The replication machinery of Clostridium difficile:a potential target for novel antimicrobials

Eijk, H.W. van

Chapter 5

Primase (CD1454) stimulates the DNA-unwinding activity of the replicative helicase (CD3657) of *Clostridium difficile*

Erika van Eijk [1]
Vasileios Paschalis [2]
Matthew Green [2]
Marilynn A. Larson [3,4]
Keith Spriggs [5]
Annemieke Friggen [1]
Panos Soultanas [2]
Wiep Klaas Smits [1]

1 Department of Medical Microbiology, Leiden University Medical Center, Leiden, the Netherlands;

2 School of Chemistry, Center for Biomolecular Sciences, University of Nottingham, United Kingdom;

3 Department of Microbiology and Pathology, University of Nebraska Medical Center, Omaha, NE 68198-6495;

4 University of Nebraska Center for Biosecurity, Omaha, NE 68198-4080;

5 School of Psychology, University of Nottingham, United Kingdom.

This chapter is published as part of van Eijk *et al.* Open Biology, 2016

Abstract

Loading of helicase is an essential step in the DNA replication process of bacteria. In the Enterobacteria, Firmicutes and Aquificae this event is facilitated by a specific loader protein, which is not conserved in bacteria outside these phyla. However, the strategy of helicase loading among bacteria that do encode a loader protein also differs. In *E. coli*, self-loading of a hexameric helicase is observed, though loading is more efficient in the presence of the loader ATPase protein. In the Gram-positive model *B. subtilis* loading of helicase from monomers occurs by the concerted actions of three proteins that includes the loader ATPase. However, in an *in vitro* system helicase activity can be observed through the action of the helicase and the loader ATPase protein alone.

Although *C. difficile* belongs to the Firmicutes like *B. subtilis*, our previous experiments have shown that loading of the helicase likely differs critically from this organism. It appears that another factor other than the putative loader ATPase is needed to reconstitute activity of the helicase.

Primase and helicase have a bidirectional effect on each other's activities in several bacteria. Here, we show that primase activates the replicative helicase and that this action is limited in the presence of putative loader protein *in vitro*. Helicase assays with Walker A mutant loader proteins suggest that ATP-binding and/or hydrolysis is required for "locking" the helicase activity by the loader protein. We find that primase preferentially initiates on CCC triplets similar to primase of the hyperthermophilic Gram-negative rod, *Aquifex aeolicus* and that the interaction of helicase with primase enhances priming at non-preferred triplets.

Thus, the interaction between helicase and primase constitutes a crucial aspect of DNA replication initiation in *C. difficile*.

Background

DNA replication in bacteria involves interactions between many different proteins. In *Escherichia coli*, for instance, it is estimated that thirty proteins are required for the replication of its chromosome [1]. Loading and activation of the helicase is a key event in replication and proteins that enable these processes can differ between species. In the Gram-negative bacterium *E. coli*, pre-formed hexamers of helicase protein (*Ec*DnaB) are capable of self-loading onto single-stranded (ss) DNA. They display *in vitro* translocation and unwinding activities, which are highly induced in the presence of a loader protein (*Ec*DnaC) [2]. This is in contrast with the Gram-positive bacterium *B. subtilis* (*Bs*DnaC), where pre-assembled hexameric helicase is inactive, irrespective of the presence of the loader protein. *In vitro*, *B. subtilis* helicase activity is only observed when the helicase protein is monomeric and the loader protein (*Bs*DnaI) is present [3]. *In vivo*, loading of the helicase in *B. subtilis* requires two additional proteins, *Bs*DnaB and *Bs*DnaD [4-6].

It has been proposed that *E. coli* helicase is loaded onto the ssDNA as a pre-formed hexameric ring, assorting it to the ring-breaker class, whereas *B. subtilis* helicase is assembled onto the DNA from monomers by its cognate loader protein in a ring-maker fashion [6-8]. Our organism of interest, *Clostridium difficile*, belongs to the Firmicutes, like *B. subtilis*. It codes for homologues of *Bs*DnaI (CD3654) and *Bs*DnaD (CD3653), but a homologue of *Bs*DnaB could not be identified in a BLAST analysis (**Chapter 4**). This suggests that loading of helicase may occur in a different manner. Indeed, our previous experiments suggest that another factor than CD3654 is needed to load and/or activate the replicative helicase CD3657 *in vitro* (**Chapter 4**).

The multi-protein primosome consists not only of helicase loader protein and helicase, but also of primase [9,10]. Primase is pivotal for the initiation of DNA-synthesis at the replication origin and remains of utmost importance during the DNA-replication process in restarting stalled replication forks as well as *de novo* priming of Okazaki fragments for lagging strand synthesis [11]. Prokaryotic primases have a three-domain structure consisting of an N-terminal zinc-binding domain (ZBD), a central RNA polymerase domain that catalyses the polymerization of ribonucleotides, and a C-terminal domain that either is responsible for the interaction with helicase (helicase interaction domain) or has helicase activity itself [12].

The latter region, also known as P16 (reflecting the approximate mass of 16 kDa), is variable in prokaryotes and seems to be crucial for direct interaction with and activation of helicase in *E. coli* [13-16]. Interestingly, the P16 domain is structurally and functionally homologous to the N-terminal domain of the replicative helicase to which it binds [14,17,18]. Depending on the bacterial species, the interaction between helicase and primase can be either transient (*E. coli*) or stable (*Geobacillus stearothermophilus*) [13,19].

Primase and helicase affect each other's activities in *E. coli* [1,13,16,19,20]. Helicase affects primase by modulating initiation specificity, stimulating primer synthesis, reducing length of primers synthesized and increasing its affinity for single stranded DNA [1,12,16,20-25]. Although increased activity of primase in the presence of helicase was shown for Firmicutes such as *Staphylococcus aureus* and *B. subtilis*, primer length was minimally or not altered in either organism [26-28]. Conversely, ATPase and helicase activities of helicase in *E. coli* are stimulated by primase due to stabilization of the helicase hexamer by this interaction [13,29-32]. In another Gram-negative bacterium, *Helicobacter pylori*, it was shown that interaction of primase with helicase resulted in dissociation of the double hexamer of helicase, thereby increasing ATP hydrolysis, DNA binding and unwinding [33].

Although no interaction of primase with loader protein has been demonstrated to date, it has been suggested that primase affects the interaction between helicase and the loader protein in *E. coli* [34]. Dissociation of loader protein from the C-terminal region of helicase in this organism is thought to be induced by primer synthesis and conformational changes resulting from primase-helicase interaction [34]. Similar observations were made in *B. subtilis*, where the loader protein was found to dissociate from a complex when primase and polymerase bind to helicase in gel-filtration experiments [28]. However, a ternary complex comprising of helicase, loader, and helicase binding domain of primase in *G. stearothermophilus* is capable of loading. These observations indicate that primosome-formation, like helicase loading, may also be species-specific [34-36].

In order to investigate the role of primase in DNA replication initiation in *C. difficile*, we identified CD1454 as the primase. Primase has a stimulatory effect on helicase activity that appears dampened in the presence of the putative loader protein, depending on the conserved ATPase motif of the loader. Primase preferen-

tially initiates on CCC triplets, like the primase from the hyperthermophilic Gram-negative rod, *Aquifex aeolicus* [37], and demonstrates more efficient initiation on the non-preferred CTA triplet in the presence of helicase. Together, our results show that interactions between helicase and primase are crucial for the functionality of both proteins in *C. difficile*.

Results

Identification of primase

A BLASTP query of the *C. difficile* genome (GenBank AM180355.1) allowed the identification of a protein homologous to *B. subtilis* primase BsDnaG (GenBank NC_009089.1). This protein, CD1454, shared 31 percent identity with its *B. subtilis* counterpart across the full length of the protein (*e*-value = 9×10^{-98}), and contains all domains expected for a primase protein (**Figure 1A**). Thus, *cd1454* probably encodes the primase for chromosomal DNA replication. We purified the *C. difficile* primase without an affinity tag to >95 percent purity. SEC-MALS analysis showed a major peak corresponding to a molecular weight (MW) of 72.9 kDa, indicating that the primase is monomeric under the conditions tested (expected MW 70 kDa) (**Figure 1B**).

C. difficile helicase is activated by primase

In previous experiments, we did not observe *in vitro* helicase activity with a combination of *C. difficile* helicase and loader ATPase protein, in contrast to what has been shown for the *B. subtilis* proteins. This suggested that another factor was needed to reconstitute the activity of helicase under these conditions (**Chapter 4**). In *E. coli*, helicase activity is stimulated by primase, presumably as a result of increased stability of the hexameric helicase secondary to the interaction between these two proteins. Therefore, we wanted to investigate if *C. difficile* helicase could be activated by primase.

