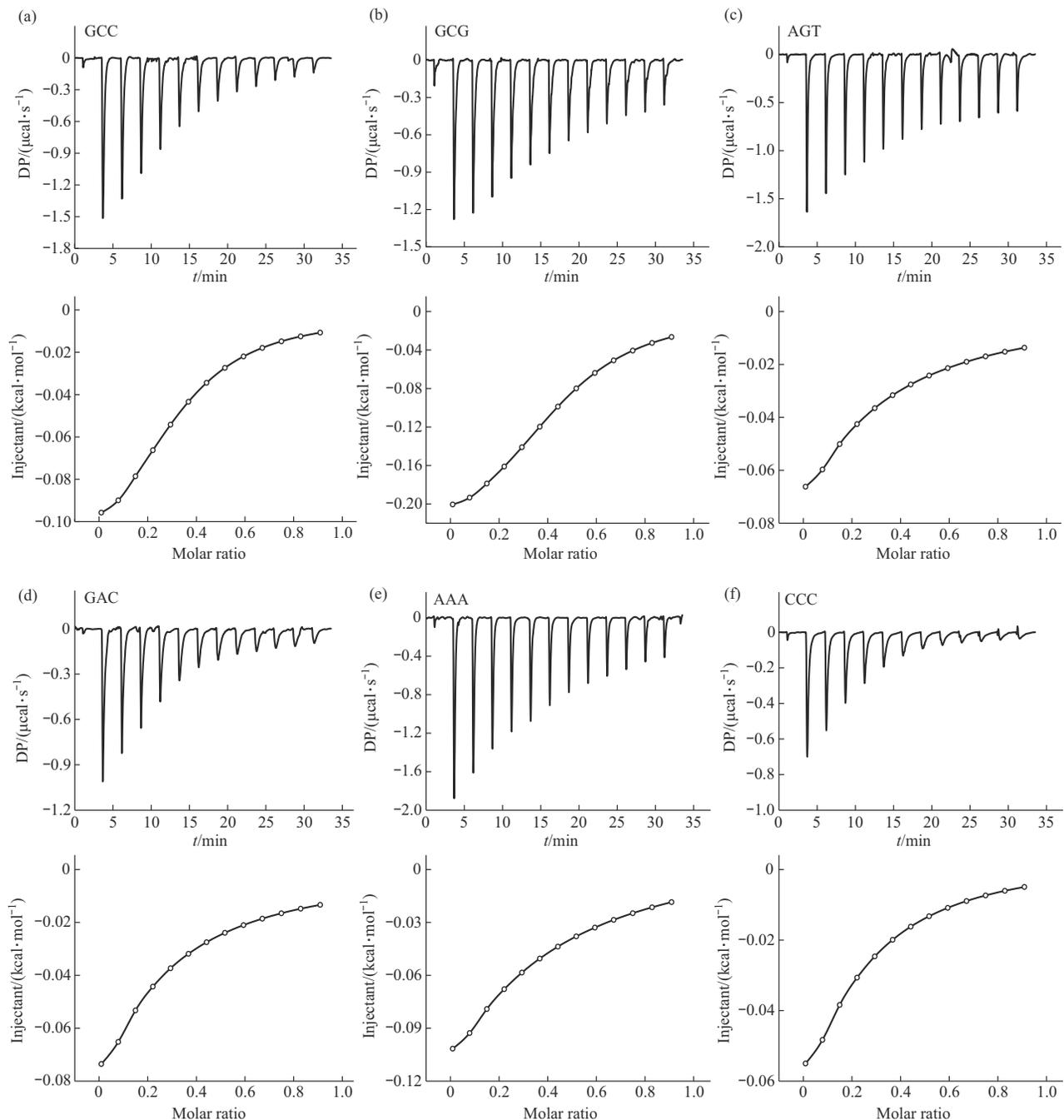


Fig. 4 ITC analysis of MtuP49 binding with 6 ssDNAs containing various trinucleotide

(a) 5'-GCC-3' site; (b) 5'-GCG-3' site; (c) 5'-AGT-3' site; (d) 5'-GAC-3' site; (e) 5'-AAA-3' site; (f) 5'-CCC-3' site. The upper panels show raw titration data from ITC experiments, and the lower panels show curve fits to the results. DP: differential power.

Table 1 Binding affinity of ssDNAs to MtuP49 from EMSA and ITC analysis

	GCC	GCG	AGT	GAC	AAA	CCC
$K_{D, EMSA}/(\mu\text{mol}\cdot\text{L}^{-1})$	4.19±0.18	6.68±0.20	12.61±0.31	11.59±0.25	nd	nd
$K_{d, ITC}/(\mu\text{mol}\cdot\text{L}^{-1})$	2.50±0.05	2.18±0.11	17.00±0.66	13.70±0.23	22.70±0.46	83.4±0.25

nd: K_D acquired from the EMSA is beyond the maximum concentration set experimentally.

