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

Eijk, H.W. van

ERIKA VAN EIJK

The **replication machinery** of *Clostridium difficile*

a potential target for novel antimicrobials

ERIKA VAN EIJK

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

Colophon

The replication machinery of *Clostridium difficile*:
a potential target for novel antimicrobials
Erika van Eijk
PhD thesis, Leiden University Medical Center, 2019

Cover Uurwerk zakhorloge M. van Eijk (1948-2017)

Cover photo Caroline de Brouwer

Design & print SW/Ralph Boer

ISBN/EAN 978-90-824257-2-7

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

Proefschrift

ter verkrijging van de graad van
Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op donderdag 16 mei 2019 klokke 16:15 uur

door

Hendrika Willemina van Eijk
geboren 22 juli 1976 te Aalten

PROMOTOR

Prof. Dr. E.J. Kuijper

CO-PROMOTOR

Dr. W.K. Smits

LEDEN PROMOTIECOMMISSIE

Prof. dr. M. Yazdanbakhsh

Prof. dr. O.P. Kuipers (Rijksuniversiteit Groningen)

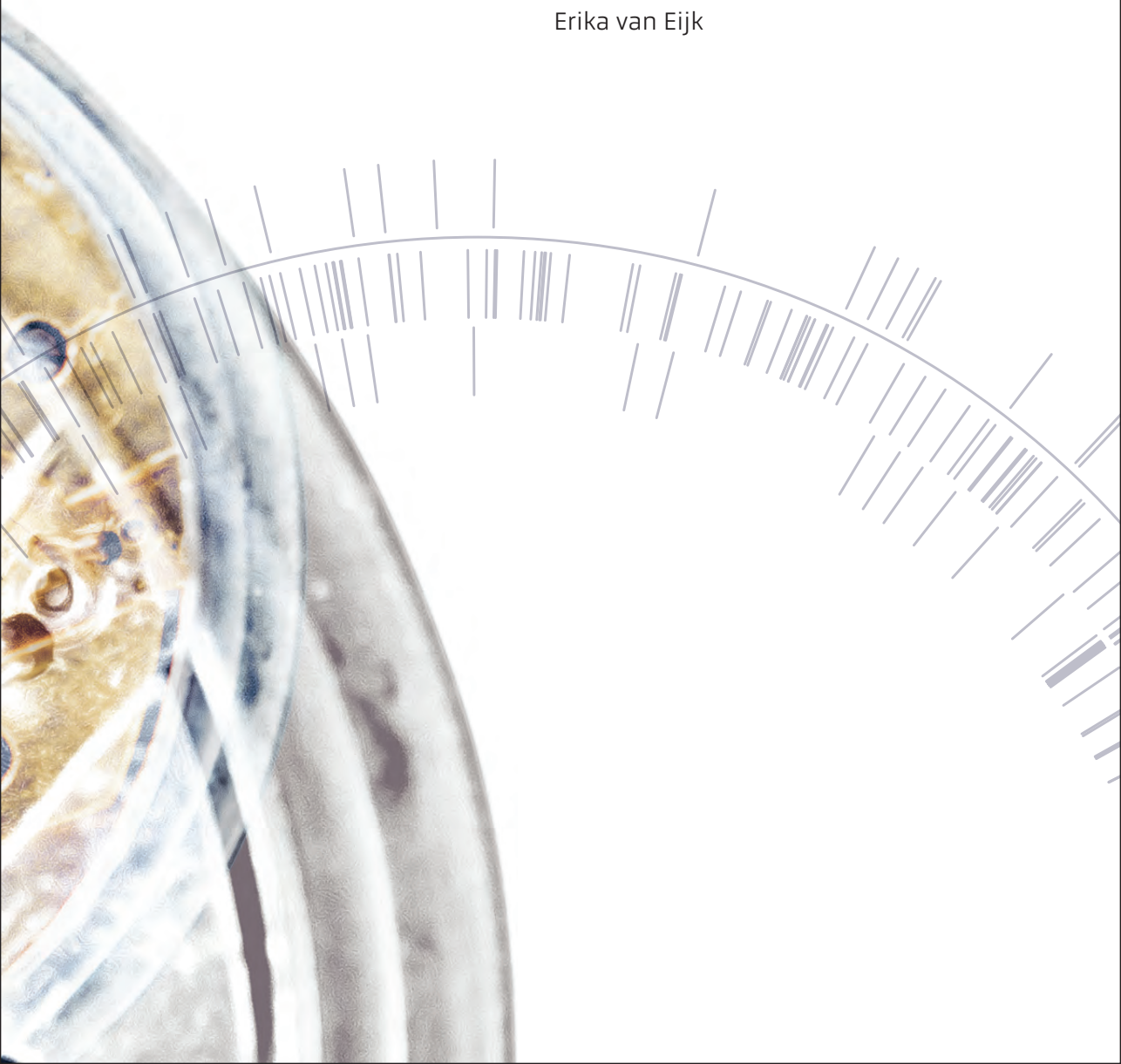
Dr. A.P. Roberts (Liverpool School of Tropical Medicine)

Table of contents

Chapter 1

General Introduction and outline thesis

Erika van Eijk



General Introduction and outline thesis

Clostridium difficile infection

Clostridium difficile (*Clostridioides difficile*) is a Gram-positive, spore-forming obligate anaerobic bacterium that can asymptomatically colonize the intestine of humans, other mammals, reptiles, birds and insects [1,2]. The bacterium is ubiquitous in the environment and particularly present in soil. Ingestion of spores and the subsequent germination into vegetative toxin-producing cells or prior colonization by *Clostridium difficile* within healthy or immunocompromised individuals, and/or elderly patients may induce *Clostridium difficile* infection (CDI) [3]. The spectrum of disease ranges from mild, self-limited diarrhoea to life-threatening pseudomembranous colitis [4,5]. Transmission of the bacterial spores occurs via the faecal-oral route and further spread is promoted by contact of healthcare workers with contaminated surfaces and infected patients. Eradication of spores in healthcare settings is extremely difficult due to their metabolically dormant state and resistance against a variety of environmental stresses such as desiccation, high temperatures, aerobic conditions and many hospital disinfectants and -cleaning agents [4,6–10]. *C. difficile* spores may survive for up to 5 months on environmental surfaces [11,12]. Moreover, spores are not only shed by symptomatic patients, but also by asymptomatic patients which hampers infection control in respect to identifying the source and implementing appropriate preventive measures.

Clostridium difficile infection (CDI) can occur when the normal protective intestinal microbiota is disrupted. Under these conditions, *C. difficile* can proliferate in the gut, causing an inflammatory response induced by the clostridial toxins, toxin A (TcdA) and toxin B (TcdB) [13]. Although the specific roles of these exotoxins are not entirely uncovered, direct effects on the integrity of the tight junctions of the colonic epithelium and apoptotic properties targeted against this tissue has been well described [14–17] and are associated with the main clinical manifestations of CDI: diarrhoea and colitis. Disruption of the colonic epithelium allows the toxins to interact with other cells such as immune cells and neurons and may stimulate, indirectly, the production of chemokines, pro-inflammatory cytokines, neuropeptides and other neuro-immune signals resulting in a systemic inflammation response [18].

Originally identified as part of the intestinal microbiota of healthy infants (by Hall and O'Toole in 1935 [19]), it took more than forty years to identify the causal relationship between *Clostridium difficile* and life-threatening pseudomembranous colitis [20,21]. Fatality of CDI infection has been dramatically demonstrated in the Stoke Mandeville Hospital outbreaks in the United Kingdom in 2004 and 2005 [22]. After recognition of European hospital outbreaks of *Clostridium difficile* infections (CDIs) associated with the emergence of PCR ribotype 027, surveillance at national level was encouraged by the European Centre for Disease Prevention and Control (ECDC) [23]. In the United States, the Center for Disease Control and Prevention (CDC) estimated on basis of a large population- and laboratory-based surveillance that the overall incidence of *C. difficile* infection in 2011 was 453.000 and 29.000 deaths were attributable to this infection [24]. Within the health care-associated population of CDI, the rate of first recurrence of infection was estimated at 20.9 percent (61.400 cases) [24], which is consistent with other reports (18-25 percent) [25-29]. In the US, the CDI incidence rose from 4.5 to 8.6 cases per 1.000 hospital discharges between 2001 and 2008. Furthermore, the overall mortality of CDI patients in the USA increased significantly from 6.5 to 7.2 percent in 2001-2010, with a total of 154.184 deaths (7.1 percent) during this time-period [30]. In a pilot study, conducted by the European *C. difficile* Infection Surveillance Network (ECDIS-NET) involving 37 European acute care hospitals, it was determined that the incidence rate of hospital-acquired CDI ranged from 0.6 to 18.5 per 10.000 patient days (median 3.7) [31]. It should be noted that the incidence rate in this study was based on aggregated hospital data. In the Netherlands, sentinel surveillance overseen by the National Reference Laboratory for *C. difficile* in 23 hospitals showed that the incidence rate per 10.000 patient days was 3.1 for the period of May 2015 to May 2016 [32]. Aside from the clinical implications, CDI also represents a substantial economic burden with estimated annual costs ranging from $ 5.4 billion [33] to $ 6.3 billion [34] in the US. The bulk of the economic burden of CDI consists of the costs of hospitalization and recurrence of infection [29,35]. In the USA, the national annual cost of recurrent CDI is estimated at $ 1.5 billion [33]. The costs of CDI case management in Europe are difficult to determine due to heterogeneous methods [35,36]. In 2006, Kuijper *et al.* roughly estimated that the potential cost of CDI in Europe was €3 billion, a result which was extrapolated from the (estimated) annual cost of management of CDI in the United Kingdom [37]. Recently, in a multicentre study that was conducted in the UK it was estimated that the median total management cost for a first episode of CDI and recurrent CDI was approximately €7.100 and

€ 8.500, respectively [38]. Data is scarce on the costs of CDI case management in the Netherlands specifically. However, a retrospective cost analysis that was conducted to gain insight on the financial burden of an outbreak of *C. difficile* in a tertiary hospital in the Netherlands provided some much-needed information [39]. It was estimated that the costs attributed to this outbreak, involving 72 patients in a time-period of one year, was € 1.222.376 [39].

Recent years have seen an increase in the incidence and severity of *C. difficile* infections (CDI) in both the United States and Europe, due to the emergence of certain PCR ribotypes (RT) [13,40]. The increased incidence and severity of the disease are associated with outbreaks of 'hyper-virulent' *C. difficile* strains, particularly PCR ribotype 027 (also known as NAP1/027/BI) and PCR-ribotype 078 [37,41-44]. The epidemic PCR ribotype 027 was first recognized in Western Europe and North America [2] and emergence of this strain has been linked to fluoroquinolone resistance [45]. *C. difficile* strains of both these PCR-ribotypes cause severe CDI with high mortality rates, though infections with *C. difficile* RT 078, contrary to RT 027, are often located outside the hospital environment and affect younger patients [46-49].

Antibiotic use is a well-established risk factor for CDI, but age of the patient and/or underlying comorbidities play an important role in both aetiology and severity of the disease [2,13,50]. The antibiotic use increases the risk for CDI during therapy and in the period of 3 months after cessation of antibiotic therapy. The highest risk in contracting CDI after antibiotic therapy was found to be in the first month following cessation [51]. Antibiotics associated with CDI risk, such as clindamycin, broad spectrum penicillins, cephalosporins and fluoroquinolones significantly deplete the Gram-negative microflora and augment colonization of *C. difficile* [4,52,53]. Patients who are treated with these antibiotics for another infection are prone to infection with this opportunistic pathogen [4]. In most cases, patients develop antibiotic-associated diarrhoea (AAD). Although *C. difficile* is not the sole microorganism implicated in AAD, it is the most common causative agent of infectious antibiotic-associated diarrhoea, responsible for 10-25 percent of the cases [54,55] and the leading cause of nosocomial infectious diarrhoea in adults [56]. Other pathogens associated with infectious AAD are *Clostridium perfringens*, *Klebsiella oxytoca*, *Staphylococcus aureus*, and *Candida albicans* (approximately 40 percent of AAD, *C. difficile* included), though in a substantial number of cases, the causative agent remains unknown [4,57].

Antimicrobial treatment and resistance

Paradoxically, CDI is generally treated using antibiotics that simultaneously prohibit regeneration of the protective gut microflora, so infection may persist and relapses may be promoted. At present three antibiotics, metronidazole, vancomycin and fidaxomicin, are commonly used to treat CDI [58,59]. The drug of choice is dependent on the severity of CDI and the risk to develop a recurrent episode [53]. Metronidazole is often administered when the infection ranges from mild to moderate, as this antibiotic has a low cost and does not increase the risk for the development of vancomycin resistance in *C. difficile* or other microorganisms (*Enterococcus spp.*) as overuse of vancomycin might [2,4,53,60]. Oral vancomycin is often indicated when patients suffer from a severe or complicated CDI and is superior to treatment with metronidazole under these circumstances [2,60]. Recently, treatment with vancomycin has been increased due to decreasing costs, lower side effects and evidence that vancomycin is more efficacious in mild to moderate infections compared to metronidazole [59,61–64]. In case of recurrent disease and patients with high relapse risk, treatment with fidaxomicin is preferred over vancomycin [2,4,60]. Another advantage of fidaxomicin over the first-line therapies metronidazole and vancomycin is that the former antibiotic agent has a minimal impact on the microbiota of the host as it targets specific anaerobic Gram-positive bacteria [4,60]. Despite the superiority over vancomycin in preventing relapses of CDI, fidaxomicin is not yet prescribed on a large scale due to the high cost of this treatment [2,4,60,65,66].

Significant resistance to the standard antimicrobial therapy has not yet occurred in the clinic, although an increase of treatment failure associated with metronidazole has been observed [63,64,67,68] and sporadic resistance has been reported [69–72]. However, in light of the development of resistance to clindamycin and fluoroquinolones through their extensive use in the past, it is most probable that resistance to the standard therapy will arise over time [4]. Indeed, selective pressure induced by increased prescription of even a narrow-spectrum antibiotic as fidaxomicin may increase the risk of development of resistant *C. difficile* strains as has been shown *in vitro* [53,73]. It should be noted that the broad-spectrum antibiotics clindamycin and fluoroquinolones are not used to treat CDI but are commonly administered to resolve other bacterial infections [53]. Nonetheless, their impact on the integrity of the microbiota is a predisposing factor in the occurrence and recurrence of CDI [13,53].

Resistance to clindamycin and fluoroquinolones are exemplary for the extensive arsenal of antimicrobial resistance of *C. difficile* [53,74]. Multidrug resistant (MDR) strains of *C. difficile* are common [3,74,75] and resistance patterns among MDR strains are very diverse, as are the mechanisms that confer resistance. For instance, in a recent study conducted in the US, it has been shown that out of 139 clinical isolates from patients diagnosed with CDI almost 60 percent of strains were resistant to three types of antibiotics or more [75]. Fluoroquinolones are inhibitors of type II topoisomerases, which include gyrase and topoisomerase IV, enzymes essential for DNA replication. *C. difficile* lacks genes encoding topoisomerase IV but does contain *gyrA* and *gyrB* genes that encode subunits of the gyrase [76]. Resistance to fluoroquinolones can occur through point mutations in the quinolone-resistance determining region (QRDR) of the DNA gyrase subunits, GyrA and/or GyrB, that decrease the affinity for fluoroquinolones [77–79]. The most frequent amino acid substitution found in *C. difficile* is a threonine to isoleucine mutation in the GyrA subunit (Thr82Ile) [79]. Interestingly, this particular GyrA amino acid substitution is found in epidemic PCR ribotype 027 strains that have emerged in the beginning of this century but are not found in historical isolates of the same ribotype [74,76,80]. In contrast, changes in the pathogenicity locus (PaLoc) of *C. difficile* RT 027 previously hypothesized to account for the hyper-virulence and transmissibility of this particular ribotype were present in both pre- and post-epidemic isolated of RT 027. Thus, the acquisition of fluoroquinolone resistance marked a pivotal point in the evolution of RT 027 [45,81]. Despite its clear link to epidemicity in PCR ribotype 027 strains, fluoroquinolone resistance is not an exclusive trait of this ribotype: it is also common in other ribotypes [74,80]. Therefore, the emergence of the epidemic RT 027 strains is most likely multifactorial. In support of this view, in a recent study it was proposed that the use of the disaccharide trehalose in the food industry significantly contributed to the emergence of RT 027 strains [82]. Sensitivity to low concentrations of trehalose was attributed to a single point mutation in the trehalose repressor (*treR*), thereby creating a fitness advantage over several other ribotypes lacking this mutation [82]. Interestingly, another epidemic strain, RT 078, also showed enhanced growth in presence of low concentrations of trehalose, although the molecular basis for the increased sensitivity to this specific carbon source differed from RT 027 [82].

In contrast to fluoroquinolone resistance, which is mediated by a chromosomal resistance determinant and non-transferable, clindamycin resistance is acquired by horizontal gene transfer of the mobilizable non-conjugative transposon Tn*5398*,

which contains two copies of erythromycin ribosomal methylase (*erm*) genes of class B [83]. In *Clostridium difficile*, ribosomal methylation of bacterial 23S rRNA caused by the products of these genes, is the most common mechanism of resistance [74], resulting in prevention of antibiotic binding, and thus, antimicrobial activity [81]. Like clindamycin resistance, tetracycline resistance in *C. difficile* is acquired through horizontal gene transfer. Transposons linked with tetracycline resistance are Tn*5397*, Tn*916*, Tn*916*-like, and Tn*6164* [84–91]. Although these elements are capable of transferring a number of the *tet* class of genes (*tetM*, *tet44* and *tetW*), *tetM* is the predominant class identified in *C. difficile* [53]. The product of the *tetM* gene, the TetM ribosome protection protein, confers resistance through binding and displacing tetracycline from its binding site on the ribosome [92]. It is noteworthy that 11 percent of the genome of the *C. difficile* reference strain 630 is comprised of mobile genetic (i.e. horizontally acquired) elements [89].

The mechanisms described above, that are specific to a particular class of antimicrobials, are not the only ones capable of conferring resistance to fluoroquinolones and tetracyclines. In many bacteria, general mechanisms exist that confer resistance to multiple classes of antimicrobials [93]. For instance, the multidrug resistant chloramphenicol-florfenical resistance (*cfr*) gene may confer resistance to different classes of clinically relevant antibiotics [53,74,94,95]. A *cfr*-like gene was also identified in *C. difficile* [96] and it was established that the product of this gene can function as a legitimate Cfr protein [94]. Similarly, active drug efflux by ATP-binding cassette (ABC) transporters can contribute to resistance to different classes of antimicrobials [97,98] and a recent study demonstrated that the ABC transporter CD2068 of *C. difficile* could potentially function as multidrug efflux transporter [99]. Moreover, the highly resistant endospores of *C. difficile* are intrinsically resistant to many antimicrobials, due to metabolic dormancy [4,100,101]. Clearly, *C. difficile* has developed multiple mechanisms to avoid the activity of many classes of antimicrobials [53,99].

The extensive arsenal of resistance mechanisms alone already underscores the need for development of new therapeutic options. But there are other reasons for pushing the development of novel antimicrobials against CDI as well. First, metronidazole and vancomycin are both broad-spectrum antibiotics targeting not only *Clostridium difficile*, but also certain bacteria that are part of the protective microbiota. Dysbiosis of the microbiota caused by these antibiotics may perpetuate CDI [13].

This may at least partly explain why the average cure rate achieved by the standard therapy recommended for CDI does not exceed 80-90 percent [102]. Other underlying causes that may lead to suboptimal results in terms of clinical cure and prevention of relapse in relation to the standard therapies are the spore-forming ability of *C. difficile* and the altered immunity of CDI patients [103]. An additional concern is that vancomycin-use leads to increased prevalence of vancomycin-resistant enterococci (VRE), other important nosocomial pathogens [104], and/or result in VRE overgrowth. Using narrow-spectrum antimicrobials limits these off-target effects and thus contributes to both the integrity of the microbiota and reduces the development and spread of antimicrobial resistance in other bacteria. Second, prescribing fidaxomicin instead of the first-line antibiotics is still considered as not favourable, in terms of cost-effectiveness [13,105]. Third, it has been reported that although administering fidaxomicin leads to a reduction of the relapse rate, sustained cure may be limited to 75 percent among treated patients who have experienced multiple CDI recurrences [103].

Novel drugs for CDI targeting DNA replication

Evidently, there is an urgent demand for more efficient therapies and tools to combat CDI [103]. In recent years, tremendous efforts have been made by pharmaceutical industry and academia to develop new treatments for CDI, which are directed at the various stages of infection [105]. The latter is reflected in diversity the strategies that have been employed such as, microbiologic approaches for treatment of CDI (faecal microbiotica transplantation (FMT), microbiota supplements), non-microbiologic approaches for CDI treatment and prevention (passive immunization and vaccination) and antibiotic inactivation for CDI prevention and alternative antibiotics (new antimicrobials, off-label and derivatives) [60,62,65,103,105-107]. However, drug discovery and development of new antibiotics against *C. difficile* is challenging in terms of the characteristics that have to be met [108]. Ideally, the new antimicrobial agents should possess specific physiochemical properties, such as low solubility, high molecular weight, high polarity and low permeability and absorption, coupled with the ability to withstand excretion by efflux pumps [108]. Furthermore, the spectrum of activity should be narrow to safeguard the integrity of the normal gut flora. As the 'one drug-one target' model has proven its limited viability in regard to emergence of resistance and recurrence, novel targets and mechanisms of action

should be explored [108,109]. Additionally, new agents should be superior or at least non-inferior to the standard therapy, in terms of efficacy and tolerability, and should display decreased permeability at the site of action [108]. Finally, the drug acquisition cost of the new agent should be reasonable, so its use will not be limited due to high pricing. Promising candidates (such as ramoplanin, cadazolid, ridinazole, and tigecycline) are in different stages of development and reviewed extensively [59,60,65,103,105,106]. Hereafter, a few interesting compounds with direct relevance for this thesis are highlighted.

Cadazolid (Actelion Pharmaceuticals) has dual mode of action directed at the 50S ribosome unit and DNA gyrase due to its chimeric structure with quinolonyl and oxazolidinone moieties [65,103,105]. This antibiotic has received a fast track status by the FDA to accelerate its development but has shown no superiority over vancomycin [110].

Another FQ-hybrid antibiotic under investigation is MBX-500 (Microbiotix) [111], which consists of a FQ-pharmacophore covalently linked to anilinouracil (AU) component [112]. Mechanistically, MBX-500 functions as a dual gyrase/topoisomerase inhibitor and DNA polymerase inhibitor [103,108]. Although *in vivo* efficacy has been demonstrated in several animal studies [111,113], no data is available for efficacy in humans [103].

The compound 362E exerts it action on DNA polymerase in a similar fashion as the AU component of MBX-500, though belongs to another class of PolC inhibitors: both inhibitors bind via specific domains to the DNA strand and DNA polymerase simultaneously leading to the formation of an inert ternary complex [114,115]. The antimicrobial activities of 362E were similar to those of the standard therapy, when tested in a small panel of *C. difficile* strains [116]. Furthermore, the specificity of 362E regarding *C. difficile* towards purified *C. difficile* PolC [115] was supported by the results of agar dilution data, where it demonstrated lower activity against the majority of a modest collection of Gram-positive anaerobes [116]. Currently, this compound is still in the phase of preclinical development.

The common factor between the agents discussed above is that they are – at least partially – directed against DNA replication proteins (**Figure 1**). DNA replication is the process in which a mother cell duplicates its DNA semi-conservatively, to ensure that upon division both daughter cells contain the same genetic information.

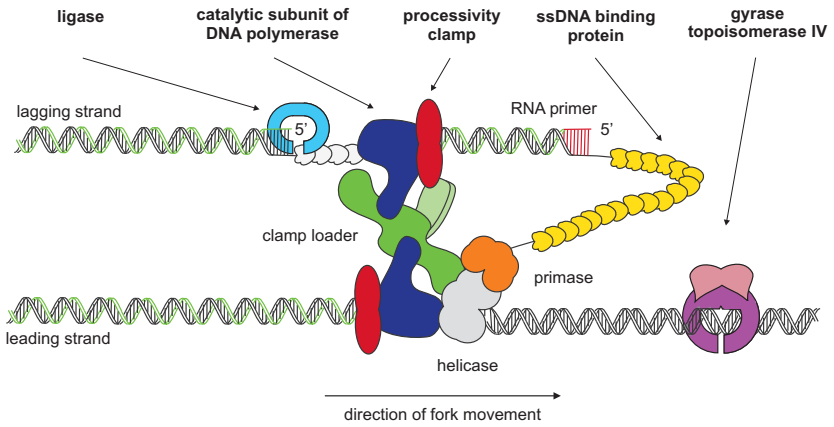


Figure 1. Schematic representation of the bacterial replisome and accessory replication proteins

Indicated is the core of the replisome and the other accessory replication proteins. For simplicity, replication initiation proteins and regulators have been omitted from this figure.

By and large, the mechanisms of DNA replication are conserved across all domains of life. Faithfull replication of the bacterial DNA is an essential requirement for viability of bacterial cells, and stringent coordination of proteins involved in this process is needed to accomplish this. In contrast to *C. difficile*, DNA replication has been extensively studied in the non-pathogenic Gram-positive model bacterium *Bacillus subtilis* (*B. subtilis*), which is a high A+T content Firmicute, like *C. difficile* [117-119]. The process of DNA replication can be divided in an initiation and an elongation phase. In this section, the description of the function of proteins involved in these particular phases is based on extensive work done in *B. subtilis*.

In most organisms, DNA replication starts by binding of a replication initiator protein at a specific chromosomal location. In bacteria with a single chromosome, such as *B. subtilis*, this protein is the highly conserved DnaA [120], and the chromosomal location is the origin of replication, *oriC*. DnaA binds to specific DnaA boxes in the *oriC* where it melts and unwinds the double-stranded (ds) DNA helix and recruits the replicative DNA helicase and other proteins [120-123]. Subsequently, the replicative helicase DnaC is loaded onto ssDNA by three proteins: DnaI, DnaB and

DnaD, which are all proven to be essential for DNA replication [124]. DnaI is the loader ATPase and is responsible for the assembly of the helicase hexamer on single-stranded (ss) DNA [124]. DnaI, DnaB and DnaD proteins are conserved in many, but not all, low G+C Gram-positive bacteria [118]. The replicative DNA helicase, encoded by the *dnaC* gene in *B. subtilis*, is responsible for unwinding of the DNA helix at the replication fork and is conserved in all prokaryotes [118,125,126]. As *B. subtilis* contains one chromosome, two replication forks are assembled at the singular origin of replication which move bi-directionally [127]. For lagging strand DNA synthesis, short RNA primers have to be synthesized by primase (DnaG) [128] that are extended by DNA polymerase into Okazaki fragments. *B. subtilis* primase interacts with the replicative DNA helicase and is thought to form a complex that modulates primase activity [129,130]. Loading of the helicase marks the start of the elongation phase, in which the leading and lagging DNA templates are replicated. This highly progressive process is carried out by the DNA polymerase III holoenzyme. The main subunits in the *B. subtilis* PolIII complex are two α subunits; PolC and DnaE. Both polymerases are essential for DNA replication in *B. subtilis*, and also for cell viability [131]. PolC possesses a proofreading exonuclease domain and is responsible for rapid elongation of both the leading strand and the lagging strand [124,131]. In contrast, *subtilis* DnaE is an inefficient and error prone polymerase, which is indispensable for synthesis of the lagging strand but not for leading strand synthesis [124,132]. The combination of PolC and DnaE is not found in Gram-negative bacteria, as they do not possess PolC [133]. Other subunits of the *B. subtilis* polymerase III holoenzyme are the β-clamp, τ (and γ) subunit, δ subunit and the δ' subunit encoded by *dnaN*, *dnaX*, *holA* and *holB* respectively. The β-clamp encircles primed DNA strands and is able to slide across the strands. It intimately links the DNA polymerases to the DNA template, thereby ensuring the high processivity of DNA replication [132,134]. Loading of the β-clamp on DNA depends on initiation of Okazaki fragment synthesis and the actual loading is performed by the clamp-loader complex, which in *B. subtilis* consists of a τ trimer, a δ and δ' monomer [129,130,135]. The β-clamp and the τ/γ subunits are conserved in most bacteria, while the δ and δ' subunits are more variable [118]. The whole complex responsible for DNA synthesis (that includes one or more DNA polymerases, proofreading enzymes and factors ensuring processivity such as the beta-clamp) is commonly termed replisome (**Figure 1**).

Other accessory proteins are also important for DNA replication, including DNA ligase, single-stranded DNA-binding protein (SSB) and DNA gyrase. The primary

function of DNA ligase is to catalyse the joining of breaks in duplex DNA that are left as a result of the discontinuous lagging strand synthesis, repair or recombination [136]. Disruption of the gene encoding ligase is detrimental for cell growth in *B. subtilis* and therefore considered to be essential [136]. SSB maintains and protects the ssDNA strands produced by the replicative helicase [137]. The unwinding and reannealing of double stranded DNA can result in topological problems that are resolved by topoisomerases. DNA gyrase is a topoisomerase type II and its main function is the introduction of negative supercoils that relax the DNA helix ahead of the replication fork [138]. Therefore, it is essential for the progression of the replication machinery, and thus for process of DNA replication as a whole.

Though this overall mechanism of DNA replication is conserved, details differ substantially between organisms, with respect to the proteins performing the various functions [117], the molecular mechanisms [139] and the way their activity is regulated [129,140–147].

Due to their essential nature, proteins involved in DNA replication are attractive targets for the development of antimicrobials. However, to date, none of the antimicrobials used in the clinical setting, target the main components of the DNA replication machinery [148] and topoisomerase II is the only replication-associated protein that is targeted by commonly used therapeutics. The development of new therapeutics is hindered by a limited characterization of the replication machinery of bacterial pathogens, such as *C. difficile*.

Outline of this thesis

Chapter 1 is this introduction.

Chapter 2 provides an overview of DNA replication proteins that potentially may serve as targets for antimicrobials in drug-resistant pathogens and includes the *in silico* identification of the replication proteins of *C. difficile*. Additionally, the mode of action and the current developmental status of the compounds directed at the core replication machinery and accessory replication proteins are discussed.

In **chapter 3** we determined the complete genome sequence of our reference laboratory strain, *C. difficile* 630Δ*erm*, and compared this sequence to the sequence of *C. difficile* 630. This revealed an unexpected transposition of the mobile genetic element CTn5. The genome sequence now offers an appropriate reference for all molecular work on this strain.

Apart from the identification and characterization of gyrase and DNA polymerase in previous studies, no information on DNA replication proteins or the mechanism of replication was available. To address this hiatus, we sought out to identify core replication proteins and to experimentally validate the functional role of these proteins. The findings of this experimental work**,** with a focus on helicase and primase, are presented in **chapter 4** and **chapter 5**, respectively.

In **chapter 6**, we tested the antimicrobial activity of the DNA polymerase inhibitor 362E against a large and diverse panel of clinical *C. difficile* isolates. Furthermore, we have determined the transcriptional response of *C. difficile* to replication inhibition by 362E.