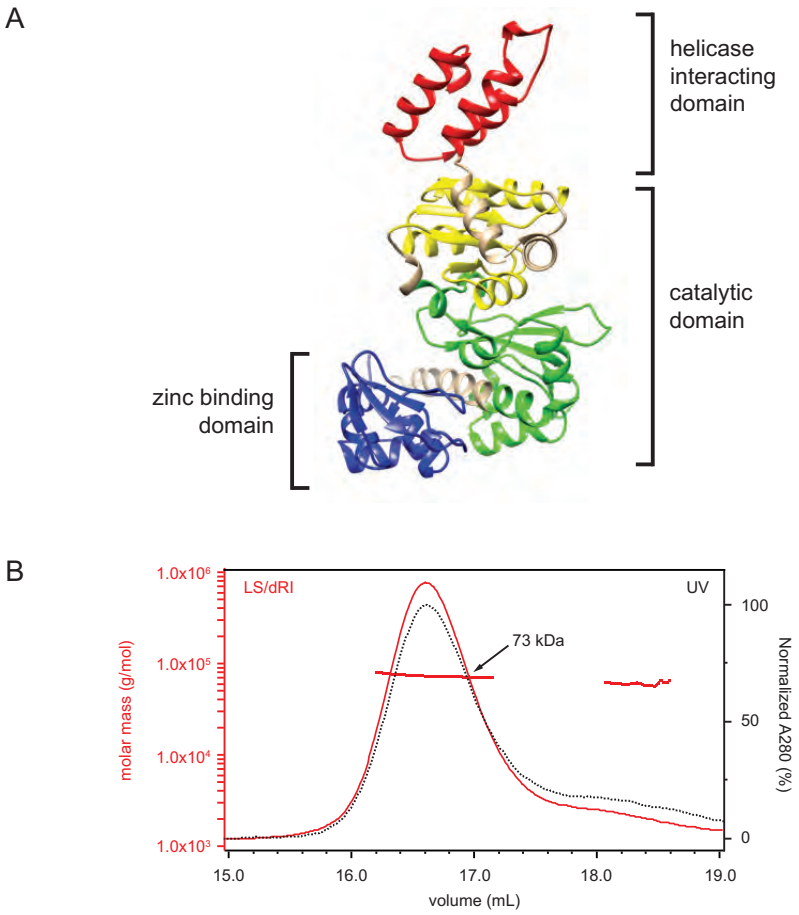


Figure 1. CD1454 is the *C. difficile* primase.

A. Phyre2 model of the CD1454 primase based on PDB 2AU3. Domains are coloured as follows. *Zinc finger* (blue; zf-CHC2/PFAM01807), *catalytic domain* (green: TOPRIM_N/PFAM8275; yellow: TOPRIM_DnaG primases/cd03364) and *helicase interacting domain* (red: DnaB_bind; PFAM10410).

B. SEC-MALS analysis of the CD1454 primase protein. Indicated molecular mass is calculated based on light scattering and dRI signals at the peak of the curve (Mp 72.9 kDa/ 0.5% uncertainty, Mw 74.1 kDa/0.9% uncertainty). The molecular mass indicates a monomeric form of the protein.

Helicase activity was assessed by quantifying the displacement of a radio-labelled (γ^{32}P-ATP) oligonucleotide (partially) annealed to the single stranded circular DNA (ssM13mp18). To enable loading of helicase, the 5' end of the oligonucleotide is a poly(dCA) tail thereby producing a forked substrate upon annealing of the complementary region to ssM13. First, we determined if the CD3657 helicase displayed activity in this particular assay in the absence of any other proteins. Approximately 20 percent of the oligonucleotide was displaced after as little as 2 minutes of incubation, and this fraction remained stable over the course of 35 minutes (**Figure 2A**). This suggests that helicase may be capable of self-loading but displays marginal DNA unwinding activity by itself.

When the CD1454 primase and CD3657 helicase were combined in equimolar amounts, a 3.5- fold increase in displacement of the oligo (up to ~80 percent) was observed after 35 minutes (**Figure 2A**). This indicates that primase has a profound stimulatory effect on helicase activity in this assay. Strikingly, strand displacement by helicase seems inhibited in the presence of primase compared to helicase alone at early time points (less than 10 minutes) (**Figure 2A**). This suggests that primase may inhibit self-loading, in addition to its role as an activator of helicase activity. Control experiments using only primase did not result in significant displacement of the oligonucleotide, demonstrating that the displacement is not the result of an inherent property of the primase protein (*data not shown*).

The results from the helicase assay suggest a functional interaction between the helicase and primase proteins. We therefore tried to validate the interaction using a bacterial two-hybrid system. Despite of our efforts, no interaction could be demonstrated between full-length primase and helicase, or the helicase interacting domain of primase and helicase in bacterial two-hybrid experiments (*data not shown*). This suggests that the interaction between primase and helicase may be very weak (below detection limit of the assay) and/or transient, concordant with observations in *E. coli* [19], or that the bacterial two-hybrid system does not allow recapitulation of the conditions necessary for the interaction between the helicase and primase proteins.

The putative loader protein CD3654 'locks' the CD3657 helicase

Next, we set out to investigate the role of the putative loader CD3654 on the helicase activity of CD3657 in the presence and absence of primase. Consistent with our previous findings (**Chapter 4**), we observed a very low level of displacement (<20 percent) in our helicase assay in the presence of both helicase and the putative loader (**Figure 2B**). Notably, within the first 10 minutes of the assay the fraction of displaced oligonucleotide did not exceed 10 percent, in contrast with the situation with helicase alone where it reached 20 percent (**Figure 2A**). At end point, the fraction was comparable between the two conditions. This suggests that – at least at early time points – the presence of the putative loader negatively affects helicase activity.

We wondered if the positive effect of the CD1454 primase on CD3657 helicase activity could be observed in the presence of both the putative loader ATPase CD3654 and helicase. From 8 minutes on, a clear stimulation of helicase activity by primase was observed, reaching ~60 percent displacement over a time course of 35 minutes (**Figure 2B**). Interestingly, the addition of the putative loader protein resulted in a 20 percent reduction in strand displacement over a time course of 35 minutes compared to results obtained with a combination of helicase and primase (~80 percent) (**Figure 2A**). We conclude that primase can activate helicase activity in the presence of the putative loader, but the loader retains a negative effect on overall helicase activity in this assay.

To exclude the possibility that the displacement observed in our helicase assays could be attributed to another protein than the CD3657 helicase, we used the previously characterized Walker A mutants of this protein (T215A and K214R) (**Chapter 4**). We observed only background levels (< 5 percent) of displacement in the presence of CD3657 T215A mutant helicase, primase and the putative loader (**Supplemental Figure 1**) in comparison to ~60 percent strand displacement for the wild type helicase in the presence of the same proteins (**Figure 2B**). Similar results were obtained with another Walker A mutant of helicase (CD3657 K214R; **Supplemental Figure 1**). This shows that the displacement observed in our experiments can be attributed to helicase alone and not some other factor.

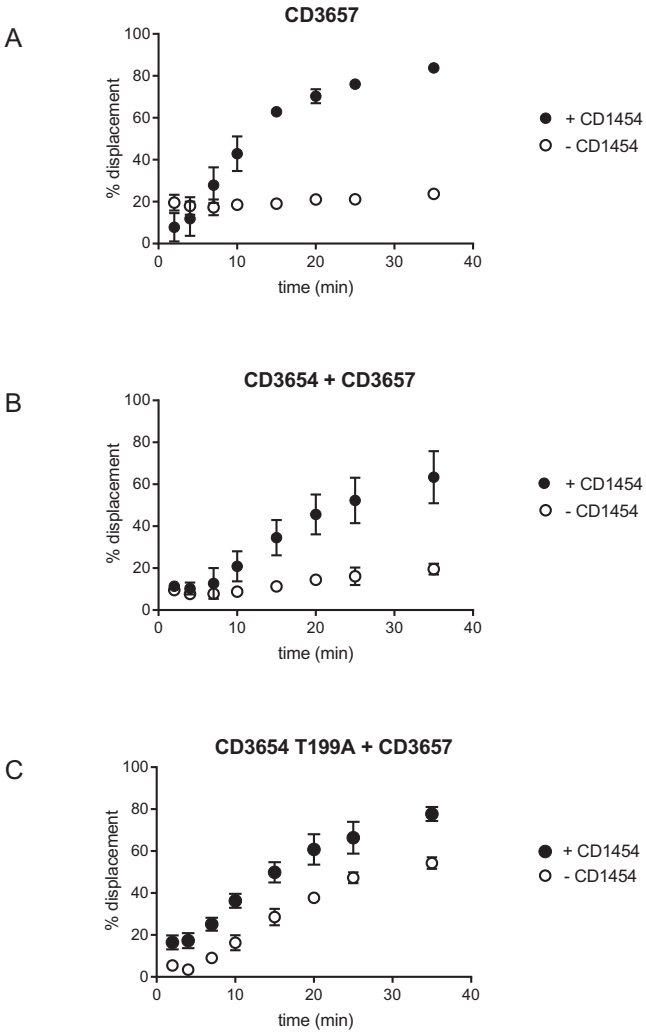


Figure 2. Helicase activity is stimulated by primase.

A. Helicase activity of the CD3657 helicase with and without the CD1454 primase.

B. Helicase activity of the CD3657 helicase in the presence of the putative loader protein CD3654, in the presence and absence of the CD1454 primase.

C. Helicase activity of the CD3657 helicase in the presence of a representative Walker A mutant (T199A) of the putative loader protein CD3654, in the presence and absence of the CD1454 primase. Walker A mutant versions of CD3657 showed negligible activity (Supplemental Figure 1). The other mutants of CD3654 (K198R, D258Q) gave similar results (Supplemental Figure 3) but have been omitted for clarity.