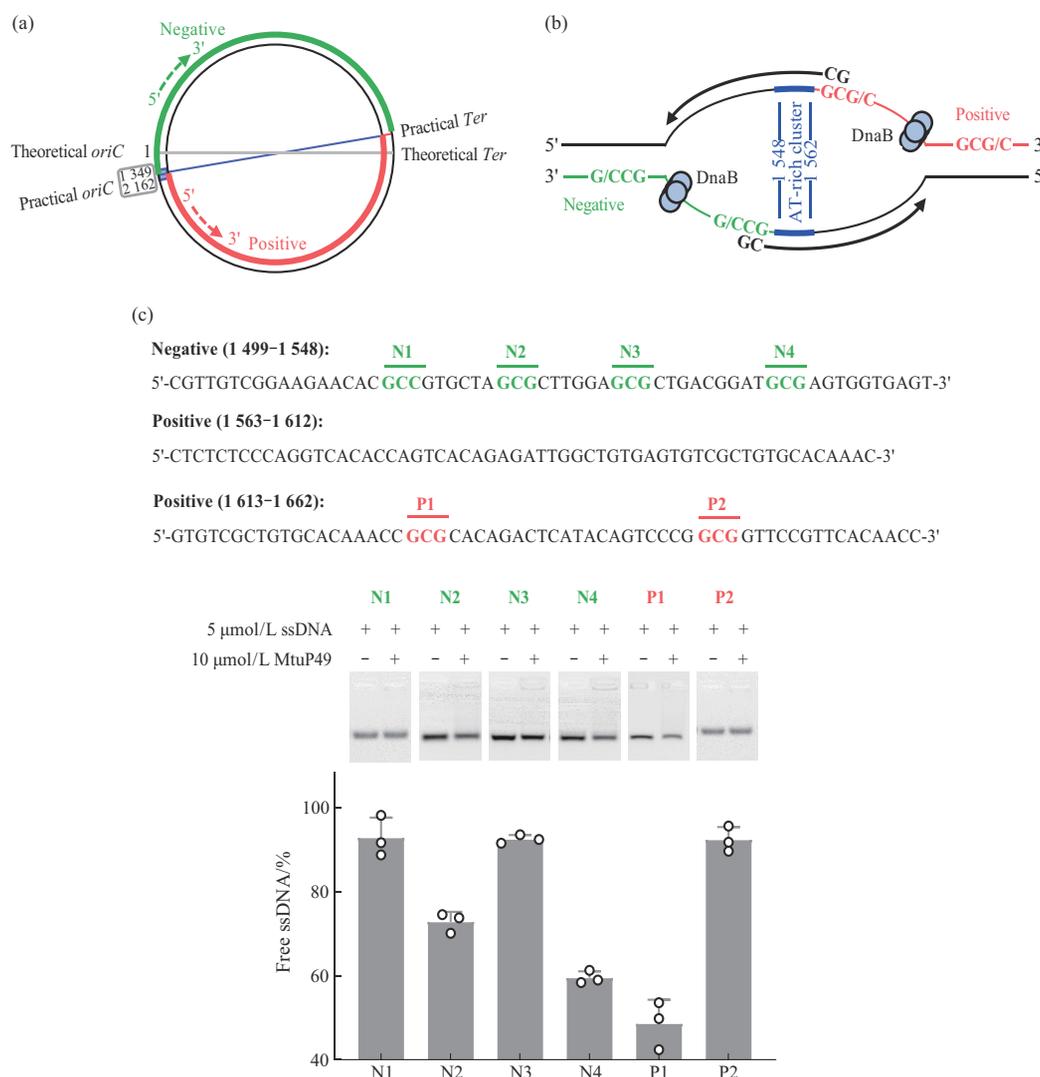
gradient EMSA and ITC were generally consistent with that from the initial screening, and suggested the reliability of rapid EMSA for screening potential SRS of MtuDnaG.

Of note, sequence analysis of ssDNAs with 5'-AGC-3' and 5'-CGC-3' sites indicated the equivalences to ssDNA with 5'-GCC-3' site, since both 5'-AGC-3' and 5'-CGC-3' are followed by (CA)₃ (Figure 2c). Therefore, the potential SRS for

MtuDnaG could be 5'-GCG/C-3'.

2.3 MtuDnaG binding sites in the *oriC* DNA of *M. tuberculosis*

The triple nucleotide site 5'-GCG/C-3' was investigated in the genome of *M. tuberculosis*. The *oriC* region (nucleotides 1 349 to 2 162) in the genome contains an AT-rich cluster (nucleotides 1 548 to 1 562) (Figure 5a). Sequences after nucleotide 1 562 on the positive strand, and before nucleotide

**Fig. 5** MtuDnaG binding sites within the *oriC* DNA of *M. tuberculosis*

(a) Location of *oriC* and *Ter* sites in the circular genome of *M. tuberculosis*. (b) AT-rich cluster and 5'-GCG/C-3' sites on the positive (red) and negative (green) DNA strands in the *oriC* region of the genome. It should be noted that, theoretically, the first synthesized primer may be responsible for replicating the 5'-sequence of *oriC* on both strands (marked in black). (c) Binding affinity of MtuP49 to ssDNAs extracted from genome region adjacent to the AT-rich cluster; ssDNAs P1 and P2 were from the positive strand, while N1-4 were from the negative strand.