Finally, in **chapter 7**, we place our findings in a broader perspective and provide an outlook for the field.

References

1 Hensgens, M. P. *et al.* Clostridium difficile infection in the community: a zoonotic disease? *Clin Microbiol Infect* **18**, 635–645, doi:10.1111/j.1469-0691.2012.03853.x (2012).

2 Leffler, D. A. & Lamont, J. T. Clostridium difficile Infection. *N Engl J Med* **373**, 287–288, doi:10.1056/NEJMc1506004 (2015).

3 Spigaglia, P., Barbanti, F., Morandi, M., Moro, M. L. & Mastrantonio, P. Diagnostic testing for Clostridium difficile in Italian microbiological laboratories. *Anaerobe* **37**, 29–33, doi:10.1016/j.anaerobe.2015.11.002 (2016).

4 Johanesen, P. A. *et al.* Disruption of the Gut Microbiome: Clostridium difficile Infection and the Threat of Antibiotic Resistance. *Genes (Basel)* **6**, 1347–1360, doi:10.3390/genes6041347 (2015).

5 Katzianer, D. S., Yano, T., Rubin, H. & Zhu, J. A high-throughput small-molecule screen to identify a novel chemical inhibitor of Clostridium difficile. *Int J Antimicrob Agents* **44**, 69–73, doi:10.1016/j.ijantimicag.2014.03.007 (2014).

6 Gerding, D. N., Muto, C. A. & Owens, R. C., Jr. Measures to control and prevent Clostridium difficile infection. *Clin Infect Dis* **46 Suppl 1**, S43–49, doi:10.1086/521861 (2008).

7 Lawley, T. D. *et al.* Use of purified Clostridium difficile spores to facilitate evaluation of health care disinfection regimens. *Appl Environ Microbiol* **76**, 6895–6900, doi:10.1128/AEM.00718-10 (2010).

8 Weber, D. J., Rutala, W. A., Miller, M. B., Huslage, K. & Sickbert-Bennett, E. Role of hospital surfaces in the transmission of emerging health care-associated pathogens: norovirus, Clostridium difficile, and Acinetobacter species. *Am J Infect Control* **38**, S25–33, doi:10.1016/j.ajic.2010.04.196 (2010).

9 Maillard, J. Y. Innate resistance to sporicides and potential failure to decontaminate. *Journal of Hospital Infection* **77**, 204–209, doi:10.1016/j.jhin.2010.06.028 (2011).

10 Rutala, W. A., Gergen, M. F. & Weber, D. J. Efficacy of different cleaning and disinfection methods against Clostridium difficile spores: importance of physical removal versus sporicidal inactivation. *Infect Control Hosp Epidemiol* **33**, 1255–1258, doi:10.1086/668434 (2012).

11 Fekety, R. *et al.* Epidemiology of antibiotic-associated colitis; isolation of Clostridium difficile from the hospital environment. *Am J Med* **70**, 906–908 (1981).

12 Claro, T., Daniels, S. & Humphreys, H. Detecting Clostridium difficile spores from inanimate surfaces of the hospital environment: which method is best? *J Clin Microbiol* **52**, 3426–3428, doi:10.1128/JCM.01011-14 (2014).

13 Smits, W. K., Lyras, D., Lacy, D. B., Wilcox, M. H. & Kuijper, E. J. Clostridium difficile infection. *Nat Rev Dis Primers* **2**, 16020, doi:10.1038/nrdp.2016.20 (2016).

14 Hecht, G., Koutsouris, A., Pothoulakis, C., LaMont, J. T. & Madara, J. L. Clostridium difficile toxin B disrupts the barrier function of T84 monolayers. *Gastroenterology* **102**, 416–423 (1992).

15 Hecht, G., Pothoulakis, C., LaMont, J. T. & Madara, J. L. Clostridium difficile toxin A perturbs cytoskeletal structure and tight junction permeability of cultured human intestinal epithelial monolayers. *J Clin Invest* **82**, 1516–1524, doi:10.1172/JCI113760 (1988).

16 Aktories, K. & Just, I. Monoglucosylation of low-molecular-mass GTP-binding Rho proteins by clostridial cytotoxins. *Trends Cell Biol* **5**, 441–443 (1995).

17 Feltis, B. A. *et al.* Clostridium difficile toxins A and B can alter epithelial permeability and promote bacterial paracellular migration through HT-29 enterocytes. *Shock* **14**, 629–634 (2000).

18 Sun, X., Savidge, T. & Feng, H. The enterotoxicity of Clostridium difficile toxins. *Toxins (Basel)* **2**, 1848–1880, doi:10.3390/toxins2071848 (2010).

19 Hall, I. C. & O'Toole, E. Intestinal flora in new-born infants: With a description of a new pathogenic anaerobe, bacillus difficilis. *American Journal of Diseases of Children* **49**, 390-402, doi:10.1001/archpedi.1935.01970020105010 (1935).

20 George, R. H. *et al.* Identification of Clostridium difficile as a cause of pseudomembranous colitis. *Br Med J* **1**, 695 (1978).

21 George, W. L., Sutter, V. L., Goldstein, E. J., Ludwig, S. L. & Finegold, S. M. Aetiology of antimicrobial-agent-associated colitis. *Lancet* **1**, 802-803 (1978).

22 Inspection, C. f. H. A. a. Investigation into outbreaks of Clostridium difficile at Stoke Mandeville Hospital, Buckinghamshire Hospitals NHS Trust. (2006).

23 Kuijper, E. J. *et al.* Update of Clostridium difficile-associated disease due to PCR ribotype 027 in Europe. *Euro Surveill* **12**, E1-2 (2007).

24 Lessa, F. C. *et al.* Burden of Clostridium difficile infection in the United States. *N Engl J Med* **372**, 825-834, doi:10.1056/NEJMoa1408913 (2015).

25 Deshpande, A. *et al.* Risk factors for recurrent Clostridium difficile infection: a systematic review and meta-analysis. *Infect Control Hosp Epidemiol* **36**, 452-460, doi:10.1017/ice.2014.88 (2015).

26 Louie, T. J. *et al.* Fidaxomicin versus vancomycin for Clostridium difficile infection. *N Engl J Med* **364**, 422-431, doi:10.1056/NEJMoa0910812 (2011).

27 Lowy, I. *et al.* Treatment with monoclonal antibodies against Clostridium difficile toxins. *N Engl J Med* **362**, 197-205, doi:10.1056/NEJMoa0907635 (2010).

28 Rodriguez-Pardo, D. *et al.* Epidemiology of Clostridium difficile infection and risk factors for unfavorable clinical outcomes: results of a hospital-based study in Barcelona, Spain. *J Clin Microbiol* **51**, 1465-1473, doi:10.1128/JCM.03352-12 (2013).

29 Zhang, D., Prabhu, V. S. & Marcella, S. W. Attributable Healthcare Resource Utilization and Costs for Patients With Primary and Recurrent Clostridium difficile Infection in the United States. *Clin Infect Dis* **66**, 1326-1332, doi:10.1093/cid/cix1021 (2018).

30 Reveles, K. R., Lee, G. C., Boyd, N. K. & Frei, C. R. The rise in Clostridium difficile infection incidence among hospitalized adults in the United States: 2001-2010. *Am J Infect Control* **42**, 1028-1032, doi:10.1016/j.ajic.2014.06.011 (2014).

31 van Dorp, S. M. *et al.* Standardised surveillance of Clostridium difficile infection in European acute care hospitals: a pilot study, 2013. *Euro Surveill* **21**, doi:10.2807/1560-7917.ES.2016.21.29.30293 (2016).

32 Tenth Annual Report of the National Reference Laboratory for Clostridium difficile and results of the sentinel surveillance, May 2015- May 2016 <https://www.rivm.nl/Documenten_en_publicaties/Algemeen_Actueel/Uitgaven/Infectieziekten/CDiffNL/Tenth_Annual_Report_of_the_National_Reference_Laboratory_for_Clostridium_difficile_and_results_of_the_sentinel_surveillance> (2016).

33 Desai, K. *et al.* Epidemiological and economic burden of Clostridium difficile in the United States: estimates from a modeling approach. *BMC Infect Dis* **16**, 303, doi:10.1186/s12879-016-1610-3 (2016).

34 Zhang, S. *et al.* Cost of hospital management of Clostridium difficile infection in United States-a meta-analysis and modelling study. *BMC Infect Dis* **16**, 447, doi:10.1186/s12879-016-1786-6 (2016).

35 Heimann, S. M., Cruz Aguilar, M. R., Mellinghof, S. & Vehreschild, M. Economic burden and cost-effective management of Clostridium difficile infections. *Med Mal Infect* **48**, 23–29, doi:10.1016/j.medmal.2017.10.010 (2018).

36 Nanwa, N. *et al.* The economic impact of Clostridium difficile infection: a systematic review. *Am J Gastroenterol* **110**, 511–519, doi:10.1038/ajg.2015.48 (2015).

37 Kuijper, E. J. *et al.* Emergence of Clostridium difficile-associated disease in North America and Europe. *Clin Microbiol Infect* **12 Suppl 6**, 2–18, doi:10.1111/j.1469-0691.2006.01580.x (2006).

38 Wilcox, M. H. *et al.* Impact of recurrent Clostridium difficile infection: hospitalization and patient quality of life. *J Antimicrob Chemother* **72**, 2647–2656, doi:10.1093/jac/dkx174 (2017).

39 van Beurden, Y. H. *et al.* Cost analysis of an outbreak of Clostridium difficile infection ribotype 027 in a Dutch tertiary care centre. *J Hosp Infect* **95**, 421–425, doi:10.1016/j.jhin.2016.12.019 (2017).

40 Bauer, M. P. *et al.* Clostridium difficile infection in Europe: a hospital-based survey. *Lancet* **377**, 63–73, doi:10.1016/S0140-6736(10)61266-4 (2011).

41 Clements, A. C., Magalhaes, R. J., Tatem, A. J., Paterson, D. L. & Riley, T. V. Clostridium difficile PCR ribotype 027: assessing the risks of further worldwide spread. *Lancet Infect Dis* **10**, 395–404, doi:10.1016/S1473-3099(10)70080-3 (2010).

42 Goorhuis, A. *et al.* Spread and epidemiology of Clostridium difficile polymerase chain reaction ribotype 027/toxinotype III in The Netherlands. *Clin Infect Dis* **45**, 695–703, doi:10.1086/520984 (2007).

43 Kuijper, E. J. *et al.* Update of Clostridium difficile infection due to PCR ribotype 027 in Europe, 2008. *Euro Surveill* **13** (2008).

44 Pepin, J. *et al.* Clostridium difficile-associated diarrhea in a region of Quebec from 1991 to 2003: a changing pattern of disease severity. *CMAJ* **171**, 466–472, doi:10.1503/cmaj.1041104 (2004).

45 He, M. *et al.* Emergence and global spread of epidemic healthcare-associated Clostridium difficile. *Nat Genet* **45**, 109–113, doi:10.1038/ng.2478 (2013).

46 Goorhuis, A. *et al.* Clostridium difficile PCR Ribotype 078: an Emerging Strain in Humans and in Pigs? *J Clin Microbiol* **46**, 1157–1158, doi:10.1128/jcm.01536-07 (2008).

47 McDonald, L. C. Clostridium difficile: responding to a new threat from an old enemy. *Infect Control Hosp Epidemiol* **26**, 672–675, doi:10.1086/502600 (2005).

48 Patterson, L. *et al.* Morbidity and mortality associated with Clostridium difficile ribotype 078: a case-case study. *J Hosp Infect* **82**, 125–128, doi:10.1016/j.jhin.2012.07.011 (2012).

49 Pepin, J., Valiquette, L. & Cossette, B. Mortality attributable to nosocomial Clostridium difficile-associated disease during an epidemic caused by a hypervirulent strain in Quebec. *CMAJ* **173**, 1037–1042, doi:10.1503/cmaj.050978 (2005).

50 Abt, M. C., McKenney, P. T. & Pamer, E. G. Clostridium difficile colitis: pathogenesis and host defence. *Nat Rev Microbiol* **14**, 609–620, doi:10.1038/nrmicro.2016.108 (2016).

51 Hensgens, M. P., Goorhuis, A., Dekkers, O. M. & Kuijper, E. J. Time interval of increased risk for Clostridium difficile infection after exposure to antibiotics. *J Antimicrob Chemother* **67**, 742–748, doi:10.1093/jac/dkr508 (2012).

52 Davies, J. & Davies, D. Origins and evolution of antibiotic resistance. *Microbiol Mol Biol Rev* **74**, 417–433, doi:10.1128/MMBR.00016-10 (2010).

53 Peng, Z. *et al.* Update on Antimicrobial Resistance in Clostridium difficile: Resistance Mechanisms and Antimicrobial Susceptibility Testing. *J Clin Microbiol* **55**, 1998–2008, doi:10.1128/JCM.02250-16 (2017).

54 Bartlett, J. G. Clostridium difficile: history of its role as an enteric pathogen and the current state of knowledge about the organism. *Clin Infect Dis* **18 Suppl 4**, S265–272 (1994).

55 Bartlett, J. G. & Gerding, D. N. Clinical recognition and diagnosis of Clostridium difficile infection. *Clin Infect Dis* **46 Suppl 1**, S12–18, doi:10.1086/521863 (2008).

56 Kelly, C. P. & LaMont, J. T. Clostridium difficile—more difficult than ever. *N Engl J Med* **359**, 1932–1940, doi:10.1056/NEJMra0707500 (2008).

57 McFarland, L. V. Update on the changing epidemiology of Clostridium difficile-associated disease. *Nat Clin Pract Gastroenterol Hepatol* **5**, 40–48, doi:10.1038/ncpgasthep1029 (2008).

58 Johnson, A. P. & Wilcox, M. H. Fidaxomicin: a new option for the treatment of Clostridium difficile infection. *J Antimicrob Chemother* **67**, 2788–2792, doi:10.1093/jac/dks302 (2012).

59 Ooijevaar, R. E. *et al.* Update of treatment algorithms for Clostridium difficile infection. *Clin Microbiol Infect* **24**, 452–462, doi:10.1016/j.cmi.2017.12.022 (2018).

60 Martin, J. & Wilcox, M. New and emerging therapies for Clostridium difficile infection. *Curr Opin Infect Dis* **29**, 546–554, doi:10.1097/QCO.0000000000000320 (2016).

61 Cohen, S. H. *et al.* Clinical practice guidelines for Clostridium difficile infection in adults: 2010 update by the society for healthcare epidemiology of America (SHEA) and the infectious diseases society of America (IDSA). *Infect Control Hosp Epidemiol* **31**, 431–455, doi:10.1086/651706 (2010).

62 Debast, S. B., Bauer, M. P., Kuijper, E. J., European Society of Clinical, M. & Infectious, D. European Society of Clinical Microbiology and Infectious Diseases: update of the treatment guidance document for Clostridium difficile infection. *Clin Microbiol Infect* **20 Suppl 2**, 1–26, doi:10.1111/1469-0691.12418 (2014).

63 Johnson, S. *et al.* Vancomycin, metronidazole, or tolevamer for Clostridium difficile infection: results from two multinational, randomized, controlled trials. *Clin Infect Dis* **59**, 345–354, doi:10.1093/cid/ciu313 (2014).

64 Pepin, J., Valiquette, L., Gagnon, S., Routhier, S. & Brazeau, I. Outcomes of Clostridium difficile-associated disease treated with metronidazole or vancomycin before and after the emergence of NAP1/027. *Am J Gastroenterol* **102**, 2781–2788, doi:10.1111/j.1572-0241.2007.01539.x (2007).

65 Roshan, N., Hammer, K. A. & Riley, T. V. Non-conventional antimicrobial and alternative therapies for the treatment of Clostridium difficile infection. *Anaerobe* **49**, 103–111, doi:10.1016/j.anaerobe.2018.01.003 (2018).

66 Surawicz, C. M. *et al.* Guidelines for diagnosis, treatment, and prevention of Clostridium difficile infections. *Am J Gastroenterol* **108**, 478–498; quiz 499, doi:10.1038/ajg.2013.4 (2013).

67 Musher, D. M. *et al.* Clostridium difficile colitis that fails conventional metronidazole therapy: response to nitazoxanide. *J Antimicrob Chemother* **59**, 705–710, doi:10.1093/jac/dkl553 (2007).

68 Musher, D. M., Nuila, F. & Logan, N. The long-term outcome of treatment of Clostridium difficile colitis. *Clin Infect Dis* **45**, 523–524, doi:10.1086/520008 (2007).

69 Barbut, F. *et al.* Antimicrobial susceptibilities and serogroups of clinical strains of Clostridium difficile isolated in France in 1991 and 1997. *Antimicrob Agents Chemother* **43**, 2607–2611 (1999).

70 Brazier, J. S., Fawley, W., Freeman, J. & Wilcox, M. H. Reduced susceptibility of Clostridium difficile to metronidazole. *J Antimicrob Chemother* **48**, 741–742 (2001).

71 Indra, A. *et al.* First isolation of Clostridium difficile PCR ribotype 027 in Austria. *Euro Surveill* **11**, E060914 060913 (2006).

72 Pelaez, T. *et al.* Metronidazole resistance in Clostridium difficile is heterogeneous. *J Clin Microbiol* **46**, 3028–3032, doi:10.1128/JCM.00524-08 (2008).

73 Leeds, J. A., Sachdeva, M., Mullin, S., Barnes, S. W. & Ruzin, A. In vitro selection, via serial passage, of Clostridium difficile mutants with reduced susceptibility to fidaxomicin or vancomycin. *J Antimicrob Chemother* **69**, 41–44, doi:10.1093/jac/dkt302 (2014).

74 Spigaglia, P. Recent advances in the understanding of antibiotic resistance in Clostridium difficile infection. *Therapeutic Advances in Infectious Disease* **3**, 23–42, doi:10.1177/2049936115622891 (2016).

75 Peng, Z., Addisu, A., Alrabaa, S. & Sun, X. Antibiotic Resistance and Toxin Production of Clostridium difficile Isolates from the Hospitalized Patients in a Large Hospital in Florida. *Front Microbiol* **8**, 2584, doi:10.3389/fmicb.2017.02584 (2017).

76 Dridi, L., Tankovic, J., Burghoffer, B., Barbut, F. & Petit, J. C. gyrA and gyrB mutations are implicated in cross-resistance to Ciprofloxacin and moxifloxacin in Clostridium difficile. *Antimicrob Agents Chemother* **46**, 3418–3421 (2002).

77 Ackermann, G. *et al.* Resistance to moxifloxacin in toxigenic Clostridium difficile isolates is associated with mutations in gyrA. *Antimicrob Agents Chemother* **45**, 2348–2353, doi:10.1128/AAC.45.8.2348-2353.2001 (2001).

78 Drudy, D., Kyne, L., O'Mahony, R. & Fanning, S. gyrA mutations in fluoroquinolone-resistant Clostridium difficile PCR-027. *Emerg Infect Dis* **13**, 504–505, doi:10.3201/eid1303.060771 (2007).

79 Spigaglia, P., Barbanti, F., Louie, T., Barbut, F. & Mastrantonio, P. Molecular analysis of the gyrA and gyrB quinolone resistance-determining regions of fluoroquinolone-resistant Clostridium difficile mutants selected in vitro. *Antimicrob Agents Chemother* **53**, 2463–2468, doi:10.1128/AAC.01252-08 (2009).

80 Spigaglia, P. *et al.* Fluoroquinolone resistance in Clostridium difficile isolates from a prospective study of C. difficile infections in Europe. *J Med Microbiol* **57**, 784–789, doi:10.1099/jmm.0.47738-0 (2008).

81 Baines, S. D. & Wilcox, M. H. Antimicrobial Resistance and Reduced Susceptibility in Clostridium difficile: Potential Consequences for Induction, Treatment, and Recurrence of C. difficile Infection. *Antibiotics (Basel)* **4**, 267–298, doi:10.3390/antibiotics4030267 (2015).

82 Collins, J. *et al.* Dietary trehalose enhances virulence of epidemic Clostridium difficile. *Nature* **553**, 291–294, doi:10.1038/nature25178 (2018).

83 Farrow, K. A., Lyras, D. & Rood, J. I. Genomic analysis of the erythromycin resistance element Tn5398 from Clostridium difficile. *Microbiology* **147**, 2717–2728, doi:10.1099/00221287-147-10-2717 (2001).

84 Brouwer, M. S., Roberts, A. P., Mullany, P. & Allan, E. In silico analysis of sequenced strains of Clostridium difficile reveals a related set of conjugative transposons carrying a variety of accessory genes. *Mob Genet Elements* **2**, 8–12, doi:10.4161/mge.19297 (2012).

85 Brouwer, M. S., Warburton, P. J., Roberts, A. P., Mullany, P. & Allan, E. 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 (2011).

86 Corver, J. *et al.* Analysis of a Clostridium difficile PCR ribotype 078 100 kilobase island reveals the presence of a novel transposon, Tn6164. *BMC Microbiol* **12**, 130, doi:10.1186/1471-2180-12-130 (2012).

87 Roberts, A. P., Johanesen, P. A., Lyras, D., Mullany, P. & Rood, J. I. Comparison of Tn5397 from Clostridium difficile, Tn916 from Enterococcus faecalis and the CW459tet(M) element from Clostridium perfringens shows that they have similar conjugation regions but different insertion and excision modules. *Microbiology* **147**, 1243-1251, doi:10.1099/00221287-147-5-1243 (2001).

88 Roberts, A. P. & Mullany, P. Tn916-like genetic elements: a diverse group of modular mobile elements conferring antibiotic resistance. *FEMS Microbiol Rev* **35**, 856-871, doi:10.1111/j.1574-6976.2011.00283.x (2011).

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

90 Spigaglia, P., Barbanti, F. & Mastrantonio, P. New variants of the tet(M) gene in Clostridium difficile clinical isolates harbouring Tn916-like elements. *J Antimicrob Chemother* **57**, 1205-1209, doi:10.1093/jac/dkl105 (2006).

91 Spigaglia, P., Carucci, V., Barbanti, F. & Mastrantonio, P. ErmB determinants and Tn916-Like elements in clinical isolates of Clostridium difficile. *Antimicrob Agents Chemother* **49**, 2550-2553, doi:10.1128/AAC.49.6.2550-2553.2005 (2005).

92 Arenz, S., Nguyen, F., Beckmann, R. & Wilson, D. N. Cryo-EM structure of the tetracycline resistance protein TetM in complex with a translating ribosome at 3.9-A resolution. *Proc Natl Acad Sci U S A* **112**, 5401-5406, doi:10.1073/pnas.1501775112 (2015).

93 Blair, J. M., Webber, M. A., Baylay, A. J., Ogbolu, D. O. & Piddock, L. J. Molecular mechanisms of antibiotic resistance. *Nat Rev Microbiol* **13**, 42-51, doi:10.1038/nrmicro3380 (2015).

94 Hansen, L. H. & Vester, B. A cfr-like gene from Clostridium difficile confers multiple antibiotic resistance by the same mechanism as the cfr gene. *Antimicrob Agents Chemother* **59**, 5841-5843, doi:10.1128/AAC.01274-15 (2015).

95 Long, K. S., Poehlsgaard, J., Kehrenberg, C., Schwarz, S. & Vester, B. The Cfr rRNA methyltransferase confers resistance to Phenicols, Lincosamides, Oxazolidinones, Pleuromutilins, and Streptogramin A antibiotics. *Antimicrob Agents Chemother* **50**, 2500-2505, doi:10.1128/AAC.00131-06 (2006).

96 Schwarz, S. & Wang, Y. Nomenclature and functionality of the so-called cfr gene from Clostridium difficile. *Antimicrob Agents Chemother* **59**, 2476-2477, doi:10.1128/AAC.04893-14 (2015).

97 Lubelski, J., Konings, W. N. & Driessen, A. J. Distribution and physiology of ABC-type transporters contributing to multidrug resistance in bacteria. *Microbiol Mol Biol Rev* **71**, 463-476, doi:10.1128/MMBR.00001-07 (2007).

98 Putman, M., Van Veen, H. W., Degener, J. E. & Konings, W. N. Antibiotic resistance: era of the multidrug pump. *Mol Microbiol* **36**, 772-773 (2000).

99 Ngernsombat, C., Sreesai, S., Harnvoravongchai, P., Chankhamhaengdecha, S. & Janvilisri, T. CD2068 potentially mediates multidrug efflux in Clostridium difficile. *Sci Rep* **7**, 9982, doi:10.1038/s41598-017-10155-x (2017).

100 Awad, M. M., Johanesen, P. A., Carter, G. P., Rose, E. & Lyras, D. Clostridium difficile virulence factors: Insights into an anaerobic spore-forming pathogen. *Gut Microbes* **5**, 579-593, doi:10.4161/19490976.2014.969632 (2014).

101 Smits, W. K. Hype or hypervirulence: a reflection on problematic C. difficile strains. *Virulence* **4**, 592–596, doi:10.4161/viru.26297 (2013).

102 Cornely, O. A. *et al.* Clinical efficacy of fidaxomicin compared with vancomycin and metronidazole in Clostridium difficile infections: a meta-analysis and indirect treatment comparison. *J Antimicrob Chemother* **69**, 2892–2900, doi:10.1093/jac/dku261 (2014).

103 Feher, C., Soriano, A. & Mensa, J. A Review of Experimental and Off-Label Therapies for Clostridium difficile Infection. *Infect Dis Ther* **6**, 1–35, doi:10.1007/s40121-016-0140-z (2017).

104 Al-Nassir, W. N. *et al.* Both oral metronidazole and oral vancomycin promote persistent overgrowth of vancomycin-resistant enterococci during treatment of Clostridium difficile-associated disease. *Antimicrob Agents Chemother* **52**, 2403–2406, doi:10.1128/AAC.00090-08 (2008).

105 Ivarsson, M. E., Leroux, J. C. & Castagner, B. Investigational new treatments for Clostridium difficile infection. *Drug Discov Today* **20**, 602–608, doi:10.1016/j.drudis.2014.12.003 (2015).

106 Basseres, E., Endres, B. T., Dotson, K. M., Alam, M. J. & Garey, K. W. Novel antibiotics in development to treat Clostridium difficile infection. *Curr Opin Gastroenterol* **33**, 1–7, doi:10.1097/MOG.0000000000000332 (2017).

107 McDonald, L. C. *et al.* Clinical Practice Guidelines for Clostridium difficile Infection in Adults and Children: 2017 Update by the Infectious Diseases Society of America (IDSA) and Society for Healthcare Epidemiology of America (SHEA). *Clin Infect Dis* **66**, 987–994, doi:10.1093/cid/ciy149 (2018).

108 Tsutsumi, L. S., Owusu, Y. B., Hurdle, J. G. & Sun, D. Progress in the discovery of treatments for C. difficile infection: A clinical and medicinal chemistry review. *Curr Top Med Chem* **14**, 152–175 (2014).

109 Worthington, R. J. & Melander, C. Combination approaches to combat multidrug-resistant bacteria. *Trends Biotechnol* **31**, 177–184, doi:10.1016/j.tibtech.2012.12.006 (2013).

110 Wilcox, M. H. *et al.* in *European Congress of Clinical Microbiology and Infectious Diseases (ECCMID).*

111 Butler, M. M. *et al.* MBX-500, a hybrid antibiotic with in vitro and in vivo efficacy against toxigenic Clostridium difficile. *Antimicrob Agents Chemother* **56**, 4786–4792, doi:10.1128/AAC.00508-12 (2012).

112 Butler, M. M. *et al.* Antibacterial activity and mechanism of action of a novel anilinouracil-fluoroquinolone hybrid compound. *Antimicrob Agents Chemother* **51**, 119–127, doi:10.1128/AAC.01311-05 (2007).

113 Steele, J. *et al.* MBX-500 is effective for treatment of Clostridium difficile infection in gnotobiotic piglets. *Antimicrob Agents Chemother* **57**, 4039–4041, doi:10.1128/AAC.00304-13 (2013).

114 Clements, J. E., D'Ambrosio, J. & Brown, N. C. Inhibition of Bacillus subtilis deoxyribonucleic acid polymerase III by phenylhydrazinopyrimidines. Demonstration of a drug-induced deoxyribonucleic acid-enzyme complex. *J Biol Chem* **250**, 522–526 (1975).

115 Torti, A. *et al.* Clostridium difficile DNA polymerase IIIC: basis for activity of antibacterial compounds. *Curr Enzym Inhib* **7**, 147–153 (2011).

116 Dvoskin, S. *et al.* A novel agent effective against Clostridium difficile infection. *Antimicrob Agents Chemother* **56**, 1624–1626, doi:10.1128/AAC.06097-11 (2012).

117 Briggs, G. S., Smits, W. K. & Soultanas, P. Chromosomal replication initiation machinery of low-G+C-content Firmicutes. *J Bacteriol* **194**, 5162–5170, doi:10.1128/JB.00865-12 (2012).

118 Lemon, K., Moriya, S., Ogasawara, N. & Grossman, A. in *Bacillus subtilis and its closest relatives* (eds AL Sonenshein, JA Hock , & R Losick) 73-86 (ASM Press, 2002).

119 Wolf, M., Muller, T., Dandekar, T. & Pollack, J. D. Phylogeny of Firmicutes with special reference to Mycoplasma (Mollicutes) as inferred from phosphoglycerate kinase amino acid sequence data. *International journal of systematic and evolutionary microbiology* **54**, 871-875, doi:10.1099/ijs.0.02868-0 (2004).

120 Messer, W. The bacterial replication initiator DnaA. DnaA and oriC, the bacterial mode to initiate DNA replication. *FEMS Microbiol Rev* **26**, 355-374 (2002).

121 Richardson, T. T., Harran, O. & Murray, H. The bacterial DnaA-trio replication origin element specifies single-stranded DNA initiator binding. *Nature* **534**, 412-416, doi:10.1038/nature17962 (2016).

122 Smits, W. K., Goranov, A. I. & Grossman, A. D. 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 (2010).

123 Soultanas, P. Loading mechanisms of ring helicases at replication origins. *Mol Microbiol* **84**, 6-16, doi:10.1111/j.1365-2958.2012.08012.x (2012).

124 Sanders, G. M., Dallmann, H. G. & McHenry, C. S. Reconstitution of the B.subtilis replisome with 13 proteins including two distinct replicases. *Molecular cell* **37**, 273-281, doi:10.1016/j.molcel.2009.12.025 (2010).

125 Sakamoto, Y., Nakai, S., Moriya, S., Yoshikawa, H. & Ogasawara, N. The Bacillus subtilis dnaC gene encodes a protein homologous to the DnaB helicase of Escherichia coli. *Microbiology* **141 (Pt 3)**, 641-644, doi:10.1099/13500872-141-3-641 (1995).

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

127 O'Donnell, M., Langston, L. & Stillman, B. Principles and concepts of DNA replication in bacteria, archaea, and eukarya. *Cold Spring Harbor perspectives in biology* **5**, doi:10.1101/cshperspect.a010108 (2013).

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

129 Noirot-Gros, M. F. *et al.* An expanded view of bacterial DNA replication. *Proc Natl Acad Sci U S A* **99**, 8342-8347, doi:10.1073/pnas.122040799 (2002).