In *E. coli*, interaction between the loader ATPase protein and helicase does not require ATP binding by the loader [2,38,39]. However, hydrolysis of ATP to ADP by the loader (which results in dissociation from the helicase-loader complex) appears crucial to lift the negative effect of the loader on helicase activity [2,38,39]. Therefore, we set out to investigate the effect of CD3654 proteins with a mutated Walker A or Walker B motif. The lysine –residue in (K198) in the Walker A motif is predicted to directly interact with ATP, and mutation of homologous residues is known to eliminate appropriate ATP binding of AAA+ proteins, resulting in inactivation [40]. The threonine residue in Walker A (T199) and the aspartic acid residue in Walker B (D258Q) are predicted to be involved in coordination of an Mg^{2+} ion within the ATP-binding site. Overall, the three mutants (K198R, T199A and D258Q) should be affected in their ability to coordinate and/or hydrolyse ATP.

We found a 2.5-fold higher displacement (~50 percent) of the oligonucleotide in the helicase assays with all three CD3654 Walker A and B mutants (**Figure 2C, Supplemental Figure 2A** and **B**) compared to the wild type (**Figure 2B**) in the absence of primase. Interestingly, helicase activity in assays combining wild type CD3657 helicase, CD1454 primase, and mutants of the putative loader was similar (~80 percent, **Figure 2C, Supplemental Figure 2A** and **B**) to the activity measured in the two-protein helicase-primase experiment (**Figure 2A**). This suggest that binding and/or hydrolysis of ATP by the putative loader is not required to deliver the CD3657 helicase to the DNA but is at least partially responsible for the negative effect ('locking') of the helicase activity.

C. difficile primase trinucleotide specificity is similar to *Aquifex aeolicus* primase

Above, we have established a crucial role for the *C. difficile* primase CD1454 in the activation of the CD3657 helicase, which has not been reported for Gram-positive bacteria before. Next, we sought to evaluate the activity of primase and to obtain initial information on the trinucleotide specificity for the enzyme. To this end, two 50-mer oligonucleotides that comprised all 64 potential trinucleotide sequences were tested using thermally denaturing high-performance liquid chromatography analysis, as previously described [37]. We confirmed enzymatic activity and *de novo* primer synthesis on specific motifs within these templates

(*data not shown*). Next, a more detailed analysis of template specificity was performed using a 23-mer ssDNA-template containing the trinucleotide of interest (**Figure 3A**). We found that the CD1454 primase preferentially initiated *de novo* RNA primer synthesis on the initial 5'-d(CCC) motif within the 23-mer ssDNA template; therefore, a 17-mer rather than a 16-mer RNA primer was produced (**Figure 3B**). As a negative control, the 5'-d(ACA)-specific 23-mer ssDNA was tested in the priming assay and no RNA primers were synthesized by *C. difficile* primase, further demonstrating a requirement for the appropriate template initiation motif (**Figure 3B**). The activity of the CD1454 was notably inhibited by Mg^{2+} concentrations more than 30 mM, and largely independent of NTP concentration (more than 1 mM) under the conditions of the assay (**Supplemental Figure 3**).

The observed template specificity for the CD1454 primase differs from most other bacterial primases. More specifically, 5'-d(CTA) only supported minimal priming and no initiation occurred on 5'-d(TTA) by the *C. difficile* primase (**Figure 3B**), whereas substantial dinucleotide polymerization occurred on these two trinucleotides by primases from other Firmicutes such as *G. stearothermophilus*, *Staphylococcus aureus*, and *Bacillus anthracis* [16,22,27].

The trinucleotide 5'-d(CTG) was also not an effective template for *C. difficile* primase (*data not shown*), whereas it is efficiently primed by primase from Gammaproteobacteria such as *E. coli*, *Pseudomonas aeruginosa*, and *Yersinia pestis* [27,41,42]. To date, only the primase from the hyperthermophile *Aquifex aeolicus* has been shown to initiate primer synthesis on the trinucleotide 5'-d(CCC) [37].

As primase presumably scans the ssDNA template in the 3' to 5' direction until an initiation trinucleotide is encountered, the influence of the 5' nucleotide adjacent to the preferred 5'-d(CCC) trinucleotide motif was evaluated. We found that CD1454 primase was most efficient at initiating *de novo* primer synthesis when the 5' nucleotide was a cytosine or – with a slightly lower efficiency – a thymine (**Figure 3C**). Considerably less priming occurred when the 5' adjacent nucleotide contained an adenine base and only marginal non-specific priming occurred when this nucleotide contained a guanine (**Figure 3C**). These results suggest that the nucleotide 5' to the preferred trinucleotide influences the catalytic activity at the active site for subsequent initiation of *de novo* primer synthesis.

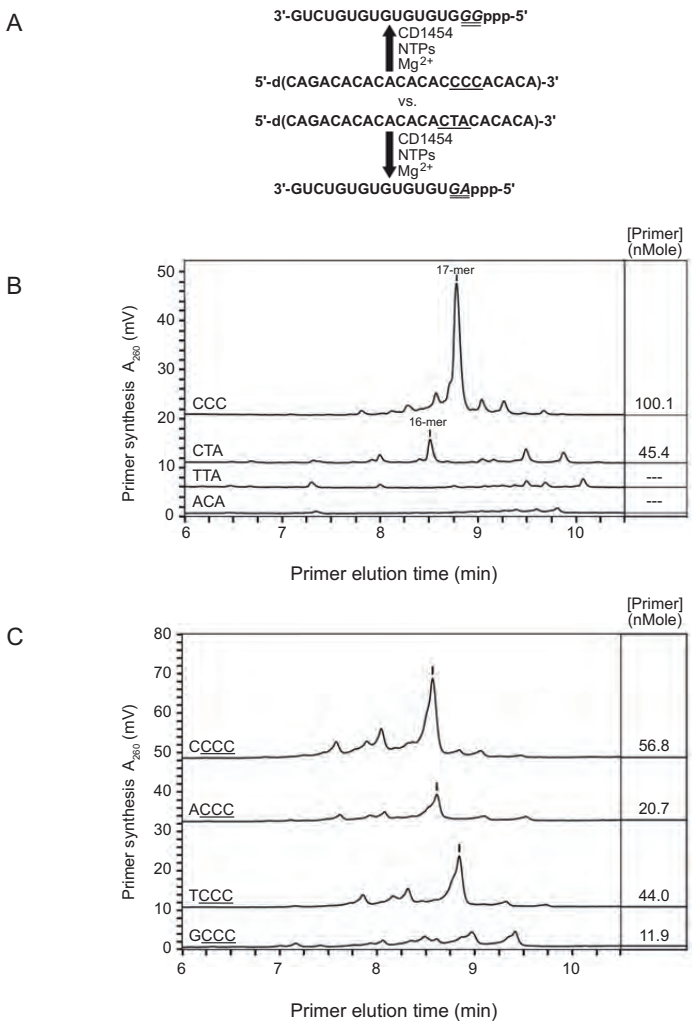


Figure 3. Trinucleotide specificity of CD1454 primase.

A. Schematic depiction of the thermally denaturing high-performance liquid chromatography analysis of primase activity.

B. Analysis of CD1454 priming activity at different trinucleotides.

C. The effect of the 5' flanking base on priming efficiency of CD1454 at the preferred trinucleotide. The numbers to the right of the chromatogram denote the total peak area for the RNA primers products synthesised and shown in the associated chromatogram to the left.

Helicase influences trinucleotide specificity of primase

The identification of the 5'd(CCC) motif as the preferred trinucleotide for initiation by the CD1454 primase was unexpected, as the *C. difficile* chromosome has a G+C content of only 29 percent [43] and other low-GC Firmicutes, such as *S. aureus* and *G. stearothermophilus* [16,22] preferentially initiate at 5'-d(CTA). Therefore, we evaluated the relative frequency of the 5'-d(CCC) motif on the plus- and minus strand of the *C. difficile* chromosome and compared it to the relative frequency of 5'-d(CTA) motif on which the CD1454 primase initiates less efficiently.

Our analysis showed that CTA triplets were on average five to ten-fold more frequent within the *C. difficile* chromosome than the preferred CCC-motif. Strikingly, there appears to be a strand bias in the occurrence of the motifs, that mirrors the GC-skew ([G-C]/[G+C]) of the chromosome (**Figure 4**). This suggests that the motifs are preferentially associated with the lagging strand where primase acts and indicate a possible role for primase in generating the strand bias.

As it has previously been shown that helicase can stimulate primase activity, affect primer length and modulate trinucleotide specificity in *E. coli* [1,20,24], we determined whether the CD3657 helicase could enhance priming at the non-preferred trinucleotide. The effect of CD3657 helicase on CD1454 primase activity was evaluated using 23-mer ssDNA templates that either contained the preferred trinucleotide 5'-d(CCC), the Firmicute-preferred trinucleotide 5'-d(CTA), or 5'-d(ACA) (negative control). RNA primer production by the primase CD1454 was strongly stimulated on 5'-d(CTA) in the presence of the CD3657 helicase (45.9 vs 25.3 nMole) at stoichiometric concentrations of $[CD3657]_6$:$[CD1454]_3$. Helicase-stimulated primase activity on 5'-d(CCC) increased RNA primer synthesis only 1.15-fold and, as expected, no stimulation of RNA primer synthesis occurred on the 5'-d(ACA) trinucleotide (**Figure 5**).

