



**Cite this article:** van Eijk E, Paschalis V, Green M, Friggen AH, Larson MA, Spriggs K, Briggs GS, Soultanas P, Smits WK. 2016 Primase is required for helicase activity and helicase alters the specificity of primase in the enteropathogen *Clostridium difficile*. *Open Biol.* **6**: 160272. <http://dx.doi.org/10.1098/rsob.160272>

Received: 28 September 2016  
Accepted: 22 November 2016

**Subject Area:**  
biochemistry/microbiology/molecular biology

**Keywords:**  
DNA replication initiation, helicase loading and activation, primase trinucleotide specificity, ATPase, *Clostridium difficile*

**Author for correspondence:**  
Wiep Klaas Smits  
e-mail: w.k.smits@lumc.nl

<sup>†</sup>These authors contributed equally to this study.

Electronic supplementary material is available online at <https://dx.doi.org/10.6084/m9.figshare.c.3588704>.

# Primase is required for helicase activity and helicase alters the specificity of primase in the enteropathogen *Clostridium difficile*

Erika van Eijk<sup>1,†</sup>, Vasileios Paschalis<sup>2,†</sup>, Matthew Green<sup>2,†</sup>, Annemieke H. Friggen<sup>1</sup>, Marilyn A. Larson<sup>4,5</sup>, Keith Spriggs<sup>3</sup>, Geoffrey S. Briggs<sup>2</sup>, Panos Soultanas<sup>2</sup> and Wiep Klaas Smits<sup>1</sup>

<sup>1</sup>Department of Medical Microbiology, Leiden University Medical Center, Leiden, The Netherlands  
<sup>2</sup>School of Chemistry, Center for Biomolecular Sciences, and <sup>3</sup>School of Pharmacy, University of Nottingham, UK  
<sup>4</sup>Department of Pathology and Microbiology, University of Nebraska Medical Center, Omaha, NE 68198-5900, USA  
<sup>5</sup>National Strategic Research Institute, Omaha, NE 68105, USA

**VP**, 0000-0003-2757-6918; **MG**, 0000-0003-2771-9894; **MAL**, 0000-0002-0633-0313; **KS**, 0000-0001-5946-1440; **GSB**, 0000-0002-4149-8752; **PS**, 0000-0003-3578-5699; **WKS**, 0000-0002-7409-2847

DNA replication is an essential and conserved process in all domains of life and may serve as a target for the development of new antimicrobials. However, such developments are hindered by subtle mechanistic differences and limited understanding of DNA replication in pathogenic microorganisms. *Clostridium difficile* is the main cause of healthcare-associated diarrhoea and its DNA replication machinery is virtually uncharacterized. We identify and characterize the mechanistic details of the putative replicative helicase (CD3657), helicase-loader ATPase (CD3654) and primase (CD1454) of *C. difficile*, and reconstitute helicase and primase activities *in vitro*. We demonstrate a direct and ATP-dependent interaction between the helicase loader and the helicase. Furthermore, we find that helicase activity is dependent on the presence of primase *in vitro*. The inherent trinucleotide specificity of primase is determined by a single lysine residue and is similar to the primase of the extreme thermophile *Aquifex aeolicus*. However, the presence of helicase allows more efficient *de novo* synthesis of RNA primers from non-preferred trinucleotides. Thus, loader–helicase–primase interactions, which crucially mediate helicase loading and activation during DNA replication in all organisms, differ critically in *C. difficile* from that of the well-studied Gram-positive *Bacillus subtilis* model.

## 1. Background

Extensive research, primarily on the model organisms *Escherichia coli* (Gram-negative) and *Bacillus subtilis* (Gram-positive), has shown that many different proteins are involved in DNA replication. Although the overall mechanism of replication is highly conserved in all domains of life and replication is initiated by the highly conserved DNA replication initiator protein DnaA in all bacteria [1], it is perhaps not surprising that details of the molecular mechanisms can vary substantially as these prokaryotes diverged more than 3 billion years ago [2].

One of the best-characterized distinctions between *B. subtilis* and *E. coli* is the mechanism of loading the replicative helicase at the origin of replication (*oriC*), an essential step in the DNA replication process of bacteria [3,4]. Helicase is required to unwind the DNA duplex at the replication fork, and during the loading step a functional helicase multimer is assembled onto the DNA. In the

Enterobacteria, Firmicutes and Aquificae helicase loading is facilitated by a specific loader protein, which is not conserved in bacteria outside these phyla [5]. However, the strategy of helicase loading among bacteria that do encode a loader protein also differs [6,7].

For historical reasons, the nomenclature for the replication proteins differs between bacterial species (e.g. *E. coli* helicase, DnaB; *B. subtilis* helicase, DnaC). For clarity, protein names hereafter will be used in conjunction with species and either written in full or abbreviated (e.g. Ec and Bs). The *E. coli* helicase (EcDnaB) is loaded by a single loader protein (EcDnaC) *in vivo* [8–10], whereas loading of the *B. subtilis* helicase (BsDnaC) requires three accessory proteins (BsDnaD, BsDnaB and BsDnaI) *in vivo* [11–14] in addition to the replication initiator DnaA that is required in both organisms. One possible explanation for the requirement of multiple proteins in *B. subtilis* may lie in the fact that *E. coli* and *B. subtilis* employ different mechanisms to deliver the replicative helicase onto the DNA [6,7]. Alternatively, it may reflect different *oriC* architectures, requiring different mechanisms of origin remodelling [15].

Replicative helicases form hexameric rings and require single-stranded DNA (ssDNA) to be threaded through the central channel of the protein to unwind the DNA duplex [7,16]. To accomplish this, it is thought that either the pre-formed ring is physically opened (ring-breaker) or that the ring is assembled from monomers at *oriC* (ring-maker) [6].

In *E. coli*, pre-formed hexamers of the helicase protein are capable of self-loading onto ssDNA. They display *in vitro* translocation and unwinding activities, which are highly induced in the presence of the loader protein [8]. This is in contrast with *B. subtilis*, 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 is present [17]. Thus, helicase loading in *E. coli* is an example of the ring-breaker mechanism, whereas the situation in *B. subtilis* exemplifies a ring-maker mechanism [6].

*Bacillus subtilis* helicase loading *in vivo* is a hierarchical process [12–14]. Initially, the double-stranded DNA (dsDNA) at *oriC* is melted into ssDNA by the initiation protein DnaA, thereby creating a substrate for primosome assembly. The BsDnaD and BsDnaB co-loader proteins, which are structural homologues (PFAM DnaB\_2), associate sequentially with the replication origin [12] and possibly contribute to origin remodelling. This ultimately enables the ATPase loader protein to load the helicase [12–15,18–22].

The replication initiation protein (BsDnaA) and helicase-loader protein (BsDnaI) belong to the AAA+ (ATPases associated with various cellular activities) family of ATPases [23–26]. These AAA+ enzymes are, in their turn, part of the additional strand catalytic glutamate (ASCE) family [15,26–30]. The BsDnaI loader protein consists of a C-terminal AAA+ domain that is necessary for nucleotide and ssDNA binding, and an N-terminal helicase-interacting domain [17]. The loader interacts with the helicase regulating its activity and is therefore pivotal in replisome assembly [4,31,32].

The BsDnaC helicase is also an ASCE protein, and belongs to RecA-type helicase Superfamily 4 (SF4), which is involved in DNA replication [28,33,34]. The SF4 superfamily of helicases is characterized by five sequence motifs: H1, H1a, H2, H3 and H4 [28,35,36]. Motifs H1 and H2 are equivalent

to the ATP-coordinating Walker A and B motifs found in many other ATPases [28].

The multi-protein primosome of *B. subtilis* consists not only of the helicase-loader protein and helicase, but also of primase [37,38]. 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 [39]. 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 [40]. The latter region, also known as P16 (reflecting the approximate mass of 16 kDa), is variable in prokaryotes and seems to be crucial for the direct interaction with and activation of the helicase [41–44]. Interestingly, the P16 domain is structurally and functionally homologous to the N-terminal domain of the replicative helicase to which it binds [43,45,46]. Depending on the bacterial species, the interaction between helicase and primase can be either transient (*E. coli*) or stable (*Geobacillus stearothermophilus*) [41,47].

Primase and helicase affect each other's activities [41,44,48–50]. Helicase affects primase by modulating initiation specificity, stimulating primer synthesis, reducing the length of primers synthesized and increasing its affinity for single-stranded DNA [40,44,48–55]. 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 [50,56,57]. Conversely, ATPase and helicase activities of helicase are stimulated by primase due to stabilization of the helicase hexamer by this interaction [41,44,58–60]. In *Helicobacter pylori*, it was shown that the interaction of primase with helicase resulted in dissociation of the double hexamer of helicase, thereby increasing ATP hydrolysis, DNA binding and unwinding [16].

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* [61]. 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 [61]. 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 [57]. However, a ternary complex comprising helicase, loader and the helicase binding domain of primase in *G. stearothermophilus* is capable of loading [62]. These observations indicate that primosome formation, like helicase loading, may also be species-specific [59,61,62].

We are interested in DNA replication in the Gram-positive enteropathogen *C. difficile*, the most common causative agent of antibiotic associated diarrhoea [63,64]. Like *B. subtilis*, *C. difficile* belongs to the low-GC content Firmicutes. This strictly anaerobic bacterium can cause potentially fatal intestinal inflammation (colitis), is resistant to multiple antibiotics and is capable of forming highly resistant endospores [63,64]. However, the *C. difficile* DNA replication machinery is virtually uncharacterized.

Herein, we address this gap by identifying and characterizing the *C. difficile* replicative helicase (CD3657), helicase loader (CD3654) and primase (CD1454), and reconstituting helicase and primase activities *in vitro*. Our results indicate that the interaction between *C. difficile* helicase and primase is crucial for the functions of both proteins and that a single residue can strikingly determine trinucleotide specificity. Our results show that helicase loading and activation, as well as primase template specificity, in *C. difficile* differs critically from the *B. subtilis* model.

## 2. Results

### 2.1. *In silico* identification of putative replication initiation proteins

In *B. subtilis*, replication initiation requires the coordinated action of multiple proteins, DnaA, DnaD, DnaB and DnaI, *in vivo* [12,15]. BLASTP queries of the genome of *C. difficile* 630 (Genbank AM180355.1) using the amino acid sequence of these proteins from *B. subtilis* subsp. *subtilis* strain 168 (GenBank NC0989.1) allowed the identification of homologues of most, but not all, proteins that were found to be essential for replication initiation in *B. subtilis*.

Homologues of the initiation protein DnaA (CD0001; *e*-value 0.0) and the replicative helicase (CD3657; *e*-value  $7 \times 10^{-713}$ ) were identified with high confidence, sharing respectively 62% and 52% identity with their *B. subtilis* counterparts across the full length of the protein. Thus, CD0001 and CD3657 are very likely to correspond to the replication initiator protein and replicative helicase of *C. difficile*, respectively. Interestingly, no homologue of BsDnaB was found using this strategy, indicating the absence of a protein with substantial primary amino acid similarity to DnaB in *C. difficile*. However, BLASTP shows that the genome of *C. difficile* does harbour two homologues of BsDnaD (CD2943: *e*-value =  $2 \times 10^{-5}$ , identity = 29%, query coverage = 32%; CD3653: *e*-value =  $4 \times 10^{-5}$ , identity 29%, query coverage 47%). As BsDnaB and BsDnaD are strictly required for replication initiation in *B. subtilis* and are structurally related despite limited amino acid sequence similarity [18], we further examined the *C. difficile* homologues of BsDnaD. BsDnaD is composed of two domains: DDBH1 (DnaD DnaB Homology 1) and DDBH2 (DnaD DnaB Homology 2, also known as PFAM DnaB\_2). BsDnaB has a DDBH1-DDBH2-DDBH2 structure [18]. The DnaD-like proteins CD3653 and CD2943 of *C. difficile* both consist of three domains (DDBH1-DDBH2-DDBH2) and therefore resemble BsDnaB in domain structure (figure 1a). CD2943 is annotated as a putative phage replication protein and is located in the approximately 50 kb prophage 2 [65]. Also in *Listeria monocytogenes*, *S. aureus* and *Lactobacillus plantarum* DDBH2-containing phage, genes have been identified [18,66]. Considering the fact that the prophage is not part of the *C. difficile* core genome, we consider a role for CD2943 in chromosomal DNA replication unlikely, but a role for CD3653 plausible.