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

131 Dervyn, E. *et al.* Two essential DNA polymerases at the bacterial replication fork. *Science (New York, N.Y.)* **294**, 1716-1719, doi:10.1126/science.1066351 (2001).

132 Bruck, I. & O'Donnell, M. The DNA replication machine of a gram-positive organism. *J Biol Chem* **275**, 28971-28983, doi:10.1074/jbc.M003565200 (2000).

133 Timinskas, K., Balvociute, M., Timinskas, A. & Venclovas, C. Comprehensive analysis of DNA polymerase III alpha subunits and their homologs in bacterial genomes. *Nucleic acids research* **42**, 1393-1413, doi:10.1093/nar/gkt900 (2014).

134 Su'etsugu, M. & Errington, J. The replicase sliding clamp dynamically accumulates behind progressing replication forks in Bacillus subtilis cells. *Molecular cell* **41**, 720-732, doi:10.1016/j.molcel.2011.02.024 (2011).

135 Afonso, J. P. *et al.* Insights into the structure and assembly of the Bacillus subtilis clamp-loader complex and its interaction with the replicative helicase. *Nucleic acids research* **41**, 5115-5126, doi:10.1093/nar/gkt173 (2013).

136 Petit, M. A. & Ehrlich, S. D. The NAD-dependent ligase encoded by yerG is an essential gene of Bacillus subtilis. *Nucleic acids research* **28**, 4642-4648 (2000).

137 Lohman, T. M. & Ferrari, M. E. Escherichia coli single-stranded DNA-binding protein: multiple DNA-binding modes and cooperativities. *Annual review of biochemistry* **63**, 527-570, doi:10.1146/annurev.bi.63.070194.002523 (1994).

138 Sissi, C. & Palumbo, M. In front of and behind the replication fork: bacterial type IIA topoisomerases. *Cellular and molecular life sciences : CMLS* **67**, 2001-2024, doi:10.1007/s00018-010-0299-5 (2010).

139 Davey, M. J. & O'Donnell, M. Replicative helicase loaders: ring breakers and ring makers. *Current Biology* **13**, R594-R596, doi:10.1016/s0960-9822(03)00523-2 (2003).

140 Goranov, A. I., Breier, A. M., Merrikh, H. & Grossman, A. D. YabA of Bacillus subtilis controls DnaA-mediated replication initiation but not the transcriptional response to replication stress. *Mol Microbiol* **74**, 454-466, doi:10.1111/j.1365-2958.2009.06876.x (2009).

141 Hayashi, M., Ogura, Y., Harry, E. J., Ogasawara, N. & Moriya, S. Bacillus subtilis YabA is involved in determining the timing and synchrony of replication initiation. *FEMS microbiology letters* **247**, 73-79, doi:10.1016/j.femsle.2005.04.028 (2005).

142 Kaguni, J. M. DnaA: controlling the initiation of bacterial DNA replication and more. *Annual review of microbiology* **60**, 351-375, doi:10.1146/annurev.micro.60.080805.142111 (2006).

143 Katayama, T., Ozaki, S., Keyamura, K. & Fujimitsu, K. Regulation of the replication cycle: conserved and diverse regulatory systems for DnaA and oriC. *Nat Rev Microbiol* **8**, 163-170, doi:10.1038/nrmicro2314 (2010).

144 Leonard, A. C. & Grimwade, J. E. Initiation of DNA Replication. *EcoSal Plus* **4**, doi:10.1128/ecosalplus.4.4.1 (2010).

145 Leonard, A. C. & Grimwade, J. E. Regulating DnaA complex assembly: it is time to fill the gaps. *Current opinion in microbiology* **13**, 766-772, doi:10.1016/j.mib.2010.10.001 (2010).

146 Merrikh, H. & Grossman, A. D. Control of the replication initiator DnaA by an anti-cooperativity factor. *Mol Microbiol* **82**, 434-446, doi:10.1111/j.1365-2958.2011.07821.x (2011).

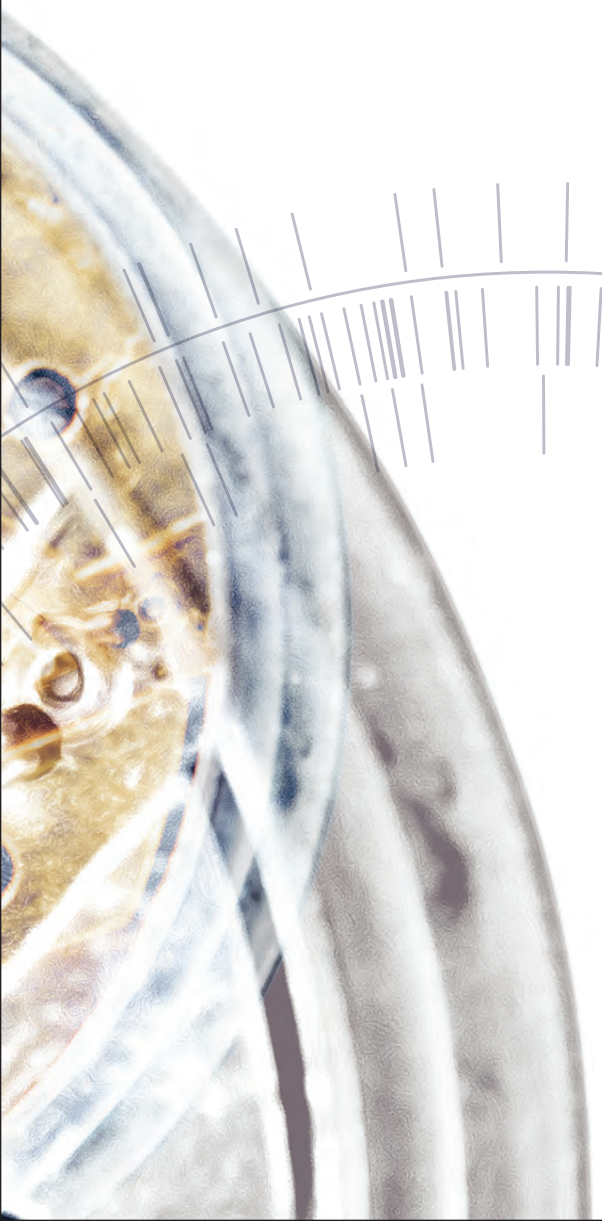
147 Mott, M. L. & Berger, J. M. DNA replication initiation: mechanisms and regulation in bacteria. *Nat Rev Microbiol* **5**, 343-354, doi:10.1038/nrmicro1640 (2007).

148 Fossum, S. *et al.* A robust screen for novel antibiotics: specific knockout of the initiator of bacterial DNA replication. *FEMS microbiology letters* **281**, 210-214, doi:10.1111/j.1574-6968.2008.01103.x (2008).

Chapter 2

DNA replication proteins as potential targets for antimicrobials in drug–resistant bacterial pathogens

Erika van Eijk [1]
Bert Wittekoek [1]
Ed J. Kuijper [1]
Wiep Klaas Smits [1]



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

Published in Journal of Antimicrobial Chemotherapy, 2017

Abstract

With the impending crisis of antimicrobial resistance, there is an urgent need to develop novel antimicrobials to combat difficult infections and multidrug resistant pathogenic microorganisms. DNA replication is essential for cell viability and is therefore an attractive target for antimicrobials. Although several antimicrobials targeting DNA replication proteins have been developed to date, gyrase/topoisomerase inhibitors are the only class widely used in the clinic. Given the numerous essential proteins in the bacterial replisome that may serve as a potential target for inhibitors and the relative paucity of suitable compounds, it is evident that antimicrobials targeting the replisome are underdeveloped so far. In this review, we report on the diversity of antimicrobial compounds targeting DNA replication and highlight some of the challenges in developing new drugs that target this process.

Introduction

The increase in multidrug resistant bacteria has resulted in limited treatment options, and therefore the development of compounds directed against these micro-organisms is of utmost importance. In recent years, the pipeline of new anti-microbials has almost dried up, apart from the approved follow-up compounds (2nd, 3r , 4th generation) which have the same mode of action as their predecessors [1]. The development of antimicrobials derived from existing scaffolds is not without risk, as these compounds may be vulnerable to the same resistance mechanisms. Therefore, exploring new potential targets and/or increasing structural diversity in the next-generation antimicrobials are paramount in minimizing the risk of rapid acquisition of antimicrobial resistance. There are several essential cellular processes that can serve as targets for novel antimicrobials and many of these are exploited by antimicrobials. Of particular interest for this review is DNA replication. Correct replication of DNA by a multi-protein complex, the replisome, and proteins associated with it (**Table 1, Figure 1**) is an essential requirement for cell viability. The 'core' replisome complex consists of helicase, primase, DNA polymerase, sliding clamp, clamp loader and single-stranded DNA binding (SSB) proteins. Stringent coordination of this complex is essential for DNA replication, and inhibition of the function of any of these proteins or their interactions in principle disrupts the process and results in cell death [2]. Other proteins that are crucial for DNA replication include topoisomerase II and DNA ligase.

Despite the potential of replication proteins to serve as a target for antimicrobial compounds, clinical use has primarily been limited to topoisomerase II inhibitors that target DNA gyrase and/or topoisomerase IV (TopoIV). In this review we will discuss inhibitors that target 'core' replisome proteins as well as associated proteins that are crucial for DNA replication. We illustrate three key challenges (antimicrobial resistance, specificity, and exploration of new targets) and potential strategies to meet these challenges using examples of novel DNA replication-targeting antimicrobials active against *Clostridium difficile* and other multidrug resistant pathogens.

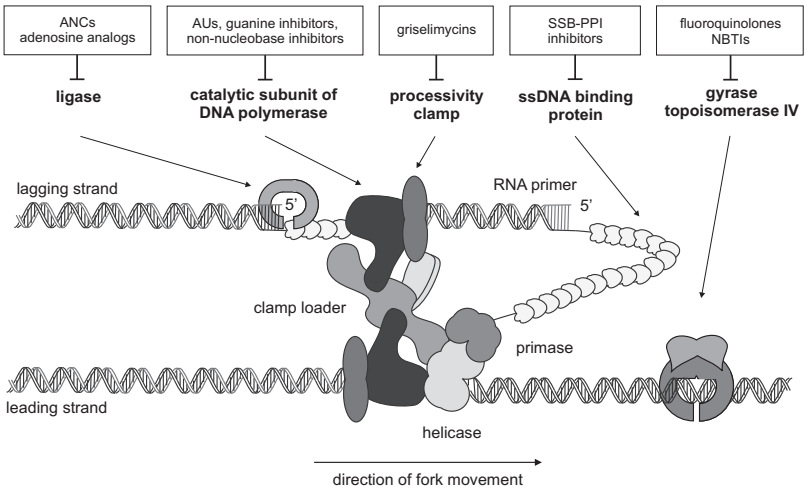


Figure 1. Schematic representation of the variety of targets of antimicrobials in the bacterial replisome.

Indicated is the core of the replisome and the other proteins that have been targeted by antimicrobial compounds. For simplicity, replication initiation proteins and regulators have been omitted from this figure. Important classes of drugs inhibiting specific proteins are boxed. The activity of all proteins is described in the main text. PPI; protein–protein interaction.

Table 1. Diversity of targets in the bacterial DNA replication machinery.

Function	*E. coli*	*B. subtilis*	*C. difficile* [a]	Activity	Inhibitors in (pre) clinical development
Chromosomal replication initiator protein	DnaA	DnaA	CD0001	Initiation of DNA replication at *oriC*	
Replicative helicase	DnaB	DnaC	CD3657 [b]	Unwinding of double stranded DNA at the replication fork	
Replicative helicase loader	DnaC	DnaI	CD3654 [b]	Required for functional loading of the replicative helicase	
Primosome protein	–	DnaB	–	Replication initiation and membrane attachment; enhancing the helicase loading process, origin remodelling	
Primosome protein	–	DnaD	CD3653	Initiation of DNA replication through interactions with other initiation proteins, origin remodelling	
Primase	DnaG	DnaG	CD1454 [b]	Synthesis of primers on the lagging strand	
Primosomal protein N′	PriA	PriA	CD2586		
Sliding clamp	β (DnaN)	DnaN	CD0002		Griselimycins
Clamp loader complex	Multiple proteins, including γ and τ (DnaX)	Multiple proteins, including DnaX	Multiple proteins, including CD0016		
DNA polymerase III α subunit DnaE	DnaE	DnaE	CD3396	Elongation of leading and lagging strand during DNA synthesis (E.coli); initial extension of the RNA primers on the lagging strand (B.subtilis)	Guanine inhibitors
DNA polymerase III α subunit PolC	–	PolC	CD1305	Elongation of both leading and lagging strand during DNA synthesis	AUs, guanine inhibitors, non-nucleobase inhibitors
DNA polymerase I	PolA	PolA	CD1128	Removal of RNA primers and gap filling	
Gyrase	GyrA, GyrB	GyrA, GyrB	CD0005, CD0006	Relaxing the DNA double helix by introducing negative supercoils, catenating and decatenating DNA rings	Fluoroquinolones, NBTIs
Topoisomerase IV	ParE, ParC	ParE, ParC	–		Fluoroquinolones, NBTIs
DNA ligase	LigA	LigA	CD3309	Joining of Okazaki fragments during DNA replication.	ANCs, adenosine analogues
Single-stranded DNA binding protein	SSB	SSB	CD3662, CD3235	Preventing degradation of ssDNA in the replication fork; protein interaction platform	PPI inhibitors

OriC, chromosomal origin of replication.

a Putative; most replication proteins of *C. difficile* are not characterized (apart from PolC [3]).

b These have been recently characterized [4] (Chapters 4 and 5).

Clinically used antimicrobials targeting DNA replication: topoisomerase II inhibitors

The two bacterial topoisomerase II enzymes – DNA gyrase and TopoIV – modify the topology of DNA during replication [5]. Gyrase and TopoIV are tetramers composed of two GyrA and GyrB subunits (encoded by *gyrA* and *gyrB*) or two ParC and ParE subunits (encoded by *parC* and *parE*), respectively [6,7]. Despite structural similarities [5], the two topoisomerase II enzymes perform distinct functions in prokaryotes. Gyrase is able to introduce negative supercoiling into DNA, thereby relaxing the DNA helix, while TopoIV is specialized in DNA decatenation and unknotting of DNA [8-10]. Some bacteria only encode gyrase in their genome; it is possible that gyrase in these bacteria can efficiently decatenate DNA without TopoIV, as was shown with *Mycobacterium smegmatis* [11].

Existing topoisomerase II inhibitors can roughly be divided into (i) ATPase inhibitors and (ii) topoisomerase II poisons that interfere with the catalytic DNA cleavage/joining process [12]. The first group is represented by the aminocoumarin class, the second group by the fluoroquinolones (FQs).

Aminocoumarins, such as the naturally occurring novobiocin, are cyclic peptides that compete with ATP to bind GyrB [6,13]. Though novobiocin has little structural similarity with ATP, the binding sites of the drug partially overlaps with the binding sites of ATP in GyrB [14]. Due to this overlap, mutations that confer aminocoumarin resistance are likely to affect enzymatic activity of the topoisomerase [12]. Novobiocin was licensed for treatment of infections by staphylococci and other susceptible organisms, but the clinical use of aminocoumarins is very limited due to poor pharmacological properties (e.g. poor solubility, poor absorption). Though improvement of pharmacological properties may yield clinical candidates, none have entered trials yet [15].

Fluoroquinolones (FQs) are the most successful class of antimicrobials targeting DNA replication and one of the most widely used antimicrobials on the market [2]. The FQ mode of action is to stabilize cleaved DNA-topoisomerase II complexes, thereby increasing the number of double-stranded DNA breaks in the bacterial cell [16,17]. Rapid cell death induced by FQs is likely the consequence of

chromosome fragmentation, while inhibition of DNA replication results in reduced cell growth instead of cell death[16]. Most FQs are able to inhibit both gyrase and TopoIV with different efficiencies, with actual target preference depending on the specific compound and the bacterial species against which it is used[16,18-20].

A major concern is the rise of FQ-resistant pathogens. Fluoroquinolones are commonly used to treat infections by Enterobacteriaceae, non-fermenting Gram-negative bacilli (in particular *Pseudomonas aeruginosa* and *Acinetobacter baumannii*) and *Mycobacterium tuberculosis* [21,22], but resistance can also occur when FQs are used to treat infections with a different pathogen. For instance, increased use of FQs and simultaneous development of fluoroquinolone resistance in clinical isolates of *C. difficile* resulted in the emergence of the epidemic PCR ribotype 027 as evidenced from whole-genome sequence data [23,24], even though FQs are not the drug of choice to treat *C. difficile* infections. FQ resistance is mainly acquired through mutations in the so-called quinolone resistance-determining regions (QRDRs) of the gyrase and/or topoIV genes [25]. In most resistant pathogens the mutations are located in the *gyrA* and/or *parC*, and rarely in *gyrB* or *parE*. In most Gram-positive bacteria, TopoIV is the primary target for fluoroquinolones and resistance mutations arise first in *parC*. In contrast, mutations in most Gram-negative bacteria occur first in *gyrA* [26]. Single-step mutations can lead to resistance and the *C. difficile* example illustrates how such a single mutation can fuel an epidemic with detrimental clinical outcome. FQ resistance can also be conferred by non-specific efflux systems that can export quinolones and other antimicrobial agents or by plasmids harbouring a quinolone resistance determinant [21].

Antimicrobials targeting DNA replication under development

There are many different compounds that have been identified as DNA replication inhibitors with potential to be used as an antimicrobial. These have been comprehensively reviewed elsewhere [2,27]. Here, we discuss several classes of DNA replication antimicrobials to highlight the diversity of replication proteins that can be exploited as targets and indicate some of the new developments.

Novel bacterial topoisomerase inhibitors (NBTIs)

Increasing FQ resistance has spurred the development of novel topoisomerase inhibitors that are active against FQ-resistant gyrase or TopoIV. Besides modifying existing FQ scaffolds [28], novel non-quinolone topoisomerase II inhibitors, which are collectively called novel bacterial topoisomerase inhibitors (NBTIs), have been developed [15]. Antimicrobials of this class target the catalytic core of topoisomerase, as do FQs, but NBTIs have a different mode of action. For example, the NBTI GKS299423 (GlaxoSmithKline) was shown to stabilize a pre-cleavage complex, in contrast to the cleavage-complex with double-stranded DNA breaks targeted by FQs [29]. Examples of other NBTI compounds are a series of pyrazole derivatives and related tetrahydroindazoles [30], NXL101 [31], NBTI 5463 [32], gyramides (N-benzyl-3-sulfonamidopyrrolidines) [33], and ACT-387042 and ACT-292706 [34]. None of these compounds have entered clinical trials yet.

DNA ligase inhibitors

The primary function of bacterial nicotinamide adenine dinucleotide (NAD+)-dependent DNA ligase (LigA) is to join the Okazaki fragments for the completion of lagging-strand DNA replication synthesis [35]. It is an attractive target for the development of novel antimicrobials as the ligase gene is present in all bacterial genomes and was found to be essential in several key pathogens [35,36]. It shares limited similarity to eukaryotic DNA ligase [36,37] and the binding of NAD+ as a substrate for ligase activity is a unique feature of bacterial ligase, limiting the potential for toxicity in humans.

All ligase inhibitors that have been developed to date are competitive inhibitors that bind a hydrophobic pocket near the binding site of the NAD+ substrate [38]. Four predominant classes of LigA inhibitors have been identified and here we will focus on the two most promising classes, the 2-amino-[1,8]-naphthyridine-3-carboxamides (ANCs) and the adenosine analogues [38].

The ANC scaffold was identified in high-throughput screening against *Escherichia coli* LigA (IC_{50} of 25 μM) but displayed better activity against *Staphylococcus aureus* LigA (IC_{50} of 2.5 μM) [39]. The compound was not active against wild-

type *E. coli* cells, however, due to poor permeability and efflux. Optimization of the ANC scaffold yielded compounds with improved activity (MIC 1-8 mg/L) against primarily Gram-positive pathogens (including MRSA and *C. difficile*), and promising results in animal models [39].

Adenosine analogues were identified in high-throughput screening for inhibitors of *Haemophilus influenzae* LigA [40,41]. Biochemical experiments and X-ray crystallography showed that this class of compounds competes with NAD+ and blocks the AMP-binding pocket of the LigA adenylation domain. Adenosine analogues displayed good antimicrobial activities against both Gram-positive and Gram-negative bacteria *in vivo* (MIC 1-8 mg/L) [40]. The initial adenosine analogues had favourable pharmacological properties (mainly good solubility) and one compound showed promise in *S. aureus* thigh infection and *Streptococcus pneumoniae* lung infection animal models [40]. However, the adenosine analogues were rapidly cleared in rats due to metabolism by cytochrome P-450s [41,42]. Despite attempts to optimize the solubility, antimicrobial activity and the clearance of these compounds, there are no reports published on further progress.

Although the hydrophobic pocket near the binding site of the NAD+ substrate enables specific inhibition of the bacterial LigA by competitive inhibitors, it is not directly engaged in interactions with the NAD+ substrate [35]. A spontaneous resistant mutant of *S. aureus* contains a single leucine-to-phenylalanine mutation in the hydrophobic pocket that does not affect the ligase activity of the LigA mutant but leads to a significant loss of target binding by adenosine analogues [35,40]. This relatively high risk of resistance against this class of compounds may have contributed to the fact that no LigA inhibitors have been entered into human trials to date.

DNA polymerase III inhibitors

The bacterial DNA polymerase III α subunit (PolIII) is an essential enzyme for DNA replication as it is responsible for the synthesis of DNA and its potential as antimicrobial target has been noted in recent *in silico* analyses of *E. coli* O157:H7 and *M. tuberculosis* [43,44]. Moreover, the distribution of two different homologues of the α subunit, PolC and DnaE, between Gram-positive and Gram-negative bacteria

presents an opportunity to develop inhibitors specific to either group [45]. The first inhibitor of PolIII, 6-(p-Hydroxyphenylazo)uracil (HPUra) was identified in 1970 [46]. The majority of PolIII inhibitors specifically target PolC of low-G+C Gram-positive bacteria. Indeed, the single DnaE-specific compound reported to date (324C) did not demonstrate any *in vivo* antimicrobial activity against *B. subtilis*, while it was highly active against purified *B. subtilis* DnaE *in vitro* [47] . The reason(s) for this is (are) unclear.

PolIII inhibitors can be categorized into three main classes: (i) the 6-anilinouracils (AU); (ii) the guanine inhibitors; and (iii) the non-nucleobase inhibitors. The AU class of PolIII inhibitors, which includes HPUra, is composed of a uracil-containing base-pairing domain that binds the DNA at cytosine bases and an aryl domain that determines the selectivity and affinity for PolC [48]. AUs competitively inhibit PolC with respect to dGTP through simultaneous binding to the cytosine of the DNA strand and near the active site of PolC, resulting in a ternary inactive complex of AU inhibitor, DNA, and PolC [48,49]. HPUra served as a scaffold for the development of numerous AUs with a broad range of antimicrobial and pharmacological properties. Two promising AUs, 6-(3-ethyl-4-methylanilino) uracil (EMAU) and 6-([3,4-trimethylene]anilino) uracil (TMAU) were highly active against PolC *in vitro*, but required optimization to increase activity against various Gram-positive bacteria, including MRSA [48,50-52]. Improvement of solubility of AUs compromises antimicrobial activity but allowed for the production of compounds that could be delivered intravenously rather than subcutaneously in animal models of infection [48,52-54]. The frequencies of mutations leading to AU resistance ranged from 3.6×10^{-10} to 1.2×10^{-8}, comparable to the frequency of ciprofloxacin resistance [55]. All mutations that conferred AU resistance were located at a specific amino acid in the presumed dNTP/AU-binding site of PolC [55,56]. Unexpectedly, however, the polymerase activity of these mutant PolC enzymes was unchanged in comparison with the wild-type *in vitro*.

The guanine inhibitors of bacterial DNA polymerase III have a purine moiety as the base-pairing domain instead of the uracil ring of the AUs. They act via the same active site -directed competitive inhibition as AUs, but are active against both PolC and DnaE [49,57,58]. Two lines of guanines inhibitors have been developed: the *N*2-(3,4-dichlorobenzyl) guanines (DCBGs) and the *N*2-(3-ethyl-4-methylphenyl)guanines (EMPGs). *N*7-substituted DCBGs and EMPGs displayed potent

in vitro antimicrobial activities against several Gram-positive pathogens but showed limited efficacy in animal models [49,58]. Similar to the AUs, the DCBGs have poor solubility in water and no attempts have been made to produce soluble analogues so far.

The non-nucleobase class of DNA polymerase III inhibitors includes anilinopyrimidinediones (APs) and quinazolin-2-ylamino-quinazolin-4-ols (BisQuinols). The APs are structural isomers of AUs and are also competitive inhibitors of dGTP [59]. They show minimal cytotoxicity and moderate antimicrobial activities (MIC ranging from 8 to 16 mg/L) [59]. In contrast to AUs and APs, BisQuinols have been suggested to compete with the DNA template, rather than nucleotides [60]. Though BisQuinol analogues were able to inhibit Gram- positive pathogens, they were unselective for mammalian Polδ, which raises concerns about cytotoxicity.

There is no information on resistance development against the guanine inhibitors or non-nucleobase inhibitors, but considering the similarities in mode of action to the AUs, caution is warranted.

Inhibitors of other replication related proteins

The sliding clamp, or A subunit of the replication machinery, is a polymerase processivity factor [61,62]. The protein can be targeted by novel griselimycins [63]. Griselimycin is a natural product of *Streptomyces griseus* with specific activity against the Corynebacterineae suborder, including *Mycobacterium* species [64]. The development of this class of compounds as anti-tuberculosis drugs was initially abandoned as rifampicin became available for treatment, but revisited in light of its activity against drug-resistant isolates [65]. Poor pharmacokinetic properties of griselimycin were addressed by the total synthesis of derivatives [63]. In particular, cyclohexylgriselimycin was highly active against *M. tuberculosis in vitro* and in a mouse model of infection, comparable to isoniazid. Evidence for the mechanism of action came from observations that the griselimycin biosynthetic operon contains a sliding clamp homologue capable of conferring resistance to a susceptible *Streptomyces* strain, the selective amplification of a *dnaN* containing chromosomal fragment in mycobacterial strains with evolved resistance (i.e. resistant strains harbouring multiple copies of the *dnaN* gene), and the crystal structure of griselimycin in complex with DnaN [63].

Single-stranded DNA at the replication fork is stabilized and protected by single-stranded binding (SSB) proteins [66]. These proteins are also an integral part of nucleoprotein complexes involved recombination and repair [67,68]. Although present in all domains of life, amino acid sequence, subunit composition and oligomeric state of these proteins differ substantially between organisms [67]. The essential role of SSB protein-protein interactions (SSB-PPI) and low sequence similarity between eukaryotic and prokaryotic SSBs allows the potential development of SBB-PPI inhibitors that could serve as novel antimicrobials [68]. Indeed, small molecules that interfere with the interaction between SSB and one or several binding partners have been identified by a high- throughput fluorescence polarization assay [68,69]. Inhibition by these compounds is based on mimicking the SSB protein C-terminus, which acts as a platform for interaction with other proteins, or more targeted inhibition of the SSB protein/Exonuclease I (binding partner) interface [68].

Challenges in developing novel antimicrobials targeting replication

We consider three main challenges in the development of novel antimicrobial compounds. First, novel antimicrobials should overcome resistance to known drugs and minimize development of resistance against the new drug. Second, they should preferably be specific to the microorganism/pathogen of interest to prevent dysbiosis of the host microbiome. The third challenge is to move away from modification and optimization of existing scaffolds that inhibit established cellular targets and explore novel targets and mechanisms of action. In the following section, we illustrate these challenges with a focus on compounds inhibiting the multidrug resistant organism *C. difficile*.

Challenge: antimicrobial resistance

The use of any antimicrobial agent exerts a selective pressure on susceptible bacterial populations, thereby creating an environment where the development of antimicrobial resistance is selected for. It can be assumed that resistance to any antimicrobial is unavoidable. Though resistance is reported to nearly all of the

discussed replication antimicrobials, the likelihood of developing resistance varies greatly between classes. For example, mutations that confer resistance to LigA inhibitors do not affect the activity of the enzyme, while most mutations leading to aminocoumarin resistance impair gyrase functioning. Both antimicrobials target the active site of their target enzymes, but the binding sites of LigA inhibitors are located in a region that is not directly involved in substrate (NAD+) binding of LigA. Antimicrobials targeting DNA replication proteins that bind their targets at the active site, such as the aminocoumarins, have a relative low risk for resistance, since mutations that affect drug-protein binding tend to lead to a non-functional protein.

LigA inhibitors, PolC inhibitors and most aminocoumarins target one specific protein and therefore, a single mutation is frequently sufficient to cause resistance. FQs are able to target both gyrase and TopoIV. As a result, a high level of FQ resistance usually requires the presence of mutations in both gyrase and TopoIV [70], which reduces the risk of resistance development. However, variations in the potency of FQs against gyrase and TopoIV still enable the emergence of resistant bacteria [16] and for organisms that only have gyrase encoded in their genome, such as *C. difficile*, a single mutation is still sufficient to cause resistance. Indeed, a single mutation in *gyrA* (Thr82Ile) was most frequently found in FQ-resistant clinical *C. difficile* isolates [71].

Two promising strategies that might reduce development of resistance (and cross-resistance) are (i) to target multiple proteins in the bacterial cell, using so-called hybrid antimicrobials, or (ii) to use multiple binding sites in a single target. Multi-targeting reduces the chance of resistance development and results in compounds that remain active against mutants resistant to either one of the parent compounds. Though polypharmacological modelling and rational design of multi-targeting drugs is a major challenge [72,73], interesting progress has been made in recent years.

In FQ hybrid antimicrobials, a FQ moiety is covalently linked to another pharmacophore, with a distinct cellular target [74]. Various FQ hybrid classes have been developed, but of particular interest here are the 6-anilinouracil-fluoroquinolone (AU-FQ) hybrids, since both moieties target DNA replication. The AU-FQ hybrid class was created by linking various FQs to the *N*3 of the PolIIIC inhibitor HB-EMAU, with different linkers to modulate antimicrobial activity and pharmacological properties [75]. AU-FQ hybrids are highly active against Gram-positive bacteria, and some compounds had moderate activity against the Gram-negative *E. coli*. MBX-500,

one of the best AU-FQ hybrids, displayed 3-fold stronger inhibition of *B. subtilis* PolC than the AU moiety alone, while it showed 5- to 10-fold less potent inhibition of *B. subtilis* TopoIV and gyrase in comparison to the FQ component [76]. Despite these differences in target inhibition, MBX-500 had strong antimicrobial activities against *Bacillus* species (MIC of 0.156 mg/L), *S. aureus* strains (MIC ranging from 0.625 to 5 mg/L) and various other Gram-positive bacteria. As expected, MBX-500 retained high antimicrobial activity against strains that were resistant to the FQ component. Resistance to MBX-500 was only found in an AU- and FQ-resistant *S. aureus* that carried mutations in both targets, thereby providing evidence that MBX-500 truly acts via a dual-targeting mechanism. Indeed, the spontaneous mutation frequency against MBX-500 was low ($<5.6x10^{-10}$ at 4 x MIC); no resistant *S. aureus* strains could be isolated after a single passage, whereas *S. aureus* did develop resistance against the individual AU and FQ components [76]. The antimicrobial properties of MBX-500 were also investigated in *C. difficile* [77], where fluoroquinolone resistance is widespread among clinical strains [71]. The compound was active against a panel of 30 *C. difficile* isolates (MIC range 1-4 mg/L), which included several multi-resistant strains and isolates from the epidemic PCR ribotype 027. Thus, dual target antimicrobials are promising in the fight against resistant pathogens, and the low mutation frequencies (e.g. no spontaneous resistance) indicate a possible reduced risk of resistance development with comparison to single-target compounds.