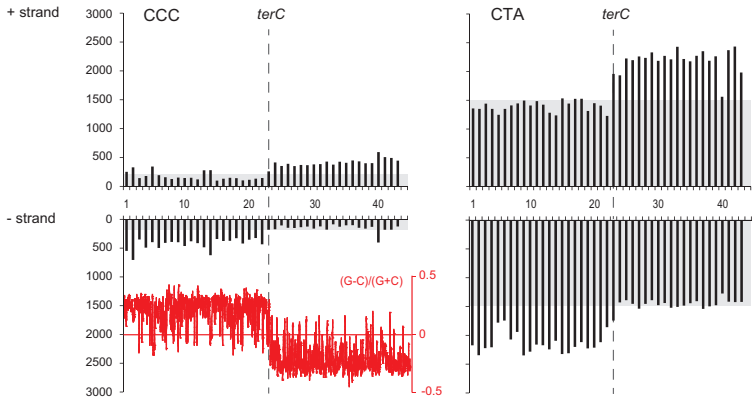


Figure 4. Trinucleotide frequency in the genome of *C. difficile*.

The number of CCC and CTA trinucleotides on the + and – strand of the *C. difficile* 630Δ*erm* was calculated in 100 000 bp bins. Skew in trinucleotide frequency is highlighted using a grey box. The position of the putative terminus of replication (*terC*) is indicated with a vertical dashed line. The red inset shows the GC skew, as calculated in Artemis with a window size of 5000.

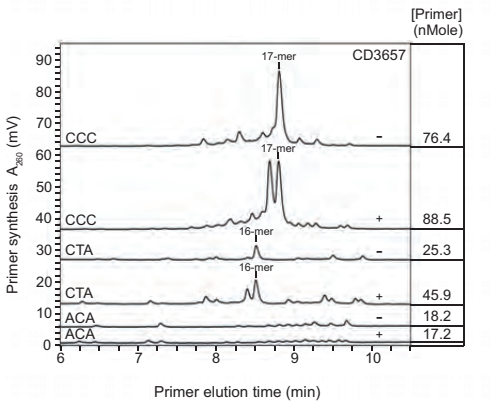


Figure 5. CD3657 helicase affects CD1454 primase activity.

Primase activity was determined by thermally denaturing HPLC analysis (see Materials and Methods). RNA primer synthesis was quantified from three different trinucleotides in the presence (**+**) and absence (**–**) of CD3657. Helicase stimulates primase activity, most notably at the non-preferred 5'-d(CTA) trinucleotide. The numbers to the right of the chromatogram denote the total peak area for the RNA primers products synthesised and shown in the associated chromatogram to the left.

We also observed that helicase stimulated primase to synthesize slightly shorter RNA polymers (**Figure 5**). The production of a shorter primer might allow faster transfer to the replicative DNA polymerase, increasing the replication speed. Collectively, our data show that the stimulatory effect of helicase on primase activity *in vitro* is clearly enhanced on the trinucleotide 5'-d(CTA) which is preferred by most Firmicute primases, whereas the effect at the *C. difficile*-preferred 5'-d(CCC) sequence was only minimal, probably due to the already high efficiency of priming on this motif by the CD1454 primase.

We conclude that through the interaction of primase with helicase and the relative abundance of the 5'd(CTA) the overall efficiency of priming is likely to be greatly enhanced.

A lysine residue contributes to trinucleotide specificity of primase

Considering that the *C. difficile* primase trinucleotide specificity is unusual for Firmicutes (**Figure 3B**) but resembles that of the hyperthermophilic *A. aeolicus* primase [37], we wanted to determine the factors that might contribute to this trinucleotide template specificity for subsequent dinucleotide polymerization.

We performed pair-wise sequence comparisons between the zinc binding domain (ZBD), that is involved in sequence specific DNA binding of primase in *C. difficile* and ten other bacterial species (*Clostridium perfringens*, *G. stearothermophilus*, *Staphylococcus aureus*, *Bacillus anthracis*, *B. subtilis*, *E. coli*, *Pseudomonas aeruginosa*, *Yersinia pestis*, *A. aeolicus*, and *Francisella tularensis*). *C. perfringens* (53.8% identity and 87.9% similarity in 91 residues) had the highest amino acid sequence homology followed by *B. anthracis* (49.5% identity and 87.9% similarity in 91 residues) and *A. aeolicus* (47.2% identity and 83.1% similarity in 89 amino acids). The *E. coli* ZBD had the lowest homology (43.2% identity and 83.0% similarity in 88 residues). The Clustal Omega alignment tool placed the *C. difficile* CD1454 ZBD closest to the ZBD of *A. aeolicus* and *G. stearothermophilus* (**Figure 6A**). The (partial) clustering of the *C. difficile* CD1454 primase with primases that have different trinucleotide specificity led us to anticipate that the composition and spatial location of specific amino acids in the ZBD relative to the RNA polymerase domain in

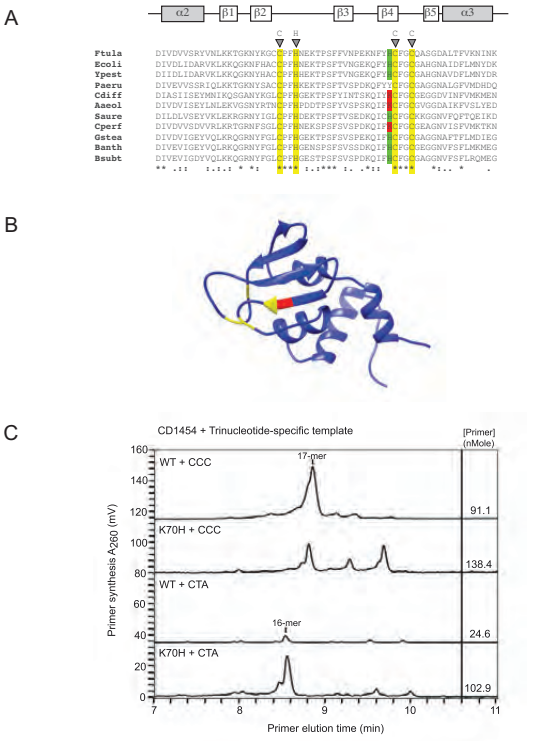


Figure 6. A lysine residue affects trinucleotide specificity.

A. Alignment of α2–α3 region of the zinc binding domain of primases with characterized trinucleotide specificity. Predicted secondary structure (alpha helix in grey, beta sheet in white) is indicated above the alignment. Ftula: *Francisella tularensis*, Ecoli: *Escherichia coli*, Ypest: *Yersinia pestis*, Paeru: *Pseudomonas aeruginosa*, Cdiff: *Clostridium difficile*, Aaeol: *Aquifex aeolicus*, Saure: *Staphylococcus aureus*, Cperf: *Clostridium perfringens*, Gstea: *Geobacillus stearothermophilus*, Banth: *Bacillus anthracis*, Bsubt: *Bacillus subtilis*. Highlighted are the zinc coordinating residues of the CHC2 zinc binding motif (yellow), and the adjacent histidine (green) or lysine (red) residues. Sequence conservation is indicated below the alignment.

B. Phyre2 model of the CD1454 primase zinc binding domain based on PDB 2AU3 (see also Figure 1) indicating the zinc coordinating residues (yellow) and the location of the lysine residue (red).

C. Thermally denaturing high-performance liquid chromatography analysis of primase activity of wild type (WT) and K70H mutant CD1454 primase protein with the CCC and CTA containing templates. The primase (WT or K70H mutant) and the initiation trinucleotide in the single-stranded 23-mer template used is shown in the upper left corner of the respective chromatogram. The numbers in the panel to the right of the chromatograms denote the total peak area for the RNA primer products synthesised and are shown in the associated chromatogram to the left. The mutation affects RNA primer synthesis from both preferred 5'-d(CCC) and non-preferred 5'-d(CTA) trinucleotides but shows the strongest increase in activity from the 5'-d(CTA) trinucleotide.

primase, rather than overall sequence similarity, must be critical for template recognition and phosphodiester bond formation.

In the multiple sequence alignment, a lysine (K) at residue 70 in the ZBD of *C. difficile* primase was unique to *C. difficile*, *C. perfringens*, and *A. aeolicus* (**Figure 6A**). Structural modelling of CD1454 with Phyre2 [44] revealed that this particular lysine residue is in close proximity of the zinc-ribbon motif that tetrahedrally coordinates a zinc ion and is essential for primase function (**Figure 6B**) [45]. Interestingly, the Protein Data Bank (PDB) template used by Phyre2 is 2AU3, the *A. aeolicus* primase [21]. The crystal structure of *G. stearothermophilus* DnaG primase contains a histidine at this position [45] and the multiple sequence alignment suggests that other Firmicutes also do. We hypothesized that the exposed lysine residue influences primase initiation specificity, as *C difficile* and *A. aeolicus* preferred the 5'-d(CCC) motif, although Firmicutes generally prefer the 5'-d(CTA) motif.

To address this experimentally, we mutated the K70 residue of the CD1454 primase to a histidine. If the lysine contributes to the unusual trinucleotide specificity of the *C. difficile* primase, we would expect the CD1454 K70H mutant to demonstrate reduced priming on the 5'-d(CCC) motif compared to the wild-type CD1454 primase. Indeed, CD1454 K70H showed substantially reduced initiation on the 5'-d(CCC) motif (**Figure 6C**). Slightly more primers were synthesized, but with relaxed template specificity, as evidenced by the synthesis of RNA polymers longer than 17 nucleotides in length (**Figure 6C**). By contrast, a substantial (~4-fold) increase in primer production was observed for the mutant primase compared to wild-type in reactions with the 5'-d(CTA) containing template and priming for this trinucleotide was highly specific (**Figure 6C**).

Importantly, these results demonstrate that modifying a single residue within the ZBD of the CD1454 primase is sufficient to alter initiation specificity and suggests that the exposed lysine residue is crucial for preferential initiation on the 5'-d(CCC) motif.