A similar argument can be made for the putative helicase loader. Two homologues of the BsDnaI protein were identified in *C. difficile* by BLASTP (CD3654: *e*-value =  $8 \times 10^{-10}$ , identity 26%, query coverage = 46%; CD0410: *e*-value =  $1 \times 10^{-18}$ , identity = 31%, query coverage = 51%). The sequence homology of these putative loader proteins with

their counterpart in *B. subtilis* is mainly confined to the C-terminal AAA+ domain that contains the Walker A and B motifs [59]. CD0410 is located on the conjugative transposon CTn2 [67] and therefore not part of the core genome of *C. difficile*; this makes a role as a chromosomal replication initiation protein unlikely. The *dnaD*-like gene CD3653 is located adjacent to the putative loader (CD3654), in the same genomic region as the replicative helicase (CD3657) (figure 1b). Of note, the *dnaB* gene of *B. subtilis* is located next to the gene helicase loader *dnaI* [68], suggesting a functional relationship between the loader ATPase and a DnaB\_2 family protein. Indeed, it has been suggested that BsDnaB is a co-loader of the BsDnaC helicase [14]. The genomic region containing CD3654 and CD3657 is conserved and essential [69], and our analyses therefore strongly suggest that CD3654 is the cognate loader protein for the *C. difficile* replicative helicase CD3657.

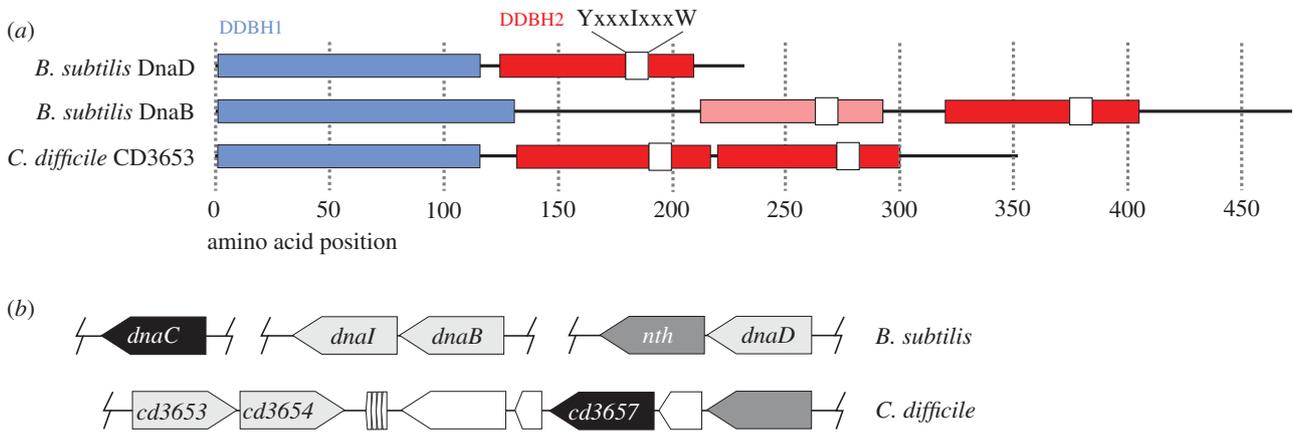
A BLASTP query identified a single protein homologous to *B. subtilis* primase BsDnaG. This protein, CD1454, shared 31% identity with its *B. subtilis* counterpart across the full length of the protein (*e*-value =  $9 \times 10^{-98}$ ), and contains all domains expected for a bacterial primase (electronic supplementary material, figure S1). Thus, CD1454 probably encodes the primase for chromosomal DNA replication.

### 2.2. Helicase can form hexamers at high concentration

A distinguishing feature of the different modes of helicase loading (ring-maker versus ring-breaker) is the multimeric state of helicase at dilute concentrations of protein [6]. Therefore, we purified recombinant *C. difficile* helicase protein and determined its multimeric state using analytical gel filtration. At concentrations below 5  $\mu$ M, we observed predominantly monomeric protein, with a fraction of the protein forming low molecular weight (MW) complexes (probably dimers or trimers) while at 10  $\mu$ M and above, the helicase formed larger oligomers, most probably hexamers (electronic supplementary material, figure S2). Thus, at physiological (nM) concentrations the *C. difficile* helicase is predominantly monomeric, suggesting it is of the ring-maker type, like *B. subtilis*. Multimerization at high concentrations of protein was independent of the presence of ATP (G.S.B., P.S., V.P. & M.G. 2014, unpublished data).

To confirm the gel-filtration data, we investigated the self-interaction of helicase in a bacterial two-hybrid system [70]. This system detects interactions between a protein fused to Zif (Zinc-finger DNA binding domain) and a protein fused to the RNA polymerase  $\omega$  subunit. Interaction between proteins of interest facilitates transcriptional activation of a Zif-dependent *lacZ* reporter gene in a dedicated reporter strain [70]. In order to quantify the interaction, *E. coli* cells containing the plasmids encoding the fusion proteins were permeabilized and assayed for  $\beta$ -galactosidase activity. We found that in the reporter strain transformed with plasmids harbouring both fusion proteins,  $\beta$ -galactosidase activity was approximately threefold higher than for the reporter strains harbouring the individual plasmids, indicating a clear self-interaction for the *C. difficile* helicase protein (figure 2a).

Similar experiments were carried out with the putative helicase loader (CD3654). Analytical gel filtration using purified loader protein (CD3654) showed that it was monomeric at all concentrations tested (P.S., V.P. & M.G. 2014,



**Figure 1.** *In silico* analysis of putative replication initiation proteins of *C. difficile*. (a) Domain structure of BsDnaB, BsDnaD and CD3653. Domain nomenclature according to Marston *et al.* [18]. Note that DDBH2 corresponds to PFAM DnaB\_2. (b) Chromosomal organization of the *dnaBI* genomic region of *B. subtilis* and the CD3653-CD3654 genomic region of *C. difficile*.

unpublished data). Consistent with this observation, no self-interaction of CD3654 was found in the bacterial two-hybrid system (figure 2b).

We conclude that helicase can form homomultimeric assemblies, whereas the putative loader is monomeric under the conditions tested.

### 2.3. Helicase and the putative helicase loader interact in an ATP-dependent manner

If CD3654 is a legitimate loader for the *C. difficile* helicase (CD3657), we expect that the proteins interact *in vivo* and *in vitro*. To determine if this is the case, we performed bacterial two-hybrid and analytical gel-filtration experiments. First, we tested if an interaction between the helicase CD3657 and the putative loader CD3654 could be demonstrated in the bacterial two-hybrid system. CD3657 was fused to the RNA polymerase  $\omega$  subunit and CD3654 was fused to Zif (Karna 2010). The  $\beta$ -galactosidase activity in the reporter strain containing both plasmids was approximately fivefold increased ( $p < 0.001$ ) compared with the reporter strains containing the individual plasmids (background) (figure 2c). The highly significant increase in  $\beta$ -galactosidase activity implies substantial interactions between the *C. difficile* helicase and putative helicase loader. Similar results were obtained when CD3657 was fused to Zif and CD3654 to the RNA polymerase  $\omega$  subunit (E.v.E., A.F. & W.K.S. 2015, unpublished data). This suggests that the combination of protein and fusion domain does not influence the results of this assay.

To exclude the possibility for false negative or false positive results as a result of the two-hybrid system, we additionally performed analytical gel-filtration experiments using purified non-tagged CD3657 and CD3654 proteins. In these experiments, *C. difficile* helicase and loader were combined in equimolar concentrations (2.21  $\mu$ M of monomer) in the presence and absence of ATP (1 mM). In the presence of ATP, the elution profile showed a major high MW peak (approx. 10 ml; approx. 500 kDa, P1) and a minor low MW peak (approx. 15 ml; approx. 40 kDa, P2) (figure 2d). In combination with a visual inspection of the fractions collected from both peaks on an Coomassie-stained SDS-PAGE gel, we believe that the major peak can be attributed to a large complex (most probably a dodecameric assembly consisting

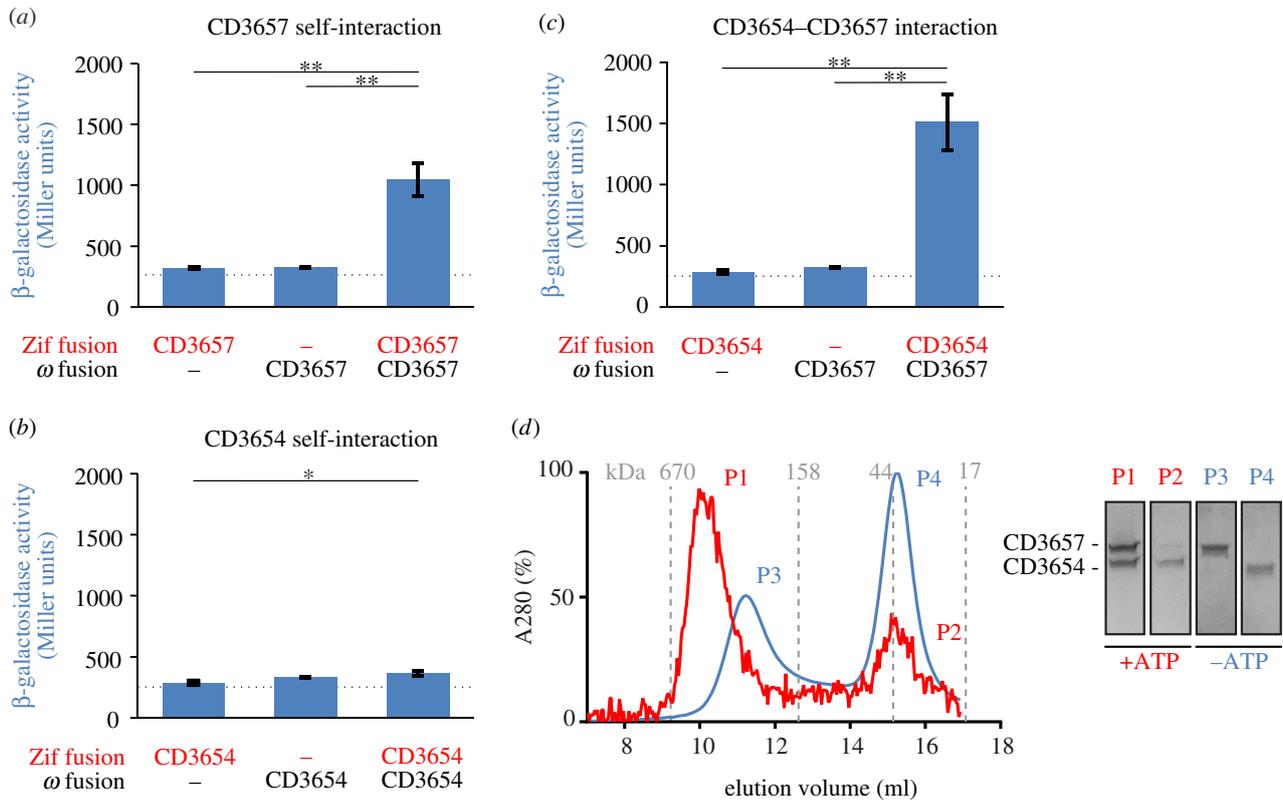
of six CD3657 monomers and six CD3654 monomers; theoretical MW 522 kDa), while the minor peak corresponds to predominantly free monomeric CD3654 (theoretical MW 38 kDa). Similar results were obtained when a high concentrations of proteins (approx. 10  $\mu$ M) were used (electronic supplementary material, figure S3), suggesting that preformed hexameric helicase retains the ability to interact with the CD3654 protein at the same stoichiometry.

The elution profile of the same concentration of proteins in the absence of ATP showed a completely different picture (figure 2d). A minor peak was observed at approximately 11 ml (approx. 300 kDa, P3) and a major second peak eluted at approximately 15 ml (approx. 40 kDa, P4). MW estimates, in combination with an evaluation of peak fractions on an SDS-PAGE gel, indicated that the first peak most probably corresponds to a complex of six monomers of CD3657 (theoretical MW 297 kDa), while the second peak corresponds to monomeric CD3654 protein.