1 548 on the negative strand, namely the 3'-sequences of *oriC* on each strand, should be the regions within which SRS are recognized by MtuDnaG (Figure 5b)^[9].

Initially, 23-nt ssDNAs containing 5'-GCG/C-3' sites and a 10-nt flanking sequence on each side were extracted from the region 50-nt adjacent to the AT-rich cluster (Figure 5c). On negative strand (nucleotides 1 499–1 548), 4 qualifying ssDNAs were identified and named N1–4, respectively (Figure 5c). On the positive strand, no 5'-GCG/C-3' site was found in the first 50-nt region (nucleotides 1 563–1 612), while two ssDNAs with 5'-GCG-3' site was found from the next 50-nt region (nucleotides 1 613–1 662), and named P1 and P2 (previous P4 and P6), respectively (Figure 5c). Thus, a total of 6 ssDNAs containing 5'-GCG/C-3' sites (N1–4, P1 & P2) were applied for binding assays by EMSA.

As shown in Figure 5c, MtuP49 bind to N4 with high affinity ((59.58±1.56)% free ssDNA after binding), suggesting that N4 might be the potential SRS on negative strand for MtuDnaG during DNA replication. Besides, MtuP49 also bind strongly to P1, thus, the potential SRS for MtuDnaG on the positive strand was also identified. In addition, the binding affinity of MtuP49 to the 6 ssDNAs template from genome was measured by gradient EMSA. The K_D values for ssDNA containing N4 site and P1 site were (7.18±2.61) and (5.91±0.9) $\mu\text{mol/L}$, respectively, lower than those of others (Figure 6). These results implied that 5'-GCG/C-3' sites in the *M. tuberculosis* genome could be recognized by MtuDnaG.

2.4 Asymmetric effect of flanking sequence on the binding affinity of MtuP49

The data above showed that MtuP49 had variable binding affinity to sequences containing 5'-GCG/C-3' sites in the *oriC* region of *M. tuberculosis* genome (Figure 5c), which might result from the sequences flanking SRS^[32]. Here, N4 was selected to investigate the effect of flanking sequences length on the binding of MtuP49 to ssDNA. As shown in Figure 7a, binding affinity of MtuP49 to N4 was gradually enhanced when the length of 3'-flanking sequence increased, especially from 3-nt to 7-nt; after that, binding affinity showed no obvious changes. In contrast, length of 5'-flanking sequence had no obvious effect on the binding ability of MtuP49 to N4 (Figure 7b). These results suggested an asymmetric effect of flanking sequence on the binding affinity of MtuP49 to ssDNA.

2.5 Residue R31 in the ZBD is crucial for ssDNA binding

It is reported that the ZBD is indispensable for primase initiating primer synthesis^[10], but it is still elusive how the ZBD interacts with a ssDNA template. Here, structure of MtuP49 was firstly predicted by AlphaFold2, and by being superimposed on the structure of ZBD-RPD didomain from *A. aeolicus* DnaG (PDB: 2au3), Zn^{2+} was modeled into it. Then, possible residues on ZBD that may be involved in ssDNA binding were predicted. Among the various residues associated with ssDNA binding given by Graphbind, iDRNA-ITF, and NCBRPred, respectively, 4 residues in ZBD of MtuP49—R31, A32, A34, and S36—were listed by all three predictions, which were selected for further experiments (Figure 8a). Of note, these 4 residues are all located in a positively charged surface patch on ZBD (Figure 8b).

As shown in Figure 8c, MtuP49 mutated in these 4 residues, respectively, showed different binding affinities to ssDNA. Specially, R31E and R31Q variants showed 10%–20% lower affinity for the ssDNA than the wild-type, while the affinity to ssDNA increased 10%–20% for the A32V and A34V variants. And mutations of S36 had no effect on ssDNA binding. Taken together, these results indicated that ZBD is involved in template DNA binding by MtuDnaG, which might be related to the residues located in its positively charged surface patch.

3 Discussion

3.1 DnaG primase in *M. tuberculosis* preferentially binds to 5'-GCG/C-3' sites

DnaG primase plays an essential role in genome maintenance. In *M. tuberculosis*, MtuDnaG is regarded as a target for novel antituberculosis drug discovery^[10]. However, the mechanism of MtuDnaG stands out as relatively underrepresented. Here, 5'-GCG/C-3' sites were screened out from 64 possible triple nucleotides as the potential SRS for MtuDnaG, which were also identified in the *oriC* region of *M. tuberculosis* genome. Yet, MtuP49 showed variable affinity to several ssDNAs containing 5'-GCG/C-3' sites within the *oriC* region of the *M. tuberculosis* genome, indicating that the SRS alone is not sufficient for MtuP49 to recognize and bind to the template ssDNA. Researches on T7 primase have