Discussion

Dedicated helicase loader proteins have not been found in the majority of bacteria. However, in the subset of bacterial species in which this protein has been identified, including the Firmicutes, their function is pivotal for DNA replication [46]. The Gram-positive model bacterium *Bacillus subtilis* belongs to the Firmicutes, like *Clostridium difficile*, and is amongst those who require a loader protein (and other accessory proteins) to load the replicative helicase [2,5,6]. Previously, we have identified the putative helicase loader of *C. difficile* (CD3654), on the basis of homology to the *Bs*DnaI loader protein, genomic context (proximity to the helicase CD3657 and the DnaB_2 family protein CD3653) and a demonstrable interaction with helicase (**Chapter 4**). The results of our helicase activity assays show an inhibitory effect of the CD3654 loader on the activity of the CD3657 helicase *in vitro*, at least in the presence of the CD1454 primase (**Figure 2B**). This effect offers an explanation for the lack of helicase activity we observed in earlier experiments (**Chapter 4**), but at first seems at odds with a role as putative helicase loader. However, in *E. coli* it has been demonstrated that ATP-bound *Ec*DnaC loader protein can act as an inhibitor of the *Ec*DnaB helicase [2,38,39]. We previously demonstrated an interaction between Walker A mutant putative loader proteins and wild-type helicase by means of bacterial two-hybrid- and analytical gel-filtration experiments (**Chapter 4**). Here, we show that helicase activity was over 2.6-fold increased when Walker A mutant putative loader proteins instead of wild-type loader protein were assayed (**Figure 2C**). Together these data suggest that the Walker A mutant of CD3654 is still capable of loading the helicase but is defective in restraining the DNA-unwinding activity of the helicase. A functional role for CD3654 in the essential process of helicase loading and/or activation is supported by the observation that no transposon insertions were obtained in the *cd3654* gene (as well as the other replication genes *cd3653*, *cd3657* and *cd1454*) *in vivo* [47].

Loading of the CD3657 helicase differs from the situation in another Firmicute, *B. subtilis*. Most notably, CD3657 seems to be capable of self-loading, has increased helicase activity in the presence of primase (CD1454) and is negatively influenced by the presence of the putative loader protein (**Figure 2A**). Instead, helicase loading in *C. difficile* is somewhat reminiscent of the situation in *H. pylori*,

where a dodecameric self-loading helicase remains inactive until activated by primase, leading to the dissociation of the dodecamer into two hexamers [33].

Earlier experiments indicated that helicase is monomeric at low-micromolar concentrations of protein (**Chapter 4**). CD3657 can form hexameric assemblies at higher concentrations (**Chapter 4**), but these pre-formed hexamers are inactive, even in the presence of primase (our unpublished observations), in contrast to the situation in *E. coli* [2]. This supports the notion that CD3657 belongs to the ring-maker class of helicases [8].

Helicase loading in *C. difficile* also seems to differ in crucial aspects from the Gram-negative model *E. coli*. Walker A mutants of the *Ec*DnaC loader protein are capable of loading the *Ec*DnaB helicase but do not sustain helicase activity, suggesting that ATP turnover by the loader is required to release the helicase [2]. Our data show that in *C. difficile*, (self-) loading of the CD3657 helicase is stimulated by Walker A mutants of CD3654 loader, but that these mutants of the loader readily release active helicase (**Figure 2C**). This suggests a role for ATP binding and/or hydrolysis in 'locking' helicase activity.

*Ec*DnaB helicase is in complex with the *Ec*DnaC loader protein throughout the loading process and remains inactive until the *Ec*DnaG primase binds to helicase, thereby releasing the loader protein [34]. Our data do not exclude a similar role for primase in helicase loading of *C. difficile* but do show that the role of the CD1454 primase is not limited to the release of the putative loader; very strong stimulation of CD3657 helicase activity is observed when CD1454 primase is added in the helicase assay in the absence of the CD3654 loader (**Figure 2A**). The CD1454 primase may still stimulate ATP turnover by the CD3654 loader protein. We consider two possible, not mutually exclusive, scenarios to explain the activation of helicase by primase in the absence of the loader. First, primase may stabilize the hexameric helicase on the DNA, which indirectly contributes to the unwinding activity. Second, primase may act as a direct activator of the DNA-unwinding activity of helicase. Our experiments do not discriminate between these possibilities, though stabilization of the hexamer or other conformational changes in the hexameric helicase induced by primase have also been observed in *E. coli* [13,29,32,34].

We found that the primase of *C. difficile* has an unusual trinucleotide specificity, with a preference for 5'-d(CCC) (**Figure 3B**), similar to *A. aeolicus* [37]. This is in contrast with primases from Firmicutes, which initiate *de novo* primer synthesis on 5'-d(CTA) and 5'-d(TTA), and Gammaproteobacteria primases, which initiate on 5'-d(CTG) and 5'-d(CTA). The first two nucleotides in all of these preferred and recognized trinucleotides are pyrimidines (C or T). As they serve as the template for the corresponding dinucleotide, the first two nucleotides in the RNA primer will be purines (G or A). The levels of both ATP and GTP directly or indirectly provide a means by which bacteria can sense the energy status of the cell [48-52]. The nucleotide preference might couple the efficiency of lagging strand DNA synthesis and nutritional status of the cell.

Priming by the primase CD1454 was highest when the nucleotide 5' adjacent to the preferred 5'-d(CCC) trinucleotide was a pyrimidine (**Figure 3C**), consistent with a previously hypothesized context-dependent enzyme activity [37]. Our results indicate that the pyrimidines, probably in part due to the smaller size compared to purines, provide the optimal context for catalysis and dinucleotide polymerization.

We probed the origin of primase trinucleotide specificity and found that a lysine (K) adjacent to the zinc-ribbon motif of CD1454 is important for the optimal physico-chemical environment for primer initiation on 5'-d(CCC) (**Figure 6C**). Based on this finding, we would predict a similar specificity for the primase of *Clostridium perfringens*. Unfortunately, we could not confirm this result with *C. perfringens* primase so far. We purified the CPF_2265 primase from *C. perfringens* ATCC 133124, but our attempts to obtain a primase with priming activity has failed thus far.

A 5'-d(CCC) trinucleotide specificity has previously only been observed for *A. aeolicus* [37]. Despite the predicted structural homology between the *A. aeolicus* and *C. difficile* primases (**Figure 6B**), this was unexpected, as *C. difficile* is a mesophilic, spore-forming, Gram-positive pathogen, whereas *A. aeolicus* is a hyperthermophilic Gram-negative bacterium. However, cladistic studies using multiple signature proteins indicate that the *Aquifex* lineage emerged from Gram-positive bacteria, prior to the split of Gram-positive and Gram-negative bacteria [53,54]. These studies concluded that the Firmicutes are presumably among the most ancient bacteria and that the Aquificales have diverged much later in evolution [53]. Indeed, an analysis of the GC-content of rRNA clusters suggests that hyperthermophilic species have

evolved from mesophilic organisms via adaptation to high temperature [55] and that the Gram-negative double membrane may have been derived from sporulating Gram-positives [56,57].

Materials and Methods

Plasmid construction

All oligonucleotides and plasmids constructed or used for this study are listed in **Supplemental Table 1** and **2**. To construct the primase (CD1454) expression plasmids, the CD1454 open reading frame was amplified with high fidelity polymerase *Pfu* via PCR from *C. difficile* strain 630Δ*erm* chromosomal DNA [58,59], using primers oWKS-1183 and oWKS-1184. The PCR product was digested with *Nde*I and *Xho*I and ligated into vector pET21b (Novagen). The reverse primer introduces a stop codon before the *Xho*I site, thereby ensuring that the protein is in its native form, when expressed (no C-terminal 6x His-tag). This resulted in the construction of a CD1454 expression vector (pEVE7).

Site-directed mutagenesis

The helicase- (CD3657) and helicase loader (CD3654) mutants were constructed according to the QuikChange protocol (Stratagene), as previously described (Chapter 4). The expression construct for the primase mutant (CD1454 K70H) was constructed according to the QuikChange protocol (Stratagene). Primers were generated with Primer X, a web-based tool for automatic design of mutagenic primers for site-directed mutagenesis. QuikChange was carried out using *Pfu* polymerase and plasmids pEVE-7 as a template. All mutant constructs were verified for the correct mutation by DNA-sequencing.

Protein purifications

Purification of helicase CD3657 and putative loader CD3654

Purification of these proteins was performed as described previously (Chapter 4). Walker A and B mutant proteins were purified in a manner identical to the wild-type proteins. Protein concentrations mentioned in this manuscript refer to concentration of the monomer of the protein.