Together, the data show that the CD3657 helicase and the loader CD3654 can form a complex in an ATP-dependent manner.

### 2.4. Mutation of the helicase Walker A motif abrogates protein–protein interactions

To address the question of which of the proteins (or whether both) require ATP to promote the formation of a CD3657-CD3654 complex, mutants in the Walker A motif of both proteins were created. The Walker A motif (GXXXXGK [T/S]) directly and indirectly interacts with ATP, and is the principal ATP-binding motif of P-loop ATPases [26]. The motif is highly conserved in both helicase and helicase-loader proteins [7]. The lysine residue (K) forms a direct interaction with the negatively charged nucleotide  $\beta$  or  $\gamma$  phosphate group, and mutation of this residue is known to abrogate nucleotide binding and lead to inactivation of P-loop ATPases. The threonine (T) residue in the Walker A motif either directly or indirectly coordinates a  $Mg^{2+}$  within the ATP-binding site, which in turn coordinates the phosphate groups of ATP. In the *G. stearothermophilus* replicative helicase, mutation of the threonine residue results in a protein that lacks ATPase and unwinding activities [59]. Based on this knowledge, the equivalent residues were identified in



**Figure 2.** CD3654 and CD3657 interact. (a) Bacterial two-hybrid analysis of CD3657 self-interaction. (b) Bacterial two-hybrid analysis of CD3654 self-interaction. (c) Bacterial two-hybrid analysis of the CD3657–CD3654 interaction. Bar graphs in (a–c) indicate average values and error bars indicate standard deviation of the measurements ( $n = 3$ ). Significance was determined using the Student's  $t$ -test ( $*p < 0.05$ ,  $**p < 0.001$ ). (d) CD3657 and CD3654 interact in an ATP-dependent manner. Analytical gel filtration was performed on a HiLoad 10/300 GL Superdex analytical grade size exclusion column with  $2.21 \mu\text{M}$  (monomer) of CD3657 and CD3654 in the presence (red) and absence (blue) of  $1 \text{ mM}$  ATP. Inset shows a Coomassie-stained SDS-PAGE gel of the numbered peak fractions. Dashed line indicates the maximum background level of  $\beta$ -galactosidase expression observed in our experimental set-up.

the *C. difficile* helicase protein. Using site-directed mutagenesis, we generated mutant helicase proteins in which the lysine at position 214 was changed into an arginine (K214R) and the threonine at position 215 was changed into an alanine (T215A).

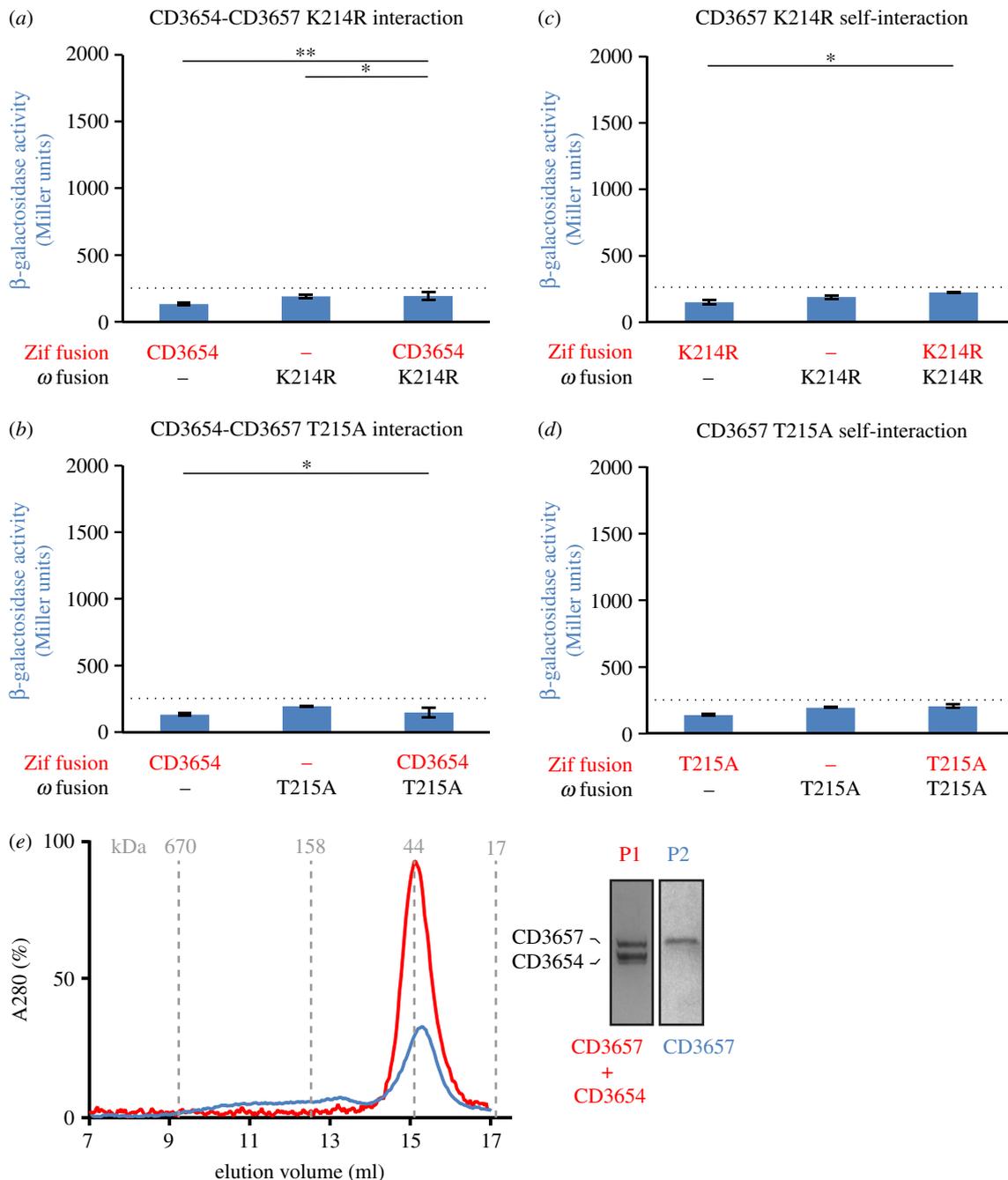
To determine whether these CD3657 proteins showed altered protein–protein interactions, we performed bacterial two-hybrid experiments. We fused the CD3657 protein to the  $\omega$ -subunit and CD3654 or CD3657 proteins to the Zif subunit and evaluated their ability to drive the expression of a transcriptional reporter. For both the CD3657 K214R and CD3657 T215A mutants, interaction with the wild-type CD3654 protein were severely reduced or completely lost (figure 3*a,b*). This prompted us to investigate if the mutant helicases still had the capacity to form homomultimeric assemblies, as it has been shown in other bacteria that oligomerization is an important step in the mechanism of action of DNA helicases [71]. We found that the CD3657 K214R also demonstrated reduced (possibly absent) self-interactions, whereas the CD3657 T215A mutant had completely lost the ability to self-interact (figure 3*c,d*). We conclude that the ability of the CD3657 helicase to coordinate ATP correlates with its ability to interact with the putative loader CD3654 and the ability to self-interact. As the most dramatic effect was observed for CD3657 T215A, we focused our further experiments on this particular mutant.

To confirm the findings from the bacterial two-hybrid experiments, we purified CD3657 T215A protein and subjected it to size exclusion chromatography (figure 3*e*). *Clostridium difficile* helicase T215A mutant ( $2.43 \mu\text{M}$ ) was

incubated in the presence of ATP ( $1 \text{ mM}$ ) and loaded onto a size exclusion column. A major peak was observed at approximately  $15 \text{ ml}$  (approx.  $40 \text{ kDa}$ ), probably corresponding to monomeric CD3657 T215A (theoretical MW  $49 \text{ kDa}$ ). We did not observe any high MW complexes under these conditions, in contrast with the wild-type CD3657 protein (figure 2*d*). Next, we combined the CD3657 T215A mutant and the wild-type CD3654 protein (both  $2.43 \mu\text{M}$ ) in the presence of ATP ( $1 \text{ mM}$ ) (figure 3*e*). A single peak was observed at approximately  $15 \text{ ml}$  (approx.  $40 \text{ kDa}$ ). Analysis of the peak fractions on SDS-PAGE demonstrated that the peak contained both CD3657 T215A and CD3654 protein, and thus corresponds to monomeric forms of both proteins (theoretical MW  $49$  and  $38 \text{ kDa}$ , respectively). Also in these experiments, no high MW complexes were found, in contrast with the wild-type CD3657 and CD3654 proteins (figure 2*d*).

We also generated constructs with Walker A mutations in CD3654 (K198R, T199A) and tested these in the bacterial two-hybrid assay and in size exclusion chromatography. The mutant loader proteins retained the ability to interact with the wild-type helicase protein (electronic supplementary material, figure S4). Self-interaction was not observed for the mutant loader proteins, concordant with the results obtained with the wild-type loader protein (electronic supplementary material, figure S5).

Overall, our data show that the Walker A mutant CD3657 T215A can no longer self-interact and has lost the capacity to interact with the putative loader protein CD3654. We conclude that the ATP requirement for the interaction between



**Figure 3.** Walker A mutants of CD3657 are defective in protein–protein interactions. (a,b) Walker A mutants of CD3657 ((a) K214R; (b) T215A) show no or severely reduced interactions with the putative loader protein CD3654 in a bacterial two-hybrid assay. (c,d) Walker A mutants of CD3657 ((c) K214R; (d) T215A) no longer self-interact in a bacterial two-hybrid assay. Bar graphs in (a–d) indicate average values and error bars indicate standard deviation of the measurements ( $n = 3$ ). Significance was determined using the Student's  $t$ -test (\* $p < 0.05$ , \*\* $p < 0.001$ ). (e) CD3657 T215A is defective in protein–protein interactions. Analytical gel filtration was performed on a HiLoad 10/300 GL Superdex 200 analytical grade size exclusion column with 2.43  $\mu$ M of CD3657 T215A in the presence (red) and absence (blue) of 2.43  $\mu$ M CD3654. Inset shows a Coomassie-stained SDS-PAGE gel of the numbered peak fractions. Dashed line indicates the maximum background level of  $\beta$ -galactosidase expression observed in our experimental set-up.

the two proteins is most likely to be the result of ATP binding to CD3657.

## 2.5. Helicase loading of *Clostridium difficile* differs from *Bacillus subtilis*

So far, our data show that the replicative helicase of *C. difficile*, CD3657, interacts in an ATP-dependent manner with the helicase-loader protein CD3654 and that loading probably occurs via a ring-maker mechanism, as for *B. subtilis*.

In *B. subtilis*, stimulation of activity of the (monomeric) helicase protein by the loader protein was clearly shown using an *in vitro* helicase activity assay [17,57]. Therefore, we set out to investigate if the activity of *C. difficile* helicase could be reconstituted in the presence of the CD3654 protein in a similar experiment. DNA-unwinding helicase activity was assayed by monitoring and quantifying the displacement of a radiolabelled oligonucleotide annealed to single-stranded circular M13mp18 DNA. To enable loading of helicase, the 5' end of the oligonucleotide contained a poly(dCA) tail that produces a forked substrate upon annealing of the

complementary region to ssM13. Wild-type CD3657 and CD3654 proteins were mixed in equimolar concentrations (monomers) in the presence of ATP and reaction buffer, and displacement of the radiolabelled oligonucleotide was monitored over time. In contrast with the *B. subtilis* proteins, helicase activity was not observed during this time course (electronic supplementary material, figure S6). This suggests that another factor is required for *in vitro* loading and/or activation of the *C. difficile* helicase.

## 2.6. *Clostridium difficile* helicase is activated by primase

Primase has been shown to interact with helicase and stimulate its activity in a variety of organisms [16,57,61]. Therefore, we wanted to investigate if *C. difficile* helicase could be activated by primase. We purified the *C. difficile* primase without an affinity tag to more than 95% purity. SEC-MALS/dRI analysis showed a single peak corresponding to a MW of 74 kDa, indicating that the primase is monomeric under the conditions tested (expected MW 70 kDa) (electronic supplementary material, figure S7). We performed helicase assays to determine the effect of primase on helicase activity.