Another approach to avoid cross-resistance and reduce the risk of resistance is to identify multiple (novel) binding sites or multiple modes of actions in a single established target, as exemplified by the novel topoisomerase II inhibitor kibdelomycin [78]. Biochemical studies in *S. aureus* showed that kibdelomycin is a potent ATPase inhibitor of both topoisomerase II enzymes, although it is >80-fold more active against the ATPase subunit of gyrase than TopoIV. The co-crystal structures of kibdelomycin bound to the N-terminal domains of *S. aureus* GyrB and ParE revealed a novel mode of 'dual' ATPase inhibition by blocking ATP binding and destabilizing GyrB/ParE subunit dimerization [79]. Kibdelomycin displays strong antimicrobial activities against predominantly Gram-positive bacteria, including MRSA and *C. difficile* [80,81]. The MIC_{90} for *C. difficile* was 0.5 mg/L, similar to the novel therapeutic fidaxomicin, but more potent than metronidazole and vancomycin [80]. Importantly, kibdelomycin was equally active against FQ-resistant and -susceptible strains and showed favourable pharmacokinetics in mice and promise in a hamster model of *C. difficile* infection [80]. Kibdelomycin also retained antimicrobial activity against

S. aureus strains resistant to novobiocin and coumermycin A1, two known ATP inhibitors of topoisomerase II, and showed a low frequency of resistance development ($<5.4x10^{-10}$) [78]. The MIC of kibdelomycin against a coumermycin-resistant *S. aureus* strain carrying three mutations in GyrB was modestly increased from 1 to 4 mg/L relative to the susceptible wild-type strain. This was attributed to one of the three mutations (Ile175Thr) in GyrB, which also interfered with kibdelomycin binding. Although the reduced susceptibility of the coumermycin A1 resistant strain warrants some caution, the data suggests that complete resistance caused by a single point mutation is not likely to occur due to the dual binding mode of this compound.

Challenge: specificity of novel antimicrobials

The normal gut microbiota is a diverse community of microbes that lives in a complex ecological system with its host [82]. With the increasing knowledge on interactions and dynamics of the microbiota with its host, it becomes evident that the human microbiome is of great importance for the prevention and treatment of infectious diseases. It can act as a protective barrier that prevents colonization of the gut by non-commensal pathogens and opportunistic pathogens that are already present. This so-called colonization resistance [83] is greatly influenced by the diversity of the gut microbiota. Many broad-spectrum antimicrobials used in the clinic are known to affect the composition of the gut microbiota [84,85]. Colonization of the gut by pathogenic *C. difficile* is believed to require the disruption of the gut microbiota [86,87], and patients with recurrent *C. difficile* infection (CDI) have a reduced diversity of their microbiota compared with healthy controls and patients with initial CDI [88]. There is a strong association between the use of broad-spectrum antimicrobials and increased risk of (recurrent) disease [89,90]. Patients treated with the narrow-spectrum antimicrobial fidaxomicin had lower *C. difficile* recurrence rates in comparison with broad-spectrum vancomycin [91], which was attributed to the minimal impact of fidaxomicin on the gut microbiota, especially on bacteria of the *Bacteroides* cluster, clostridial cluster XIVa and bifidobacteria [91,92]. The example of *C. difficile* underscores that it is desirable for novel antimicrobials to limit the impact on the human microbiota as a whole.

Two strategies to generate or identify compounds with a narrow spectrum are (i) to test derivatives of an existing inhibitor for increased specificity and (ii) to screen for compounds with a selective inhibition of a specific pathogen.

The first strategy has been used to identify and develop PolC inhibitors with increased species-specificity. Many PolC inhibitors have a broad Gram-positive spectrum and are therefore likely to significantly affect the composition of the microbiota. However, the isolation and purification of the *C. difficile* PolC as well as polymerases from various other organisms enabled the identification of a series of novel 7-substituted DCBG inhibitors that showed improved potency and specificity against *C. difficile* [3]. One of these compounds, 362E, showed potent antimicrobial activity against *C. difficile* strains, similar to the first-line therapeutics metronidazole and vancomycin (MIC_{90} of 4 mg/L) [3,93]. Limited *in vitro* tests indicated that 362E is inactive against certain other Gram-positive anaerobes [3], but the impact of 362E on the gut microbiota has not been reported to date. 362E protected hamsters from death due to *C. difficile* similar to vancomycin and, though recurrent CDI was observed in both treatment groups when animals were treated for 3 days, prolonged treatment with 362E resulted in a reduced recurrence rate [93]. Further experimental evidence is needed to confirm the selectivity of this compound for *C. difficile* during *in vivo* treatment.

A second approach is to screen compounds *de novo* for specific activity against certain pathogens and couple this to an assay to determine the desired mode of action. SMT19969 (ridinilazole) is a bis(4-pyridyl)bibenzimidazole that has been specifically developed as an novel front-line antimicrobial for CDI treatment (MIC 0.06 – 0.5 mg/L) [94,95] with a low chance of resistance development [96]. Although the mode of action awaits validation, compounds of the bisbenzimidazole class were shown to bind the minor groove of the DNA duplex and are thought to inhibit DNA replication through inhibition of DNA helicases and/or topoisomerases [97]. SMT19969 showed high selectivity for *C. difficile* relative to other intestinal isolates during *in vitro* susceptibility testing. In a Phase I clinical trial, treatment with SMT19969 resulted in minimal changes in the composition of the faecal microbiota of healthy human subjects, except for a significant reduction in total clostridial count [98] and this was confirmed in an *in vitro* gut model of CDI [99]. SMT19969 was superior to vancomycin and, in a subset of strains, to fidaxomicin in a hamster model of recurrent CDI [95,98]. A clinical Phase II trial is currently ongoing where the efficacy of SMT19969 treatment will be compared with vancomycin treatment in CDI patients.

In another example, as discussed above, kibdelomycin showed potent antimicrobial activities against a large panel *C. difficile* isolates [80]. In contrast to metronidazole, kibdelomycin is mostly inactive against Gram-negative anaerobes, including

the *Bacteroides fragilis* group, similar to the narrow-spectrum drug fidaxomicin [80]. The narrow spectrum of fidaxomicin is believed to contribute to its favourable characteristics in treating recurrent CDI [91], but the effect of kibdelomycin on recurrent disease has not been investigated yet. MBX-500 also had little activity against most Gram-negative anaerobes and was selectively active against *C. difficile* strains among Gram-positive anaerobes [77]. MBX-500 was shown to be efficacious in a gnotobiotic piglet model of acute CDI, with 100 percent survival rate and only mild CDI symptoms [100]. Though the efficacy of MBX-500 was comparable to vancomycin in a hamster model of CDI, it was superior in a murine model of recurrent CDI and associated with improved weight gain in both animal models [77]. The weight gain of infected animals treated with MBX-500 and the results in the murine model suggests that this compound might have a low impact on the gut microbiota.

Narrow-spectrum antimicrobials are useful to prevent or treat opportunistic infections such as CDI that are associated with dysbiosis of the microbiome. Careful assessment of the effects on microbiota should be considered for all antimicrobials developed in the future.

Challenge: identifying novel targets in the bacterial replisome

Despite the success of the FQs, the number of compounds targeting DNA replication in clinical use or development is relatively low compared to inhibitors of other cellular processes such as protein synthesis. Why is this group of antimicrobials underdeveloped, considering that there are many more proteins in the bacterial replisome?

A major issue is the poor characterization of the DNA replication machinery of important pathogens. For instance, the DNA replication proteins in *C. difficile* are merely predicted through homology with proteins in other organisms (**Table 1**) but functional evidence is limited to *C. difficile* PolC so far [3]. It can be expected that the extensive biochemical and structural characterization of replisome proteins of drug-resistant pathogens might lead to the identification of unique features that can be used to design antimicrobials that specifically target this specific species.

Another possible explanation for the underdevelopment of antimicrobials targeting DNA replication is that screens for replication inhibitors have been notori-

ously difficult. Most inhibitors are evaluated through *in vitro* biochemical assays that measure inhibition of either enzymatic activity or DNA replication and these are not always suitable for high-throughput screening [27]. Moreover, compounds identified this way may display potent *in vitro* activity, while they have no activity against bacterial cells *in vivo* due to undesirable pharmacokinetic properties. Besides some of the examples mentioned (aminocoumarins, ANCs and PolIII inhibitors), this is illustrated by two studies on inhibitors of the bacterial primosome [101,102]. Inhibitor peptides that target the DnaB-DnaG helicase-primase complex in *Bacillus stearothermophilus* were able to inhibit purified proteins *in vitro* [101], but were inactive *in vivo*. A highly active triaminotriazine inhibitor of the *Pseudomonas aeruginosa* replicative helicase DnaB did not show antimicrobial activity towards wild-type *P. aeruginosa* or *E. coli*, showed poor activity against *S. aureus* and showed problematic toxicity towards mammalian cells [102]. To circumvent the limitations of biochemical assays, screening could be performed using a whole-cell assay. By limiting the levels of a certain replisome protein or affecting its function in cells [103,104], screens can be enriched for compounds that target DNA replication. A successful application of a whole-cell screen is the identification of vibrepin as an inhibitor of DNA replication in *Vibrio* [104]. *Vibrio* species, including the causative agent of cholera (*Vibrio cholerae*) have a bipartite genome, where replication of the second chromosome is dependent on the replication initiator RctB. In the screen, the growth of an *E. coli* strain harbouring an RctB-dependent plasmid with an antimicrobial resistance determinant was evaluated in the presence and absence of the antimicrobial and inhibition of growth was indicative of an RctB inhibitor. Vibrepin was found to interfere with chromosome II origin opening by RctB *in vitro* and inhibited various *Vibrio* species (MIC 0.4-2.0 mg/L), but also inhibited various other species via an RctB-independent mechanism [104]. This example also illustrates how differences between replication machineries can be exploited to work towards species-specific antimicrobial compounds.

Outlook

Despite the challenges that are discussed with respect to antimicrobials targeting DNA replication, we feel it is important to focus research on the identification of novel therapeutic targets in the bacterial replication machinery, whether they are

individual proteins or protein-protein interactions [27], for several reasons. First, DNA replication is essential for cell viability and inhibiting any essential protein would be detrimental for cell survival. Second, most components of the bacterial replisome are substantially different from their eukaryote counterparts and can therefore be exploited to develop compounds with minimal cytotoxicity. Third, antimicrobials directed at novel targets are likely to avoid existing resistance mechanisms. And finally, some components are well conserved between most bacteria, such as the DNA initiator protein DnaA, while others are more restricted to certain species. Therefore, the bacterial replisome can be used to develop both broad- and narrow-spectrum antimicrobials. Recent insights on the structure of several DNA replication proteins and their inhibitors and the development of new assays that enable high-throughput screening for inhibitors of DNA replication are expected to influence the rate of success of this class of antimicrobials considerably [2].

Acknowledgements

The authors thank Jeroen Corver and Adam P. Roberts for helpful comments on the manuscript.

Funding

Work in the group of WKS is supported by a Gisela Their Fellowship of the Leiden University Medical Center and VIDI Fellowship 864.13.003 from the Netherlands Organisation for Scientific Research.

Transparency declarations

EVE and WKS have performed research for Cubist. EJK has performed research for Cubist, Novartis and Qiagen, and has participated in advisory forums of Astellas, Optimer, Actelion, Pfizer, Sanofi Pasteur and Seres Therapeutics. The companies, and the funding agencies above, had no role writing this manuscript. BW: none to declare.

References

1 Walsh, C. T. & Wencewicz, T. A. Prospects for new antibiotics: a molecule-centered perspective. *The Journal of antibiotics* **67**, 7–22, doi:10.1038/ja.2013.49 (2014).

2 Sanyal, G. & Doig, P. Bacterial DNA replication enzymes as targets for antibacterial drug discovery. *Expert opinion on drug discovery* **7**, 327–339, doi:10.1517/17460441.2012.660478 (2012).

3 Torti, A. *et al.* Clostridium difficile DNA polymerase IIIC: basis for activity of antibacterial compounds. *Current enzyme inhibition* **7**, 147–153 (2011).

4 van Eijk, E. *et al.* Primase is required for helicase activity and helicase alters the specificity of primase in the enteropathogen Clostridium difficile. *Open biology* **6**, doi:10.1098/rsob.160272 (2016).

5 Sissi, C. & Palumbo, M. In front of and behind the replication fork: bacterial type IIA topoisomerases. *Cellular and molecular life sciences : CMLS* **67**, 2001–2024, doi:10.1007/s00018-010-0299-5 (2010).

6 Mizuuchi, K., Mizuuchi, M., O'Dea, M. H. & Gellert, M. Cloning and simplified purification of Escherichia coli DNA gyrase A and B proteins. *The Journal of biological chemistry* **259**, 9199–9201 (1984).

7 Kato, J. *et al.* New topoisomerase essential for chromosome segregation in E. coli. *Cell* **63**, 393–404 (1990).

8 Deibler, R. W., Rahmati, S. & Zechiedrich, E. L. Topoisomerase IV, alone, unknots DNA in E. coli. *Genes & development* **15**, 748–761, doi:10.1101/gad.872301 (2001).

9 Gellert, M., Mizuuchi, K., O'Dea, M. H. & Nash, H. A. DNA gyrase: an enzyme that introduces superhelical turns into DNA. *Proceedings of the National Academy of Sciences of the United States of America* **73**, 3872–3876 (1976).

10 Zechiedrich, E. L., Khodursky, A. B. & Cozzarelli, N. R. Topoisomerase IV, not gyrase, decatenates products of site-specific recombination in Escherichia coli. *Genes & development* **11**, 2580–2592 (1997).

11 Manjunatha, U. H. *et al.* Functional characterisation of mycobacterial DNA gyrase: an efficient decatenase. *Nucleic acids research* **30**, 2144–2153 (2002).

12 Collin, F., Karkare, S. & Maxwell, A. Exploiting bacterial DNA gyrase as a drug target: current state and perspectives. *Applied microbiology and biotechnology* **92**, 479–497, doi:10.1007/s00253-011-3557-z (2011).

13 Sugino, A., Higgins, N. P., Brown, P. O., Peebles, C. L. & Cozzarelli, N. R. Energy coupling in DNA gyrase and the mechanism of action of novobiocin. *Proceedings of the National Academy of Sciences of the United States of America* **75**, 4838–4842 (1978).

14 Lewis, R. J. *et al.* The nature of inhibition of DNA gyrase by the coumarins and the cyclothialidines revealed by X-ray crystallography. *The EMBO journal* **15**, 1412–1420 (1996).

15 Ehmann, D. E. & Lahiri, S. D. Novel compounds targeting bacterial DNA topoisomerase/DNA gyrase. *Current opinion in pharmacology* **18**, 76–83, doi:10.1016/j.coph.2014.09.007 (2014).

16 Drlica, K. *et al.* Quinolones: action and resistance updated. *Current topics in medicinal chemistry* **9**, 981–998 (2009).

17 Wohlkonig, A. *et al.* Structural basis of quinolone inhibition of type IIA topoisomerases and target-mediated resistance. *Nature structural & molecular biology* **17**, 1152–1153, doi:10.1038/nsmb.1892 (2010).

18 Pan, X. S. & Fisher, L. M. Targeting of DNA gyrase in Streptococcus pneumoniae by sparfloxacin: selective targeting of gyrase or topoisomerase IV by quinolones. *Antimicrobial agents and chemotherapy* **41**, 471–474 (1997).

19 Fournier, B., Zhao, X., Lu, T., Drlica, K. & Hooper, D. C. Selective targeting of topoisomerase IV and DNA gyrase in Staphylococcus aureus: different patterns of quinolone-induced inhibition of DNA synthesis. *Antimicrobial agents and chemotherapy* **44**, 2160-2165 (2000).

20 Takei, M., Fukuda, H., Kishii, R. & Hosaka, M. Target preference of 15 quinolones against Staphylococcus aureus, based on antibacterial activities and target inhibition. *Antimicrobial agents and chemotherapy* **45**, 3544-3547, doi:10.1128/aac.45.12.3544-3547.2001 (2001).

21 Redgrave, L. S., Sutton, S. B., Webber, M. A. & Piddock, L. J. Fluoroquinolone resistance: mechanisms, impact on bacteria, and role in evolutionary success. *Trends in microbiology* **22**, 438-445, doi:10.1016/j.tim.2014.04.007 (2014).

22 Zumla, A. I. *et al.* New antituberculosis drugs, regimens, and adjunct therapies: needs, advances, and future prospects. *The Lancet. Infectious diseases* **14**, 327-340, doi:10.1016/s1473-3099(13)70328-1 (2014).

23 McDonald, L. C. *et al.* An epidemic, toxin gene-variant strain of Clostridium difficile. *The New England journal of medicine* **353**, 2433-2441, doi:10.1056/NEJMoa051590 (2005).

24 He, M. *et al.* Emergence and global spread of epidemic healthcare-associated Clostridium difficile. *Nature genetics* **45**, 109-113, doi:10.1038/ng.2478 (2013).

25 Aldred, K. J., Kerns, R. J. & Osheroff, N. Mechanism of quinolone action and resistance. *Biochemistry* **53**, 1565-1574, doi:10.1021/bi5000564 (2014).

26 Jacoby, G. A. Mechanisms of resistance to quinolones. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **41 Suppl 2**, S120-126, doi:10.1086/428052 (2005).

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

28 Wiles, J. A., Bradbury, B. J. & Pucci, M. J. New quinolone antibiotics: a survey of the literature from 2005 to 2010. *Expert opinion on therapeutic patents* **20**, 1295-1319, doi:10.1517/13543776.2010.505922 (2010).

29 Bax, B. D. *et al.* Type IIA topoisomerase inhibition by a new class of antibacterial agents. *Nature* **466**, 935-940, doi:10.1038/nature09197 (2010).

30 Gomez, L. *et al.* Novel pyrazole derivatives as potent inhibitors of type II topoisomerases. Part 1: synthesis and preliminary SAR analysis. *Bioorganic & medicinal chemistry letters* **17**, 2723-2727, doi:10.1016/j.bmcl.2007.03.003 (2007).

31 Black, M. T. *et al.* Mechanism of action of the antibiotic NXL101, a novel nonfluoroquinolone inhibitor of bacterial type II topoisomerases. *Antimicrobial agents and chemotherapy* **52**, 3339-3349, doi:10.1128/aac.00496-08 (2008).

32 Dougherty, T. J. *et al.* NBTI 5463 is a novel bacterial type II topoisomerase inhibitor with activity against gram-negative bacteria and in vivo efficacy. *Antimicrobial agents and chemotherapy* **58**, 2657-2664, doi:10.1128/aac.02778-13 (2014).

33 Foss, M. H. *et al.* N-Benzyl-3-sulfonamidopyrrolidines are a New Class of Bacterial DNA Gyrase Inhibitors. *ACS medicinal chemistry letters* **2**, 289-292, doi:10.1021/ml1002822 (2011).

34 Lepak, A. J. *et al.* In Vivo Pharmacodynamic Target Investigation of Two Bacterial Topoisomerase Inhibitors, ACT-387042 and ACT-292706, in the Neutropenic Murine Thigh Model against Streptococcus pneumoniae and Staphylococcus aureus. *Antimicrobial agents and chemotherapy* **60**, 3626-3632, doi:10.1128/aac.00363-16 (2016).

35 Jahic, H. *et al.* The kinetic mechanism of S. pneumoniae DNA ligase and inhibition by adenosine-based antibacterial compounds. *Biochemical pharmacology* **84**, 654-660, doi:10.1016/j.bcp.2012.06.017 (2012).

36 Shuman, S. DNA ligases: progress and prospects. *The Journal of biological chemistry* **284**, 17365-17369, doi:10.1074/jbc.R900017200 (2009).

37 Korycka-Machala, M. *et al.* Evaluation of NAD(+)-dependent DNA ligase of mycobacteria as a potential target for antibiotics. *Antimicrobial agents and chemotherapy* **51**, 2888-2897, doi:10.1128/aac.00254-07 (2007).

38 Swift, R. V. & Amaro, R. E. Discovery and design of DNA and RNA ligase inhibitors in infectious microorganisms. *Expert opinion on drug discovery* **4**, 1281-1294, doi:10.1517/17460440903373617 (2009).

39 Surivet, J. P. *et al.* Structure-guided design, synthesis and biological evaluation of novel DNA ligase inhibitors with in vitro and in vivo anti-staphylococcal activity. *Bioorganic & medicinal chemistry letters* **22**, 6705-6711, doi:10.1016/j.bmcl.2012.08.094 (2012).

40 Mills, S. D. *et al.* Novel bacterial NAD+-dependent DNA ligase inhibitors with broad-spectrum activity and antibacterial efficacy in vivo. *Antimicrobial agents and chemotherapy* **55**, 1088-1096, doi:10.1128/aac.01181-10 (2011).

41 Stokes, S. S. *et al.* Discovery of bacterial NAD+-dependent DNA ligase inhibitors: optimization of antibacterial activity. *Bioorganic & medicinal chemistry letters* **21**, 4556-4560, doi:10.1016/j.bmcl.2011.05.128 (2011).

42 Stokes, S. S. *et al.* Discovery of bacterial NAD(+)-dependent DNA ligase inhibitors: improvements in clearance of adenosine series. *Bioorganic & medicinal chemistry letters* **22**, 85-89, doi:10.1016/j.bmcl.2011.11.071 (2012).

43 Mondal, S. I. *et al.* Identification of potential drug targets by subtractive genome analysis of Escherichia coli O157:H7: an in silico approach. *Advances and applications in bioinformatics and chemistry : AABC* **8**, 49-63, doi:10.2147/aabc.s88522 (2015).

44 Jadaun, A., Sudhakar, D. R., Subbarao, N. & Dixit, A. In silico screening for novel inhibitors of DNA polymerase III alpha subunit of Mycobacterium tuberculosis (MtbDnaE2, H37Rv). *PloS one* **10**, e0119760, doi:10.1371/journal.pone.0119760 (2015).

45 Timinskas, K., Balvociute, M., Timinskas, A. & Venclovas, C. Comprehensive analysis of DNA polymerase III alpha subunits and their homologs in bacterial genomes. *Nucleic acids research* **42**, 1393-1413, doi:10.1093/nar/gkt900 (2014).

46 Brown, N. C. 6-(p-hydroxyphenylazo)-uracil: a selective inhibitor of host DNA replication in phage-infected Bacillus subtilis. *Proceedings of the National Academy of Sciences of the United States of America* **67**, 1454-1461 (1970).

47 Barnes, M. H., Butler, M. M., Wright, G. E. & Brown, N. C. Antimicrobials targeted to the replication-specific DNA polymerases of gram-positive bacteria: target potential of dnaE. *Infectious disorders drug targets* **12**, 327-331 (2012).

48 Tarantino, P. M., Jr., Zhi, C., Wright, G. E. & Brown, N. C. Inhibitors of DNA polymerase III as novel antimicrobial agents against gram-positive eubacteria. *Antimicrobial agents and chemotherapy* **43**, 1982-1987 (1999).

49 Wright, G. E. *et al.* Active site directed inhibitors of replication-specific bacterial DNA polymerases. *Bioorganic & medicinal chemistry letters* **15**, 729-732, doi:10.1016/j.bmcl.2004.11.016 (2005).

50 Daly, J. S. *et al.* In vitro antimicrobial activities of novel anilinouracils which selectively inhibit DNA polymerase III of gram-positive bacteria. *Antimicrobial agents and chemotherapy* **44**, 2217–2221 (2000).

51 Zhi, C. *et al.* Synthesis of substituted 6-anilinouracils and their inhibition of DNA polymerase IIIC and Gram-positive bacterial growth. *Journal of medicinal chemistry* **46**, 2731–2739, doi:10.1021/jm020591z (2003).

52 Zhi, C. *et al.* Synthesis and antibacterial activity of 3-substituted-6-(3-ethyl-4-methylanilino)uracils. *Journal of medicinal chemistry* **48**, 7063–7074, doi:10.1021/jm050517r (2005).

53 Kuhl, A. *et al.* Biological characterization of novel inhibitors of the gram-positive DNA polymerase IIIC enzyme. *Antimicrobial agents and chemotherapy* **49**, 987–995, doi:10.1128/aac.49.3.987-995.2005 (2005).

54 Svenstrup, N., Ehlert, K., Ladel, C., Kuhl, A. & Habich, D. New DNA polymerase IIIC inhibitors: 3-subtituted anilinouracils with potent antibacterial activity in vitro and in vivo. *ChemMedChem* **3**, 1604–1615, doi:10.1002/cmdc.200800117 (2008).

55 Butler, M. M. *et al.* Low frequencies of resistance among Staphylococcus and Enterococcus species to the bactericidal DNA polymerase inhibitor N(3)-hydroxybutyl 6-(3'-ethyl-4'-methylanilino) uracil. *Antimicrobial agents and chemotherapy* **46**, 3770–3775 (2002).

56 Barnes, M. H., Hammond, R. A., Kennedy, C. C., Mack, S. L. & Brown, N. C. Localization of the exonuclease and polymerase domains of Bacillus subtilis DNA polymerase III. *Gene* **111**, 43–49 (1992).

57 Butler, M. M., Dudycz, L. W., Khan, N. N., Wright, G. E. & Brown, N. C. Development of novel inhibitor probes of DNA polymerase III based on dGTP analogs of the HPUra type: base, nucleoside and nucleotide derivatives of N2-(3,4-dichlorobenzyl)guanine. *Nucleic acids research* **18**, 7381–7387 (1990).

58 Xu, W. C. *et al.* 7-Alkyl-N(2)-substituted-3-deazaguanines. Synthesis, DNA polymerase III inhibition and antibacterial activity. *Bioorganic & medicinal chemistry letters* **21**, 4197–4202, doi:10.1016/j.bmcl.2011.05.093 (2011).

59 Rose, Y. *et al.* Novel non-nucleobase inhibitors of Staphylococcus aureus DNA polymerase IIIC. *Bioorganic & medicinal chemistry letters* **16**, 891–896, doi:10.1016/j.bmcl.2005.11.009 (2006).

60 Guiles, J. *et al.* Quinazolin-2-ylamino-quinazolin-4-ols as novel non-nucleoside inhibitors of bacterial DNA polymerase III. *Bioorganic & medicinal chemistry letters* **19**, 800–802, doi:10.1016/j.bmcl.2008.12.038 (2009).

61 Kelch, B. A. Review: The lord of the rings: Structure and mechanism of the sliding clamp loader. *Biopolymers* **105**, 532–546, doi:10.1002/bip.22827 (2016).

62 Hedglin, M., Kumar, R. & Benkovic, S. J. Replication clamps and clamp loaders. *Cold Spring Harbor perspectives in biology* **5**, a010165, doi:10.1101/cshperspect.a010165 (2013).

63 Kling, A. *et al.* Antibiotics. Targeting DnaN for tuberculosis therapy using novel griselimycins. *Science (New York, N.Y.)* **348**, 1106–1112, doi:10.1126/science.aaa4690 (2015).

64 Terlain, B. & Thomas, J. P. [Structure of griselimycin, polypeptide antibiotic extracted Streptomyces cultures. I. Identification of the products liberated by hydrolysis]. *Bulletin de la Societe chimique de France* **6**, 2349–2356 (1971).

65 Toyohara, M. Aspects of the antituberculous activity of 27753-RP, a new semisynthetic derivative of griselimycine. *Annales de l'Institut Pasteur. Microbiology* **138**, 737–744 (1987).

66 Shereda, R. D., Kozlov, A. G., Lohman, T. M., Cox, M. M. & Keck, J. L. SSB as an organizer/mobilizer of genome maintenance complexes. *Critical reviews in biochemistry and molecular biology* **43**, 289–318, doi:10.1080/10409230802341296 (2008).

67 Marceau, A. H. Functions of single-strand DNA-binding proteins in DNA replication, recombination, and repair. *Methods in molecular biology (Clifton, N.J.)* **922**, 1–21, doi:10.1007/978-1-62703-032-8_1 (2012).

68 Marceau, A. H. *et al.* Protein interactions in genome maintenance as novel antibacterial targets. *PloS one* **8**, e58765, doi:10.1371/journal.pone.0058765 (2013).

69 Lu, D., Bernstein, D. A., Satyshur, K. A. & Keck, J. L. Small-molecule tools for dissecting the roles of SSB/protein interactions in genome maintenance. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 633–638, doi:10.1073/pnas.0909191107 (2010).

70 Ruiz, J. Mechanisms of resistance to quinolones: target alterations, decreased accumulation and DNA gyrase protection. *The Journal of antimicrobial chemotherapy* **51**, 1109–1117, doi:10.1093/jac/dkg222 (2003).

71 Spigaglia, P. *et al.* Fluoroquinolone resistance in Clostridium difficile isolates from a prospective study of C. difficile infections in Europe. *Journal of medical microbiology* **57**, 784–789, doi:10.1099/jmm.0.47738-0 (2008).

72 Hopkins, A. L. Network pharmacology: the next paradigm in drug discovery. *Nature chemical biology* **4**, 682–690, doi:10.1038/nchembio.118 (2008).

73 Reddy, M. R. Advances in rational drug design. *Current pharmaceutical design* **19**, 4673 (2013).

74 Pokrovskaya, V. & Baasov, T. Dual-acting hybrid antibiotics: a promising strategy to combat bacterial resistance. *Expert opinion on drug discovery* **5**, 883–902, doi:10.1517/17460441.2010.508069 (2010).

75 Zhi, C. *et al.* Hybrid antibacterials. DNA polymerase-topoisomerase inhibitors. *Journal of medicinal chemistry* **49**, 1455–1465, doi:10.1021/jm0510023 (2006).

76 Butler, M. M. *et al.* Antibacterial activity and mechanism of action of a novel anilinouracil-fluoroquinolone hybrid compound. *Antimicrobial agents and chemotherapy* **51**, 119–127, doi:10.1128/aac.01311-05 (2007).

77 Butler, M. M. *et al.* MBX-500, a hybrid antibiotic with in vitro and in vivo efficacy against toxigenic Clostridium difficile. *Antimicrobial agents and chemotherapy* **56**, 4786–4792, doi:10.1128/aac.00508-12 (2012).

78 Phillips, J. W. *et al.* Discovery of kibdelomycin, a potent new class of bacterial type II topoisomerase inhibitor by chemical-genetic profiling in Staphylococcus aureus. *Chemistry & biology* **18**, 955–965, doi:10.1016/j.chembiol.2011.06.011 (2011).