Purification of primase CD1454 and CD1454 K70H

Overexpression *C. difficile* CD1454 proteins (wild type and K70H) was carried out in *E. coli* BL21 (DE3) from the pEVE7 plasmid. The growth medium consisting of 2XYT broth (1.2 L), carbenicillin (50 µg/mL) and antifoam 204 (Sigma-Aldrich) was inoculated with a pre-culture (10 mL). The cell culture was incubated at 37°C with mechanical shaking, at 180 rpm, until an optical density (600 nm) of 0.70-0.85 was reached (after approximately 3 h). Protein expression was induced via the addition of IPTG (1 mM final concentration) and the culture was incubated at 30°C for 3 h. The cells were harvested by centrifugation (3000 g, 15 min, 4°C) and resulting cell paste was stored at -80°C. CD1454 cell paste, prepared from 1.2 L cell culture, was re-suspended in 30 mL TED0 sonication buffer (Tris pH 7.5 50 mM, EDTA 1 mM, DTT 1 mM) with PMSF (1 mM). The bacterial cells were lysed by sonication and crude lysate was clarified by centrifugation (42,000 g, 30 min, 4°C). The resulting supernatant was separated from the cell debris using a 0.22 µm pore filter and loaded onto a 2x 5 ml HiTrap Q HP column in series with a 5ml heparin column, equilibrated in TED0 (50 mM Tris pH7.5, 1 mM EDTA, 1 mM DTT). After loading, the columns were separated, and the heparin column was reconnected to the FPLC system, washed extensively with TED0, step-washed with 10% TED1000 and gradient eluted with TED1000 (50 mM Tris pH7.5, 1 mM EDTA, 1 mM DTT, 1000 mM NaCl). Primase eluted at 30mS-45mS. Fractions containing primase were pooled, diluted with TED0 to 5-10 mS and loaded onto an 8 mL MonoS column equilibrated in TED0. The protein was eluted with a gradient of TED1000 and eluted at 15-20mS. Fractions containing primase were pooled.

The collected protein was loaded onto a Hiload 26/60 Superdex 200 gel filtration column equilibrated in TED100G buffer (50 mM Tris pH7.5, 1 mM EDTA, 1 mM DTT, 100 mM NaCl, 10% glycerol). Protein aliquots were stored at -80°C. Protein purity (all >95%) was estimated by SDS-PAGE electrophoresis and

concentration was determined spectrophotometrically using extinction coefficients calculated using the ExPASy ProtParam tool (http://web.expasy.org/protparam).

Size exclusion chromatography coupled with multiangle laser light scattering analysis

The oligomeric state of CD1454 protein was assessed using a 1260 Infinity HPLC system (Agilent), with a miniDAWN Treos (Wyatt Technologies) 3-angle static light scattering detector and ERC RefractoMax 521 UV_{280nm} and refractive index detector (ThermoScientific), connected downstream of a Superose 6 10/300 gel filtration column (GE Health Care) equilibrated in 50 mM Tris, pH 7.5, 50 mM NaCl, 1 mM EDTA, 1 mM DTT, 1 mM ATP, 2 mM $MgCl_2$. The data was processed using the Astra™ package (Wyatt Technologies).

RNA primer synthesis assay and thermally denaturing HPLC analysis

RNA priming assays and denaturing HPLC analyses were conducted as was previously described for other bacterial primases[27]. Initially, two 50-mer oligonucleotides that comprised all 64 trinucleotide sequences were used to assess template specificity, as previously described[37]. Confirmation of the preferred initiation motif was obtained by using the 23-mer trinucleotide-specific template 5'-CAGA(CA)5XYZ(CA)3-1,3-propanediol, whereby XYZ was the trinucleotide of interest. The purified oligonucleotide templates used in this study were synthesized by Integrated DNA Technologies, Inc. (Coralville, IA). Templates were quantified by spectrophotometry at 260 nm using the extinction coefficients obtained for each oligonucleotide from the online OligoTools Calculator from Integrated DNA Technologies, Inc. All RNA primer synthesis reactions were carried out in 50 µL nuclease-free water containing 50 mM HEPES pH 7.5, 100 mM potassium glutamate, 10 mM DTT, 2 µM ssDNA template, 30 mM magnesium acetate, and 1.2 mM of each NTP, unless otherwise specified. Priming reactions were incubated at 30°C for 1 h, desalted in a Sephadex G-25 spin column, and dried using a speed vacuum. The pellet was suspended in water to $1/5^{th}$ of the original volume of the sample and 8 µL of that sample was analysed by HPLC under thermally-denaturing conditions at 80°C.

For the denaturing HPLC analyses, a gradient 0–8.8% v/v acetonitrile over 16 min was used to obtain optimal separation of primer products and ssDNA templates on a DNA Sep column. The WAVE HPLC Nucleic Acid Fragment Analysis System, HPLC Buffer A (0.1 M triethylammonium acetate, pH 7.0), HPLC Buffer B (0.1 M triethylammonium acetate, 25% acetonitrile v/v), and the DNA Sep HPLC column were obtained from Transgenomic (Omaha, NE). Primer products and ssDNA templates were detected by UV absorbance at 260 nm and elution times of the nucleic acids were correlated to the elution profile of the appropriate standard to confirm composition and length. To normalize variability introduced during sample preparation and injection into the HPLC column, primer abundance was determined by using the ssDNA template as an internal standard. The moles of RNA primers synthesized were quantified using the relative extinction coefficient for the oligonucleotides and reported as the sum of moles for all primer lengths, as previously described [26].

Helicase assays

Helicase assays were performed as described before (Chapter 4). In short, activity was assayed by monitoring (and quantifying) the displacement of a radiolabelled (γ^{32}P-ATP) oligonucleotide oVP-1 (partially) annealed to the single stranded circular DNA m13mp18 (ssM13; Affymetrix). The 105-mer oligonucleotide was radiolabelled at the 5' end using γ^{32}ATP and T4 polynucleotide kinase (New England Biolabs) and subsequently purified through an S-200 mini-spin column (GE Healthcare). All reactions, containing 0.658 nM radiolabelled DNA substrate, were initiated by the addition of 2.5 mM ATP and carried out at 37˚C in buffer containing 20mM HEPES-NaOH (pH 7.5), 50 mM NaCl, 10 mM $MgCl_2$ and 1 mM DTT for various times. The reactions were terminated by adding 5x SDS-STOP buffer (100mM Tris pH8.0, 200mM EDTA, 2.5% (w/v) SDS, 50% (v/v) glycerol, 0.15% (w/v) bromophenol blue).

To investigate the effect of the putative helicase loader (CD3654) on the activity of the helicase (CD3657), the proteins were mixed in equimolar concentrations (1μM) and incubated for 10 minutes at 37˚C prior to adding reaction buffer. The buffer with CD3657 was preincubated for 5 mins before adding CD3654, incubated for 5 more mins after which the reaction was initiated with 2.5mM ATP (final concentration). Stop buffer was added to terminate the reactions (1% v/w SDS,

40 mM EDTA, 8% v/v glycerol, 0.1% w/v bromophenol blue). Reaction samples (10 µl) were loaded on a 10% non-denaturing polyacrylamide gel, run in 1xTBE (89 mM Tris, 89 mM boric acid, 2 mM EDTA) at 150V, 40mA/gel for 60 mins. The gel was dried, scanned and analysed using a molecular imager and associated software (Biorad). Experiments were carried out in triplicate, and data analysis was performed using Prism 6 (GraphPad Software).

Acknowledgements

This work was supported, in part, by a VENI and a VIDI fellowship from the Netherlands Organization for Scientific Research and a Gisela Thier Fellowship from the Leiden University Medical Center to WKS and a Biotechnology and Biological Sciences Research Council (UK) grant (BB/K021540) to PS. We thank Sjaak van Voorden for performing the site-directed mutagenesis on our constructs, Roger Scherrers from Wyatt Technology for the SEC-MALS analysis of the CD1454 protein, and Danuta Tomkiewicz for help with preparing Figure 1B.

Supplemental information

Primase (CD1454) stimulates the DNA-unwinding activity of the replicative helicase (CD3657) of *Clostridium difficile*

Erika van Eijk [1], Vasileios Paschalis [2], Matthew Green[2], Marilynn A. Larson [3,4], Keith Spriggs [5], Annemieke Friggen [1], Panos Soultanas [2], Wiep Klaas Smits [1]

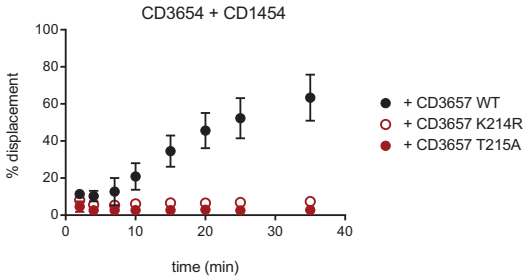
1 Department of Medical Microbiology, Leiden University Medical Center, Leiden, the Netherlands;

2 School of Chemistry, Center for Biomolecular Sciences, University of Nottingham, United Kingdom;

3 Department of Microbiology and Pathology, University of Nebraska Medical Center, Omaha, NE 68198-6495,

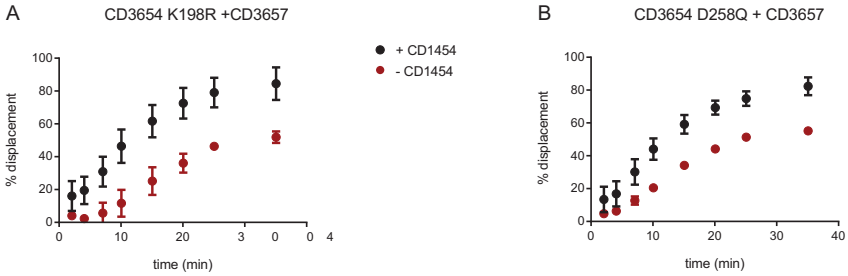
4 University of Nebraska Center for Biosecurity, Omaha, NE 68198-4080,

5 School of Psychology, University of Nottingham, United Kingdom.



Supplemental Figure 1. Helicase activity is abrogated in a Walker A mutant of CD3657.