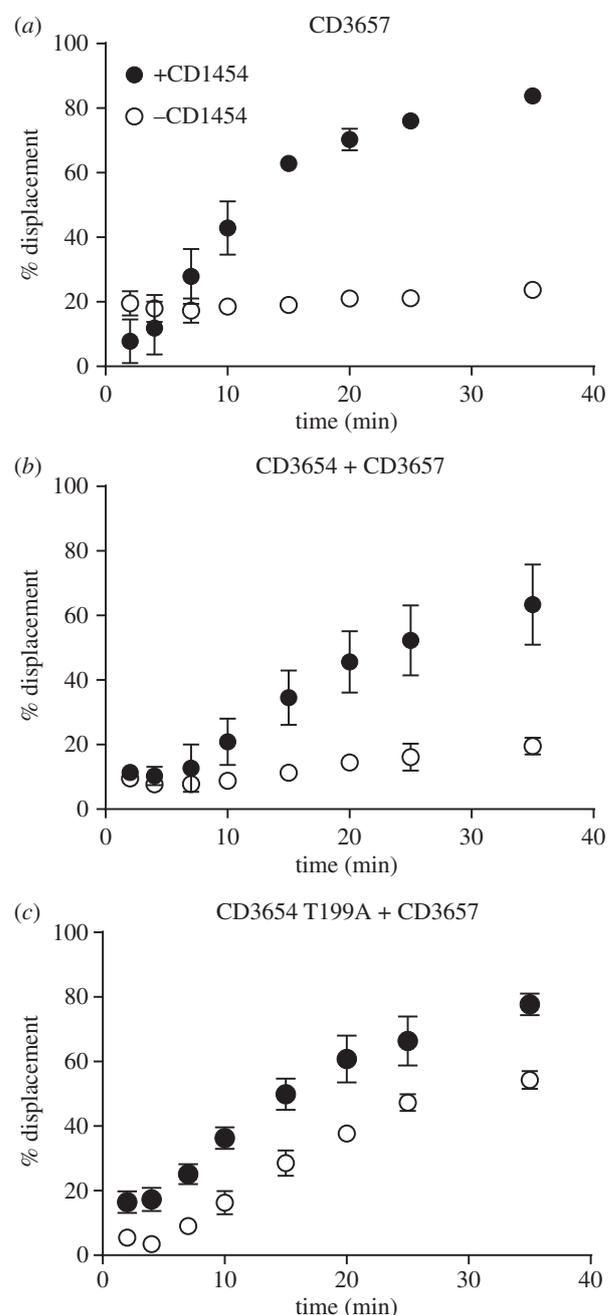
First, we determined if the CD3657 helicase displayed activity in this particular assay in the absence of any other proteins. Approximately 20% of the oligonucleotide was displaced after as little as 2 min of incubation, and this fraction remained stable over the course of 35 min (figure 4a). This suggests 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 of monomer, a 3.5-fold increase in displacement of the oligonucleotide (up to approx. 80%) was observed after 35 min (figure 4a). 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 with helicase alone at early time points (less than 10 min). 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 (P.S., V.P. & M.G. 2015, unpublished data).

The results from the helicase assay suggest a functional interaction between the helicase and primase proteins. We tried to validate the interaction using a bacterial two-hybrid system. Despite our efforts, no interaction could be demonstrated between the full-length primase and helicase proteins, or the helicase-interacting domain of primase and helicase in bacterial two-hybrid experiments (E.v.E., A.F. & W.K.S. 2015, unpublished data). 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* [47], or that the bacterial two-hybrid system does not allow recapitulation of the conditions necessary for the interaction between the helicase and primase proteins.

## 2.7. 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



**Figure 4.** Helicase activity is stimulated by primase. Helicase activity was assayed by quantifying the displacement of a radiolabelled ( $\gamma^{32}\text{P}$ -ATP) oligonucleotide partially annealed to the single stranded circular DNA m13mp18. Per cent displaced signal from the helicase assays in time. (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. The other mutants of CD3654 (K198R, D258Q) gave similar results (electronic supplementary material, figure S7) but have been omitted for clarity. Error bars indicate standard deviation ( $n = 3$ ).

absence of primase. We observed a very low level of displacement (less than 20%) in our helicase assay in the presence of both helicase and the putative loader (figure 4b). Notably, within the first 10 min of the assay the fraction of displaced oligonucleotide did not exceed 10%, in contrast with the situation with helicase alone where it reached 20% (figure 4a). At endpoint, 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 min on, a clear stimulation of helicase activity by primase was observed, reaching approximately 60% displacement over a time course of 35 min (figure 4b). Interestingly, the addition of the putative loader protein resulted in a 20% reduction in strand displacement over a time course of 35 min compared with results obtained with a combination of helicase and primase (approx. 80%; figure 4a). We conclude that primase can activate helicase activity in the presence of the putative loader, but that 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 CD3657 K214R and T215A mutants. We observed only background levels (less than 5%) of displacement in the presence of CD3657 mutant helicases, primase and the putative loader (electronic supplementary material, figure S8) in comparison with approximately 60% strand displacement for the wild-type helicase in the presence of the same proteins (figure 4b). This shows that the displacement observed in our experiments can be attributed to helicase alone and not some other factor.

In *E. coli*, interaction between the loader ATPase protein and helicase does not require ATP binding by the loader [8,72,73]. 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 [8,72,73]. Therefore, we set out to investigate the effect of CD3654 proteins with a mutated Walker A or Walker B motif on helicase activity. 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<sup>+</sup> proteins, resulting in inactivation [26]. 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<sup>2+</sup> 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 (approx. 50%) of the oligonucleotide in the helicase assays with all three CD3654 Walker A and B mutants (figure 4c; electronic supplementary material, figure S9) compared with the wild-type (figure 4b) 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 (approx. 80%; figure 4c) to the activity measured in the two protein helicase–primase experiment (figure 4a). This suggests 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.

## 2.8. *Clostridium 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,

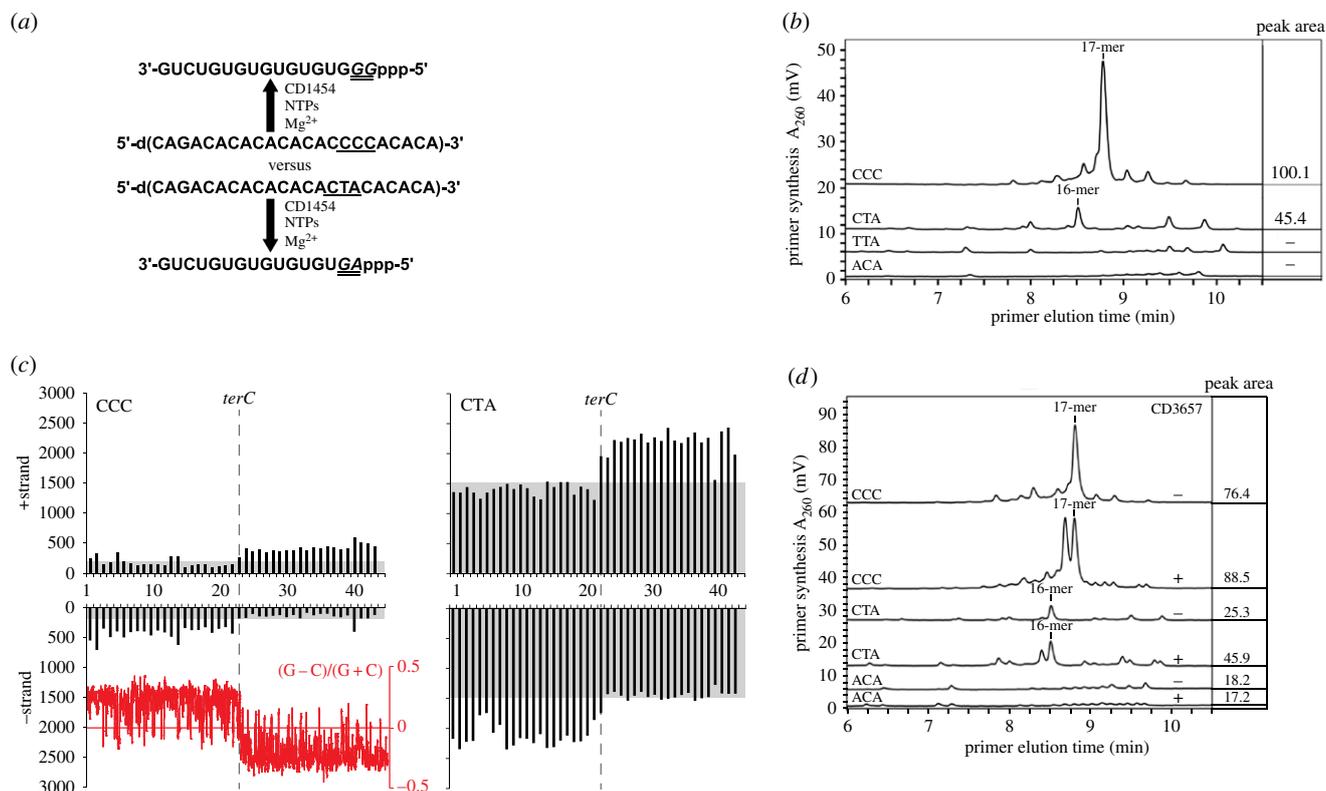
which has not been reported for Gram-positive bacteria before. Next, we sought to evaluate primase activity. To confirm enzymatic activity and identify the template initiation motif, two 50-mer oligonucleotides that comprised all 64 potential trinucleotide sequences were used in a priming assay, as previously described [74]. These results confirmed enzymatic activity and revealed likely trinucleotide motifs that supported de novo primer synthesis (M.A.L. 2014, unpublished data). Next, a more detailed analysis of template specificity was performed using a 23-mer ssDNA-template containing the trinucleotide of interest (figure 5a). We found that the CD1454 primase preferentially initiated de novo RNA primer synthesis on the initial 5'-d(CCC) motif encountered by this enzyme within the 23-mer ssDNA template; therefore, a 17-mer rather than an anticipated 16-mer RNA primer was produced (figure 5b). 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 5b). The activity of the CD1454 was notably inhibited by Mg<sup>2+</sup> concentrations more than 30 mM, and largely independent of NTP concentration (more than 1 mM) under the conditions of the assay (electronic supplementary material, figure S10).

The observed template specificity for the CD1454 primase differs from most other bacterial primases evaluated to date. More specifically, 5'-d(CTA) only supported minimal priming and no initiation occurred on 5'-d(TTA) by the *C. difficile* primase (figure 5b), whereas substantial dinucleotide polymerization occurred on these two trinucleotides by primases from other Firmicutes such as *G. stearothermophilus*, *S. aureus* and *Bacillus anthracis* [44,53,76]. The trinucleotide 5'-d(CTG) was also not an effective template for *C. difficile* primase (M.A.L. 2014, unpublished data), whereas it is efficiently primed by primase from Gammaproteobacteria such as *E. coli*, *Pseudomonas aeruginosa* and *Yersinia pestis* [76–78]. To date, only the primase from the hyperthermophile *Aquifex aeolicus* has been shown to initiate primer synthesis on the trinucleotide 5'-d(CCC) [74].

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 (electronic supplementary material, figure S11). 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 (electronic supplementary material, figure S11). 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.

## 2.9. 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% [65], and other low-GC Firmicutes, such as



**Figure 5.** CD3657 affects trinucleotide specificity of CD1454. (a) Schematic depiction of the primase activity assay using two different single-stranded DNA templates, along with the resulting full-length RNA primer. (b) Analysis of CD1454 priming activity at different trinucleotides. (c) 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 [75] with a window size of 5000. (d) CD3657 affects CD1454 primase activity. The amount of primer produced by CD1454 in the presence (+) or absence (-) of CD3657 on ssDNA templates with different trinucleotides at stoichiometric concentrations of [CD3657]<sub>6</sub>:[CD1454]<sub>3</sub>. The numbers in the panel to the right of the chromatograms in (b) and (d) denote the total peak area for the RNA primer products synthesized and shown in the associated chromatogram to the left.