79 Lu, J. *et al.* Structures of kibdelomycin bound to Staphylococcus aureus GyrB and ParE showed a novel U-shaped binding mode. *ACS chemical biology* **9**, 2023–2031, doi:10.1021/cb5001197 (2014).

80 Miesel, L. *et al.* Kibdelomycin is a potent and selective agent against toxigenic Clostridium difficile. *Antimicrobial agents and chemotherapy* **58**, 2387–2392, doi:10.1128/aac.00021-14 (2014).

81 Singh, S. B. *et al.* Kibdelomycin is a bactericidal broad-spectrum aerobic antibacterial agent. *Antimicrobial agents and chemotherapy* **59**, 3474–3481, doi:10.1128/aac.00382-15 (2015).

82 Ley, R. E., Peterson, D. A. & Gordon, J. I. Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* **124**, 837–848, doi:10.1016/j.cell.2006.02.017 (2006).

83 Vollaard, E. J. & Clasener, H. A. Colonization resistance. *Antimicrobial agents and chemotherapy* **38**, 409–414 (1994).

84 Rea, M. C. *et al.* Effect of broad- and narrow-spectrum antimicrobials on Clostridium difficile and microbial diversity in a model of the distal colon. *Proceedings of the National Academy of Sciences of the United States of America* **108 Suppl 1**, 4639–4644, doi:10.1073/pnas.1001224107 (2011).

85 Sullivan, A., Edlund, C. & Nord, C. E. Effect of antimicrobial agents on the ecological balance of human microflora. *The Lancet. Infectious diseases* **1**, 101–114, doi:10.1016/s1473-3099(01)00066-4 (2001).

86 Pham, T. A. & Lawley, T. D. Emerging insights on intestinal dysbiosis during bacterial infections. *Current opinion in microbiology* **17**, 67–74, doi:10.1016/j.mib.2013.12.002 (2014).

87 Theriot, C. M. *et al.* Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to Clostridium difficile infection. *Nature communications* **5**, 3114, doi:10.1038/ncomms4114 (2014).

88 Chang, J. Y. *et al.* Decreased diversity of the fecal Microbiome in recurrent Clostridium difficile-associated diarrhea. *The Journal of infectious diseases* **197**, 435–438, doi:10.1086/525047 (2008).

89 Drekonja, D. M. *et al.* Antimicrobial use and risk for recurrent Clostridium difficile infection. *The American journal of medicine* **124**, 1081.e1081–1087, doi:10.1016/j.amjmed.2011.05.032 (2011).

90 Smits, W. K., Lyras, D., Lacy, D. B., Wilcox, M. H. & Kuijper, E. J. Clostridium difficile infection. *Nature reviews. Disease primers* **2**, 16020, doi:10.1038/nrdp.2016.20 (2016).

91 Louie, T. J. *et al.* Fidaxomicin versus vancomycin for Clostridium difficile infection. *The New England journal of medicine* **364**, 422–431, doi:10.1056/NEJMoa0910812 (2011).

92 Tannock, G. W. *et al.* A new macrocyclic antibiotic, fidaxomicin (OPT-80), causes less alteration to the bowel microbiota of Clostridium difficile-infected patients than does vancomycin. *Microbiology (Reading, England)* **156**, 3354–3359, doi:10.1099/mic.0.042010-0 (2010).

93 Dvoskin, S. *et al.* A novel agent effective against Clostridium difficile infection. *Antimicrobial agents and chemotherapy* **56**, 1624–1626, doi:10.1128/aac.06097-11 (2012).

94 Goldstein, E. J., Citron, D. M. & Tyrrell, K. L. Comparative in vitro activities of SMT19969, a new antimicrobial agent, against 162 strains from 35 less frequently recovered intestinal Clostridium species: implications for Clostridium difficile recurrence. *Antimicrobial agents and chemotherapy* **58**, 1187–1191, doi:10.1128/aac.02184-13 (2014).

95 Weiss, W., Pulse, M. & Vickers, R. In vivo assessment of SMT19969 in a hamster model of clostridium difficile infection. *Antimicrobial agents and chemotherapy* **58**, 5714–5718, doi:10.1128/aac.02903-14 (2014).

96 Vickers , R. *et al.* in *51st Interscience Conf on Antimicrobial Agents and Chemotherapy (ICAAC)* (Chicago , United States, 2011).

97 Baraldi, P. G. *et al.* DNA minor groove binders as potential antitumor and antimicrobial agents. *Medicinal research reviews* **24**, 475–528, doi:10.1002/med.20000 (2004).

98 Best, E., Vickers , R. & Wilcox , M. in *53rd Intersci. Conf. Antimicrob. Agents Chemother.* (2013).

99 Baines, S. D. *et al.* SMT19969 as a treatment for Clostridium difficile infection: an assessment of antimicrobial activity using conventional susceptibility testing and an in vitro gut model. *The Journal of antimicrobial chemotherapy* **70**, 182–189, doi:10.1093/jac/dku324 (2015).

100 Steele, J. *et al.* MBX–500 is effective for treatment of Clostridium difficile infection in gnotobiotic piglets. *Antimicrobial agents and chemotherapy* **57**, 4039–4041, doi:10.1128/aac.00304–13 (2013).

101 Gardiner, L., Coyle, B. J., Chan, W. C. & Soultanas, P. Discovery of antagonist peptides against bacterial helicase–primase interaction in B. stearothermophilus by reverse yeast three–hybrid. *Chemistry & biology* **12**, 595–604, doi:10.1016/j.chembiol.2005.04.007 (2005).

102 McKay, G. A. *et al.* Triaminotriazine DNA helicase inhibitors with antibacterial activity. *Bioorganic & medicinal chemistry letters* **16**, 1286–1290, doi:10.1016/j.bmcl.2005.11.076 (2006).

103 Fossum, S. *et al.* A robust screen for novel antibiotics: specific knockout of the initiator of bacterial DNA replication. *FEMS microbiology letters* **281**, 210–214, doi:10.1111/j.1574–6968.2008.01103.x (2008).

104 Yamaichi, Y., Duigou, S., Shakhnovich, E. A. & Waldor, M. K. Targeting the replication initiator of the second Vibrio chromosome: towards generation of vibrionaceae–specific antimicrobial agents. *PLoS pathogens* **5**, e1000663, doi:10.1371/journal.ppat.1000663 (2009).

Chapter 3

Complete genome sequence of the *Clostridium difficile* laboratory strain 630Δ*erm* reveals differences from strain 630, including translocation of the mobile element CTn*5*

Erika van Eijk [1,†]
Seyed Yahya Anvar [2,3,†]
Hilary P Browne [4]
Wai Yi Leung [5]
Jeroen Frank [3]
Arnoud M. Schmitz [3]
Adam P. Roberts [6]
Wiep Klaas Smits [1]

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

2 Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands;

3 Leiden Genome Technology Center, Human and Clinical Genetics, Leiden University Medical Center, Leiden, the Netherlands;

4 Wellcome Trust Sanger Institute, Hinxton, UK;

5 Sequence Analysis Support Core, Leiden University Medical Center, Leiden, the Netherlands;

6 Department of Microbial Diseases, UCL Eastman Dental Institute, London, United Kingdom.

† Authors contributed equally

Published in BMC Genomics, 2015

Abstract

Background: ***Clostridium difficile*** **strain 630Δ*erm* is a spontaneous erythromycin sensitive derivative of the reference strain 630 obtained by serial passaging in antibiotic-free media. It is widely used as a defined and tractable *C. difficile* strain. Though largely similar to the ancestral strain, it demonstrates phenotypic differences that might be the result of underlying genetic changes. Here, we performed a *de novo* assembly based on single-molecule real-time sequencing and an analysis of major methylation patterns.**

Results: **In addition to single nucleotide polymorphisms and various indels, we found that the mobile element CTn*5* is present in the gene encoding the methyltransferase *rumA* rather than adhesin CD1844 where it is located in the reference strain.**

Conclusions: **Together, the genetic features identified in this study may help to explain at least part of the phenotypic differences. The annotated genome sequence of this lab strain, including the first analysis of major methylation patterns, will be a valuable resource for genetic research on *C. difficile*.**

Keywords: **Genome sequence, Conjugative transposon, Integrative and conjugative element, Single-molecule real-time sequencing.**

Background

Clostridium difficile is a Gram-positive, anaerobic bacterium that can asymptomatically colonize the intestine of humans and other mammals. It was originally identified as part of the intestinal microbiota of healthy infants [1]. However, when the normal flora is disturbed – for instance as a result of antibiotic treatment – *C. difficile* can overgrow and cause potentially fatal disease [2,3]. The main virulence factors are toxins A and B, that are encoded on a chromosomal region called the pathogenicity locus (PaLoc) [4], but other factors are also likely to play a role [5]. Recent years have seen an increase in the incidence and severity of *C. difficile* infections, for reasons that are only partially understood [6,7].

In 2006, the first genome sequence of a *C. difficile* strain was published [8]. This multi-resistant strain, designated 630, was isolated from a patient with severe pseudomembranous colitis and caused an outbreak of diarrhoeal disease in a Swiss hospital [9]. Analysis of the 630 genome sequence revealed that approximately 11 percent consists of mobile genetic elements [8]. The majority of these elements are conjugative transposons of the Tn*916* and Tn*1549* families called CTns, which have the ability to excise from their genomic target sites and transpose intra- or intercellularly [8,10]. Exchange of mobile elements occurs frequently and contributes to the plasticity of the genome of *C. difficile* [8,11,12]. Functions encoded on conjugative transposons can contribute to environmental adaptation and antimicrobial resistance [10,13]. In *C. difficile*, transfer of the conjugative elements CTn*1*, CTn*2*, CTn*4*, CTn*5* and CTn*7* from strain 630 into a non-toxigenic strain has been shown [10]. Transfer of CTn*3* (Tn*5397*), harbouring a tetracycline resistance gene, has been demonstrated between species [14,15]. CTn*1*, CTn*3*, CTn*6* and CTn*7* are related to Tn*916*, based on their conjugation module [8,13]. CTn*2*, CTn*4* and CTn*5* are all part of the Tn*1549* family, based on DNA sequence homology, and their accessory modules code for uncharacterized ABC-transporters [8,10]. Recently it has been shown that these CTns may also be responsible for transfer of the PaLoc on large chromosomal fragments [16].

After the demonstration of conjugative transfer from DNA from *Escherichia coli* to *C. difficile* [17], genetic tools were developed for *C. difficile*. To facilitate the genetic manipulation, an erythromycin sensitive variant was derived from strain 630 by serial passaging [18]. This strain is particularly useful for generation of insertional mutants

using ClosTron that employs a retrotransposition activated erythromycin resistance marker (*erm*RAM [19]). Recently, allelic exchange methods have been developed for *C. difficile* [20,21]. The efficiency of both methods depends on the accuracy of the genome sequence for selection of target sites and recombination events. However, no comprehensive mapping of differences between the lab- and reference strains has been published to date.

The most notable phenotypic difference between 630 and 630Δ*erm*, erythromycin resistance, was found to be the result of a 2.4 kb deletion in the mobile genetic element Tn*5398* that eliminates an *ermB* gene [18]. This explains at least in part the different behaviour of the two strains in a Golden Syrian hamster model of acute disease [22], as animals are generally sensitized to *C. difficile* with a clindamycin treatment (*ermB* is an rRNA adenine N-6-methyltransferase that also confers resistance to clindamycin). At a genetic level, another difference between the two strains reported to date is a duplication in the master regulator of sporulation, *spo0A*, that is apparently without phenotypic consequences [23].

In another Gram-positive bacterium, *Bacillus subtilis*, phenotypic differences between the ancestral strain NCIB3610 and widely used laboratory strains have been linked to specific genetic differences [24-26]. A detailed map of the genetic differences between the *C. difficile* strains 630 and 630Δ*erm* could therefore not only facilitate genetic manipulation, but also form the basis for the investigation of phenotypic differences between these strains.

Results and discussion

Reference assembly of the 630Δ*erm* genome reveals four breakpoints

We set out to investigate differences between the laboratory strain 630Δ*erm* and reference strain 630 by performing short-read next generation sequencing on the Illumina HiSeq platform. Based on the report that the erythromycin sensitivity of strain 630Δ*erm* is due to a 2.4 kb deletion in Tn*5398*, we examined this region of the reference alignment. The analysis revealed the absence of reads mapping to the CD2007A and CD2008 genes which are located in the expected deletion [18]. Reads

that mapped to CD2007 (*erm2(B)/ermB1*), the main erythromycin resistance determinant in strain 630 [18] are likely due to the fact that this gene shares 100 percent nucleotide identity with CD2010 (*erm1(B)/ermB*), which is still present. This is supported by the observation that the coverage of both these genes is approximately 2-fold lower than the immediate surrounding regions (**Figure 1A**). Notably, the reference assembly failed to identify the previously identified duplication in *spo0A* [23] (*data not shown*).

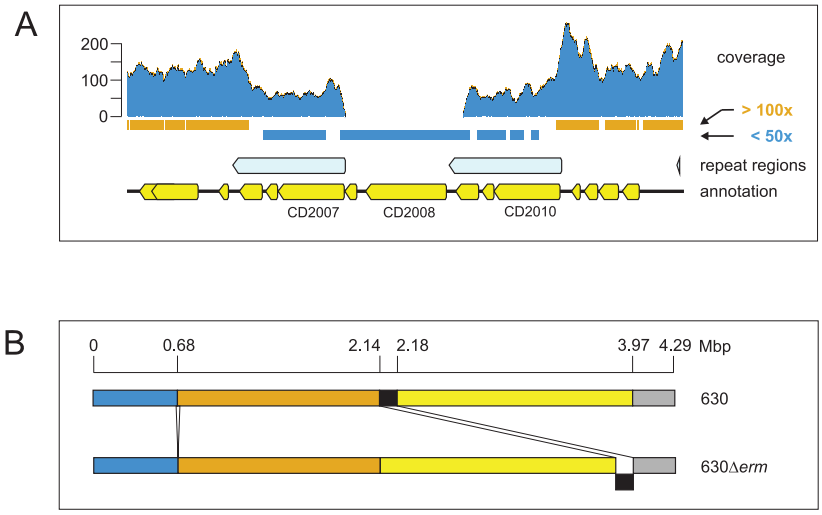


Figure 1. Results of short read next generation sequencing of *C. difficile* 630Δ*erm*.

A. Coverage of the region of Tn*5398* harbouring the two erythromycin resistance genes (CD2007 and CD2010). Bars underneath the graph indicate a greater than 100-fold (orange) and lesser than 50-fold (blue) coverage, respectively. Reference assembly was performed using Geneious 7.1 software (Biomatters, http://www.geneious.com).

B. Schematic representation of the breakpoint analysis (for details see Methods). Segments between breakpoints are indicated with different colours. The putative transposed element is indicated in black.

A further analysis of the reference assembly against a linearized 630 genome revealed four breakpoints (regions with discordantly mapped read-pairs). The first breakpoint is consistent with a deletion of ~70 bp. The remaining breakpoints are consistent with a transposition event, in which the transposed sequence is re-inserted elsewhere in the genome and in the inverse orientation compared to the reference (**Figure 1B**).

De novo assembly of the 630Δ*erm* genome using third generation sequencing

Based on the identification of a potential transposition event, and our previous finding that indels may have occurred that are difficult to detect using short reads, we decided to perform an unbiased, *de novo*, assembly of the 630Δ*erm* genome using single-molecule real-time sequencing. The Pacific Biosciences RSII system is capable of generating large reads, and with sufficient coverage, can generate high quality single contigs for bacterial genome sequences. We sequenced a genomic library of strain 630Δ*erm* on two SMRT cells and validated the resulting single contig with a third SMRT cell. The resulting genome consists of 4.293.049 base pairs, with an average GC content of 29.08 percent and an estimated coverage of 158× (**Figure 2A**). We generated an annotated version of this genome by transferring the most recent version of the 630 annotations [EMBL:AM180355][27], updating it with recent gene annotations from literature and incorporating qualifiers in the file to indicate specific features of 630Δ*erm*. The annotated sequence has been deposited under accession number EMBL:LN614756. Satisfyingly, our unbiased approach identified the 18-bp duplication in the *spo0A* gene, encoding the master regulator of sporulation, which we previously found [23] (**Figure 2B**). This demonstrates that the third-generation sequencing approach is superior to Illumina in identifying this type of difference. In addition, we could confirm the expected 2.4 kb deletion in Tn*5398* (**Figure 2C**).

The sequence of Tn*5398*ΔE which we determined shows 4 Single Nucleotide Polymorphisms (SNPs) compared to an *in silico* generated theoretical sequence of Tn*5398*ΔE (based on Hussain *et al.*) [18]. As a result of these differences, a progressive MAUVE[28] alignment of the Tn*5398*ΔE element from our strain with Tn*5398* of strain 630 demonstrates the deletion of CD2010 (*ermB1/erm(1)B*), CD2009A (ORF3), CD2009 (fragment of a putative topoisomerase), CD2008

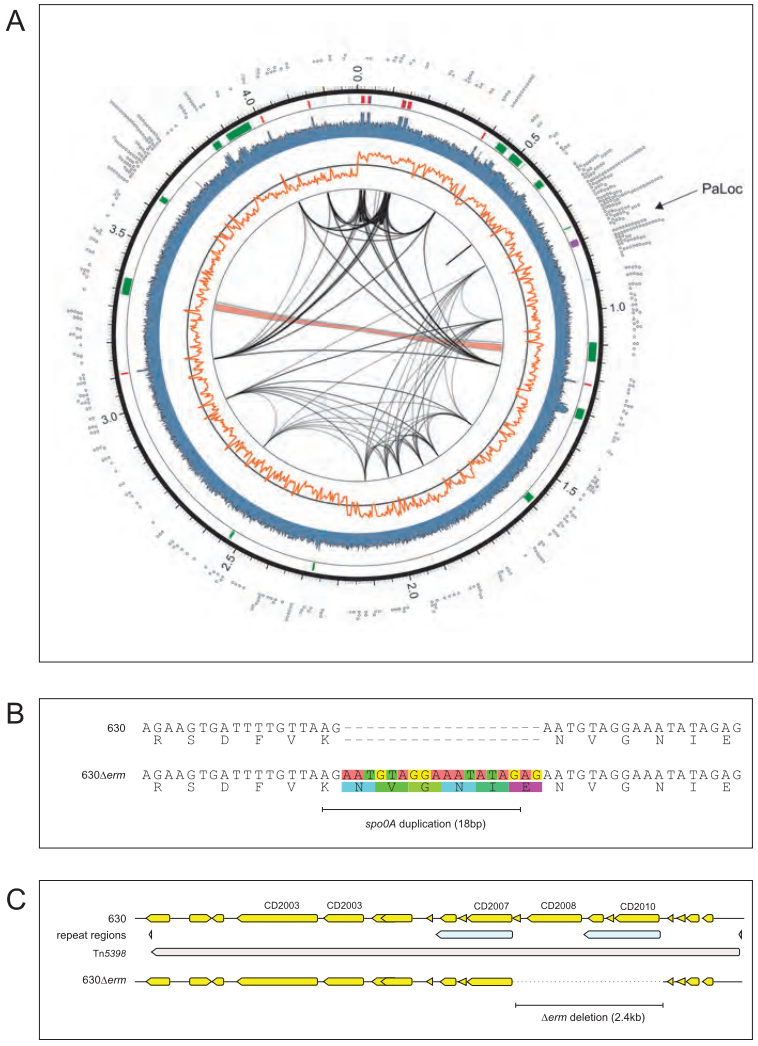


Figure 2. The complete genome of *C. difficile* 630Δ*erm*.

A. Overview of genomic features. Indicated are (from outside to inside); Short Tandem Repeats <500 bp (dots); rRNA (red), tRNA (blue), mobile genetic elements (green) and the PaLoc (purple); GC content per 1 kb window; GC skew (orange line) in a 5 kb sliding window; grey links represent repeats (193 repeats identified with Blast2Seq) having >95% identity and an alignment length of >500 bp; red links indicate an alignment length >2 kb.

B. Confirmation of the 18 bp duplication in *spo0A* resulting in a 6 amino acid direct repeat [23].

C. Confirmation of the 2.4 kb Δ*erm* deletion [18]. Open reading frames are indicated as yellow arrows, repeat elements in blue.

(ORF298) and most of CD2007A. This effectively removes the region between the two copies of *ermB*. The most likely scenario by which this occurred is through recombination between the two *ermB* genes or their immediate surrounding region; the sequence information is unable to determine the exact site of recombination, as these regions are identical, and the copies of *ermB* and ORF3 in 630Δ*erm* may therefore represent hybrids of CD2007/CD2010 or CD2006A/CD2009A, respectively. To reflect the results of the alignment as well as the mechanism described above, we have chosen to rename the *ermB* gene of strain 630Δ*erm* CD2007B/*ermB* (locus tag: CD630Derm_20072) and ORF3 as CD2006B (locus tag: CD630Derm_20062). The resulting arrangement suggests that CD2007B is potentially expressed, as it is fused to the promoter region of CD2010/*ermB1* at the exact same location, though the strain remains erythromycin sensitive. This discrepancy has been noted since the isolation of 630Δ*erm* [18] and cannot be resolved using the sequence information from our study.

We also identified short tandem repeats (>90 percent nucleotide identity) up to 500 bp. Strikingly, the genome analysis revealed two regions of high repeat density (**Figure 2A**). The first region (approximately 0.6 Mb-0.9 Mb) includes the PaLoc that encodes toxins A and B. This region was found to be capable of transfer by a conjugation like mechanism [16] and it is tempting to speculate that the high repeat density may contribute to this phenomenon. The second region (approximately 3.6 Mb-3.75 Mb) contains many genes involved in sugar metabolism but does not seem to be associated with annotated or characterized mobile elements. Large repeats (>95 percent identity and >500 bp in length) generally coincide with regions of high-GC content, and mainly reflect ribosomal gene clusters.

Analysis of ^{m6}A and ^{m4}C methylation patterns of *C. difficile*

In bacteria, post-replicative addition of a methyl group to a base by a DNA methyltransferase can result in the formation of *N*6-methyladenine (^{m6}A), *C*5-methylcytosine (^{m5}C) and *N*4-methylcytosine (^{m4}C) [29,30]. These modified bases play a role in restriction/modification systems or may regulate cellular processes (reviewed in [30-33]).

There is little information on methylation of chromosomal DNA in *C. difficile*. Five methylases have been identified in *C. difficile* 630 [34], but *in vivo* methylation

patterns have not been characterized. We took advantage of the pulse profiles of the Pacific Biosciences RSII reads that hold information about base modifications[35,36] to generate the first comprehensive analysis of methylation patterns in *C. difficile* (**Figure 3A**).

^{m6}A modifications can be identified with high confidence and the vast majority of these modifications (7288/7687 = 95 percent) were associated with the motif CAAAA**A**, in which the last adenine residue is modified (**Figure 3B**). Previous studies identified a single methylase, M.*Cdi*25 (corresponding to CD2758) with homology to adenine specific methylases, but failed to identify its target site in restriction protection experiments[34] We postulate that CD2758 recognizes and methylates the last adenine residue the CAAAAA motif and that this is possibly the only adenine-methylase in *C. difficile* 630Δ*erm*.

The pulse profiles of the Pacific Biosciences RSII reads also identify modified cytosines. Only a fraction of these are positively identified as ^{m4}C, in part due to the effect of modifications that are in close proximity to each other on the pulse profiles[36,37]. We did not further investigate ^{m5}C modifications, as they can only reliably be detected on the Pacific Biosciences platform after Tet1-treatment, by preparation of shorter library fragments that are not ideal for genome *de novo* assembly, and with much higher coverage than obtained in our experiment [36]. Unspecified modifications may therefore represent ^{m4}C, and possibly ^{m5}C or other modifications.

The SMRT Portal identified the motif G**C**AGCAGC, in which the first cytosine residue is modified, as overrepresented in the methylcytosine dataset (**Figure 3B**). This motif is remarkably similar to the GCWGC motif identified for the M.*Cdi*1226 methylase (CD3147)[34]. We could identify 146 instances of ^{m4}C methylation and 16 of those contained the motif (11 percent). When a DREME search was performed [38] using 41 bp sequences centred on ^{m4}C only, a highly similar motif (GCAGCR) was found in 33 instances. Moreover, none of the other motifs (see below) were specifically linked to ^{m4}C modifications, suggesting that many if not all of the ^{m4}C modifications are due to CD3147.

^{m4}C and ^{m6}A methylations that were not associated with the overrepresented motifs seemed to correspond to regions of high GC-content, including the mobile elements CTn*1*, CTn*2* and CTn*4* (**Figure 3**).

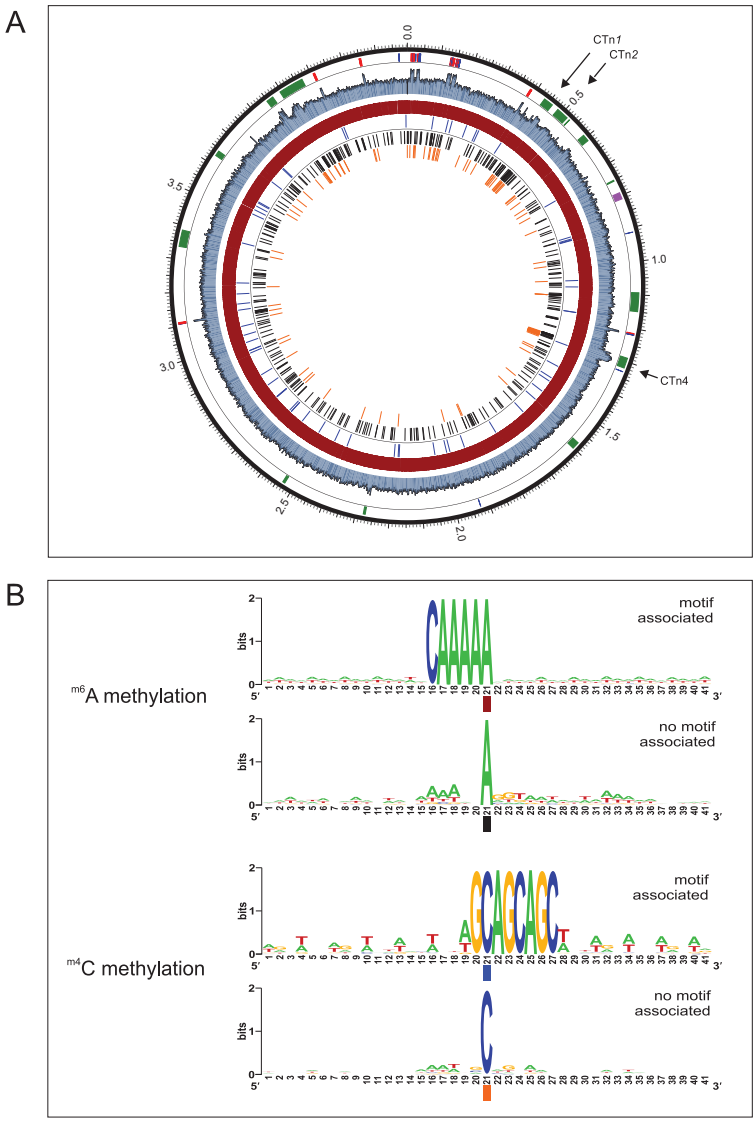


Figure 3. Methylation patterns in *C. difficile* 630Δ*erm*.

A. Genome wheel showing motif-associated ^{m6}A methylation (red), motif associated ^{m4}C methylation (blue), ^{m6}A methylation not associated with a motif (black), and ^{m4}C methylation events not associated with a motif (orange) in relation to GC content (per 1 kb window), rRNA (red), tRNA (blue) and mobile genetic elements (green).

B. Sequence logos for the ^{m6}A methylated sequences and sequence logos for the ^{m4}C methylated sequences.

We also evaluated motifs previously identified as putative target sites for the other three cytosine specific methylases of *C. difficile*, M.*Cdi*633 (CD0935), M.*Cdi*587 (CD0927) and M.*Cdi*824 (CD1109)[34]. CD0935 conferred partial protection against digestion with *Ball* (target site: TGGCCA). Our data did not show any modifications on cytosine or adenine residues of this motif anywhere in the genome (n = 396). Considering that we cannot reliably detect ^{m5}C modifications in our setup, it is possible that M.*Cdi*633is an ^{m5}C specific methylase. CD0927 could confer protection against *Sau*96I (target site: GGNCC) in *E. coli*, but *C. difficile* chromosomal DNA is only partially resistant to *Sau*96I digestion[34]. We found only very low levels (~0.1 percent) of modified cytosines for this motif (n = 3824) in 630Δ*erm*, which together with the earlier observations suggests that CD0927 is either minor ^{m4}C or a ^{m5}C methylase. CD1109 conferred protection against *Sma*I (which recognizes CCCGGG). We found that 6/60 (10 percent) of the motifs contained a modified cytosine at the third position. These modifications are likely ^{m4}C's that cannot be positively identified as ^{m4}C due to adjacent modified bases.

C. difficile chromosomal DNA is wholly resistant to *Tse*I (target site: GCWGC) and *Sma*I (target site (CCCGGG), though we only detected modifications for ~10 percent of the occurrences of these motifs. This may be due to only a fraction of the methylcytosine modifications being called by the Pacific Biosciences SMRT platform in our analyses.

The function of the methylases of *C. difficile* is unknown. None seem associated with an endonuclease, indicating they are not likely to be part of a restriction-modification system. Consistent with this, no effect on conjugation efficiency was observed[34]. CD0927 and CD0935 are part of prophage 1, and CD1109 is present on the CTn*4* element, suggesting they may play a role in the biology of mobile elements.