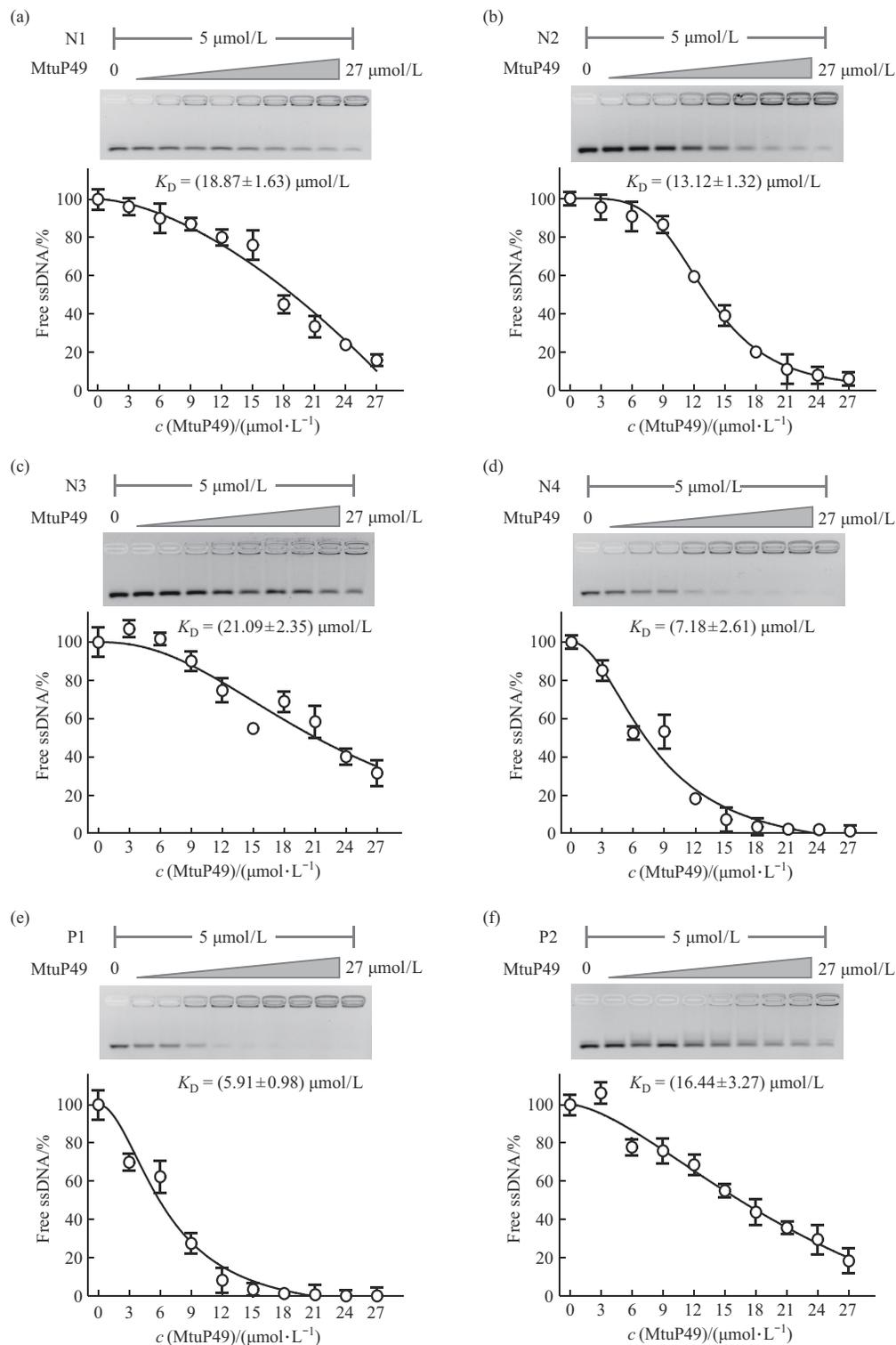


Fig. 6 Gradient EMSA analysis of MtuP49 binding with ssDNAs extracted from genome region adjacent to the AT-rich cluster

(a) N1 site; (b) N2 site; (c) N3 site; (d) N4 site; (e) P1 site; (f) P2 site. The upper panels show representative migration of the ssDNAs in EMSA; the lower panels show nonlinear curve fits to the data, with the proportion of free ssDNA as a function of the concentration of MtuP49.

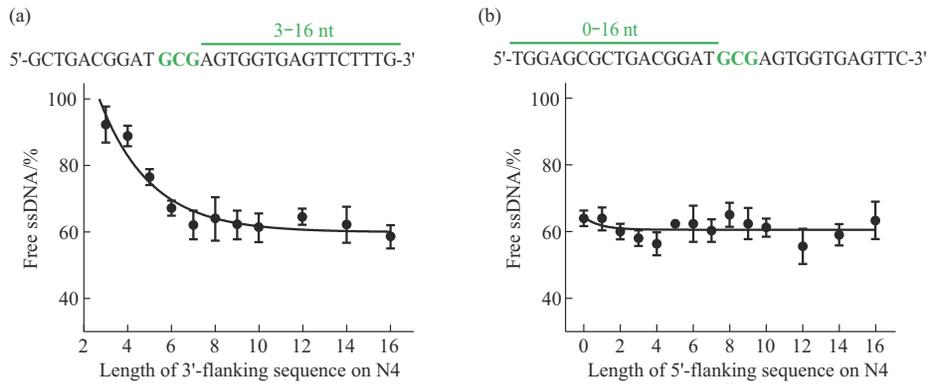


Fig. 7 Influence of flanking sequence length on the binding affinity of MtuP49 to the specific recognition site of ssDNA N4 (a) 3'-flanking sequence; (b) 5'-flanking sequence.