*S. aureus* and *G. stearothermophilus* [44,53], preferentially initiate at 5'-d(CTA). Therefore, we evaluated the relative frequency of the CCC motif on the plus and minus strand of the *C. difficile* chromosome and compared it with the relative frequency of CTA motif on which the CD1454 initiates less efficiently (figure 5c). Our analysis showed that CTA trinucleotides were on average fivefold to 10-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, which mirrors the GC-skew ( $[G - C]/[G + C]$ ) of the chromosome (figure 5c). This suggests that the motifs are preferentially associated with the lagging strand where primase acts and indicates 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* [48,49,51], 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 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 versus 25.3 nMole) at stoichiometric concentrations of [CD3657]<sub>6</sub>:[CD1454]<sub>3</sub>. Helicase-stimulated primase activity on 5'-d(CCC) increased RNA primer synthesis by only 1.15-fold and, as expected, no stimulation of RNA primer synthesis occurred on the 5'-d(ACA) trinucleotide

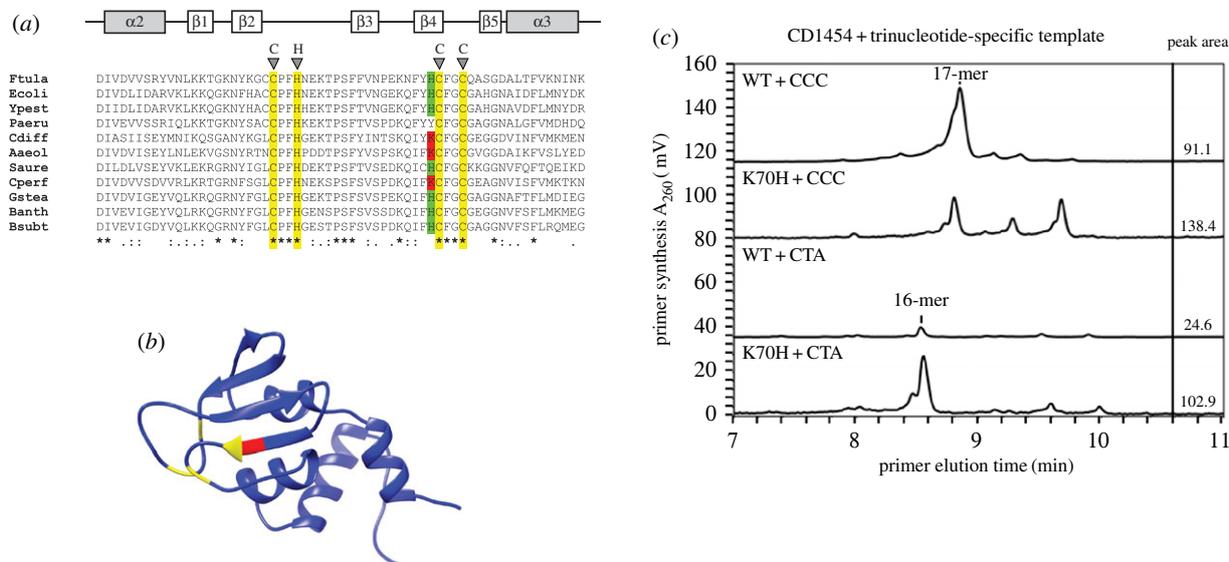
(figure 5d). We also observed that helicase stimulated primase to synthesize slightly shorter RNA polymers (figure 5d). The production of a shorter primer might allow faster transfer to the replicative polymerase, increasing the replication speed. Collectively, our data show that the stimulatory effect of helicase on primase activity *in vitro* is clearly enhanced on the trinucleotide 5'-d(CTA), which is preferred by other Firmicute primases studied to date, 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.

## 2.10. A lysine residue contributes to trinucleotide specificity of primase

Considering that the *C. difficile* primase trinucleotide specificity (figure 5b) is unusual for Firmicutes, but resembles that of the hyperthermophilic *A. aeolicus* primase [74], 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 ZBD that is involved in sequence specific DNA binding of primase in *C. difficile* and 10 other bacterial species (*Clostridium perfringens*, *G. stearothermophilus*, *S. aureus*, *B. anthracis*,



**Figure 6.** A lysine residue affect trinucleotide specificity of CD1454. (a) Alignment of the  $\alpha 2$ – $\alpha 3$  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*; *Faeru*, *Pseudomonas aeruginosa*; *Cdiff*, *Clostridium difficile*; *Aeol*, *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) Ribbon representation of the CD1454 primase zinc-binding domain 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 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 synthesized and 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.

*B. subtilis*, *E. coli*, *P. aeruginosa*, *Y. pestis*, *A. aeolicus* and *Francisella tularensis*). *Clostridium 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 (<http://www.ebi.ac.uk/Tools/msa/clustalo>) 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 other bacterial 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 primase 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 [79] revealed that this particular lysine residue is in close proximity to the zinc-ribbon motif that tetrahedrally coordinates a zinc ion and is essential for primase function (figure 6b; electronic supplementary material, figure S1) [80]. Interestingly, the PDB template used by Phyre2 is 2AU3, the *A. aeolicus* primase [54]. The crystal structure of *G. stearothermophilus* DnaG primase contains a histidine at this position [50,80] and the multiple sequence alignment suggests that other Firmicutes also do. We hypothesized that the exposed lysine residue influences primase initiation specificity. 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 reduced priming on the 5'-d(CCC) motif by the CD1454 K70H mutant compared with the wild-type CD1454 primase. Indeed, CD1454 K70H showed substantially reduced initiation on the 5'-d(CCC) motif and enhanced priming on the 5'-d(CTA) motif (figure 6c). The CD1454 K70H synthesized slightly more primers than wild-type primase due to relaxed template specificity, as evidenced by the synthesis of RNA polymers longer than 17 nucleotides in length (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.

### 3. Discussion

*In silico* analysis of the *C. difficile* genome failed to identify a homologue of DnaB of *B. subtilis*, which is generally considered the model for Gram-positive bacteria. BsDnaB, together with BsDnaD and BsDnaI, is strictly required for helicase loading in *B. subtilis in vivo* [12–14]. A homologue of BsDnaD was identified that may be involved in DNA replication (CD3653; *e*-value  $4 \times 10^{-5}$ ), although query coverage (47%) and identity (29%) were low. This situation is reminiscent of that in some Mollicutes, where also only a *DnaD*-like gene was identified [15]. Despite a lack of clear homology at the primary amino acid sequence level, BsDnaB and BsDnaD are structural homologues [18]. Fusions of these proteins, are found in phage-related replication proteins and it was suggested that in the absence of DnaB, a single fusion protein may couple or combine both functions [15,66]. Nevertheless, the situation in

*C. difficile* differs from those in phage and Mollicutes. Structure predictions reveal that the phage-related and the Mollicutes DnaD-like proteins have a two-domain structure containing one copy of the DDBH1 and DDBH2 domain [18], and the proposed hybrid function of phage proteins is based on limited local amino acid sequence similarity in the DDBH2 domain only [66]. CD3653 on the other hand has a three-domain structure with a single DDBH1 and two DDBH2 domains, like DnaB, despite the lack of clear sequence similarity to this protein (figure 1). It is tempting to speculate that CD3653 in *C. difficile* may perform functions similar to both DnaD and DnaB in *B. subtilis*, which include origin remodelling and contributing to the helicase loading process [14,15,81]. DDBH2 domains are characterized by an YxxxIxxxW motif [18]. In BsDnaB, this motif is degenerate in the first DDBH2 domain. By contrast, the motif is readily identified in both DDBH2 domains of CD3653 (figure 1). Our data are consistent with a model where an ancestral three-domain DnaD-like protein was duplicated and subsequently diverged in certain Firmicutes like *B. subtilis*.

DnaB-like helicases (note that this nomenclature is based on the *E. coli* protein name) belong to the superfamily IV of DNA helicases, and the functional unit of this protein is a hexamer [26–28,34]. In *E. coli*, the helicase is found to be a stable hexamer over a broad protein concentration range of 0.1–10  $\mu$ M [82] and is active as a pre-formed multimer. Helicases belonging to the ring-maker class, such as BsDnaC of *B. subtilis*, can occur in a low oligomeric or monomeric state under dilute conditions [6,14]. Our experiments indicated that CD3657 is monomeric in the low micromolar or nanomolar range (electronic supplementary material, figure S2), which is likely to be reflective of the intracellular concentration of protein [72]. *Clostridium difficile* CD3657 can form hexameric assemblies at higher concentrations (electronic supplementary material, figures S2 and S3), but these pre-formed hexamers are inactive, even in the presence of primase (G.S.B. & P.S. 2012, unpublished data), in contrast with the situation in *E. coli*. Our data are therefore consistent with the notion that CD3657 belongs to the ring-maker class of helicases [6].

The interaction between putative helicase and loader protein was found to be ATP dependent (figures 2 and 3). Mutations in the conserved Walker A and B motifs of the putative loader protein did not fully abrogate the interaction with the wild-type helicase (electronic supplementary material, figure S4). Similarly, in *E. coli*, nucleotide binding to the helicase loader was not a pre-requisite for association with helicase [8,72,83]. Instead, our data indicate that the association of ATP with helicase is crucial for the interaction with the loader protein (figure 3). Notably, there is a correlation between the ability of the helicase to interact with the loader protein and form homohexamers, because a T215A (Walker A) mutant of helicase is defective for both, at least under dilute concentrations of helicase (figure 3). By contrast, the equivalent mutation in *G. stearothermophilus* helicase (T217A) does not affect its ability to form hexamers [59], and the interaction of this protein with *B. subtilis* DnaI readily occurs in the absence of ATP [84].

We attribute the effect on protein–protein interactions of the Walker A mutants of CD3657 to defects in ATP binding rather than hydrolysis (figure 3); a Walker B mutant (CD3657 D318A), but not a mutant that is predicted to be able to bind but not hydrolyse ATP (CD3657 E239A), mirrors our findings with the Walker A mutants in a bacterial two-hybrid assay (E.v.E., A.H.F. & W.K.S. 2015, unpublished data). Binding of

the nucleotide to helicase is associated with conformational changes; the N-terminal collar domain constricts upon nucleotide binding in *A. aeolicus* and to a lesser extent *E. coli*. This constricted conformation is believed to favour an interaction with the loader protein [85] and it is therefore conceivable that in the absence of ATP, the *C. difficile* helicase adopts a (dilated) conformation that is incompatible with a functional interaction with the putative loader protein.

Despite substantial bioinformatic and biochemical evidence that CD3654 is indeed the *C. difficile* helicase loader, we found an inhibitory effect of CD3654 on the activity of CD3657 *in vitro*, at least in the presence of CD1454 (figure 4). At first, the negative effect of CD3654 seems at odds with its proposed role as loader. However, in *E. coli* ATP-bound EcDnaC loader protein can act as an inhibitor of the EcDnaB helicase [8,72,73]. We also find that mutants of CD3654 appear capable of loading and/or activating the helicase, but are defective in restraining the DNA-unwinding activity of the helicase like wild-type CD3654 (figure 4c; electronic supplementary material, figure S9). Moreover, 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 homologue of *cd3654* in an epidemic strain of *C. difficile* (R20291\_3513) [69].

Our data strongly suggest that the presence of loader protein alone is not sufficient to load and activate the helicase of *C. difficile*, and that at least one other factor is needed to reconstitute its activity. Indeed, primase is required as an activator of helicase activity in the presence of the loader (figure 4).

Helicases are complex proteins, and their properties can both alter and be altered by other replication factors [85]. DnaB-like helicases consist of two-tiered homo-hexameric rings, one assembled from six subunits of the C-terminal domain and the other formed by the N-terminal domains. The helicase loader interacts with the C-terminal ATPase domain [62,72,86], and the same domain is required for the interaction with the  $\tau$  subunit of the clamp loader in *E. coli* [87] and *B. subtilis* [88,89]. Strikingly, BsDnaC T217A Walker A mutant fails to form a complex with  $\tau$  [89]. This finding is very similar to our observations for the interaction between the *C. difficile* helicase and loader.