Comparison of the complete genome of 630Δ*erm* with strain 630 reveals SNPs, indels and rearrangements

It is likely that more than the two previously identified differences (Δ*erm* deletion and 18 bp duplication in *spo0A*) exist between strain 630 and strain 630Δ*erm*. We therefore compared our *de novo* assembled genome to the reference sequence. We identified 71 differences between the two strains. These encompass 8 deletions

Table 1. Structural variants associated with coding sequences

AM183055		630*erm*		Type	Description	Region	Gene name	Function	Details
Start	End	Start	End						
84143		89438		Substitution	C > T	CD630_00580	*tuf1*	Elongation factor EFTu/EF1A	Synonymous
84227		89522		substitution	C > T	CD630_00580	*tuf1*	Elongation factor EFTu/EF1A	Synonymous
103225		108520		substitution	G > T	CD630_00730	*rplC*	50S ribosomal protein L3	Synonymous
610336		615631		substitution	G > A	CD630_05140	*cwpV*	Cell surface protein	Val > Ile
610480		615775		substitution	C > T	CD630_05140	*cwpV*	Cell surface protein	Synonymous
610563	610564	615859	615861	insertion	610563_610564ins	CD630_05140	*cwpV*	Cell surface protein	In frame Ala insertion
610570		615868		substitution	A > G	CD630_05140	*cwpV*	Cell surface protein	Ile > Val
610638		615936		substitution	C > T	CD630_05140	*cwpV*	Cell surface protein	Synonymous
610752		616050		substitution	G > A	CD630_05140	*cwpV*	Cell surface protein	Synonymous
610840		616138		substitution	C > T	CD630_05140	*cwpV*	Cell surface protein	Synonymous
610875		616173		substitution	C > T	CD630_05140	*cwpV*	Cell surface protein	Synonymous
755776	755800	760995	760996	deletion	755776_755800del	CD630_06320		Conserved hypothetical protein	In frame 8aa deletion in repeat region
1000995		1006274		substitution	A > G	CD630_08260		Putative ferric-uptake regulator	Thr > Ala
1391850		1397129		substitution	T > C	CD630_11900		Putative acyl-CoA N-acyltransferase	Phe > Leu
1413060	1413077	1418339	1418354	duplication	1413060_1413077dup	CD630_12140	*spo0A*	Stage 0 sporulation protein A	6aa (NVGNIE) duplication

Table 1. (continued) Structural variants associated with coding sequences

AM183055		630Δerm		Type	Description	Region	Gene name	Function	Details
Start	End	Start	End						
1607458	1607459	1612756		insertion	1607458_1607459insT	CD630_13880		Putative transcriptional regulator	Restores transcriptional regulator
2044514		2049813		substitution	C > G	CD630_17670	*gapB*	Glyceraldehyde-3-phosphate dehydrogenase GAPDH	Pro > Ala
2137467	2183040	2142764	2142765	deletion	2137467_2183040del	CD630_18440		Putative adhesin	Translocation of CTn5, CD1844 restored
2209236		2168961		substitution	G > A	CD630_19070	*eutG*	Ethanolamine iron-dependent Alcohol dehydrogenase	Gly > Glu
2924655		2881973		substitution	C > T	CD630_25320		Aminotransferase, alanine—glyoxylate transaminase	Synonymous
3034953		2992271		substitution	C > A	CD630_26270		Conserved hypothetical protein	Gly > Cys
3080703		3038021		substitution	C > T	CD630_26670	*ptsG -BC*	PTS system, glucose-specific IIBC component	Val > Ile
3686534	3686535	3643756	3643756	insertion	3686534_3686535insA	CD630_31561		Conserved hypothetical protein	Restores conserved hypothetical protein
3967522	3967523	3924743	3970315	insertion	3967522_3967523insAM180355: g.2137467_2183040	CD630_33930	*rumA*	23S rRNA (uracil-5-)-methyltransferase	Translocation of CTn5; fuses rumA (CD3393) to CD1844A
4166495		4169292		substitution	G > A	CD630_35650		Transcriptional regulator, GntR family	Ala > Val
12347	12348	12348	17642	insertion	12347_12348ins	multiple			rRNA/tRNA cluster
2317627	2320033	2277358	2277359	deletion	2317633_2320041del	multiple			Loss of erythromycin resistance (Δerm)

(including the Δ*erm* mutation)[18], 10 insertions (including the duplication in *spo0A*[23]), 2 insertion-deletions, 50 substitutions and 1 region of complex structural variation (**Additional file 1**). Of these, 23 were located intergenically. This includes a 102 bp deletion which likely corresponds to the breakpoint at 0.68 Mb identified in the short read next generation sequencing (**Figure 1B**). A complete list of identified structural variants is available as Supplemental Material (**Additional file 1**).

Twenty-three of the identified differences are associated with rRNA sequences. We found that strain 630Δ*erm* has acquired an extra ~5 kb rRNA/tRNA cluster that is inserted between CD0011 and CD0012 compared to strain 630 (**Table 1, Figure 4**). Copy number variations in rRNA operons have previously been noted for *C. difficile*[39] and may reflect an adaptation to favourable growth conditions in the laboratory. Similar to rRNA operon 6, this operon contains $tRNA^{Leu}$ and $tRNA^{Met}$ genes downstream of the 23S rRNA gene, but the intergenic spacer region (ISR) between the 16S and 23S rRNA genes does not contain a $tRNA^{Ala}$. A detailed comparison of the ISRs of the different rRNA operons is provided as **Additional file 2**. A striking number of differences were found in rRNA operon 11 (**Figure 4**). As observed previously[40], the sequence variations cluster in the 3' region of the 16S rRNA and 5' of the 23S rRNA genes.

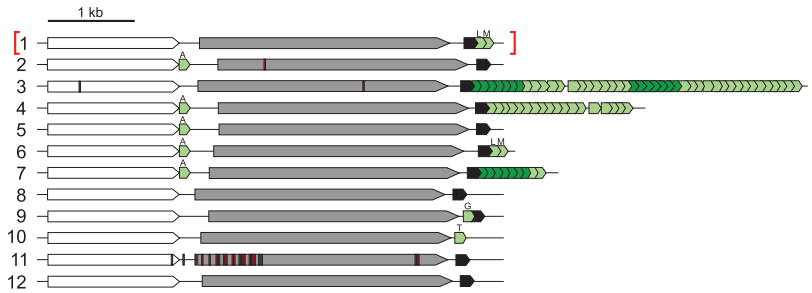


Figure 4. Schematic representation of the rRNA operons and associated tRNA clusters of *C. difficile* 630Δ*erm*.

Operons are numbered from 1–12 in the order they appear in the genome sequence. 16S rRNA, 23S rRNA and 5S rRNA genes are indicated with white, grey and black arrow shapes, respectively. tRNAs are indicated by green arrow shapes. SNPs between the rRNA clusters of strains 630Δ*erm* and 630 are indicated in red (for details see Additional file 1). Brackets indicate that operon 1 is unique to strain 630Δerm. A cluster of tRNAs that is found multiple times associated with rRNAs ($tRNA^{Asn}$–$tRNA^{Leu}$–$tRNA^{Met}$–$tRNA^{Glu}$–$tRNA^{Gly}$–$tRNA^{Val}$–$tRNA^{Asp}$) is indicated in dark green. A = $tRNA^{Ala}$, L = $tRNA^{Leu}$, M = $tRNA^{Met}$, G = $tRNA^{Gly}$, T = $tRNA^{Thr}$. Figure is approximately to scale.

We focused our further analysis on the 26 variants that are associated with annotated pseudogenes or open reading frames (**Table 1**). A 24 bp deletion in CD0632, a conserved protein of unknown function, shortens the arginine-alanine repeat in this protein by 8 amino acids. In two cases, a single base pair insertion restores a pseudogene (CD1388 and CD3156A). This was confirmed by assembling the short-read Illumina sequences against both the 630 reference genome and the *de novo* assembled 630Δ*erm* genome, as a variant was identified in the former but not the latter. CD1388 encodes a putative regulatory protein with a helix-turn-helix motif and CD3156A a conserved protein of unknown function. Interestingly, both proteins encoded by these genes were previously identified in a proteomic analysis [27], indicating that they are expressed in strain 630Δ*erm*. Two in-frame insertions were identified (an extra alanine residue in CD0514 and the published duplication in *spo0A*/CD1214). Out of 18 identified nucleotide substitutions, 9 were synonymous. These include SNPs in the gene encoding elongation factor Tu (*tuf1*/CD0058), ribosomal protein L50 (*rplC*/CD0073) and the putative aminotransferase CD2532. Strikingly, the CD0514 gene, encoding the cell wall protein *cwpV* [41,42], contains an unusually high density of mutations. In addition to the insertion and 5 synonymous mutations, it contains 2 non-synonymous but conservative mutations.

Other non-synonymous mutations are located in the putative ferric uptake regulator CD0826, the putative acyl-CoA N-acyltransferase CD1190, predicted glyceraldehyde-phosphate dehydrogenase CD1767 (*gapB*), ethanolamine utilization protein CD1907 (*eutG*), the hypothetical protein CD2627, the phosphotransferase system protein CD2667 (*ptsG-BC*) and the transcriptional regulator CD3565. In all these cases, the *de novo* assembly of the 630Δ*erm* genome was clearly supported by the short-read Illumina data.

CTn*5* is present in the *rumA* gene in both 630Δ*erm* (LUMC) and 630Δ*erm* (UCL)

In an attempt to visualize the proposed transposition event **(Figure 1B)**, we generated a dot plot of the genome sequence of our strain versus the reference **(Figure 5A)**. It is immediately evident that the CTn*5* element seems to have excised from its original location in CD1844 (encoding a putative cell wall adhesin) and has inserted in an inverted manner in *rumA* (CD3393) in our isolate of 630Δ*erm*, for clarity hereafter referred to as 630Δ*erm* (LUMC).

To exclude that the finding represents a misassembly error in the original 630 genome sequence and confirm the presence of CTn*5* in *rumA* in 630Δ*erm* (LUMC), we performed various control PCRs **(Figure 5B)**. In strain 630, we found CTn*5* inserted in CD1844 and confirmed an intact *rumA* gene. In contrast, in 630Δ*erm* (LUMC), we detected no product for the left and right junctions of CTn*5* in CD1844/CD1878A, indicating that the element is not present at this location. We readily amplified fragments corresponding to the left and right junction of CTn*5* when inserted in *rumA* in *C. difficile* 630Δ*erm* (LUMC), but not 630, chromosomal DNA. Interestingly, we observed a faint band corresponding to intact *rumA* even in strain 630Δ*erm* (LUMC). This indicates that a subpopulation of cells does not contain CTn*5* at this location, either because it has not inserted yet, or retains the ability to excise spontaneously as previously observed for 630 [8].

The CTn*5* insertion site identified here is located immediately downstream of CTn*7*. A similar tandem arrangement has previously been observed in two clinical PCR ribotype 001 isolates [10,43]. In another clinical isolate (RT027), which lacked a CTn*7*-like element, a CTn*5*-like element was found to be integrated at a site homologous to the target site of CTn*7* in 630 [43]. The annotation of CD3393 as *rumA* in *C. difficile* is based on homology of the predicted protein to *E. coli* RumA (also known as RlmD). This enzyme methylates a uracil nucleotide of the ribosomal RNA [44-46]. *E. coli rumA* mutants perform similarly compared with the wild type strain, in terms of cell growth, antibiotic resistance, and fidelity of translation. However, Δ*rumA* cells are outcompeted by wild-type cells in growth competition assays, which may imply that ribosome function is moderately affected [46].

The translocation of CTn*5* to *rumA* has two major consequences. First, the CD1844 gene, encoding a putative adhesin is restored. Second, the *rumA* open reading frame is fused to the CD1844A open reading frame resulting into a hybrid protein (CD3393A). CD1844A shows very high similarity (e-value: 1×10^{-62}, 97 percent identity) to the C-terminus of an *Enterococcus faecalis rumA* homologue [EMBL:EOK00135.1]. However, the homology of *C. difficile rumA* to this gene is limited to the N-terminal TrmA-like domain (COG2265) **(Figure 5B)**. Thus, a link between these open reading frames is also found in other organisms than *C. difficile*. In order to determine what the phenotypic consequences are of the transposition of CTn*5* further experiments are required.

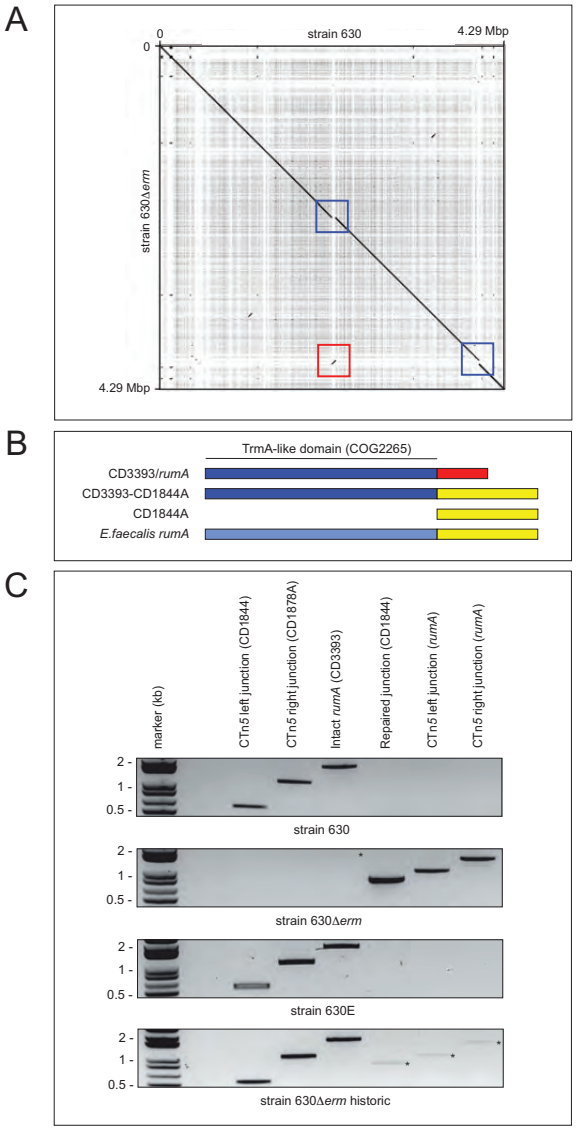


Figure 5. CTn5 is present in *rumA* in 630Δ*erm* but not 630 or 630E.

A. Dotplot of the reference sequence for *C. difficile* 630 (x-axis) versus the de novo assembled 630Δ*erm* sequence (y-axis), indicating the location of CD1844 and *rumA* (boxed in blue), and the CTn5 element (boxed in red). Note the inverted orientation of the mobile element.

B. Schematic representation of the *rumA*-CD1844A hybrid protein (CD3393A).

C. PCR confirmation of the transposition event. For primers used see Methods and Table 2.

To further our understanding of the origin of the transposition event, we compared the location of CTn*5* by PCR in different related strains; a non-passaged isolate of the original 630Δ*erm* [18], hereafter referred to as 630Δ*erm* (UCL), and another erythromycin sensitive derivative of 630, 630E/JIR8094 [47]. We found that in strain 630E the element is present in CD1844/CD1878A, identical to the reference strain, suggesting that the transposition event is not linked to the loss of erythromycin resistance. The 630Δ*erm* (UCL) strain shows prominent bands corresponding to CTn*5* at its CD1844/CD1878A location, but also a weak signal for CTn*5* at *rumA* **(Figure 5C)**. Therefore, this isolate likely contains a subpopulation of cells with the transposition identified in this study. It is possible that CTn*5* is stable at either location and the stock of the 630Δ*erm* (UCL) is non-clonal, or that CTn*5* in 630Δ*erm* (UCL) is highly mobile. During redistribution of the strain, isolates with either insertion could have been selected.

In summary, our data show that integration of CTn*5* can occur in at least two different sites in the *C. difficile* 630Δ*erm* genome, and that the element can switch between these locations during repeated passaging.

Conclusions

The work presented here provides the first reference genome for the widely used *C. difficile* laboratory strain 630Δ*erm*, including the first analysis of major methylation patterns for any *C. difficile* strain. Our work reveals that in addition to insertion, deletions and SNPs, the CTn*5* element has moved from its original location within CD1844 to the *rumA* gene in our isolate. The observation of such a dramatic rearrangement has important implications for the redistribution of strains with highly mobile genomes and argues for complete resequencing of common lab strains in each laboratory.

Methods and Materials

Bacterial strains and growth conditions

Our isolate of strain 630Δ*erm* was initially obtained from the Minton lab (University of Nottingham, Nottingham, UK), that in turn received it from the Mullany lab in which it was generated. For the purpose of resequencing the strain was cultured on prereduced CLO plates (Biomerieux), after which it was ented to BHI medium (Oxoid) supplemented with 0.5% yeast extract (Fluka).

Strain 630 was originally obtained from the Mastrantonio lab (Instituto Superiore di Sanità, Rome, Italy) and its use in our lab has been described before[48]. The 630Δ*erm* strain from the Mullany lab (UCL Eastman Dental Institute, London, UK), 630Δ*erm*(UCL), was transported as a glycerol stock on dry ice. Strain 630E was a kind gift of Robert Britton (Michigan State University, East Lansing, MI, USA). All strains were cultured as described for our isolate of strain 630Δ*erm*, which is referred to as 630Δ*erm*(LUMC) where appropriate.

Isolation of chromosomal DNA

For PCR analysis, chromosomal DNA was isolated using the QiaAmp Blood&Tissue kit (Qiagen) according to the manufacturer's instructions from growth obtained after streaking out the strain directly from the glycerol stock onto CLO plates (Biomerieux). For SMRT sequencing, high molecular weight DNA was isolated from 30 mL of an overnight culture, using the Qiagen GenomicTip 500/G, according to the manufacturer's instructions. The quality of the DNA was checked on a Nanodrop ND-200 machine (ThermoFisher), the integrity by agarose gel electrophoresis, and the DNA was quantified on a Qubit instrument (Invitrogen).

Illumina sequencing and analysis

For Illumina sequencing, chromosomal DNA was isolated by Baseclear (Leiden, The Netherlands) from a pellet of bacterial cells derived from 50 mL culture.

Data from 50 cycle 500 Mb paired-end read was delivered by Baseclear as 2 fastq files. Sequence reads have been deposited in the ENA Sequence Read Archive (EMBL:ERS550098). A preliminary analysis of the data was performed by aligning the paired-end reads to the reference genome of *C. difficile* strain 630 [GenBank:AM180355] using Geneious R7 (Biomatters, http://www.geneious.com). A more detailed analysis was performed using Stampy [49] and BWA [50]. In a routine quality control (QC) procedure on verifying the alignment, QC metrics including insert-sizes, mapped reads, unmapped reads and reads that align with a deviated pattern (DP; discordant read alignments) were examined. The case where a significant number of reads cannot align to the reference genome indicates an undefined sequence region in strain 630Δ*erm* or a contamination of the library. In our case, a few regions with discordantly mapped read pairs (DP > 9) were identified (Additional file 3) and validated automatically (Additional file 4). Of the validated breakpoints, the first has matches with the end of the reference assembly and is therefore an artefact of assembling the reads against a linearized genome. This was confirmed by artificially breaking the circular chromosome at a different position and repeating the procedure. Visual inspection in the Integrative Genome Viewer tool [51] on the alignment track (BAM file) was used to determine the nature of the Structural Variations).

Pacific biosciences RSII sequencing and de novo assembly

For single molecule real-time sequencing, a SMRTbell DNA template library with an insert size of ~20 kb was prepared according to the manufacturer's specification. To this end, chromosomal DNA was fragmented with G-tubes (Covaris). Subsequently, fragmented DNA was end-repaired and ligated to hairpin adapters. SMRT sequencing was carried out on the Pacific Biosciences RSII machine according to standard protocols (Magbead loading, 1×180 min). Sequence reads have been deposited in the ENA Sequence Read Archive (EMBL:ERS550016). Sequencing reads were corrected using the HGAP pipeline [52]. Assembly was performed using Celera Assembler 8.1. We observed unbalanced coverage of two regions of approximately 18.5 kb of the reference genome. These regions were found to be nearly identical phages [16], and the unbalanced coverage therefore likely represents an artefact of the unsupervised assembly procedure using the default settings. To correct for this, the assembly was artificially broken into three contigs at these

regions and was rejoined using the gap closure software PBJelly [53]. The edited assembly was then validated using reads from a third SMRT cell and polished using Quiver, a consensus algorithm that is part of the SMRT Portal. Subsequently, the consensus sequence was circularized based on the reference sequence of the ancestral 630 strain. We noted that the Pacific Biosciences consensus caller struggles with homopolymeric stretches of adenines and thymines. Therefore, a correction was carried out by performing a reference assembly of the short reads from the Illumina sequencing against the reclosed genome, yielding the final genome sequence. This sequence is available from EMBL (EMBL: LN614756).

In silico analysis of the 630Δ*erm* genome sequence

To annotate the *de novo* assembled genome sequence, we first updated the most recent version of the *C. difficile* 630 genome sequence [EMBL:AM180355.1] [27] in Artemis [54,55]. Next, we imported the flat genome sequence of strain 630Δ*erm* into Geneious R7 (Biomatters, http://www.geneious.com) and transferred the annotation using the "Live Annotate and Predict" function. The annotation track was manually curated to remove duplicate or missed annotations. The resulting file was saved as a GenBank file, further polished in a text editor and Artemis and submitted to the ENA archive. Genome wheel representations were prepared using Circos [56]. Indels and single nucleotide polymorphisms were identified using the Pacific Biosciences variant caller using the genome of *C. difficile* strain 630 [8] as a reference and further validated by MUMmer 3.0 [57] and progressiveMAUVE [28]. Subsequently a list of detected structural variants was manually curated (consensus between the alignment of Illumina and PacBio reads to the reference strain and the variants identified by MUMmer and progressiveMAUVE) as concordant description of differences in complex genomic regions could not be achieved by different methods. In addition, for all large structural variants dotplots were generated using Gepard 1.30 [58] using FASTA formatted genome sequences of strains 630 and 630Δ*erm*.

To identify modified bases, kinetic signals were processed for all genomic positions after aligning sequencing reads to the final single chromosome sequence of strain 630Δ*erm*. In order to accurately identify the methylated bases, a threshold of 45 for log-transformed *P* values was used after optimizing according to its distribution and minimizing the false positive rate. Genomic positions and identity of the

modifications were exported as a GFF file and imported as a separate track in the genome sequence in Geneious R7. Subsequently, the identification of sequence motifs was performed using the SMRT Portal and sequence logos were prepared using Weblogo (http://weblogo.berkeley.edu/)[59] with 20 bp sequence flanking the modified base.

Analysis of CTn*5* translocation

Translocation of CTn*5* was confirmed by PCR using primers (**Table 2)** designed to amplify the left and right junctions of CTn*5* as present in the *C. difficile* strain 630, as well as the *rumA* gene (**Table 1**) using Q5 polymerase (New England Biolabs). Cycling conditions were: initial denaturation 98°C 30 sec, 25 cycles 98°C 10 sec/60°C 30 sec/72°C 1 min 30 sec, and a final extension 72°C for 2 mins. Products were purified (GeneJet PCR purification kit, ThermoScientific) and run on a 0.5xTAE/1.2% agarose gel with a 1 kb + ladder (Fermentas). After staining with ethidium bromide, the DNA bands were visualized on a Geldoc system (Biorad).

Table 2. Oligonucleotides used in this study

Name	Sequence (5' – 3')	Description
oWKS-1467	CGCACCAGAATGGAAAGAAG	Left junction CTn5 [a]
oWKS-1468	AGGCGTACACTGTGGGATAG	Left junction CTn5 [b]
oWKS-1469	TAGATGATGCCGTTGCTGAG	Right junction CTn5 [b]
oWKS-1470	AAGGTTTGGGTCTGCTGTAG	Right junction CTn5 [a]
oWKS-1471	CCGTTACCGTCTGTAATG	*rumA* gene [b]
oWKS-1472	AGGGCCTATAAGGTAAGC	*rumA* gene [b]

a The repaired junction (CTn*5* excised from CD1844) is detected with oWKS-1467 and oWKS-1470.

b The insertion of CTn*5* into *rumA* is detected by primer combination oWKS-1468/oWKS-1472 and/or oWKS-1469/oWKS-1471.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

We thank Nigel Minton, Robert Britton and Paolo Mastrantonio for strains and Leon Mei of the Sequence Analysis Support Core (LUMC) for facilitating the initial Illumina analysis. Furthermore, we want to thank the Geneious team for helpful discussions. This work was supported, in part, by a Veni and a Vidi fellowship from the Netherlands Organization for Scientific Research and a Gisela Thier Fellowship from the Leiden University Medical Center to WKS.

Additional files

Additional files associated with this article can be found in the online version at https://bit.ly/2Lf9W8y

Additional_file_1 as XLSX

Additional file 1 Table summarizing structural variants identified between strain 630 and strain 630Δ*erm* (LUMC).

Additional_file_2 as ZIP

Additional file 2 ClustalW alignment of the 16S-23S regions in the 630Δ*erm* (LUMC) genome.

Additional_file_3 as XLSX

Additional file 3 Table summarizing discordantly mapped read-pairs in the Illumina HiSeq reference alignment of *C. difficile* 630Δ*erm* (LUMC) versus 630.

Additional_file_4 as XLSX

Additional file 4 Table summarizing validated discordantly mapped read-pairs in the Illumina HiSeq reference alignment of *C. difficile* 630Δ*erm* (LUMC) versus 630.

References

1 Hall, I. C. & O'Toole, E. in *Am. J. Child. Dis* Vol. 49 390–402 (1935).

2 Rupnik, M., Wilcox, M. H. & Gerding, D. N. Clostridium difficile infection: new developments in epidemiology and pathogenesis. *Nat. Rev. Microbiol* **7**, 526–536, doi:nrmicro2164 [pii];10.1038/nrmicro2164 [doi] (2009).

3 Viswanathan, V. K., Mallozzi, M. J. & Vedantam, G. Clostridium difficile infection: An overview of the disease and its pathogenesis, epidemiology and interventions. *Gut Microbes* **1**, 234–242, doi:10.4161/gmic.1.4.12706 [doi] (2010).

4 Shen, A. Clostridium difficile toxins: mediators of inflammation. *J. Innate. Immun* **4**, 149–158, doi:000332946 [pii];10.1159/000332946 [doi] (2012).

5 Vedantam, G. *et al.* Clostridium difficile infection: toxins and non-toxin virulence factors, and their contributions to disease establishment and host response. *Gut Microbes* **3**, 121–134, doi:19399 [pii];10.4161/gmic.19399 [doi] (2012).

6 He, M. *et al.* Emergence and global spread of epidemic healthcare-associated Clostridium difficile. *Nat. Genet* **45**, 109–113, doi:ng.2478 [pii];10.1038/ng.2478 [doi] (2013).

7 Smits, W. K. Hype or hypervirulence: A reflection on problematic C. difficile strains. *Virulence* **4**, doi:26297 [pii] (2013).

8 Sebaihia, M. *et al.* The multidrug-resistant human pathogen Clostridium difficile has a highly mobile, mosaic genome. *Nat. Genet* **38**, 779–786, doi:ng1830 [pii];10.1038/ng1830 [doi] (2006).

9 Wust, J., Sullivan, N. M., Hardegger, U. & Wilkins, T. D. Investigation of an outbreak of antibiotic-associated colitis by various typing methods. *J. Clin. Microbiol* **16**, 1096–1101 (1982).

10 Brouwer, M. S., Warburton, P. J., Roberts, A. P., Mullany, P. & Allan, E. 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 [doi];PONE-D-11-06984 [pii] (2011).

11 Stabler, R. A. *et al.* Comparative phylogenomics of Clostridium difficile reveals clade specificity and microevolution of hypervirulent strains. *J. Bacteriol* **188**, 7297–7305 (2006).

12 He, M. *et al.* Evolutionary dynamics of Clostridium difficile over short and long time scales. *Proc. Natl. Acad. Sci. U. S. A* **107**, 7527–7532, doi:0914322107 [pii];10.1073/pnas.0914322107 [doi] (2010).

13 Roberts, A. P. & Mullany, P. Tn916-like genetic elements: a diverse group of modular mobile elements conferring antibiotic resistance. *FEMS Microbiol. Rev* **35**, 856–871, doi:10.1111/j.1574-6976.2011.00283.x [doi] (2011).

14 Mullany, P. *et al.* Genetic analysis of a tetracycline resistance element from Clostridium difficile and its conjugal transfer to and from Bacillus subtilis. *J. Gen. Microbiol* **136**, 1343–1349 (1990).

15 Jasni, A. S., Mullany, P., Hussain, H. & Roberts, A. P. Demonstration of conjugative transposon (Tn5397)-mediated horizontal gene transfer between Clostridium difficile and Enterococcus faecalis. *Antimicrobial agents and chemotherapy* **54**, 4924–4926, doi:10.1128/aac.00496-10 (2010).

16 Brouwer, M. S. *et al.* Horizontal gene transfer converts non-toxigenic Clostridium difficile strains into toxin producers. *Nat. Commun* **4**, 2601, doi:ncomms3601 [pii];10.1038/ncomms3601 [doi] (2013).

17 Purdy, D. *et al.* Conjugative transfer of clostridial shuttle vectors from Escherichia coli to Clostridium difficile through circumvention of the restriction barrier. *Mol. Microbiol* **46**, 439–452, doi:3134 [pii] (2002).

18 Hussain, H. A., Roberts, A. P. & Mullany, P. Generation of an erythromycin-sensitive derivative of Clostridium difficile strain 630 (630Deltaerm) and demonstration that the conjugative transposon Tn916DeltaE enters the genome of this strain at multiple sites. *J. Med. Microbiol* **54**, 137–141 (2005).

19 Heap, J. T., Pennington, O. J., Cartman, S. T., Carter, G. P. & Minton, N. P. The ClosTron: a universal gene knock-out system for the genus Clostridium. *J. Microbiol. Methods* **70**, 452–464, doi:S0167-7012(07)00208-4 [pii];10.1016/j.mimet.2007.05.021 [doi] (2007).

20 Ng, Y. K. *et al.* Expanding the repertoire of gene tools for precise manipulation of the Clostridium difficile genome: allelic exchange using pyrE alleles. *PLoS. One* **8**, e56051, doi:10.1371/journal.pone.0056051 [doi];PONE-D-12-24523 [pii] (2013).

21 Cartman, S. T., Kelly, M. L., Heeg, D., Heap, J. T. & Minton, N. P. Precise manipulation of the Clostridium difficile chromosome reveals a lack of association between the tcdC genotype and toxin production. *Appl. Environ. Microbiol* **78**, 4683–4690, doi:AEM.00249-12 [pii];10.1128/AEM.00249-12 [doi] (2012).

22 Bakker, D. *et al.* The HtrA-like protease CD3284 modulates virulence of Clostridium difficile. *Infect Immun* **82**, 4222–4232, doi:10.1128/iai.02336-14 (2014).

23 Rosenbusch, K. E., Bakker, D., Kuijper, E. J. & Smits, W. K. C. difficile 630Deltaerm Spo0A regulates sporulation, but does not contribute to toxin production, by direct high-affinity binding to target DNA. *PLoS. One* **7**, e48608, doi:10.1371/journal.pone.0048608 [doi];PONE-D-12-15030 [pii] (2012).

24 Zeigler, D. R. *et al.* The origins of 168, W23, and other Bacillus subtilis legacy strains. *Journal of bacteriology* **190**, 6983–6995, doi:10.1128/jb.00722-08 (2008).

25 Srivatsan, A. *et al.* High-precision, whole-genome sequencing of laboratory strains facilitates genetic studies. *PLoS genetics* **4**, e1000139, doi:10.1371/journal.pgen.1000139 (2008).

26 McLoon, A. L., Kolodkin-Gal, I., Rubinstein, S. M., Kolter, R. & Losick, R. Spatial regulation of histidine kinases governing biofilm formation in Bacillus subtilis. *Journal of bacteriology* **193**, 679–685, doi:10.1128/jb.01186-10 (2011).

27 Pettit, L. J. *et al.* Functional genomics reveals that Clostridium difficile Spo0A coordinates sporulation, virulence and metabolism. *BMC. Genomics* **15**, 160, doi:1471-2164-15-160 [pii];10.1186/1471-2164-15-160 [doi] (2014).

28 Darling, A. E., Mau, B. & Perna, N. T. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS. One* **5**, e11147, doi:10.1371/journal.pone.0011147 [doi] (2010).