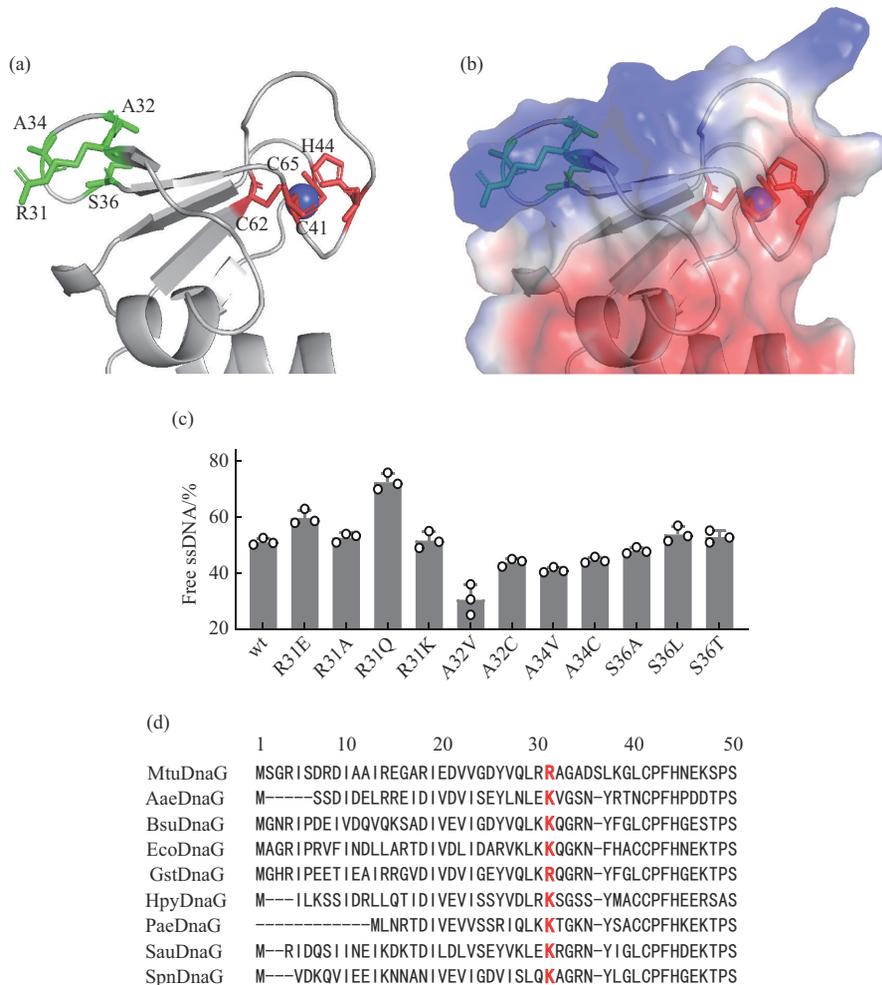


Fig. 8 Residues on ZBD affect the binding affinity of MtuP49 to template ssDNA

(a) Cartoon view of the MtuP49-ZBD predicted by AlphaFold2, the residues predicted for ssDNA binding are marked in green, the residues forming zinc finger were colored in red, and the Zn²⁺ was presented as sphere and colored in blue. (b) Surface electrostatic potential of the ZBD domain. (c) Binding ability of MtuP49 mutants to the ssDNA template (5'-GGATGCGAGTGGTGAGTTC-3'). (d) Sequence alignment of DnaG primases. The alignment includes DnaG from *A. aeolicus* (AaeDnaG), *B. subtilis* (BsuDnaG), *E. coli* (EcoDnaG), *Geobacillus stearothermophilus* (GstDnaG), *Helicobacter pylori* (HpyDnaG), *Pseudomonas aeruginosa* (PaeDnaG), *Staphylococcus aureus* (SauDnaG), and *Streptococcus pneumoniae* (SpnDnaG). Residue 31 (MtuDnaG residue numbering) is highlighted in red.

demonstrated that the flanking sequence of SRS makes undeniable contributions to the interaction between MtuP49 and ssDNA. By using a workflow called high-throughput primase profiling, the G/T-rich flanks was found to significantly increase the binding specificity as well as the functional activity of T7 primase^[32]. Therefore, although the ssDNA from *M. tuberculosis* genome all contains the SRS, the distinct flanking sequences may result in the significant differences in their affinity to MtuP49.

Besides, we found that length of 3'-sequence flanking SRS exhibited a dominant effect on the binding affinity of MtuP49 to ssDNA, as evidenced by the dramatically decreased binding affinity when the flanking sequence was shorter than 7 nt. In contrast, length of 5' flanks of SRS showed no obvious effect on the binding. This asymmetric effect of flanking sequences was also observed in other DNA binding proteins, such as for the transcription factor Egr-1, in which small changes in the DNA sequence fine tune the transcriptional regulation^[33-34]. In DNA replication, the unwound ssDNA is usually covered by a single-strand DNA binding protein and only a limited length of ssDNA adjacent to the DnaB helicase is accessible to DnaG^[9]. Therefore, it seems reasonable that the 3'-flanking sequence of SRS is vital for DnaG binding, while the 5'-flanking sequence determines the nucleotide composition of the synthesized primer. Taken together, the above evidence may imply that during the DnaG-ssDNA interaction, SRS of the template ssDNA is the core motif, while sequences flanking it are the auxillary elements, both of which are essential to set fine tune on DnaG-ssDNA binding.