The N-terminal domain of helicase forms a platform for the interaction with primase in both *E. coli* and *G. stearothermophilus* [62,85]. Unlike the helicase loader, binding of primase to helicase is promoted by a dilated conformation of the N-terminal domain that exposes the interaction surface [58,85]. The helicase–primase interaction is mutually stimulatory, with distinct but overlapping networks of residues in helicase responsible for the modulation of either helicase or primase activity [44,90,91]. Primase binding counteracts the binding of the loader protein in *E. coli* [61], and the loader from *B. subtilis* was found to dissociate from the complex when primase and polymerase bind to helicase [57]. We find that CD1454 stimulates CD3657 helicase activity, independent from the loader (figure 4).

Loading of the CD3657 helicase differs from the situation in other Gram-positive bacteria. 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). *Geobacillus stearothermophilus* helicase demonstrates significant helicase activity by itself [59], and the *B. subtilis* helicase is strongly activated by its cognate loader

[17,57]. Instead, helicase loading in *C. difficile* is somewhat reminiscent of the situation in the Gram-negative *H. pylori*, where a dodecameric self-loading helicase remains inactive until activated by primase, leading to the dissociation of the dodecamer into two hexamers [16]. However, we did not find any evidence for dodecameric assemblies for CD3657.

Helicase loading in *C. difficile* also seems to differ in crucial aspects from the Gram-negative model *E. coli*. Walker A mutants of the EcDnaC loader protein are capable of loading the EcDnaB helicase but do not sustain helicase activity, suggesting that ATP turnover by the loader is required to release the helicase [8]. Our data show that in *C. difficile*, (self) loading of CD3657 is stimulated by Walker A mutants of CD3654, 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.

Further, EcDnaB helicase is in complex with the EcDnaC loader protein throughout the loading process and remains inactive until the EcDnaG primase binds to helicase, thereby releasing the loader protein [61]. 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 by CD1454 is observed in the absence of CD3654 (figure 4a). 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, but stabilization of the hexamer or other conformational changes in the hexameric helicase induced by primase have also been observed in *E. coli* [41,58,60,61].

We found that the primase of *C. difficile* has an unusual trinucleotide specificity, with a preference for 5'-d(CCC) (figure 5), which is similar to primase of *A. aeolicus* [74] but different from the specificity of primases from other Firmicutes or Gammaproteobacteria (5'-d(CTA), 5'-d(TTA), 5'-d(CTG)) [50]. The first two nucleotides in all of these trinucleotides are pyrimidines (C or T) and as a result 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 [92–95]. The nucleotide preference might couple the efficiency of lagging strand DNA synthesis and the 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 (electronic supplementary material, figure S11), consistent with a previously hypothesized context-dependent enzyme activity [74]. Our results indicate that pyrimidines, probably in part due to the smaller size compared with 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 6). Based on this finding, we would predict a similar specificity for the primase of *C. perfringens*, but our attempts to obtain priming activity with the CPF\_2265 primase from *C. perfringens* ATCC133124 have failed thus far.

A 5'-d(CCC) trinucleotide specificity has previously only been observed for *A. aeolicus* [74]. 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 [96,97]. These studies concluded that the Firmicutes are presumably among the most ancient bacteria and that the Aquificales have diverged much later in evolution [96]. Indeed, an analysis of the GC content of rRNA clusters suggests that hyperthermophilic species have evolved from mesophilic organisms via adaptation to high temperature [98] and that the Gram-negative double membrane may have been derived from sporulating Gram-positives [99,100]. Thus, the study of *C. difficile* may also contribute to a better understanding of the evolution of the bacterial DNA replication machinery and facilitate the development of antibiotics that target these essential proteins in pathogenic bacteria.

## 4. Material and methods

### 4.1. Plasmid construction and site-directed mutagenesis

All oligonucleotides and plasmids constructed for this study are listed in electronic supplementary material, tables S1 and S2.

To construct the CD3654, CD3657 and CD1454 expression plasmids, the open reading frames were cloned into suitable pET vectors (Novagen) to yield pEVE24, pEVE87 and pEVE7, respectively.

Construction of the plasmids for the bacterial two-hybrid system was performed with Gateway cloning technology (Invitrogen). To construct the CD3654 and CD3657 entry plasmids, the CD3654 and CD3657 open reading frames were introduced into donor vector pDonR<sup>TM</sup>201. Bacterial two-hybrid constructs were made by sub-cloning the genes of interest from the entry plasmids into the destination plasmids pKEK1286 (Zif fusion plasmid) or pKEK1287 ( $\omega$  fusion plasmid) in an LR reaction [70], to yield pEVE118, pEVE120, pEVE122, pEVE123, pEVE124 and pEVE125.

Mutations in plasmids carrying CD3654, CD3654 and CD1454 mutants were constructed according to the Quik-Change protocol (Stratagene). Primers were generated with Primer X, a web-based tool for automatic design of mutagenic primers for site-directed mutagenesis.

All constructs were verified by DNA sequencing.

### 4.2. Purification of proteins

Overexpression of all proteins was carried out in *E. coli* BL21 (DE3) from the pET-based vectors pEVE12, pEVE90, pEVE92, pEVE24, pEVE59, pEVE60, pEVE203, pEVE7 and pEVE201 (electronic supplementary material, appendix table S2). Detailed description of the protein purifications is available as the electronic supplementary material. Below, we provide a short summary of the purification process.

Hexameric CD3657 was purified on a Q sepharose column, a heparin sepharose column and a Hiload 26/60 Superdex 200 gel-filtration column. Monomers were recovered

after addition of guanidinium chloride solution in incremental steps up to 2M, followed by separation of monomers and multimers on a Hiload 26/60 Superdex 200 column and stored in TED50G20 buffer (Tris pH 7.5 50 mM, EDTA 1 mM, DTT 1 mM, NaCl 50 mM, 20% v/v glycerol) at  $-80^{\circ}\text{C}$  after extensive dialysis.

CD3654 was purified on from the flow-through of a Q sepharose and 5 ml SP sepharose columns connected in series, followed by a heparin sepharose and a Hiload 26/60 Superdex 200 gel-filtration column. Purified CD3654 was stored in TED50G10 buffer (Tris pH 7.5 50 mM, EDTA 1 mM, DTT 1 mM, NaCl 50 mM, 10% v/v glycerol) at  $-80^{\circ}\text{C}$ .

CD1454 was purified on HiTrap Q HP column in series with a 5 ml heparin sepharose column, followed by a MonoS column and a Hiload 26/60 Superdex 200 column. Protein was stored in TED100G buffer (50 mM Tris pH 7.5, 1 mM EDTA, 1 mM DTT, 100 mM NaCl, 10% v/v glycerol) at  $-80^{\circ}\text{C}$ .

Proteins were quantified by UV spectrophotometry and stored at  $-80^{\circ}\text{C}$ . Protein purity (more than 95%) was estimated by SDS-PAGE electrophoresis and concentration was determined spectrophotometrically using extinction coefficients calculated by the ExPASy ProtParam tool (<http://web.expasy.org/protparam>).

Protein concentrations mentioned in this manuscript refer to concentration of the monomer of the protein.

### 4.3. Gel-filtration experiments

Purified proteins were studied in the presence and absence of 1 mM ATP on a Hiload 10/300 GL Superdex 200 analytical grade size exclusion column at a flow rate of  $0.5\text{ ml min}^{-1}$ . The elution profiles from each experiment were monitored at 280 nm and plotted as a function of the elution volume. To assess interactions between CD3657 and CD3654, purified proteins were mixed in a 1 : 1 stoichiometry (based on monomer concentrations). Samples from fractions were analysed by SDS-PAGE and Coomassie Blue staining to verify the identity of the proteins. Gel filtration standards (Bio-Rad) were run to calculate MW estimates.

### 4.4. Bacterial two-hybrid assays

To determine (self-) interactions, pKEK1286- and pKEK1287-derived constructs were subsequently transformed in to the *E. coli* reporter strain KDZif1 $\Delta$ Z [70,101] and assayed for  $\beta$ -galactosidase activity. Experiments were performed in triplicate.

### 4.5. Helicase assays

Helicase activity was assayed by monitoring (and quantifying) the displacement of the radiolabelled ( $\gamma^{32}\text{P}$ -ATP) 105-mer oligonucleotide oVP-1 (partially) annealed to the single-stranded circular DNA m13mp18 (ssM13; Affymetrix)

essentially as previously described [57]. All reactions, containing 0.658 nM radiolabelled DNA substrate, were initiated by the addition of 2.5 mM ATP and carried out at  $37^{\circ}\text{C}$  in buffer containing 20 mM HEPES-NaOH (pH 7.5), 50 mM NaCl, 10 mM  $\text{MgCl}_2$  and 1 mM DTT for various lengths of time. In these reactions, proteins are in large excess (more than 250 $\times$ ) and thus not rate limiting. The reactions were terminated by adding 5 $\times$  SDS-STOP buffer (100 mM Tris pH 8.0, 200 mM EDTA, 2.5% w/v SDS, 50% v/v glycerol, 0.15% (w/v) bromophenol blue). Proteins were added sequentially, with a 5 min preincubation after each addition. The gel was dried, scanned and analysed using a molecular imager and associated software (Bio-Rad). Experiments were carried out in triplicate, and data analysis was performed using PRISM 6 (GraphPad Software).

## 4.6. RNA primer synthesis assay and thermally denaturing HPLC analysis

RNA priming assays and denaturing HPLC analyses were conducted as was previously described for other mesophilic bacterial primases [76]. The purified oligonucleotide templates used in this study were synthesized by Integrated DNA Technologies (Coralville, IA, USA). All RNA primer synthesis reactions were carried out in 50  $\mu\text{l}$  nuclease-free water containing 50 mM HEPES pH 7.5, 100 mM potassium glutamate, 10 mM DTT, 2  $\mu\text{M}$  ssDNA template, 30 mM magnesium acetate and 1.2 mM of each NTP, unless otherwise specified. 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 WAVE HPLC nucleic acid fragment analysis system equipped with a DNA Sep HPLC column (Transgenomic; Omaha, NE, USA). The moles of RNA primers synthesized were quantified as previously described [56].

**Competing interests.** We have no competing interests.

**Funding.** This work was supported by the Netherlands Organisation for Scientific Research (Veni fellowship 016.116.043 and Vidi fellowship 864.13.003; <http://www.nwo.nl>) to W.K.S.; Leiden University Medical Center (Gisela Thier Fellowship; <http://www.lumc.nl>) to W.K.S.; Biotechnology and Biological Sciences Research Council (BB/K021540/1; <http://www.bbsrc.ac.uk>) to P.S.; a University of Nottingham (<http://www.nottingham.ac.uk>) Vice Chancellor's Excellence award to V.P.; Leiden University Fund/van Walsem (<http://www.luf.nl>) to E.v.E. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Acknowledgements.** Plasmids pKEK1286 and pKEK1287 were kindly provided by K. E. Klose (University of Texas at San Antonio). D. Jex and J. Gibson are acknowledged for exploratory helicase activity assays and analytical gel filtrations. We thank J. B. van Voorden for performing the site-directed mutagenesis on our constructs, R. Scherrers from Wyatt Technology for the SEC-MALS analysis of the CD1454 protein, and D. Tomkiewicz for help with preparing electronic supplementary material, figure S7.

## References

1. Kaguni JM. 2006 DnaA: controlling the initiation of bacterial DNA replication and more. *Annu. Rev. Microbiol.* **60**, 351–371. (doi:10.1146/annurev.micro.60.080805.142111)
2. Battistuzzi FU, Feijao A, Hedges SB. 2004 A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. *BMC Evol. Biol.* **4**, 44. (doi:10.1186/1471-2148-4-44)
3. Mott ML, Berger JM. 2007 DNA replication initiation: mechanisms and regulation in bacteria.