29 Marinus, M. G. & Casadesus, J. Roles of DNA adenine methylation in host-pathogen interactions: mismatch repair, transcriptional regulation, and more. *FEMS Microbiol Rev* **33**, 488–503, doi:10.1111/j.1574-6976.2008.00159.x (2009).

30 Collier, J. Epigenetic regulation of the bacterial cell cycle. *Current opinion in microbiology* **12**, 722–729, doi:10.1016/j.mib.2009.08.005 (2009).

31 Ratel, D., Ravanat, J. L., Berger, F. & Wion, D. N6-methyladenine: the other methylated base of DNA. *BioEssays : news and reviews in molecular, cellular and developmental biology* **28**, 309–315, doi:10.1002/bies.20342 (2006).

32 Wion, D. & Casadesus, J. N6-methyl-adenine: an epigenetic signal for DNA-protein interactions. *Nat Rev Microbiol* **4**, 183–192, doi:10.1038/nrmicro1350 (2006).

33 Lobner-Olesen, A., Skovgaard, O. & Marinus, M. G. Dam methylation: coordinating cellular processes. *Current opinion in microbiology* **8**, 154-160, doi:10.1016/j.mib.2005.02.009 (2005).

34 Herbert, M., O'Keeffe, T. A., Purdy, D., Elmore, M. & Minton, N. P. Gene transfer into Clostridium difficile CD630 and characterisation of its methylase genes. *FEMS Microbiol. Lett* **229**, 103-110, doi:S037810970300795X [pii] (2003).

35 Flusberg, B. A. *et al.* Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods* **7**, 461-465, doi:10.1038/nmeth.1459 (2010).

36 Anonymous. Detecting DNA Base Modifications: SMRT Analysis of Microbial Methylomes, <https://github.com/PacificBiosciences/Bioinformatics-Training/wiki/Methylome-Analysis-Technical-Note>

37 Biosciences, P. Detecting DNA Base Modifications using Single Molecule, Real-Time Sequencing (SMART), <https://www.pacb.com/wp-content/uploads/2015/09/WP_Detecting_DNA_Base_Modifications_Using_SMRT_Sequencing.pdf> (2012).

38 Bailey, T. L. DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics* **27**, 1653-1659, doi:10.1093/bioinformatics/btr261 (2011).

39 Sadeghifard, N., Gurtler, V., Beer, M. & Seviour, R. J. The mosaic nature of intergenic 16S-23S rRNA spacer regions suggests rRNA operon copy number variation in Clostridium difficile strains. *Applied and environmental microbiology* **72**, 7311-7323, doi:10.1128/aem.01179-06 (2006).

40 Gurtler, V. & Grando, D. New opportunities for improved ribotyping of C. difficile clinical isolates by exploring their genomes. *Journal of microbiological methods* **93**, 257-272, doi:10.1016/j.mimet.2013.02.013 (2013).

41 Reynolds, C. B., Emerson, J. E., de la Riva, L., Fagan, R. P. & Fairweather, N. F. The Clostridium difficile cell wall protein CwpV is antigenically variable between strains, but exhibits conserved aggregation-promoting function. *PLoS pathogens* **7**, e1002024, doi:10.1371/journal.ppat.1002024 (2011).

42 Emerson, J. E. *et al.* A novel genetic switch controls phase variable expression of CwpV, a Clostridium difficile cell wall protein. *Mol. Microbiol* **74**, 541-556, doi:MMI6812 [pii];10.1111/j.1365-2958.2009.06812.x [doi] (2009).

43 Brouwer, M. S., Roberts, A. P., Mullany, P. & Allan, E. In silico analysis of sequenced strains of Clostridium difficile reveals a related set of conjugative transposons carrying a variety of accessory genes. *Mob. Genet. Elements* **2**, 8-12, doi:10.4161/mge.19297 [doi];2011MGE0079R [pii] (2012).

44 Agarwalla, S., Kealey, J. T., Santi, D. V. & Stroud, R. M. Characterization of the 23 S ribosomal RNA m5U1939 methyltransferase from Escherichia coli. *The Journal of biological chemistry* **277**, 8835-8840, doi:10.1074/jbc.M111825200 (2002).

45 Madsen, C. T., Mengel-Jorgensen, J., Kirpekar, F. & Douthwaite, S. Identifying the methyltransferases for m(5)U747 and m(5)U1939 in 23S rRNA using MALDI mass spectrometry. *Nucleic Acids Res* **31**, 4738-4746 (2003).

46 Persaud, C. *et al.* Mutagenesis of the modified bases, m(5)U1939 and psi2504, in Escherichia coli 23S rRNA. *Biochemical and biophysical research communications* **392**, 223-227, doi:10.1016/j.bbrc.2010.01.021 (2010).

47 O'Connor, J. R. *et al.* Construction and analysis of chromosomal Clostridium difficile mutants. *Mol. Microbiol* **61**, 1335-1351, doi:MMI5315 [pii];10.1111/j.1365-2958.2006.05315.x [doi] (2006).

48 van den Berg, R. J., Schaap, I., Templeton, K. E., Klaassen, C. H. & Kuijper, E. J. Typing and subtyping of Clostridium difficile isolates by using multiple-locus variable-number tandem-repeat analysis. *J. Clin. Microbiol* **45**, 1024-1028, doi:JCM.02023-06 [pii];10.1128/JCM.02023-06 [doi] (2007).

49 Lunter, G. & Goodson, M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res* **21**, 936-939, doi:10.1101/gr.111120.110 (2011).

50 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).

51 Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics* **14**, 178-192, doi:10.1093/bib/bbs017 (2013).

52 Chin, C. S. *et al.* Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* **10**, 563-569, doi:nmeth.2474 [pii];10.1038/nmeth.2474 [doi] (2013).

53 English, A. C. *et al.* Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PloS one* **7**, e47768, doi:10.1371/journal.pone.0047768 (2012).

54 Rutherford, K. *et al.* Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944-945 (2000).

55 Carver, T. *et al.* Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics* **24**, 2672-2676, doi:btn529 [pii];10.1093/bioinformatics/btn529 [doi] (2008).

56 Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res* **19**, 1639-1645, doi:10.1101/gr.092759.109 (2009).

57 Kurtz, S. *et al.* Versatile and open software for comparing large genomes. *Genome Biol* **5**, R12, doi:10.1186/gb-2004-5-2-r12 (2004).

58 Krumsiek, J., Arnold, R. & Rattei, T. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics* **23**, 1026-1028, doi:10.1093/bioinformatics/btm039 (2007).

59 Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res* **14**, 1188-1190, doi:10.1101/gr.849004 (2004).

Chapter 4

The replicative helicase CD3657 of *Clostridium difficile* interacts with the putative loader protein CD3654

Erika van Eijk [1]
Annemieke Friggen [1]
Vasileios Paschalis [2]
Matthew Green [2]
Geoffrey S. Briggs [2]
James Gibson [2]
Panos Soultanas [2]
Wiep Klaas Smits [1]

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

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

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

Abstract

Clostridium difficile **is the main cause of health-care associated diarrhoea. Limited treatment options and reports of reduced susceptibility to current treatment emphasize the necessity for the development of novel antimicrobials. DNA replication is an essential and conserved process in all domains of life and may therefore serve as a target for the development of new antimicrobial therapeutics.**

Compared to its well characterized relative, *Bacillus subtilis*, knowledge of the molecular biology and genetics of *Clostridium difficile* is still in its infancy and there is a gap in our knowledge of the replication mechanisms in this organism.

Here, we identified several *C. difficile* genes with homology to *B. subtilis* replication initiation proteins and set out to characterize the replicative helicase (CD3657) and its putative loader protein (CD3654). Intra- and intermolecular protein-protein interactions were assessed by bacterial two-hybrid and analytical gel-filtration experiments. Helicase can form hexamers at high concentrations and interacts with the putative loader protein in an ATP-dependent manner. Binding of ATP to helicase is pivotal for the formation of a stable homo-hexamer and its interaction with the putative loader protein. Despite the formation of a helicase-helicase loader protein complex, no helicase activity could be demonstrated *in vitro*. Our data suggest that the *C. difficile* helicase is of the ring-maker class and that *cd3654* encodes the helicase loader, but critical aspects of helicase loading and activation differ from the Gram-positive model *Bacillus subtilis*.

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, it is perhaps not surprising that details of the molecular mechanisms can vary substantially as these prokaryotes diverged more than 3 billion years ago [1].

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 [2–4]. 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 code for a loader protein also differs [2,6]. In addition to its role in the helicase-loading process, the loader protein regulates the helicase through interactions and is therefore pivotal in replisome assembly [3,7,8]. For historical reasons, the nomenclature for the replication proteins differs between bacterial species (e.g. *E. coli* helicase; DnaB and *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 (*Ec*DnaB) is loaded by a single loader protein (*Ec*DnaC) *in vivo* [9–11], whereas loading of the *B. subtilis* helicase (*Bs*DnaC) requires three accessory proteins (*Bs*DnaD, *Bs*DnaB and *Bs*DnaI) *in vivo* [12–14] 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 [2,6]. 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 [6,16]. To accomplish this, it is thought that either a pre-formed ring is physically opened (ring-breaker) or that the ring is assembled from monomers at *oriC*

(ring-maker) [2]. In *E. coli*, preformed 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 [10]. This 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 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.

B. subtilis helicase loading *in vivo* is a hierarchical process [13,14,18]. 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 *Bs*DnaD and *Bs*DnaB co-loader proteins, which are structural homologues (PFAM DnaB_2), associate sequentially with the replication origin [14], and possibly contribute to origin remodelling. This ultimately enables the ATPase loader protein to load the helicase [13-15,18-23].

The replication initiation protein (*Bs*DnaA) and helicase-loader protein (*Bs*DnaI) belong to the AAA+ (ATPases associated with various cellular activities) family of ATPases [10,24-26]. These AAA+ enzymes are, in their turn, part of the additional strand catalytic glutamate (ASCE) family [15,25,27-30]. The *Bs*DnaI 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 *Bs*DnaC helicase is also an ASCE protein and belongs to RecA-type helicase Superfamily 4 (SF4), which is involved in DNA replication [29,31,32]. The SF4 superfamily of helicases is characterized by five sequence motifs; H1, H1a, H2, H3, and H4 [29,33,34]. Motifs H1 and H2 are equivalent to the ATP-coordinating Walker A and B motifs found in many other ATPases [29].

We are interested in DNA replication in the Gram-positive bacterium *C. difficile*, the most common causative agent of antibiotic-associated diarrhoea [35,36]. In this study, we employed an *in silico* analysis to identify homologues of replication initiation proteins from *B. subtilis*. From thereon, we focused on the helicase and a putative helicase loader. Our data show that helicase loading and activation in *C. difficile* may differ critically from the *B. subtilis* model.

Results

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* [14]. BLASTP queries of the genome of *C. difficile* 630 (GenBank AM180355.1) using the amino acid sequence of these proteins from *Bacillus subtilis* subsp. *subtilis* strain 168 (GenBank NC0989.1) allowed the identification 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 putative helicase (CD3657; *e*-value = 7×10^{-713}) were identified with high confidence, sharing respectively 62 and 52 percent identity with their *B. subtilis* counterparts across the full length of the protein. Interestingly, no homologue of *Bs*DnaB was found using this strategy. However, BLASTP shows that the genome of *C. difficile* does harbour two homologues of *Bs*DnaD (CD2943: *e*-value = 2×10^{-5}, identity = 29%, query coverage = 32%; CD3653: *e*-value = 4×10^{-5}, identity 29%, query coverage 47%). As *Bs*DnaB and *Bs*DnaD are strictly required for replication initiation in *B. subtilis* [18,37] and are structurally related despite limited amino acid sequence similarity [19], we further examined the *C. difficile* homologues of *Bs*DnaD.

*Bs*DnaD is composed of two domains; DDBH1 and DDBH2, whereas *Bs*DnaB has a DDBH1-DDBH2-DDBH2 structure [19]. The DnaD-like proteins CD3653 and CD2943 of *C. difficile* both consist of three domains (DDBH1-DDBH2-DDBH2) and therefore resemble *Bs*DnaB in domain structure (**Figure 1A**). CD2943 is annotated as a putative phage replication protein and is located in the ~50kb prophage 2 [38]. Also, in *Listeria monocytogenes*, *Staphylococcus aureus* and *Lactobacillus plantarum* DDBH2-containing phage genes have been identified [19,39].

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.

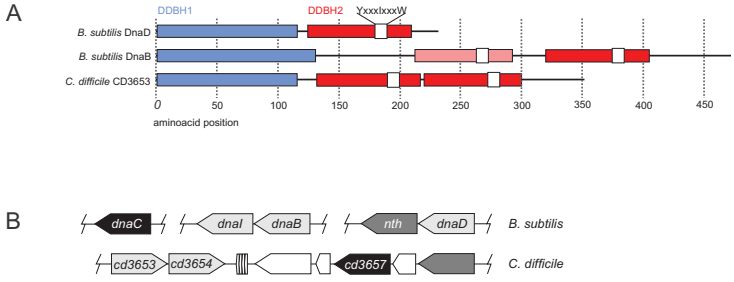


Figure 1. ***In silico*** **analysis of putative replication initiation proteins of** ***C. difficile*****.**

A. Domain structure of BsDnaD, BsDnaD and CD3653. Domain nomenclature according to Marston *et al.* [19]. 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*.

A similar argument can be made for the putative helicase loader. Two homologues of the *Bs*DnaI protein were identified in *C. difficile* by BLASTP (CD3654: *e*-value = 8×10^{-10}, identity = 26%, query coverage = 46%; CD0410: *e*-value = 1×10^{-18}, identity = 31%, query coverage = 51%). The sequence homology of putative loader proteins with their counterpart in *B. subtilis* is mainly confined to the C-terminal AAA+ domain that contain the Walker A and B motifs [40]. CD0410 is located on the conjugative transposon CTn*2* [41], and therefore not part of the core genome of *C. difficile*. 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* [37], suggesting a functional relationship between the loader ATPase and a DnaB_2 family protein. Indeed, it has been suggested that *Bs*DnaB is a co-loader of the *Bs*DnaC helicase [18]. Taken together, our analyses strongly suggest that CD3654 is the cognate loader protein for the *C. difficile* replicative helicase (CD3657).

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 [2]. Therefore, we purified recombinant *C. difficile* helicase protein and deter-

mined its multimeric state using analytical gel-filtration. At concentrations below 5 μM we observed predominantly monomeric protein, with a fraction of the protein forming low molecular weight (MW) complexes (probably dimers or trimers) while at 10 μM and above, the helicase formed hexamers (**Figure 2**). 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 (*data not shown*).

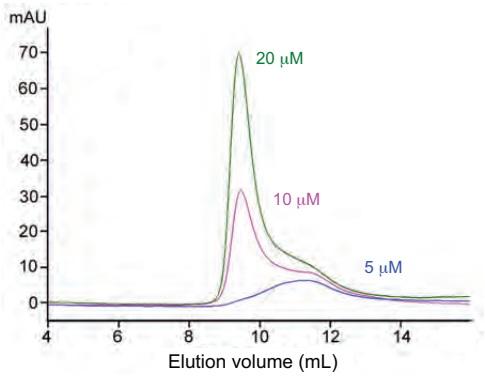


Figure 2. The helicase CD3657 demonstrates concentration dependent hexamerisation.

Analytical gel- filtration was performed in buffer B (see Methods) on a HiLoad 10/300 GL Superdex 200 analytical grade size exclusion column with the indicated concentration of CD3657 protein.

To confirm the gel-filtration data, we investigated the self-interaction of helicase in a bacterial two-hybrid system based on Gateway cloning [42]. This system detects interactions between a protein fused to Zif (Zinc-finger DNA binding domain) and a protein fused to the RNA polymerase ω subunit. Interaction between proteins of interest facilitates transcriptional activation of a Zif-dependent *lacZ* reporter gene in a dedicated reporter strain [42]. In order to quantify the interaction, *E. coli* cells containing the plasmids encoding the fusion proteins were permeabilized and assayed for β-galactosidase activity [43]. We found that in the reporter strain transformed with plasmids harbouring both fusion proteins, β-galactosidase activity was ~3-fold higher than for the reporter strains harbouring the individual plasmids, indicating a clear self-interaction for the *C. difficile* helicase protein (**Figure 3A**).

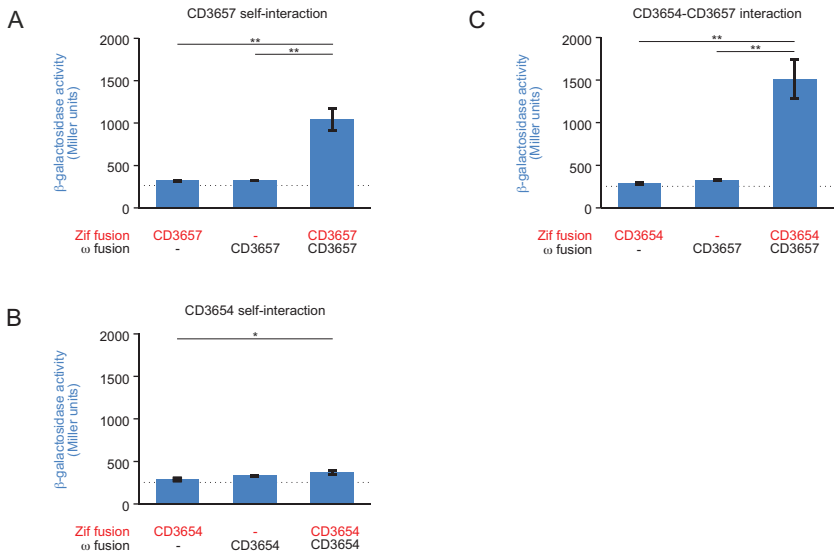


Figure 3. Bacterial two-hybrid analysis of the helicase CD3657 and putative loader protein CD3654.

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. Bars indicate average values and error-bars indicate standard deviation of the measurements (n=3). Dashed line indicates the maximum background level of β-galactosidase expression observed in our experimental set-up. Significance was determined using the Student's t-test (* $p < 0.05$, ** $p < 0.001$).

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 (*data not shown*). Consistent with this observation, no self- interaction of CD3654 was found in the bacterial two-hybrid system (**Figure 3B**).

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

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 ω subunit and CD3654 was fused to Zif [42]. The β-galactosidase activity in the reporter strain containing both plasmids was ~5 -fold increased (p <0.01) compared to the reporter strains containing the individual plasmids (background) (**Figure 3C**). The high β-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 ω subunit (*data not shown*). 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 μM) in the presence and absence of ATP (1 mM). In the presence of ATP, the elution profile showed a major high molecular weight (MW) peak (~10 mL; ~500 kDa, P1) and a minor low MW peak (~15 mL; ~40 kDa, P2) (**Figure 4**, red profile). In combination with a visual inspection of the fractions collected from both peaks on a 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), whilst the minor peak corresponds to predominantly free monomeric CD3654 (theoretical MW 38 kDa). Similar results were obtained when high concentrations of proteins (~10 μM) were used (**Supplemental Figure 1**), suggesting that pre-formed hexameric helicase retains the ability to interact with the CD3654 protein at the same stoichiometry.

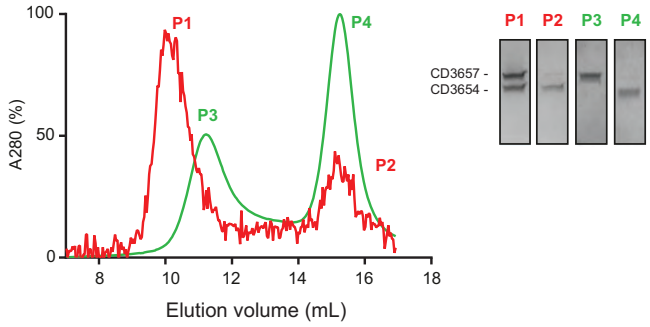


Figure 4. The helicase CD3657 and the putative loader CD3654 interact in an ATP-dependent manner.

Analytical gel filtration was performed on a HiLoad 10/300 GL Superdex analytical grade size exclusion column with 2.21 µM (monomer) of CD3657 and CD3654 in the presence (red) and absence (green) of 1 mM ATP. Inset shows a Coomassie-stained SDS-PAGE gel of the numbered peak fractions.

The elution profile of the same concentration of proteins in the absence of ATP showed a completely different picture (**Figure 4**, green profile). A minor peak was observed at ~11 mL (~300 kDa, P3) and a major second peak eluted at ~15 mL (~40 kDa, P4). Molecular weight estimates, in combination with an evaluation of peak fractions on an SDS-PAGE gel, indicated that the first peak most likely corresponds to a complex of six monomers of CD3657 (theoretical MW 297 kDa), whilst the second peak corresponds to monomeric CD3654 protein. Together, the data shows that the CD3657 helicase and the putative loader CD3654 can form a complex in an ATP-dependent manner.

Mutation of the helicase Walker A motif abrogates protein-protein interactions

To address the question which of the proteins (or whether both) requires 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 [25]. The motif is highly conserved in both helicase and helicase loader

proteins [6]. The lysine residue (K) forms a direct interaction with the negatively charged nucleotide β or γ 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 an Mg^{2+} ion within the ATP-binding site, which in turn coordinates the phosphate groups of ATP. In the *Geobacillus stearothermophilus* replicative helicase, mutation of the threonine residue results in a protein that lacks ATPase and unwinding activities [44]. Based on this knowledge, the equivalent residues were identified in 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). We performed filter-binding assays to determine if the mutant helicase proteins demonstrated altered ATP binding. We found that the K214R mutant bound ATP 2-fold less than wild type, whereas the T215A mutant demonstrated 5-fold less ATP binding at a concentration of 40 nM ATP; at lower ATP concentrations, the difference was even more pronounced (**Figure 5**).

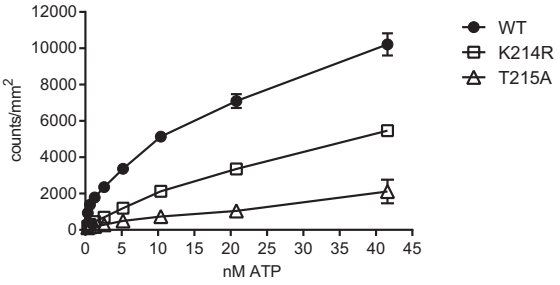


Figure 5. Walker A mutants of the helicase CD3657 show reduced binding of ATP.

2 pmol mM of CD3657 protein was incubated with the indicated amount of α-32P -radiolabelled ATP. The amount of radioactivity that remained associated with the protein was determined by autoradiography. CD3657 K214R and T215A are Walker A mutants.

To determine whether these CD3657 proteins showed altered protein-protein interactions, we performed bacterial two-hybrid experiments. We fused the CD3657 protein to the ω 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 6C** and **D**). This prompted us to investigate if the mutant helicases still had the capacity to homo-multimeric assemblies, as it has been shown in other bacteria that oligomerization is an important step in the mechanism of action of DNA helicases [45]. We found that the CD3657 K214R also demonstrated reduced (probably absent) self-interactions (**Figure 6A**), whereas the CD3657 T215A mutant had completely lost the ability to self-interact (**Figure 6B**). 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.

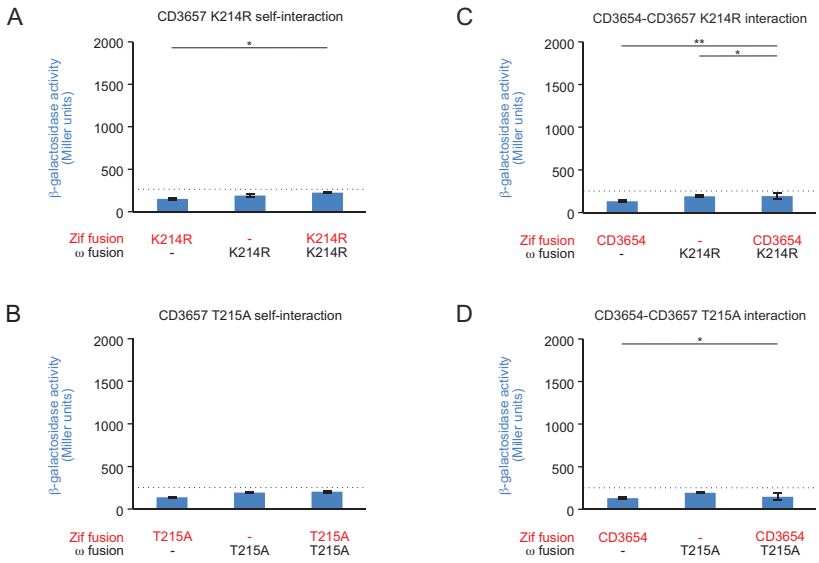


Figure 6. Walker A mutants of CD3657 are defective in protein-protein interactions.

Walker A mutants of CD3657 K214R (**A**);
T215A (**B**) no longer self-interact in a bacterial two-hybrid assay. Walker A mutants of CD3657 K214R (**C**);
T215A (**D**) show no or severely reduced interactions with the putative loader protein CD3654 in a bacterial two-hybrid assay. Bar graphs (A-D) indicate average values and error-bars indicate standard deviation of the measurements (n=3). Dashed line indicates the maximum background level of β-galactosidase expression observed in our experimental set-up. Significance was determined using the Student's t-test (* $p < 0.05$, ** $p < 0.001$).

To confirm the findings from the bacterial two-hybrid experiments, we purified CD3657 T215A protein and subjected it to size exclusion chromatography (**Figure 7**, blue line). *C. difficile* helicase T215A mutant (2.43 µM) was incubated in the presence of ATP (1 mM) and loaded onto a size exclusion column. A major peak was observed at ~15 mL (~40 kDa), probably corresponding to monomeric CD3657 T215A (theoretical MW 49 kDa) (**Figure 7**, P2 SDS-PAGE analysis). We did not observe any high MW complexes under these conditions, in contrast to the wild-type CD3657 protein (**Figure 2** and **4**). Next, we combined the CD3657 T215A mutant and the wild-type CD3654 protein (both 2.43 µM) in the presence of ATP (1 mM) (**Figure 7**, green line). A single peak was observed at ~15 mL (~40 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 kDa, respectively) (**Figure 7**, P1 SDS-PAGE analysis). Also, in these experiments no high MW complexes were found, in contrast to the wild-type CD3657 and CD3654 proteins (**Figure 4**).

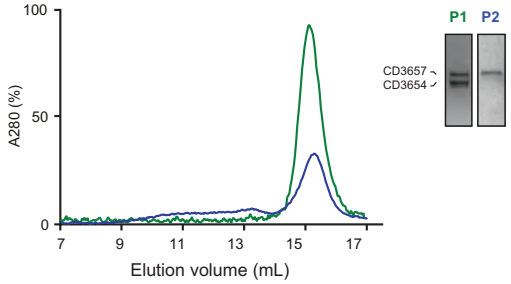


Figure 7. CD3657 T215A no longer has the ability to self-interact or interact with the putative loader CD3654.

Analytical gel filtration was performed in buffer A (see Methods) on a HiLoad 10/300 GL Superdex 200 analytical grade size exclusion column with 2.43 µM of CD3657 T215A in the presence (green) and absence (blue) of 2.43 µM CD3654. Inset shows a Coomassie-stained SDS-PAGE gel of the peak fractions.

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 (**Supplemental Figure 2**). Self-interaction was not observed for the

mutant loader proteins in a bacterial two-hybrid assay, concordant with the results for wild-type loader protein (**Supplemental Figure 3**). For this reason, we did not investigate the loader mutants further in this study.

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 the two proteins is most likely the result of ATP binding to CD3657 and not to CD3654.

Helicase loading of *C. difficile* differs from *B. subtilis*

So far, our data shows that the replicative helicase of *C. difficile*, CD3657, interacts in an ATP- dependent manner with the putative 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 [17,46]. Therefore, we set out to investigate if the activity of *C. difficile* helicase could be reconstituted in the presence of CD3654 protein in a similar experiment. DNA-unwinding helicase activity was assayed by monitoring and quantifying the displacement of a radiolabelled oligonucleotide (partially) annealed to single-stranded circular M13 DNA (**Figure 8A**). 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 to the *B. subtilis* proteins, helicase activity was not observed during this time course (**Figure 8B**). This may suggest that another factor is required for *in vitro* loading and/or activation of the *C. difficile* helicase.

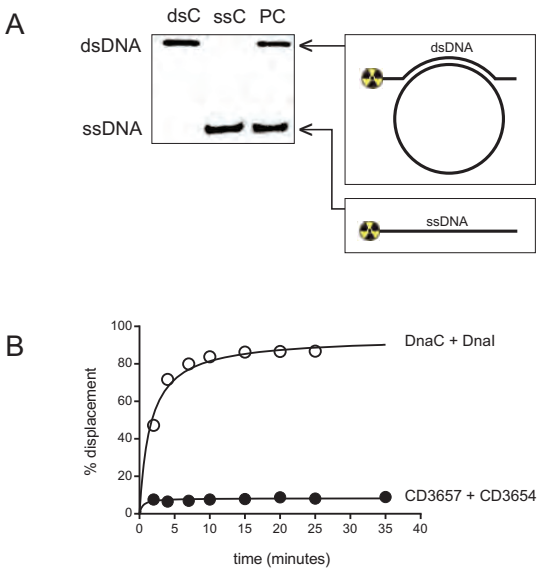


Figure 8. Helicase activity in the presence of helicase and (putative) loader proteins of *B. subtilis* and *C. difficile*.

A. Helicase activity was assayed by quantifying the displacement of a radiolabelled (γ^{32}P–ATP) oligonucleotide partially annealed to the single stranded circular DNA m13mp18. dsC: double-stranded control; ssC: single-stranded control; PC: positive control.

B. Percent displaced signal from the helicase assays in time for the *B. subtilis* proteins or *C. difficile* proteins.

Discussion

In silico analysis of the *C. difficile* genome by BLASTP identified homologues of most proteins that are involved in the DNA replication process of *B. subtilis*, which is generally considered the model for Gram-positive bacteria. However, *C. difficile* does not encode a *B. subtilis* DnaB homologue. This protein, together with *Bs*DnaD and *Bs*DnaI loader, is strictly required for helicase loading in *B. subtilis in vivo* [13,14,18]. A homologue of *Bs*DnaD was identified that may be involved in DNA replication (CD3653; *e*-value = 4×10^{-5}), although query coverage (47 percent) and identity

(29 percent) 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, *Bs*DnaB and *Bs*DnaD are structural homologues [19]. 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 [15,39]. 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 [19], and the proposed hybrid function of phage proteins is based on limited local amino acid sequence similarity in the DDBH2 domain only [39]. CD3653 on the other hand has a three-domain structure with a single DDBH1 and two DDBH2 domains, like DnaB, despite the lack of sequence similarity to this protein (**Figure 1A**). 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 [15,18,47].