Of note, the interaction between DnaG primase and ssDNA template seems to be relatively weak. In fact, the ssDNA binding affinity of primases from T7 phage & human *in vitro* are also in the micromolar range, and the full-length primases bind DNA more tightly than either the truncated fragment or the subunit^[34-35]. Due to this, the coordination with partners may enhance the DNA binding affinity of primases in physiological process. For example, in human, primase is a heterodimeric enzyme containing an RNA polymerase domain in the small subunit (p48) and a [4Fe4S] cluster-containing C-terminal domain of the large subunit (p58C)^[36]. The p48/p58C binding of ssDNA is also modulated by a redox switch in the [4Fe4S] cluster. The regulations of other

factors will facilitate more accurate priming process^[35].

3.2 Rearrangement of ZBD is involved in template ssDNA binding

It is believed that the integrity of ZBD and RPD domains is essential for DnaG primase binding to template ssDNA^[11]. While numerous studies are focused on RPD, as well as collaborations with substrates such as ssDNA and NTPs, few attentions are paid on ZBD. For example, previous studies showed that template ssDNA might bind to the RPD of DnaG primase in an "L" shape, and the nucleoside triphosphates (NTPs) were incorporated in the catalytic center of RPD^[11, 37-38]. Here, we found that substitution of residue R31 of ZBD significantly affected the affinity of MtuP49 to ssDNA, suggesting its role in DnaG-ssDNA interaction. Besides, sequence alignment illustrated that R31 is highly conserved in DnaG primases (Figure 8d), indicating that the role of ZBD in DnaG-ssDNA interaction may be universal. However, in the predicted structure of MtuP49, residue R31 is in a positively charged surface patch and distant from the catalytic center on the RPD, implying that ZBD might undergo conformational change to be accessible to the template ssDNA binding in the catalytic center. This is consistent with results from small-angle X-ray scattering analysis of *Bacillus subtilis* DnaG primases, where continuous conformational change of ZBD and HBD relative to RPD was observed^[39]. Besides, for the ZBD-RPD didomain in *A. aeolicus* DnaG, transition between an extended form and a compact state was reported^[40]. Moreover, in the structure of the DnaB/ssDNA complex, compared with apo DnaB structure, the C-terminal domain of DnaB rotated about 40° to be accessible to the ssDNA^[41]. Therefore, structural rearrangement of the ZBD of DnaG may facilitate its interaction with template ssDNA.

Notably, previous study proposed an intermolecular interaction between ZBD and RPD, which controls the initiation frequency, processivity, and Okazaki fragment synthesis^[40]. However, the helical structure of DnaB/ssDNA complex suggested that the dynamic DnaB helicase might couple with only one DnaG primase at the replication fork^[41]. These emphasize that the coordination between ZBD and RPD might occur within the same DnaG primase molecule, involving an intramolecular domain

rearrangement.

4 Conclusion

The present study characterized the interaction between MtuDnaG and template ssDNA, from the aspects of ssDNA sequence and ZBD structural characteristics. The primase MtuDnaG specifically recognized the 5'-GCG/C-3' sites in ssDNA template and the ZBD domain coordinated with the RPD domain in the initiation of priming. Further studies, especially on function and structure of MtuDnaG incorporated with ssDNA template, are required to demonstrate the mechanism of MtuDnaG priming, which could make further contributions to the development of antituberculosis drugs targeting MtuDnaG.