- Nat. Rev. Microbiol.* **5**, 343–354. (doi:10.1038/nrmicro1640)
4. Li Y, Araki H. 2013 Loading and activation of DNA replicative helicases: the key step of initiation of DNA replication. *Genes Cells* **18**, 266–277. (doi:10.1111/gtc.12040)
  5. Robinson A, Causer RJ, Dixon NE. 2012 Architecture and conservation of the bacterial DNA replication machinery, an underexploited drug target. *Curr. Drug Targets* **13**, 352–372. (doi:10.2174/138945012799424598)
  6. Davey MJ, O'Donnell M. 2003 Replicative helicase loaders: ring breakers and ring makers. *Curr. Biol.* **13**, R594–R596. (doi:10.1016/S0960-9822(03)00523-2)
  7. Soultanas P. 2012 Loading mechanisms of ring helicases at replication origins. *Mol. Microbiol.* **84**, 6–16. (doi:10.1111/j.1365-2958.2012.08012.x)
  8. Davey MJ, Fang L, McInerney P, Georgescu RE, O'Donnell M. 2002 The DnaC helicase loader is a dual ATP/ADP switch protein. *EMBO J.* **21**, 3148–3159. (doi:10.1093/emboj/cdf308)
  9. Kornberg A, Baker TA. 2005 *DNA replication*. Sausalito, CA: University Science Books.
  10. Arias-Palomo E, O'Shea VL, Hood IV, Berger JM. 2013 The bacterial DnaC helicase loader is a DnaB ring breaker. *Cell* **153**, 438–448. (doi:10.1016/j.cell.2013.03.006)
  11. Bruand C, Farache M, McGovern S, Ehrlich SD, Polard P. 2001 DnaB, DnaD and DnaI proteins are components of the *Bacillus subtilis* replication restart primosome. *Mol. Microbiol.* **42**, 245–256. (doi:10.1046/j.1365-2958.2001.02631.x)
  12. Smits WK, Goranov AI, Grossman AD. 2010 Ordered association of helicase loader proteins with the *Bacillus subtilis* origin of replication *in vivo*. *Mol. Microbiol.* **75**, 452–461. (doi:10.1111/j.1365-2958.2009.06999.x)
  13. Bruand C, Velten M, McGovern S, Marsin S, Serena C, Ehrlich SD, Polard P. 2005 Functional interplay between the *Bacillus subtilis* DnaD and DnaB proteins essential for initiation and re-initiation of DNA replication. *Mol. Microbiol.* **55**, 1138–1150. (doi:10.1111/j.1365-2958.2004.04451.x)
  14. Velten M, McGovern S, Marsin S, Ehrlich SD, Noirot P, Polard P. 2003 A two-protein strategy for the functional loading of a cellular replicative DNA helicase. *Mol. Cell* **11**, 1009–1020. (doi:10.1016/S1097-2765(03)00130-8)
  15. Briggs GS, Smits WK, Soultanas P. 2012 Chromosomal replication initiation machinery of low-G+C-content Firmicutes. *J. Bacteriol.* **194**, 5162–5170. (doi:10.1128/JB.00865-12)
  16. Bazin A, Cherrier MV, Gutsche I, Timmins J, Terradot L. 2015 Structure and primase-mediated activation of a bacterial dodecameric replicative helicase. *Nucleic Acids Res.* **43**, 8564–8576. (doi:10.1093/nar/gkv792)
  17. Ioannou C, Schaeffer PM, Dixon NE, Soultanas P. 2006 Helicase binding to DnaI exposes a cryptic DNA-binding site during helicase loading in *Bacillus subtilis*. *Nucleic Acids Res.* **34**, 5247–5258. (doi:10.1093/nar/gkl690)
  18. Marston FY, Grainger WH, Smits WK, Hopcroft NH, Green M, Hounslow AM, Grossman AD, Craven CJ, Soultanas P. 2010 When simple sequence comparison fails: the cryptic case of the shared domains of the bacterial replication initiation proteins DnaB and DnaD. *Nucleic Acids Res.* **38**, 6930–6942. (doi:10.1093/nar/gkq465)
  19. Schneider S, Zhang W, Soultanas P, Paoli M. 2008 Structure of the N-terminal oligomerization domain of DnaD reveals a unique tetramerization motif and provides insights into scaffold formation. *J. Mol. Biol.* **376**, 1237–1250. (doi:10.1016/j.jmb.2007.12.045)
  20. Zhang W, Machon C, Orta A, Phillips N, Roberts CJ, Allen S, Soultanas P. 2008 Single-molecule atomic force spectroscopy reveals that DnaD forms scaffolds and enhances duplex melting. *J. Mol. Biol.* **377**, 706–714. (doi:10.1016/j.jmb.2008.01.067)
  21. Zhang W, Allen S, Roberts CJ, Soultanas P. 2006 The *Bacillus subtilis* primosomal protein DnaD untwists supercoiled DNA. *J. Bacteriol.* **188**, 5487–5493. (doi:10.1128/JB.00339-06)
  22. Zhang W, Carneiro MJ, Turner IJ, Allen S, Roberts CJ, Soultanas P. 2005 The *Bacillus subtilis* DnaD and DnaB proteins exhibit different DNA remodelling activities. *J. Mol. Biol.* **351**, 66–75. (doi:10.1016/j.jmb.2005.05.065)
  23. Mott ML, Erzberger JP, Coons MM, Berger JM. 2008 Structural synergy and molecular crosstalk between bacterial helicase loaders and replication initiators. *Cell* **135**, 623–634. (doi:10.1016/j.cell.2008.09.058)
  24. Davey MJ, Jeruzalimi D, Kuriyan J, O'Donnell M. 2002 Motors and switches: AAA+ machines within the replisome. *Nat. Rev. Mol. Cell Biol.* **3**, 826–835. (doi:10.1038/nrm949)
  25. Caruthers JM, McKay DB. 2002 Helicase structure and mechanism. *Curr. Opin. Struct. Biol.* **12**, 123–133. (doi:10.1016/S0959-440X(02)00298-1)
  26. Hanson PI, Whiteheart SW. 2005 AAA+ proteins: have engine, will work. *Nat. Rev. Mol. Cell Biol.* **6**, 519–529. (doi:10.1038/nrm1684)
  27. Erzberger JP, Berger JM. 2006 Evolutionary relationships and structural mechanisms of AAA+ proteins. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 93–114. (doi:10.1146/annurev.biophys.35.040405.101933)
  28. Singleton MR, Dillingham MS, Wigley DB. 2007 Structure and mechanism of helicases and nucleic acid translocases. *Annu. Rev. Biochem.* **76**, 23–50. (doi:10.1146/annurev.biochem.76.052305.115300)
  29. Duderstadt KE, Chuang K, Berger JM. 2011 DNA stretching by bacterial initiators promotes replication origin opening. *Nature* **478**, 209–213. (doi:10.1038/nature10455)
  30. Snider J, Thibault G, Houry WA. 2008 The AAA+ superfamily of functionally diverse proteins. *Genome Biol.* **9**, 216. (doi:10.1186/gb-2008-9-4-216)
  31. Kaguni JM. 2011 Replication initiation at the *Escherichia coli* chromosomal origin. *Curr. Opin. Chem. Biol.* **15**, 606–613. (doi:10.1016/j.cbpa.2011.07.016)
  32. Remus D, Diffley JF. 2009 Eukaryotic DNA replication control: lock and load, then fire. *Curr. Opin. Cell Biol.* **21**, 771–777. (doi:10.1016/j.ceb.2009.08.002)
  33. Itsathitphaisarn O, Wing RA, Eliason WK, Wang J, Steitz TA. 2012 The hexameric helicase DnaB adopts a nonplanar conformation during translocation. *Cell* **151**, 267–277. (doi:10.1016/j.cell.2012.09.014)
  34. Berger JM. 2008 SnapShot: nucleic acid helicases and translocases. *Cell* **134**, 888. (doi:10.1016/j.cell.2008.08.027)
  35. Ilyina TV, Gorbalenya AE, Koonin EV. 1992 Organization and evolution of bacterial and bacteriophage primase-helicase systems. *J. Mol. Evol.* **34**, 351–357. (doi:10.1007/BF00160243)
  36. Leipe DD, Aravind L, Grishin NV, Koonin EV. 2000 The bacterial replicative helicase DnaB evolved from a RecA duplication. *Genome Res.* **10**, 5–16. (doi:10.1101/gr.10.1.5)
  37. Marsin S, McGovern S, Ehrlich SD, Bruand C, Polard P. 2001 Early steps of *Bacillus subtilis* primosome assembly. *J. Biol. Chem.* **276**, 45 818–45 825. (doi:10.1074/jbc.M101996200)
  38. Bruand C, Bidnenko V, Ehrlich SD. 2001 Replication mutations differentially enhance RecA-dependent and RecA-independent recombination between tandem repeats in *Bacillus subtilis*. *Mol. Microbiol.* **39**, 1248–1258. (doi:10.1111/j.1365-2958.2001.02312.x)
  39. Kurth I, O'Donnell M. 2009 Replisome dynamics during chromosome duplication. *EcoSal Plus* **3**. (doi:10.1128/ecosalplus.4.4.2)
  40. Frick DN, Richardson CC. 2001 DNA primases. *Annu. Rev. Biochem.* **70**, 39–80. (doi:10.1146/annurev.biochem.70.1.39)
  41. Bird LE, Pan H, Soultanas P, Wigley DB. 2000 Mapping protein-protein interactions within a stable complex of DNA primase and DnaB helicase from *Bacillus stearothermophilus*. *Biochemistry* **39**, 171–182. (doi:10.1021/bi9918801)
  42. Tougu K, Peng H, Marians KJ. 1994 Identification of a domain of *Escherichia coli* primase required for functional interaction with the DnaB helicase at the replication fork. *J. Biol. Chem.* **269**, 4675–4682.
  43. Syson K, Thirlway J, Hounslow AM, Soultanas P, Waltho JP. 2005 Solution structure of the helicase-interaction domain of the primase DnaG: a model for helicase activation. *Structure* **13**, 609–616. (doi:10.1016/j.str.2005.01.022)
  44. Thirlway J, Soultanas P. 2006 In the *Bacillus stearothermophilus* DnaB-DnaG complex, the activities of the two proteins are modulated by distinct but overlapping networks of residues. *J. Bacteriol.* **188**, 1534–1539. (doi:10.1128/JB.188.4.1534-1539.2006)
  45. Soultanas P. 2005 The bacterial helicase-primase interaction: a common structural/functional module. *Structure* **13**, 839–844. (doi:10.1016/j.str.2005.04.006)
  46. Chintakayala K, Larson MA, Grainger WH, Scott DJ, Griep MA, Hinrichs SH, Soultanas P. 2007 Domain swapping reveals that the C- and N-terminal domains of DnaG and DnaB, respectively, are functional homologues. *Mol. Microbiol.* **63**, 1629–1639. (doi:10.1111/j.1365-2958.2007.05617.x)
  47. Wu CA, Zechner EL, Marians KJ. 1992 Coordinated leading- and lagging-strand synthesis at the