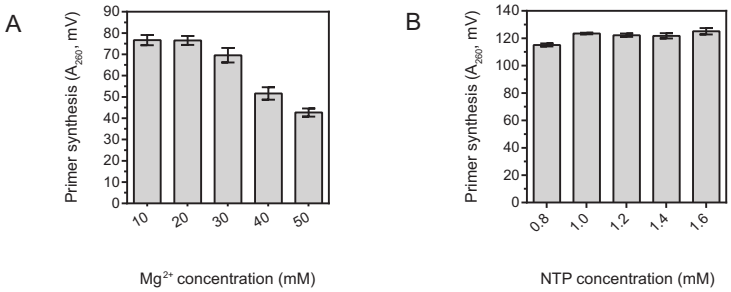
Helicase activity was assayed by quantifying the displacement of a radiolabelled (γ^{32}P–ATP) oligonucleotide partially annealed to the single stranded circular DNA m13mp18. Percent displaced signal from the helicase assays in time. Helicase activity of the wild type (WT; data series depicted in black) and mutant (K214R and T215A; data series depicted in red–edged and solid red circles, respectively) CD3657 proteins in the presence of the CD1454 primase. Error bars indicate standard deviation (*n*=3).



Supplemental Figure 2. Helicase activity is not inhibited in the presence of mutant CD3654 loader proteins, in the presence and absence of the CD1454 primase.

A. Helicase activity of CD3657 in the presence of a Walker A mutant CD3654 protein (K198R), with (black data series) and without (red data series) the CD1454 primase protein.

B. Helicase activity of CD3657 in the presence of Walker B mutant CD3654 protein (D258Q), with (black data series) and without (red data series) the CD1454 primase protein. Error bars indicate standard deviation (*n*=3).



Supplemental Figure 3. Effect of Mg^{2+} and NTP concentration as determined by thermally denaturing high-performance liquid chromatography analysis of primase activity.

A. Primase activity as a function of magnesium ion (Mg^{2+}) concentration. Shown are the relative levels of RNA primers synthesized by *C. difficile* primase (1.8 μM) in reactions with the 5′-d(CCC)-containing ssDNA template and the indicated concentration of magnesium acetate.

B. Primase activity as a function of NTP concentration. Shown are the relative levels of RNA primers synthesized by C. difficile primase (1.8 μM) in reactions with 20 mM magnesium acetate, the 23-mer 5′-d(CCC)-containing ssDNA template, and the indicated concentration of each ribonucleotide.

Supplemental Table 1. Oligonucleotides used in this study

Name	Sequence (5' – 3')	Description
oWKS-1183	TAGAATACATATGTTAACACAAAAAATTACACCAG	Forward primer on CD1454 ORF
oWKS-1184	CCGCTCGAGTTACTACAAGCTCTTTAAAATTTTATTTATC	Reverse primer on CD1454 ORF
oWKS-1272	GGTTCTACTGGACTAGGAAGGACCTATATGTGCAATTG	Forward CD3654 K198R Quikchange (QC)
oWKS-1273	CAATTGCACATATAGGTCCTTCCTAGTCCAGTAGAACC	Reverse CD3654 K198R (QC)
oWKS-1274	GGTTCTACTGGACTAGGAAAGGCCTATATGTGCAATTGTATTG	Forward CD3654 T199A (QC)
oWKS-1275	CAATACAATTGCACATATAGGCCTTTCCTAGTCCAGTAGAACC	Reverse CD3654 T199A (QC)
oEVE-69	CTTGTTTGATTGTGATTTATTAATAATACAAGACCTTGGAAC	Forward CD3654 D258Q (QC)
oEVE-70	GTTCCAAGGTCTTGTATTATTAATAAATCACAATCAAACAAG	Reverse CD3654 D258Q (QC)
oEVE-77	CGTCAAAACAGATTTATCACTGTTTTGGTTGTGGTG	Forward CD1454 K70H (QC)
oEVE-78	CACCACAACCAAAACAGTGATAAATCTGTTTTGACG	Reverse CD1454 K70H (QC)

Supplemental Table 2. Plasmids used in this study

Plasmid	Description	Reference
pEVE-87	pET21b-CD3657	Chapter 4
pEVE-92	pET21b-CD3657-T215A	Chapter 4
pEVE-24	pET28b-CD3654	Chapter 4
pEVE-59	pET28b-CD3654-K198R	This study
pEVE-60	pET28b-CD3654-T199A	This study
pEVE-203	pET28b-CD3654-D258Q	This study
pEVE-7	pET21b-CD1454	This study
pEVE-201	pET21b-CD1454-K70H	This study

References

1 Bhattacharyya, S. & Griep, M. A. DnaB helicase affects the initiation specificity of Escherichia coli primase on single-stranded DNA templates. *Biochemistry* **39**, 745-752 (2000).

2 Davey, M. J., Fang, L., McInerney, P., Georgescu, R. E. & O'Donnell, M. The DnaC helicase loader is a dual ATP/ADP switch protein. *The EMBO journal* **21**, 3148-3159, doi:10.1093/emboj/cdf308 (2002).

3 Ioannou, C., Schaeffer, P. M., Dixon, N. E. & Soultanas, P. Helicase binding to DnaI exposes a cryptic DNA-binding site during helicase loading in Bacillus subtilis. *Nucleic acids research* **34**, 5247-5258, doi:10.1093/nar/gkl690 (2006).

4 Bruand, C. *et al.* Functional interplay between the Bacillus subtilis DnaD and DnaB proteins essential for initiation and re-initiation of DNA replication. *Molecular microbiology* **55**, 1138-1150, doi:10.1111/j.1365-2958.2004.04451.x (2005).

5 Smits, W. K., Goranov, A. I. & Grossman, A. D. Ordered association of helicase loader proteins with the Bacillus subtilis origin of replication in vivo. *Molecular microbiology* **75**, 452-461, doi:10.1111/j.1365-2958.2009.06999.x (2010).

6 Velten, M. *et al.* A two-protein strategy for the functional loading of a cellular replicative DNA helicase. *Molecular cell* **11**, 1009-1020 (2003).

7 Arias-Palomo, E., O'Shea, V. L., Hood, I. V. & Berger, J. M. The bacterial DnaC helicase loader is a DnaB ring breaker. *Cell* **153**, 438-448, doi:10.1016/j.cell.2013.03.006 (2013).

8 Davey, M. J. & O'Donnell, M. Replicative helicase loaders: ring breakers and ring makers. *Current biology : CB* **13**, R594-596 (2003).

9 Bruand, C., Farache, M., McGovern, S., Ehrlich, S. D. & Polard, P. DnaB, DnaD and DnaI proteins are components of the Bacillus subtilis replication restart primosome. *Molecular microbiology* **42**, 245-255 (2001).

10 Marsin, S., McGovern, S., Ehrlich, S. D., Bruand, C. & Polard, P. Early steps of Bacillus subtilis primosome assembly. *The Journal of biological chemistry* **276**, 45818-45825, doi:10.1074/jbc.M101996200 (2001).

11 Kurth, I. & O'Donnell, M. Replisome Dynamics during Chromosome Duplication. *EcoSal Plus* **3**, doi:10.1128/ecosalplus.4.4.2 (2009).

12 Frick, D. N. & Richardson, C. C. DNA primases. *Annual review of biochemistry* **70**, 39-80, doi:10.1146/annurev.biochem.70.1.39 (2001).

13 Bird, L. E., Pan, H., Soultanas, P. & Wigley, D. B. Mapping protein-protein interactions within a stable complex of DNA primase and DnaB helicase from Bacillus stearothermophilus. *Biochemistry* **39**, 171-182 (2000).

14 Syson, K., Thirlway, J., Hounslow, A. M., Soultanas, P. & Waltho, J. P. Solution structure of the helicase-interaction domain of the primase DnaG: a model for helicase activation. *Structure (London, England : 1993)* **13**, 609-616, doi:10.1016/j.str.2005.01.022 (2005).

15 Tougu, K. & Marians, K. J. The extreme C terminus of primase is required for interaction with DnaB at the replication fork. *The Journal of biological chemistry* **271**, 21391-21397 (1996).

16 Thirlway, J. & Soultanas, P. In the Bacillus stearothermophilus DnaB-DnaG complex, the activities of the two proteins are modulated by distinct but overlapping networks of residues. *Journal of bacteriology* **188**, 1534-1539, doi:10.1128/jb.188.4.1534-1539.2006 (2006).

17 Chintakayala, K. *et al.* Domain swapping reveals that the C- and N-terminal domains of DnaG and DnaB, respectively, are functional homologues. *Molecular microbiology* **63**, 1629-1639, doi:10.1111/j.1365-2958.2007.05617.x (2007).

18 Soultanas, P. The bacterial helicase-primase interaction: a common structural/functional module. *Structure (London, England : 1993)* **13**, 839-844, doi:10.1016/j.str.2005.04.006 (2005).

19 Wu, C. A., Zechner, E. L. & Marians, K. J. Coordinated leading- and lagging-strand synthesis at the Escherichia coli DNA replication fork. I. Multiple effectors act to modulate Okazaki fragment size. *The Journal of biological chemistry* **267**, 4030-4044 (1992).

20 Johnson, S. K., Bhattacharyya, S. & Griep, M. A. DnaB helicase stimulates primer synthesis activity on short oligonucleotide templates. *Biochemistry* **39**, 736-744 (2000).