References

- [1] World Health Organization. Global Tuberculosis Report 2022. Geneva: World Health Organization, 2022. <https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2022>
- [2] Hou C, Biswas T, Tsodikov O V. Structures of the catalytic domain of bacterial primase DnaG in complexes with DNA provide insight into key priming events. *Biochemistry*, 2018, **57**(14): 2084-2093
- [3] Sotgiu G, Centis R, D'Ambrosio L, *et al.* Tuberculosis treatment and drug regimens. *Cold Spring Harb Perspect Med*, 2015, **5**(5): a017822
- [4] Fernandes G F S, Thompson A M, Castagnolo D, *et al.* Tuberculosis drug discovery: challenges and new horizons. *J Med Chem*, 2022, **65**(11): 7489-7531
- [5] Duderstadt K E, Berger J M. A structural framework for replication origin opening by AAA+ initiation factors. *Curr Opin Struct Biol*, 2013, **23**(1): 144-153
- [6] Leonard A C, Grimwade J E. The orisome: structure and function. *Front Microbiol*, 2015, **6**: 545
- [7] Chodavarapu S, Kaguni J M. Replication initiation in bacteria. *Enzymes*, 2016, **39**: 1-30
- [8] Katayama T, Kasho K, Kawakami H. The DnaA cycle in *Escherichia coli*: activation, function and inactivation of the initiator protein. *Front Microbiol*, 2017, **8**: 2496
- [9] Oakley A J. A structural view of bacterial DNA replication. *Protein Sci*, 2019, **28**(6): 990-1004
- [10] Ilic S, Cohen S, Singh M, *et al.* DnaG primase-a target for the development of novel antibacterial agents. *Antibiotics*, 2018, **7**(3): 72
- [11] Zhou Y, Luo H, Liu Z, *et al.* Structural insight into the specific DNA template binding to DnaG primase in bacteria. *Sci Rep*, 2017, **7**(1): 659
- [12] Greci M D, Bell S D. Archaeal DNA replication. *Annu Rev Microbiol*, 2020, **74**: 65-80
- [13] Singh M, Ilic S, Tam B, *et al.* Dual-acting small-molecule inhibitors targeting mycobacterial DNA replication. *Chemistry*, 2020, **26**(47): 10849-10860
- [14] Hakeem S, Singh I, Sharma P, *et al.* *In silico* screening and molecular dynamics simulations study to identify novel potent inhibitors against *Mycobacterium tuberculosis* DnaG primase. *Acta Trop*, 2019, **199**: 105154
- [15] Frick D N, Richardson C C. DNA primases. *Annu Rev Biochem*, 2001, **70**: 39-80
- [16] Sharma D P, Vijayan R, Rehman S A A, *et al.* Structural insights into the interaction of helicase and primase in *Mycobacterium tuberculosis*. *Biochem J*, 2018, **475**(21): 3493-3509
- [17] 刘文林, 罗昊, 罗爽, 等. 结核分枝杆菌DNA复制解旋酶与引物酶的相互作用. *应用与环境生物学报*, 2019, **25**(6): 1486-1491
Liu W L, Luo H, Luo S, *et al.* *Chin J Appl Environ Biol*, 2019, **25**(6): 1486-1491
- [18] Larson M A, Bressani R, Sayood K, *et al.* Hyperthermophilic *Aquifex aeolicus* initiates primer synthesis on a limited set of trinucleotides comprised of cytosines and guanines. *Nucleic Acids Res*, 2008, **36**(16): 5260-5269
- [19] Swart J R, Griep M A. Primase from *Escherichia coli* primes single-stranded templates in the absence of single-stranded DNA-binding protein or other auxiliary proteins. Template sequence requirements based on the bacteriophage G4 complementary strand origin and Okazaki fragment initiation sites. *J Biol Chem*, 1993, **268**(17): 12970-12976
- [20] Haines A M, Tobe S S, Kobus H J, *et al.* Properties of nucleic acid staining dyes used in gel electrophoresis. *Electrophoresis*, 2015, **36**(6): 941-944
- [21] Sapia R J, Campbell C, Reed A J, *et al.* Interaction of GelRed™ with single-stranded DNA oligonucleotides: preferential binding to thymine-rich sequences. *Dyes Pigm*, 2021, **188**: 109209
- [22] Ream J A, Lewis L K, Lewis K A. Rapid agarose gel electrophoretic mobility shift assay for quantitating protein: RNA interactions. *Anal Biochem*, 2016, **511**: 36-41
- [23] Biswas T, Resto-Roldán E, Sawyer S K, *et al.* A novel non-radioactive primase-pyrophosphatase activity assay and its application to the discovery of inhibitors of *Mycobacterium tuberculosis* primase DnaG. *Nucleic Acids Res*, 2013, **41**(4): e56
- [24] Traczyk A, Liew C W, Gill D J, *et al.* Structural basis of G-quadruplex DNA recognition by the yeast telomeric protein Rap1. *Nucleic Acids Res*, 2020, **48**(8): 4562-4571
- [25] Seo M, Lei L, Egli M. Label-free electrophoretic mobility shift assay (EMSA) for measuring dissociation constants of protein-RNA complexes. *Curr Protoc Nucleic Acid Chem*, 2019, **76**(1): e70
- [26] Qin M H, Madiraju M V, Rajagopalan M. Characterization of the functional replication origin of *Mycobacterium tuberculosis*. *Gene*, 1999, **233**(1-2): 121-130
- [27] Mirdita M, Schütze K, Moriwaki Y, *et al.* ColabFold: making