- Escherichia coli* DNA replication fork. I. Multiple effectors act to modulate Okazaki fragment size. *J. Biol. Chem.* **267**, 4030–4044.
48. Johnson SK, Bhattacharyya S, Griep MA. 2000 DnaB helicase stimulates primer synthesis activity on short oligonucleotide templates. *Biochemistry* **39**, 736–744. (doi:10.1021/bi9915541)
  49. Bhattacharyya S, Griep MA. 2000 DnaB helicase affects the initiation specificity of *Escherichia coli* primase on single-stranded DNA templates. *Biochemistry* **39**, 745–752. (doi:10.1021/bi991555d)
  50. Larson MA, Griep MA, Bressani R, Chintakayala K, Soultanas P, Hinrichs SH. 2010 Class-specific restrictions define primase interactions with DNA template and replicative helicase. *Nucleic Acids Res.* **38**, 7167–7178. (doi:10.1093/nar/gkq588)
  51. Lu YB, Ratnakar PV, Mohanty BK, Bastia D. 1996 Direct physical interaction between DnaG primase and DnaB helicase of *Escherichia coli* is necessary for optimal synthesis of primer RNA. *Proc. Natl Acad. Sci. USA* **93**, 12 902–12 907. (doi:10.1073/pnas.93.23.12902)
  52. Mitkova AV, Khopde SM, Biswas SB. 2003 Mechanism and stoichiometry of interaction of DnaG primase with DnaB helicase of *Escherichia coli* in RNA primer synthesis. *J. Biol. Chem.* **278**, 52 253–52 261. (doi:10.1074/jbc.M308956200)
  53. Koepsell SA, Larson MA, Griep MA, Hinrichs SH. 2006 *Staphylococcus aureus* helicase but not *Escherichia coli* helicase stimulates *S. aureus* primase activity and maintains initiation specificity. *J. Bacteriol.* **188**, 4673–4680. (doi:10.1128/JB.00316-06)
  54. Corn JE, Pease PJ, Hura GL, Berger JM. 2005 Crosstalk between primase subunits can act to regulate primer synthesis in trans. *Mol. Cell* **20**, 391–401. (doi:10.1016/j.molcel.2005.09.004)
  55. Kuchta RD, Stengel G. 2010 Mechanism and evolution of DNA primases. *Biochim. Biophys. Acta* **1804**, 1180–1189. (doi:10.1016/j.bbapap.2009.06.011)
  56. Koepsell SA, Larson MA, Frey CA, Hinrichs SH, Griep MA. 2008 *Staphylococcus aureus* primase has higher initiation specificity, interacts with single-stranded DNA stronger, but is less stimulated by its helicase than *Escherichia coli* primase. *Mol. Microbiol.* **68**, 1570–1582. (doi:10.1111/j.1365-2958.2008.06255.x)
  57. Rannou O, Le CE, Larson MA, Nouri H, Dalmais B, Laughton C, Janniere L, Soultanas P. 2013 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 Res.* **41**, 5303–5320. (doi:10.1093/nar/gkt207)
  58. Bailey S, Eliason WK, Steitz TA. 2007 Structure of hexameric DnaB helicase and its complex with a domain of DnaG primase. *Science* **318**, 459–463. (doi:10.1126/science.1147353)
  59. Soultanas P, Wigley DB. 2002 Site-directed mutagenesis reveals roles for conserved amino acid residues in the hexameric DNA helicase DnaB from *Bacillus stearothermophilus*. *Nucleic Acids Res.* **30**, 4051–4060. (doi:10.1093/nar/gkf527)
  60. Wang G, Klein MG, Tokonzaba E, Zhang Y, Holden LG, Chen XS. 2008 The structure of a DnaB-family replicative helicase and its interactions with primase. *Nat. Struct. Mol. Biol.* **15**, 94–100. (doi:10.1038/nsmb1356)
  61. Makowska-Grzyska M, Kaguni JM. 2010 Primase directs the release of DnaC from DnaB. *Mol. Cell* **37**, 90–101. (doi:10.1016/j.molcel.2009.12.031)
  62. Liu B, Eliason WK, Steitz TA. 2013 Structure of a helicase-helicase loader complex reveals insights into the mechanism of bacterial primosome assembly. *Nat. Commun.* **4**, 2495. (doi:10.1038/ncomms3495)
  63. Smits WK, Lyras D, Lacy DB, Wilcox MH, Kuijper EJ. 2016 *Clostridium difficile* infection. *Nat. Rev. Dis. Primers* **2**, 16020. (doi:10.1038/nrdp.2016.20)
  64. Rupnik M, Wilcox MH, Gerding DN. 2009 *Clostridium difficile* infection: new developments in epidemiology and pathogenesis. *Nat. Rev. Microbiol.* **7**, 526–536. (doi:10.1038/nrmicro2164)
  65. Sebahia M *et al.* 2006 The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nat. Genet.* **38**, 779–786. (doi:10.1038/ng1830)
  66. Rokop ME, Auchtung JM, Grossman AD. 2004 Control of DNA replication initiation by recruitment of an essential initiation protein to the membrane of *Bacillus subtilis*. *Mol. Microbiol.* **52**, 1757–1767. (doi:10.1111/j.1365-2958.2004.04091.x)
  67. Brouwer MS, Warburton PJ, Roberts AP, Mullany P, Allan E. 2011 Genetic organisation, mobility and predicted functions of genes on integrated, mobile genetic elements in sequenced strains of *Clostridium difficile*. *PLoS ONE* **6**, e23014. (doi:10.1371/journal.pone.0023014)
  68. Bruand C, Ehrlich SD. 1995 The *Bacillus subtilis* dnaL gene is part of the dnaB operon. *Microbiology* **141**, 1199–1200. (doi:10.1099/13500872-141-5-1199)
  69. Dembek M, Barquist L, Boinett CJ, Cain AK, Mayho M, Lawley TD, Fairweather NF, Fagan RP. 2015 High-throughput analysis of gene essentiality and sporulation in *Clostridium difficile*. *MBio* **6**, e02383. (doi:10.1128/mBio.02383-14)
  70. Karna SL, Zogaj X, Barker JR, Seshu J, Dove SL, Klose KE. 2010 A bacterial two-hybrid system that utilizes Gateway cloning for rapid screening of protein-protein interactions. *Biotechniques* **49**, 831–833. (doi:10.2144/000113539)
  71. Biswas EE, Chen PH, Biswas SB. 2002 Modulation of enzymatic activities of *Escherichia coli* DnaB helicase by single-stranded DNA-binding proteins. *Nucleic Acids Res.* **30**, 2809–2816. (doi:10.1093/nar/gkf384)
  72. Galletto R, Jezewska MJ, Bujalowski W. 2003 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. *J. Mol. Biol.* **329**, 441–465. (doi:10.1016/S0022-2836(03)00435-2)
  73. Wahle E, Lasken RS, Kornberg A. 1989 The dnaB-dnaC replication protein complex of *Escherichia coli*. II. Role of the complex in mobilizing dnaB functions. *J. Biol. Chem.* **264**, 2469–2475.
  74. Larson MA, Bressani R, Sayood K, Corn JE, Berger JM, Griep MA, Hinrichs SH. 2008 Hyperthermophilic *Aquifex aeolicus* initiates primer synthesis on a limited set of trinucleotides comprised of cytosines and guanines. *Nucleic Acids Res.* **36**, 5260–5269. (doi:10.1093/nar/gkn461)
  75. Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. 2012 Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* **28**, 464–469. (doi:10.1093/bioinformatics/btr703)
  76. Larson AM, Fung AM, Fang FC. 2010 Evaluation of tcdB real-time PCR in a three-step diagnostic algorithm for detection of toxigenic *Clostridium difficile*. *J. Clin. Microbiol.* **48**, 124–130. (doi:10.1128/JCM.00734-09)
  77. Yoda K, Okazaki T. 1991 Specificity of recognition sequence for *Escherichia coli* primase. *Mol. Gen. Genet.* **227**, 1–8. (doi:10.1007/BF00260698)
  78. Swart JR, Griep MA. 1995 Primer synthesis kinetics by *Escherichia coli* primase on single-stranded DNA templates. *Biochemistry* **34**, 16 097–16 106. (doi:10.1021/bi00049a025)
  79. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. 2015 The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858. (doi:10.1038/nprot.2015.053)
  80. Pan H, Wigley DB. 2000 Structure of the zinc-binding domain of *Bacillus stearothermophilus* DNA primase. *Structure* **8**, 231–239. (doi:10.1016/S0969-2126(00)00101-5)
  81. Grainger WH, Machon C, Scott DJ, Soultanas P. 2010 DnaB proteolysis in vivo regulates oligomerization and its localization at oriC in *Bacillus subtilis*. *Nucleic Acids Res.* **38**, 2851–2864. (doi:10.1093/nar/gkp1236)
  82. Bujalowski W, Klonowska MM, Jezewska MJ. 1994 Oligomeric structure of *Escherichia coli* primary replicative helicase DnaB protein. *J. Biol. Chem.* **269**, 31 350–31 358.
  83. Biswas SB, Biswas-Fiss EE. 2006 Quantitative analysis of binding of single-stranded DNA by *Escherichia coli* DnaB helicase and the DnaB × DnaC complex. *Biochemistry* **45**, 11 505–11 513. (doi:10.1021/bi060118d)
  84. Soultanas P. 2002 A functional interaction between the putative primosomal protein DnaL and the main replicative DNA helicase DnaB in *Bacillus*. *Nucleic Acids Res.* **30**, 966–974. (doi:10.1093/nar/30.4.966)
  85. Strycharska MS, Arias-Palomo E, Lyubimov AV, Erzberger JP, O'Shea VL, Bustamante CJ, Berger JM. 2013 Nucleotide and partner-protein control of bacterial replicative helicase structure and function. *Mol. Cell* **52**, 844–854. (doi:10.1016/j.molcel.2013.11.016)
  86. Chodavarapu S, Jones AD, Feig M, Kaguni JM. 2016 DnaC traps DnaB as an open ring and remodels the domain that binds primase. *Nucleic Acids Res.* **44**, 210–220. (doi:10.1093/nar/gkv961)
  87. Gao D, McHenry CS. 2001 tau binds and organizes *Escherichia coli* replication proteins through distinct

- domains. Domain IV, located within the unique C terminus of tau, binds the replication fork, helicase, DnaB. *J. Biol. Chem.* **276**, 4441–4446. (doi:10.1074/jbc.M009830200)
88. Haroniti A, Anderson C, Doddridge Z, Gardiner L, Roberts CJ, Allen S, Soultanas P. 2004 The clamp-loader-helicase interaction in *Bacillus*: atomic force microscopy reveals the structural organisation of the DnaB-tau complex in *Bacillus*. *J. Mol. Biol.* **336**, 381–393. (doi:10.1016/j.jmb.2003.12.043)
89. Afonso JP, Chintakayala K, Suwannachart C, Sedelnikova S, Giles K, Hoyes JB, Soultanas P, Rafferty JB, Oldham NJ. 2013 Insights into the structure and assembly of the *Bacillus subtilis* clamp-loader complex and its interaction with the replicative helicase. *Nucleic Acids Res.* **41**, 5115–5126. (doi:10.1093/nar/gkt173)
90. Chang P, Mariani KJ. 2000 Identification of a region of *Escherichia coli* DnaB required for functional interaction with DnaG at the replication fork. *J. Biol. Chem.* **275**, 26 187–26 195. (doi:10.1074/jbc.M001800200)
91. Stordal L, Maurer R. 1996 Defect in general priming conferred by linker region mutants of *Escherichia coli* dnaB. *J. Bacteriol.* **178**, 4620–4627. (doi:10.1128/jb.178.15.4620-4627.1996)
92. Sonenshein AL. 2005 CodY, a global regulator of stationary phase and virulence in Gram-positive bacteria. *Curr. Opin. Microbiol.* **8**, 203–207. (doi:10.1016/j.mib.2005.01.001)
93. Gaal T, Bartlett MS, Ross W, Turnbough CL Jr, Gourse RL. 1997 Transcription regulation by initiating NTP concentration: rRNA synthesis in bacteria. *Science* **278**, 2092–2097. (doi:10.1126/science.278.5346.2092)
94. Murray HD, Schneider DA, Gourse RL. 2003 Control of rRNA expression by small molecules is dynamic and nonredundant. *Mol. Cell* **12**, 125–134. (doi:10.1016/S1097-2765(03)00266-1)
95. Schneider DA, Gourse RL. 2003 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. *J. Bacteriol.* **185**, 6185–6191. (doi:10.1128/JB.185.20.6185-6191.2003)
96. Griffiths E, Gupta RS. 2004 Signature sequences in diverse proteins provide evidence for the late divergence of the Order Aquificales. *Int. Microbiol.* **7**, 41–52.
97. Skophammer RG, Servin JA, Herbold CW, Lake JA. 2007 Evidence for a gram-positive, eubacterial root of the tree of life. *Mol. Biol. Evol.* **24**, 1761–1768. (doi:10.1093/molbev/msm096)
98. Galtier N, Tourasse N, Gouy M. 1999 A nonhyperthermophilic common ancestor to extant life forms. *Science* **283**, 220–221. (doi:10.1126/science.283.5399.220)
99. Tocheva EI, Matson EG, Morris DM, Moussavi F, Leadbetter JR, Jensen GJ. 2011 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)
100. Tocheva EI, Ortega DR, Jensen GJ. 2016 Sporulation, bacterial cell envelopes and the origin of life. *Nat. Rev. Microbiol.* **14**, 535–542. (doi:10.1038/nrmicro.2016.85)
101. Vallet-Gely I, Donovan KE, Fang R, Joung JK, Dove SL. 2005 Repression of phase-variable cup gene expression by H-NS-like proteins in *Pseudomonas aeruginosa*. *Proc. Natl Acad. Sci. USA* **102**, 11 082–11 087. (doi:10.1073/pnas.0502663102)