21 Corn, J. E., Pease, P. J., Hura, G. L. & Berger, J. M. Crosstalk between primase subunits can act to regulate primer synthesis in trans. *Molecular cell* **20**, 391-401, doi:10.1016/j.molcel.2005.09.004 (2005).

22 Koepsell, S. A., Larson, M. A., Griep, M. A. & Hinrichs, S. H. Staphylococcus aureus helicase but not Escherichia coli helicase stimulates S. aureus primase activity and maintains initiation specificity. *Journal of bacteriology* **188**, 4673-4680, doi:10.1128/jb.00316-06 (2006).

23 Kuchta, R. D. & Stengel, G. Mechanism and evolution of DNA primases. *Biochimica et biophysica acta* **1804**, 1180-1189, doi:10.1016/j.bbapap.2009.06.011 (2010).

24 Lu, Y. B., Ratnakar, P. V., Mohanty, B. K. & Bastia, D. Direct physical interaction between DnaG primase and DnaB helicase of Escherichia coli is necessary for optimal synthesis of primer RNA. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 12902-12907 (1996).

25 Mitkova, A. V., Khopde, S. M. & Biswas, S. B. Mechanism and stoichiometry of interaction of DnaG primase with DnaB helicase of Escherichia coli in RNA primer synthesis. *The Journal of biological chemistry* **278**, 52253-52261, doi:10.1074/jbc.M308956200 (2003).

26 Koepsell, S. A., Larson, M. A., Frey, C. A., Hinrichs, S. H. & Griep, M. A. Staphylococcus aureus primase has higher initiation specificity, interacts with single-stranded DNA stronger, but is less stimulated by its helicase than Escherichia coli primase. *Molecular microbiology* **68**, 1570-1582, doi:10.1111/j.1365-2958.2008.06255.x (2008).

27 Larson, M. A. *et al.* Class-specific restrictions define primase interactions with DNA template and replicative helicase. *Nucleic acids research* **38**, 7167-7178, doi:10.1093/nar/gkq588 (2010).

28 Rannou, O. *et al.* Functional interplay of DnaE polymerase, DnaG primase and DnaC helicase within a ternary complex, and primase to polymerase hand-off during lagging strand DNA replication in Bacillus subtilis. *Nucleic acids research* **41**, 5303-5320, doi:10.1093/nar/gkt207 (2013).

29 Bailey, S., Eliason, W. K. & Steitz, T. A. Structure of hexameric DnaB helicase and its complex with a domain of DnaG primase. *Science (New York, N.Y.)* **318**, 459-463, doi:10.1126/science.1147353 (2007).

30 Soultanas, P. & Wigley, D. B. Site-directed mutagenesis reveals roles for conserved amino acid residues in the hexameric DNA helicase DnaB from Bacillus stearothermophilus. *Nucleic acids research* **30**, 4051-4060 (2002).

31 Thirlway, J. *et al.* DnaG interacts with a linker region that joins the N- and C-domains of DnaB and induces the formation of 3-fold symmetric rings. *Nucleic acids research* **32**, 2977-2986, doi:10.1093/nar/gkh628 (2004).

32 Wang, G. *et al.* The structure of a DnaB-family replicative helicase and its interactions with primase. *Nature structural & molecular biology* **15**, 94-100, doi:10.1038/nsmb1356 (2008).

33 Bazin, A., Cherrier, M. V., Gutsche, I., Timmins, J. & Terradot, L. Structure and primase-mediated activation of a bacterial dodecameric replicative helicase. *Nucleic acids research* **43**, 8564-8576, doi:10.1093/nar/gkv792 (2015).

34 Makowska-Grzyska, M. & Kaguni, J. M. Primase directs the release of DnaC from DnaB. *Molecular cell* **37**, 90–101, doi:10.1016/j.molcel.2009.12.031 (2010).

35 Liu, B., Eliason, W. K. & Steitz, T. A. Structure of a helicase–helicase loader complex reveals insights into the mechanism of bacterial primosome assembly. *Nature communications* **4**, 2495, doi:10.1038/ncomms3495 (2013).

36 Soultanas, P. A functional interaction between the putative primosomal protein DnaI and the main replicative DNA helicase DnaB in Bacillus. *Nucleic acids research* **30**, 966–974 (2002).

37 Larson, M. A. *et al.* Hyperthermophilic Aquifex aeolicus initiates primer synthesis on a limited set of trinucleotides comprised of cytosines and guanines. *Nucleic acids research* **36**, 5260–5269, doi:10.1093/nar/gkn461 (2008).

38 Galletto, R., Jezewska, M. J. & Bujalowski, W. Interactions of the Escherichia coli DnaB helicase hexamer with the replication factor the DnaC protein. Effect of nucleotide cofactors and the ssDNA on protein–protein interactions and the topology of the complex. *Journal of molecular biology* **329**, 441–465 (2003).

39 Wahle, E., Lasken, R. S. & Kornberg, A. The dnaB-dnaC replication protein complex of Escherichia coli. II. Role of the complex in mobilizing dnaB functions. *The Journal of biological chemistry* **264**, 2469–2475 (1989).

40 Hanson, P. I. & Whiteheart, S. W. AAA+ proteins: have engine, will work. *Nature reviews. Molecular cell biology* **6**, 519–529, doi:10.1038/nrm1684 (2005).

41 Swart, J. R. & Griep, M. A. Primer synthesis kinetics by Escherichia coli primase on single-stranded DNA templates. *Biochemistry* **34**, 16097-16106 (1995).

42 Yoda, K. & Okazaki, T. Specificity of recognition sequence for Escherichia coli primase. *Molecular & general genetics : MGG* **227**, 1–8 (1991).

43 Sebaihia, M. *et al.* The multidrug-resistant human pathogen Clostridium difficile has a highly mobile, mosaic genome. *Nature genetics* **38**, 779–786, doi:10.1038/ng1830 (2006).

44 Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols* **10**, 845–858, doi:10.1038/nprot.2015.053 (2015).

45 Pan, H. & Wigley, D. B. Structure of the zinc-binding domain of Bacillus stearothermophilus DNA primase. *Structure (London, England : 1993)* **8**, 231–239 (2000).

46 Robinson, A., Causer, R. J. & Dixon, N. E. Architecture and conservation of the bacterial DNA replication machinery, an underexploited drug target. *Current drug targets* **13**, 352–372 (2012).

47 Dembek, M. *et al.* High-throughput analysis of gene essentiality and sporulation in Clostridium difficile. *mBio* **6**, e02383, doi:10.1128/mBio.02383-14 (2015).

48 Gaal, T., Bartlett, M. S., Ross, W., Turnbough, C. L., Jr. & Gourse, R. L. Transcription regulation by initiating NTP concentration: rRNA synthesis in bacteria. *Science (New York, N.Y.)* **278**, 2092–2097 (1997).

49 Murray, H. D., Schneider, D. A. & Gourse, R. L. Control of rRNA expression by small molecules is dynamic and nonredundant. *Molecular cell* **12**, 125–134 (2003).

50 Schneider, D. A., Gaal, T. & Gourse, R. L. NTP-sensing by rRNA promoters in Escherichia coli is direct. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 8602–8607, doi:10.1073/pnas.132285199 (2002).

51 Schneider, D. A. & Gourse, R. L. Changes in Escherichia coli rRNA promoter activity correlate with changes in initiating nucleoside triphosphate and guanosine 5′ diphosphate 3′-diphosphate concentrations after induction of feedback control of ribosome synthesis. *Journal of bacteriology* **185**, 6185–6191 (2003).

52 Sonenshein, A. L. CodY, a global regulator of stationary phase and virulence in Gram-positive bacteria. *Current opinion in microbiology* **8**, 203-207, doi:10.1016/j.mib.2005.01.001 (2005).

53 Griffiths, E. & Gupta, R. S. Signature sequences in diverse proteins provide evidence for the late divergence of the Order Aquificales. *International microbiology : the official journal of the Spanish Society for Microbiology* **7**, 41-52 (2004).

54 Skophammer, R. G., Servin, J. A., Herbold, C. W. & Lake, J. A. Evidence for a gram-positive, eubacterial root of the tree of life. *Molecular biology and evolution* **24**, 1761-1768, doi:10.1093/molbev/msm096 (2007).

55 Galtier, N., Tourasse, N. & Gouy, M. A nonhyperthermophilic common ancestor to extant life forms. *Science (New York, N.Y.)* **283**, 220-221 (1999).

56 Tocheva, E. I. *et al.* Peptidoglycan remodeling and conversion of an inner membrane into an outer membrane during sporulation. *Cell* **146**, 799-812, doi:10.1016/j.cell.2011.07.029 (2011).

57 Tocheva, E. I., Ortega, D. R. & Jensen, G. J. Sporulation, bacterial cell envelopes and the origin of life. *Nature reviews. Microbiology* **14**, 535-542, doi:10.1038/nrmicro.2016.85 (2016).

58 Hussain, H. A., Roberts, A. P. & Mullany, P. Generation of an erythromycin-sensitive derivative of Clostridium difficile strain 630 (630Deltaerm) and demonstration that the conjugative transposon Tn916DeltaE enters the genome of this strain at multiple sites. *Journal of medical microbiology* **54**, 137-141, doi:10.1099/jmm.0.45790-0 (2005).

59 van Eijk, E. *et al.* Complete genome sequence of the Clostridium difficile laboratory strain 630Deltaerm reveals differences from strain 630, including translocation of the mobile element CTn5. *BMC genomics* **16**, 31, doi:10.1186/s12864-015-1252-7 (2015).