- protein folding accessible to all. *Nat Methods*, 2022, **19**(6): 679-682
- [28] Xia Y, Xia C Q, Pan X, *et al.* GraphBind: protein structural context embedded rules learned by hierarchical graph neural networks for recognizing nucleic-acid-binding residues. *Nucleic Acids Res*, 2021, **49**(9): e51
- [29] Wang N, Yan K, Zhang J, *et al.* iDRNA-ITF: identifying DNA- and RNA-binding residues in proteins based on induction and transfer framework. *Brief Bioinform*, 2022, **23**(4): bbac236
- [30] Zhang J, Chen Q, Liu B. NCBRPred: predicting nucleic acid binding residues in proteins based on multilabel learning. *Brief Bioinform*, 2021, **22**(5): bbaa397
- [31] Munteanu B, Braun M, Boonrod K. Improvement of PCR reaction conditions for site-directed mutagenesis of big plasmids. *J Zhejiang Univ Sci B*, 2012, **13**(4): 244-247
- [32] Afek A, Ilic S, Horton J, *et al.* DNA sequence context controls the binding and processivity of the T7 DNA primase. *iScience*, 2018, **2**: 141-147
- [33] Rudnizky S, Khamis H, Malik O, *et al.* Single-molecule DNA unzipping reveals asymmetric modulation of a transcription factor by its binding site sequence and context. *Nucleic Acids Res*, 2018, **46**(3): 1513-1524
- [34] Lee S J, Zhu B, Akabayov B, *et al.* Zinc-binding domain of the bacteriophage T7 DNA primase modulates binding to the DNA template. *J Biol Chem*, 2012, **287**(46): 39030-39040
- [35] O'Brien E, Holt M E, Thompson M K, *et al.* The [4Fe4S] cluster of human DNA primase functions as a redox switch using DNA charge transport. *Science*, 2017, **355**(6327): eaag1789
- [36] O'Brien E, Holt M E, Salay L E, *et al.* Substrate binding regulates redox signaling in human DNA primase. *J Am Chem Soc*, 2018, **140**(49): 17153-17162
- [37] Rymer R U, Solorio F A, Tehranchi A K, *et al.* Binding mechanism of metal · NTP substrates and stringent-response alarmones to bacterial DnaG-type primases. *Structure*, 2012, **20**(9): 1478-1489
- [38] Corn J E, Pelton J G, Berger J M. Identification of a DNA primase template tracking site redefines the geometry of primer synthesis. *Nat Struct Mol Biol*, 2008, **15**(2): 163-169
- [39] Luo H, Liu W L, Zhou Y Q, *et al.* The study on the solution structure of full length primase from *Bacillus subtilis*. *Prog Biochem Biophys*, 2019, **46**(11): 1101-1109
- [40] Corn J E, Pease P J, Hura G L, *et al.* Crosstalk between primase subunits can act to regulate primer synthesis in trans. *Mol Cell*, 2005, **20**(3): 391-401
- [41] Itsathitphaisarn O, Wing R A, Eliason W K, *et al.* The hexameric helicase DnaB adopts a nonplanar conformation during translocation. *Cell*, 2012, **151**(2): 267-277

结核分枝杆菌DnaG引物酶与ssDNA模板相互作用的研究*

陈江^{1,2,3)} 罗昊^{1,3)} 张知明^{1,3)} 宋旭²⁾ 王刚刚^{1)**}

¹⁾ 中国科学院成都生物研究所四川省环境微生物重点实验室, 环境与应用微生物学重点实验室, 成都 610041;

²⁾ 四川大学生命科学学院, 成都 610064; ³⁾ 中国科学院大学, 北京 100049)

摘要 目的 结核分枝杆菌 (*Mycobacterium tuberculosis*) 引物酶 DnaG (MtuDnaG) 在其基因组 DNA 复制中起着至关重要的作用, 因而被认为是抗结核药物研发的新靶点。然而 MtuDnaG 起始引物合成的机制尚不清楚, 这阻碍了 MtuDnaG 抑制剂的筛选。本研究将鉴定 MtuDnaG 结合模板的特异性识别位点, 探讨 MtuDnaG 与 ssDNA 模板之间的相互作用。**方法** 本研究以 MtuDnaG 的双结构域蛋白 MtuP49 (包含了锌指结合结构域和 RNA 聚合酶结构域) 为研究对象, 利用生物化学和生物物理学方法, 研究 MtuDnaG 与含不同三联体的 ssDNA 模板之间的相互作用, 鉴定 MtuDnaG 的特异性识别位点。**结果** 5'-GCG/C-3' 三联体可能是 MtuDnaG 结合模板 ssDNA 的特异性识别位点。此外, 在结核分枝杆菌基因组的复制起始点附近也存在可以与 MtuP49 特异结合的 5'-GCG/C-3' 位点, 其 3' 端侧翼序列显著影响 ssDNA 与 MtuP49 的亲合力。突变实验表明, 位于锌指结合结构域中的 Arg31 对 MtuP49 结合 ssDNA 的活性具有重要贡献。基于预测的 MtuP49 结构, 推测在 MtuP49 结合模板 ssDNA 过程中, 锌指结合结构域会发生分子内重排。**结论** 本研究首次鉴定了 MtuDnaG 结合模板 ssDNA 的特异性识别位点, 揭示了影响 MtuDnaG 与模板 ssDNA 相互作用的主要因素。本文的研究结果不但有助于阐明 MtuDnaG 起始引物合成的机制, 也为靶向 DnaG 的新型抗结核药物开发提供了新的信息。

关键词 引物酶 DnaG, 结核分枝杆菌, 特异性识别位点, 结构域重排, DNA 复制

中图分类号 Q71, Q937

DOI: 10.16476/j.pibb.2023.0445

* 四川省科技厅 (2022YFSY0028), 国家自然科学基金 (31470742, U1432102, 31270783) 和中国科学院百人计划资助项目。

** 通讯联系人。

Tel: 028-82890828, E-mail: wanggg@cib.ac.cn

收稿日期: 2023-11-13, 接受日期: 2024-02-26