DDBH2 domains are characterized by an YxxxIxxxW motif [19]. In *Bs*DnaB, this motif is degenerate in the first DDBH2 domain. By contrast, this motif is readily identified in both DDBH2 domains of CD3653 (**Figure 1A**). 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 the nomenclature is based on the *E. coli* protein name) belong to the superfamily 4 of DNA helicases (SF4), and the functional unit of this protein is a hexamer [25,28,29,31]. In *E. coli*, the helicase is found to be a stable hexamer over a broad protein concentration range of 0,1 to 10 μM [48] and is active as a pre-formed multimer. Helicases belonging to the ring-maker class, such as *B. subtilis*, can occur in low-oligomeric or monomeric state under dilute conditions [2,18]. Our experiments indicated that CD3657 is monomeric in the low micromolar or nanomolar range (**Figure 2**), which is likely to be reflective of the intracellular concentration of protein [49]. *Clostridium difficile* CD3657 can form hexameric assemblies at higher concentrations (**Figure 2** and **Supplemental Figure 1**), but these pre-formed hexamers are inactive (*data not shown*), 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 [2].

Though the addition of ATP was not strictly required for hexamerization of CD3657 at high concentrations of protein (**Figure 2**), the interaction between putative helicase and loader protein was found to be ATP-dependent (**Figure 4**). Mutations in the conserved Walker A and Walker B motifs of the putative loader protein did not abrogate the interaction with the wild-type helicase (**Supplemental Figure 2**). Similarly, in *E. coli*, nucleotide binding to the helicase loader was not a prerequisite for association with helicase [10,45,49]. Instead, our data indicate that association of ATP with helicase is crucial for the interaction with the loader protein (**Figure 6** and **7**). Notably, there is a correlation between the ability of the helicase interact with loader and to form homohexamers, as a T215A (Walker A) mutant of helicase is defective for both, at least under dilute concentrations of helicase (**Figure 6** and **7**). By contrast, the equivalent mutation in *G. stearothermophilus* helicase (T217A) does not affect its ability to form hexamers [44], and the interaction of this protein with *B. subtilis* DnaI readily occurs in the absence of ATP [40].

Both Walker A mutants of CD3657 demonstrate similar effects on the protein-protein interactions, that we attribute to defects in ATP binding rather than hydrolysis. Both mutants show reduced binding of ATP (**Figure 5**); a Walker B mutant (CD3657 D318A), mirrors our findings with the Walker A mutants in a bacterial two-hybrid assay. But a Walker B mutant that is predicted to be able to bind ATP but not hydrolyse (CD 3657 E239A) does not (*our unpublished observations*).

Binding of the nucleotide to helicase is associated with conformational changes; the N-terminal collar domain constricts upon nucleotide binding in *Aquifex aeolicus* and to a lesser extent *E. coli*. This constricted conformation is believed to favour an interaction with the loader protein [50] 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 a helicase loader (**Figure 1** and **4**), we did not observe helicase activity in an *in vitro* assay with purified helicase alone (*data not shown*) or in combination with the loader protein (**Figure 8B**). These findings are in contrast with two other Gram-positive replicative helicases; *G. stearothermophilus* helicase demonstrates significant helicase activity by itself [44], and the *B. subtilis* helicase is strongly activated by its cognate loader [46]. However, genes encoding homologues of CD3654 (R20291_3513)

and CD3657 (R20291_3516) were found to be essential in an epidemic strain, supporting their identification as DNA replication initiation proteins [51]. Together, these data strongly suggest that the presence of loader protein alone is not sufficient to activate the helicase in *C. difficile*, and that at least one other factor is needed to reconstitute its activity.

Helicases are complex proteins, and their properties can both alter and be altered by other replication factors [50]. 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 domain. The helicase loader interacts with the C-terminal ATPase domain [49,52,53], and the same domain is required for the interaction with the τ subunit of the clamp loader protein in *E. coli* [54], and *B. subtilis* [55,56]. Strikingly, *B. subtilis* helicase T217A Walker mutant fails to form a complex with τ [55]. 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* [50,53]. 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 [50,57]. 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 [58-60]. Primase binding counteracts the binding of the loader protein in *E. coli* [61]. Similarly, helicase loader protein from *B. subtilis* was found to dissociate from the complex when primase and polymerase bind to helicase [46]. It is unknown if *C. difficile* primase can exert a similar function.

In summary, our data show that the mechanisms of loading and activation of the replicative helicase of *C. difficile* are likely critically different from the Gram-positive model *Bacillus subtilis*. It is tempting to speculate that CD3653, primase or other protein factors will allow reconstitution of helicase activity *in vitro*.

Materials and Methods

Plasmid construction

All oligonucleotides and plasmids constructed for this study are listed in **Supplemental Table 1** and **2**. To construct the helicase loader (CD3654) and helicase (CD3657) expression plasmids, the open reading frames were amplified with high fidelity polymerase *Pfu* via PCR from *C. difficile* strain 630Δ*erm* chromosomal DNA [62,63], using primers oEVE-4 and oEVE-6 for CD3654 and oWKS-1185 and oWKS-1367 for CD3657. The reverse primer of both genes introduces a stop codon before the *Xho*I site, thereby ensuring that the protein is in its native form, when expressed (no C-terminal 6x His-tag). The CD3654 PCR product was digested with *Nco*I and *Xho*I restriction nucleases and ligated into vector pET28b (Novagen) to yield pEVE-24. The CD3657 PCR product was digested with *Nde*I and *Xho*I and ligated into vector pET21b (Novagen) to yield pEVE-87. The DNA sequence of the constructs (pEVE-24 and pEVE-87) was verified by sequencing.

Construction of the plasmids for the bacterial two-hybrid system was performed with Gateway cloning technology (Invitrogen), which is based on phage λ site- specific recombination. To construct the CD3654 and CD3657 Entry plasmids, the CD3654/CD3657 open reading frame was amplified with high fidelity polymerase *Pfu* via PCR from *C. difficile* strain 630Δ*erm* chromosomal DNA, using primers oAF-26 and oAF-27 for CD3654 and oAF-28 and oAF-29 for CD3657. This resulted in attB-flanked PCR products (1 µl) that could be recombined into donor vector pDonR™201 (1 µl, 50 ng/ µl) with BP Clonase II enzyme mix (0.5 µl). The reaction was incubated at 25°C for 1.5 hours and transformed into chemically competent *E. coli* DH5α cells by heat shock. After overnight incubation on LB plates at 37˚C, kanamycin resistant colonies were selected. 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 (ω fusion plasmid) in an LR reaction [42]. In brief, the Entry clones (1 µl, 50 ng/ µl) were mixed with one of the pKEK1286 or pKEK1287 (1 µl, 50 ng/ µl) destination vectors and LR Clonase II enzyme mix (0.5 µl). After the reaction was incubated at 25°C for 1.5 hours, the formulation was transformed as a whole into chemically competent *E. coli* DH5α cells by heat shock. Resulting Expression clones were selected with tetracycline for the Zif fusion plasmid

or ampicillin for the ω fusion plasmid. The DNA sequence of all constructs (pEVE-122, pEVE123, pEVE124 and pEVE125) was verified by sequencing.

Site-directed mutagenesis

Walker A mutant constructs of CD3657 and CD3654 mutants were constructed according to the QuikChange protocol (Stratagene). Primers were generated with Primer X, a web-based tool for automatic design of mutagenic primers for site-directed mutagenesis. QuikChange was carried out using Pfu polymerase and plasmids pEVE-24, pEVE-87, pEVE-122, pEVE123, pEVE124 and pEVE125 as templates. All mutant constructs were verified for the correct mutation by DNA-sequencing.

Purification of the helicase CD3657

Overexpression *C. difficile* CD3657 was carried out in *E. coli* BL21 (DE3) from the pEVE-12 plasmid. The growth medium consisting of 2XYT broth (1.2 L), carbenicillin (50 µg/mL) and antifoam 204 (Sigma-Aldrich) was inoculated with a pre-culture (10 mL). The cell culture was incubated at 37°C with mechanical shaking at 180rpm, until an optical density (600 nm) of 0.70-0.85 was reached (after approximately 3 h). Protein expression was induced via the addition of IPTG (1 mM) and the culture was incubated at 30°C for 3 h. The cells were harvested by centrifugation (3000 g, 15 min, 4°C) and resulting cell paste was stored at -80°C. CD3657 cell paste, prepared from 1.2 L cell culture, was re-suspended in 30 mL TED50 buffer (Tris pH 7.5 50 mM, EDTA 1 mM, DTT 1 mM, NaCl 50 mM) with PMSF (1 mM). The bacterial cells were lysed by sonication and crude lysate was clarified by centrifugation (35,000 g, 30 min, 4°C). The resulting supernatant was separated from the cell debris using a 0.22 µm pore filter before a 50% ammonium sulphate precipitation, followed by clarification by centrifugation (35,000 g, 30 min, 4°C). The ammonium sulphate precipitated pellet was suspended in TED50 buffer and loaded onto a 5mL Q sepharose column, equilibrated in TED50 buffer. The protein was eluted using a gradient of 20 to 100% TED1000 buffer (Tris pH 7.5 50 mM, EDTA 1 mM, DTT 1 mM, NaCl 1000 mM) over 15 CV. The fractions containing the protein of interest were pooled (12 mL, 25 mS) and diluted with 33 mL of TED50 buffer (Tris pH 7.5 50 mM, EDTA 1 mM, DTT 1 mM) to give an adjusted volume of 45 mL and conductivity of

10.1 mS. This solution was loaded onto a 5 mL heparin column equilibrated in TED50 buffer; the protein eluted in the flow through. After a further ammonium sulphate precipitation to concentrate the sample, the collected protein was loaded onto a Hiload 26/60 Superdex 200 gel filtration column equilibrated in TED50 buffer, yielding hexameric CD3657 oligomer. A further ammonium sulphate precipitation was used to concentrate the sample in a reduced volume of 4mL in TED50 buffer. Guanidinium chloride solution (8 M) was added portion wise (30 x 44.4 µL) to the protein solution with rapid stirring, to give a final concentration of 2 M guanidinium chloride. The protein was then loaded onto a Hiload 26/60 Superdex 200 gel filtration column equilibrated in TED50-GC buffer (Tris pH 7.5 50 mM, EDTA 1 mM, DTT 1 mM, NaCl 50 mM, guanidinium chloride 2 M), to give the CD3657 monomer. The protein was collected and the buffer was exchanged by dialysis against 2 L storage buffer (Tris pH 7.5 50 mM, EDTA 1 mM, DTT 1 mM, NaCl 50 mM, glycerol 20% v/v) for 18 h at 4°C. A second dialysis step was performed against 1 L storage buffer for 2 h at 4°C, to remove any remaining guanidinium chloride. The protein was quantified by UV spectrophotometry and stored at -80°C. The mutant proteins CD3657 K214R and CD3657 T215A were prepared in an identical manner. Expression of CD3657 E239A required 0.5mM IPTG induction with 16 h expression at 25°C. The purification protocol for CD3657 E239A was the same as described above for the other helicase proteins. Protein purity (all >95%) was estimated by SDS-PAGE electrophoresis and concentration was determined spectrophotometrically using extinction coefficients calculated using the ExPASy ProtParam tool (http://web.expasy.org/protparam).

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

Purification of the putative loader protein CD3654

C. difficile CD3654 was expressed from the pEVE-24 plasmid in *E. coli* BL21 (DE3). The growth medium consisting of 2xYT broth (1 L), kanamycin (30 µg/mL) and antifoam 204 (Sigma-Aldrich) was inoculated with a pre-culture (10 mL). The cell culture was incubated at 37°C with mechanical shaking, until an optical density (600 nm) of 0.62-0.65 was reached (after approximately 3 h). Protein expression was induced via the addition of IPTG (1 mM) and the culture was incubated at 30°C for 3 h. The cells were harvested by centrifugation (3000 g, 15 min, 4°C and

the resulting cell paste was stored at -80°C. The bacterial cell paste, prepared from 1L cell culture, was suspended in 25 mL TED50 buffer with PMSF (1 mM) and protease inhibitor cocktail (100 µL). The cells were lysed by sonication, clarified by centrifugation (40,000 g, 30 min, 4°C) and resulting supernatant separated from the cell debris using a 0.22 µm pore filter. Ammonium sulphate (7.32 g) was added slowly to the supernatant (25 mL) with stirring at 4°C, to achieve 49% saturation, the mixture was stirred for 10 min to give a white suspension. The protein pellet was collected by centrifugation (40,000 g, 30 min, 4°C) and washed with TED20 buffer (Tris pH 7.5 50 mM, EDTA 1 mM, DTT 1 mM, NaCl 20 mM) (2 x 4 mL). The precipitate was suspended in TED20 buffer (15 mL) with gentle mechanical shaking (30 min, 4°C). The suspension buffer was exchanged by dialysis against 1 L TED20 buffer for 2 h at 4°C, giving a solution with conductivity of 9.5 mS. The protein solution (~15 mL) was loaded onto combined 5 mL Q sepharose and 5 mL SP sepharose columns connected in series and equilibrated in TED20 buffer, the protein of interest eluted in the flow through. The collected protein was loaded onto a 5 mL heparin sepharose column equilibrated in TED20 buffer and eluted with a step to 15% TED1000 buffer. The collected protein was loaded onto a Hiload 26/60 Superdex 200 gel filtration column equilibrated in storage buffer (Tris pH 7.5 50 mM, EDTA 1 mM, DTT 1 mM, NaCl 50 mM, glycerol 10% v/v). The protein was quantified by UV spectrophotometry and stored at -80°C. The mutant proteins CD3654 K198R and CD3654 T199A were prepared in an identical manner. Protein purity (all >95%) was estimated by SDS-PAGE electrophoresis and concentration was determined spectrophotometrically using extinction coefficients calculated using the ExPASy ProtParam tool (http://web.expasy.org/protparam). Protein concentrations mentioned in this manuscript refer to concentration of the monomer of the protein.

ATP binding assay

The ATP binding assay [64] was performed in a 40 µl reaction containing 2 pmol wild type or mutant CD3657 protein, 0,65-41,6 nM [α-32P]ATP (3000Ci/mmol; Perkin Elmer) and ATP binding buffer (50 mM Tricine-KOH (pH 8.25), 0.5 mM magnesium acetate, 1 µM EDTA, 7 mM DTT, 0,007% Triton X-100 and 5% glycerol). The reaction was incubated for 15 min at room temperature. The samples were transferred to a Bio-Dot apparatus (BioRad) and passed through a 0.45 µm nitrocellulose membrane pre-soaked in wash buffer (50 mM Tricine-KOH (pH 8.25), 0.5 mM

magnesium acetate, 1 µM EDTA, 5 mM DTT, 10 mM ammonium sulphate, 0.005% Triton X-100 and 5% glycerol). The membrane was washed with cold buffer, first in the Bio-Dot apparatus and subsequently after removal from the Bio-Dot apparatus. The air-dried membrane was exposed to a storage phosphor screen for one hour. Binding was quantified by scanning the screen with the Typhoon 9410 imager and using QuantityOne software. Reactions without protein provided a background value that was subtracted.

Gel-filtration experiments

Self-interaction of the CD3657 and CD3654 proteins were studied in the presence and absence of ATP. In brief, purified CD3657 (or mutant) or CD3654 (or mutant) was incubated for 10 min at room temperature with MgCl2 (2 mM) in their storage buffer and ATP (1 mM). The mixture (500 µL) was loaded onto a Hiload 10/300 GL Superdex 200 analytical grade size exclusion column equilibrated in buffer A (with ATP: Tris pH 7.5 50 mM, EDTA 1 mM, DTT 1 mM, glycerol 10% v/v, MgCl2 2 mM, ATP 1 mM) or buffer B (Tris pH 7.5 50 mM, EDTA 1 mM, DTT 1 mM, glycerol 10% v/v, MgCl2 2 mM) at a flow rate of 0.5 mL/min. The elution profiles from each experiment were monitored at 280 nm and plotted as a function of the elution volume. Samples from fractions were analysed by SDS-PAGE and Coomassie Blue staining to verify the identity of the proteins. To assess interactions between CD3657 and CD3654, purified proteins were mixed in a 1:1 stoichiometry in the presence of MgCl2 (2 mM) and ATP (1 mM) and incubated for 10 min at room temperature. The mixture (500 µL) was loaded onto a Hiload 10/300 GL Superdex 200 analytical grade size exclusion column equilibrated in buffer A (Tris pH 7.5 50 mM, EDTA 1 mM, DTT 1 mM, glycerol 10% v/v, MgCl2 2 mM, ATP 1 mM) at a flow rate of 0.5 mL/min. For the experiments without ATP, buffer B was used (Tris pH 7.5 50 mM, EDTA 1 mM, DTT 1 mM, glycerol 10% v/v, MgCl2 2 mM).

Bacterial two-hybrid assays

To determine (self-)interaction, both expression constructs were subsequently transformed in to the *E. coli* reporter strain KDZif1ΔZ[65]. In order to control for background due to (possible) differences in expression of the constructs, single

expression plasmids (Zif- or ω fusion) containing the gene of interest were transformed into the reporter strain. After overnight incubation at 30˚C, three colonies per assay were cultured overnight in LB broth (30°C) in the presence of 1mM IPTG, tetracycline (selects Zif fusion plasmid) and/or ampicillin (selects ω fusion plasmid). Bacterial cells were permeabilized with SDS and chloroform and assayed for β- galactosidase activity according to the method of Miller [43]. In short, cells were diluted in Z buffer (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, 50 mM β-mercaptoethanol, pH7) to 1 ml and permeabilized with 50 µl 0,1% w/v SDS and 100 µl chloroform. After 5 minutes of equilibration, 200 µl of o-nitrophenyl-ß-D-galactopyranoside was added to each tube and incubated at room temperature until yellow colour developed. The reaction was stopped with 0,5 ml 1M Na_2CO_3 and measured at OD_{420} and OD_{550} to calculate the β- galactosidase activity in Miller Units. Experiments were performed in triplicate.

Helicase assays

Helicase activity was assayed by monitoring (and quantifying) the displacement of a radiolabelled (γ^{32}P-ATP) oligonucleotide oVP-1 (partially) annealed to the single stranded circular DNA m13mp18 (ssM13; Affymetrix) essentially as previously described [46]. In short, the 105-mer oligonucleotide was radiolabelled at the 5'end using γ^{32}ATP and T4 polynucleotide kinase (New England Biolabs) and subsequently purified through an S-200 mini-spin column (GE Healthcare). All reactions, containing 0.658 nM radiolabelled DNA substrate, were initiated by the addition of 2.5 mM ATP and carried out at 37˚C in buffer containing 20mM HEPES-NaOH (pH 7.5), 50 mM NaCl, 10 mM $MgCl_2$ and 1 mM DTT for various times. The reactions were terminated by adding 5x SDS-STOP buffer (100mM Tris pH8.0, 200mM EDTA, 2.5% (w/v) SDS, 50% (v/v) glycerol, 0.15% (w/v) bromophenol blue).

To investigate the effect of the putative helicase loader (CD3654) on the activity of the helicase (CD3657), the proteins were mixed in equimolar concentrations (1µM) and incubated for 10 minutes at 37˚C prior to adding reaction buffer. The buffer with CD3657 was preincubated for 5 mins before adding CD3654, incubated for 5 more mins after which the reaction was initiated with 2.5mM ATP (final concentration). Stop buffer was added to terminate the reactions (1% v/w SDS, 40 mM EDTA, 8% v/v glycerol, 0.1% w/v bromophenol blue). Reaction samples

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

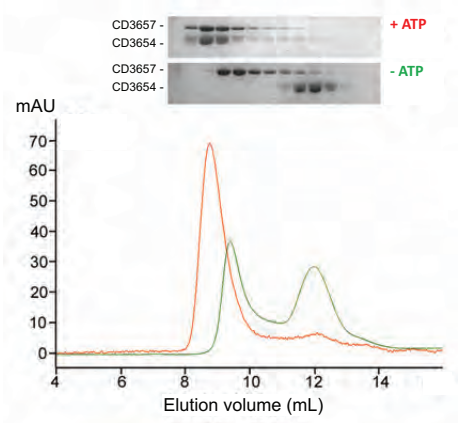
Supplemental information

The replicative helicase CD3657 of *Clostridium difficile* interacts with the putative loader protein CD3654

Erika van Eijk [1], Annemieke Friggen [1], Vasileios Paschalis [2], Matthew Green [2], Geoffrey S. Briggs [2], James Gibson [2], Panos Soultanas [2], Wiep Klaas Smits [1]

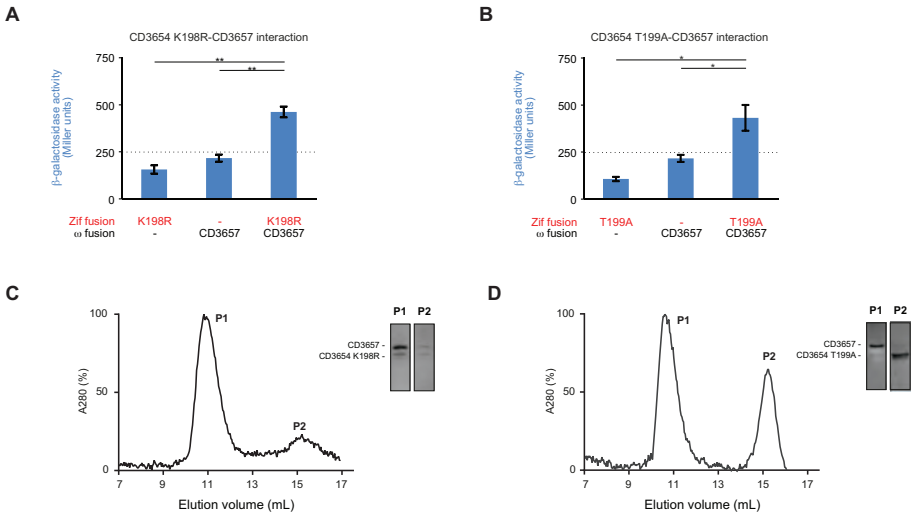
1 Department of Medical Microbiology, Leiden University Medical Center, Leiden, the Netherlands;

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



Supplemental Figure 1. ATP dependent interaction of the helicase CD3657 and the putative loader CD3654 at high concentrations of proteins.

The helicase CD3657 and the putative loader CD3654 interact in an ATP-dependent manner. Analytical gel filtration was performed in the presence (red) of absence (green) of 1mM ATP on a Hiload 10/300 GL Superdex 200 analytical grade size exclusion column. Inset shows a Coomassie-stained SDS-PAGE gels of of sampled fractions taken during both gel filtration experiments.



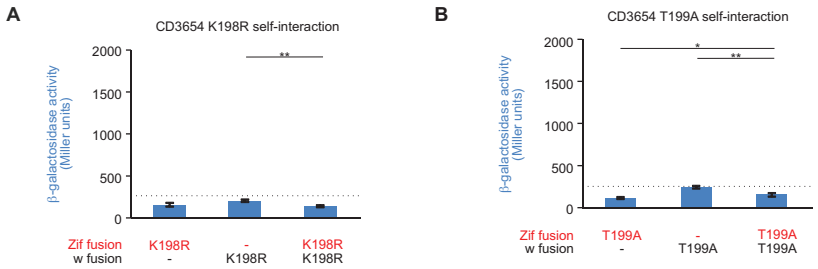
Supplemental Figure 2. Walker A mutants of the putative loader protein CD3654 retain the ability to interact with CD3657.

A. Bacterial two hybrid analysis of the interaction between CD3657 and CD3654 K198R.

B. Bacterial two hybrid analysis of the interaction between CD3657 and CD3654 T199A. Bar graphs in A–B indicate average values and error-bars indicate standard deviation of the measurements (n=3). Dashed line indicates the maximum background level of β-galactosidase expression observed in our experimental set-up. Significance was determined using the Student's t-test (* $p < 0.05$, ** $p < 0.001$).

C. Analytical gel filtration analysis of the interaction between CD3657 and CD3654 K198R.

D. Analytical gel filtration analysis of the interaction between CD3657 and CD3654 T199A. Analytical gel filtration was performed in buffer B (see Methods) with 3.10 uM proteins in the presence of 1mM ATP. Inset in C–D shows Coomassie-stained SDS-PAGE gels of the numbered peak fractions.



Supplemental Figure 3. Walker A mutants of CD3654 are monomeric.

A. Bacterial two-hybrid analysis of the self-interaction of CD3654 K198R.

B . Bacterial two hybrid analysis of the self-interaction of CD3654 T199A.
Bar graphs in A-B 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).

Supplemental Table 1. Oligonucleotides used in this study

Name	Sequence (5' – 3')	Description
oAF-26	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGAATGAGGATAAAATAAGAAAAATAC	Forward CD3654 (B2H) ★
oAF-27	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTTAAATCTTTCCCATCTCAAATCATC	Reverse CD3654 (B2H)
oAF-28	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGAAGATATGACGAGAATTCCTC	Forward CD3657 (B2H)
oAF-29	GGGGACCACTTTGTACAAGAAAGCTGGGTCTAGGTCTCTCAACTTATCCCC	Reverse CD3657 (B2H)
oEVE-4	GCTTCCATGGATGAGGATAAAATAAGAAAAATACTTGC	Forward CD3654 pET
oEVE-6	GACGCTCGAGTTATTTAAATCTTTCCCATCTCAAATCATCTCC	Reverse CD3654 pET
oWKS-1185	GTGATTCATATGGAAGATATGACGAG	Forward CD3657 pET
oWKS-1367	CCGCTCGAGTTATAGGTCTCTCAACTTATCCCC	Reverse CD3657 pET
oWKS-1272	GGTTCTACTGGACTAGGAAGGACCTATATGTGCAATTG	Forward CD3654 K198R (QC) ★★
oWKS-1273	CAATTGCACATATAGGTCCTTCCTAGTCCAGTAGAACC	Reverse CD3654 K198R (QC)
oWKS-1274	GGTTCTACTGGACTAGGAAAGGCCTATATGTGCAATTGTATTG	Forward CD3654 T199A (QC)
oWKS-1275	CAATACAATTGCACATATAGGCCTTTCCTAGTCCAGTAGAACC	Reverse CD3654 T199A (QC)
oEVE-69	CTTGTTTGATTGTGATTTATTAATAATACAAGACCTTGGAAC	Forward CD3654 D258Q (QC)
oEVE-70	GTTCCAAGGTCTTGTATTATTAATAAATCACAATCAAACAAG	Reverse CD3654 D258Q (QC)
oWKS-1276	GCTAGACCAGCAATGGGTAGAACTGCCTTTTCTTTAAAC	Forward CD3657 K214R (QC)
oWKS-1277	GTTTAAAGAAAAGGCAGTTCTACCCATTGCTGGTCTAGC	Reverse CD3657 K214R (QC)
oWKS-1278	CTAGACCAGCAATGGGTAAAGCTGCCTTTTCTTTAAACTTGG	Forward CD3657 T215A (QC)
oWKS-1279	CCAAGTTTAAAGAAAAGGCAGCTTTACCCATTGCTGGTCTAG	Reverse CD3657 T215A (QC)
oVP-1	CACACACACACACACACACACACACACACACACACACACACACACACACACACACACACCCC CTTTAAAAAAAAAAAGCCAAAAGCAGTGCCAAGCTTGCATGCC	Helicase assay

★ Bacterial two-hybrid (B2H); ★★ QuikChange (QC)

Supplemental Table 2. **Plasmids used in this study**

Plasmid	Description	Reference
pEVE24	pET28b-CD3654	This study
pEVE59	pET28b-CD3654 K198R	This study
pEVE60	pET28b-CD3654 T199A	This study
pEVE-203	pET28b-CD3654-D258Q	This study
pEVE87	pET21b-CD3657	This study
pEVE90	pET21b-CD3657 K214R	This study
pEVE92	pET21b-CD3657 T215A	This study
pEVE118	pENTRY-CD3654	This study
pEVE120	pENTRY-CD3657	This study
pEVE122	pKEK1286-CD3654	This study
pEVE123	pKEK1287-CD3654	This study
pEVE124	pKEK1286-CD3657	This study
pEVE125	pKEK1287-CD3657	This study
pEVE167	pKEK1286-CD3654 K198R	This study
pEVE168	pKEK1286-CD3654 T199A	This study
pEVE169	pKEK1287-CD3654 K198R	This study
pEVE170	pKEK1287-CD3654 T199A	This study
pEVE171	pKEK1286-CD3657 K214R	This study
pEVE172	pKEK1286-CD3657 T215A	This study
pEVE173	pKEK1287-CD3657 K214R	This study
pEVE174	pKEK1287-CD3657 T215A	This study

References

1 Battistuzzi, F. U., Feijao, A. & Hedges, S. B. A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. *BMC evolutionary biology* **4**, 44, doi:10.1186/1471-2148-4-44 (2004).

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

3 Li, Y. & Araki, H. Loading and activation of DNA replicative helicases: the key step of initiation of DNA replication. *Genes to cells : devoted to molecular & cellular mechanisms* **18**, 266-277, doi:10.1111/gtc.12040 (2013).

4 Mott, M. L. & Berger, J. M. DNA replication initiation: mechanisms and regulation in bacteria. *Nature reviews. Microbiology* **5**, 343-354, doi:10.1038/nrmicro1640 (2007).

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

6 Soultanas, P. Loading mechanisms of ring helicases at replication origins. *Molecular microbiology* **84**, 6-16, doi:10.1111/j.1365-2958.2012.08012.x (2012).

7 Kaguni, J. M. Replication initiation at the Escherichia coli chromosomal origin. *Current opinion in chemical biology* **15**, 606-613, doi:10.1016/j.cbpa.2011.07.016 (2011).

8 Remus, D. & Diffley, J. F. Eukaryotic DNA replication control: lock and load, then fire. *Current opinion in cell biology* **21**, 771-777, doi:10.1016/j.ceb.2009.08.002 (2009).

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

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

11 Kornberg, A. & Baker, T. A. *DNA Replication*. (University Science Books, 2005).

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

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

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

15 Briggs, G. S., Smits, W. K. & Soultanas, P. Chromosomal replication initiation machinery of low-G+C-content Firmicutes. *Journal of bacteriology* **194**, 5162-5170, doi:10.1128/jb.00865-12 (2012).

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

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

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

19 Marston, F. Y. *et al.* When simple sequence comparison fails: the cryptic case of the shared domains of the bacterial replication initiation proteins DnaB and DnaD. *Nucleic acids research* **38**, 6930–6942, doi:10.1093/nar/gkq465 (2010).

20 Schneider, S., Zhang, W., Soultanas, P. & Paoli, M. Structure of the N-terminal oligomerization domain of DnaD reveals a unique tetramerization motif and provides insights into scaffold formation. *Journal of molecular biology* **376**, 1237–1250, doi:10.1016/j.jmb.2007.12.045 (2008).

21 Zhang, W., Allen, S., Roberts, C. J. & Soultanas, P. The Bacillus subtilis primosomal protein DnaD untwists supercoiled DNA. *Journal of bacteriology* **188**, 5487–5493, doi:10.1128/jb.00339-06 (2006).

22 Zhang, W. *et al.* The Bacillus subtilis DnaD and DnaB proteins exhibit different DNA remodelling activities. *Journal of molecular biology* **351**, 66–75, doi:10.1016/j.jmb.2005.05.065 (2005).

23 Zhang, W. *et al.* Single-molecule atomic force spectroscopy reveals that DnaD forms scaffolds and enhances duplex melting. *Journal of molecular biology* **377**, 706–714, doi:10.1016/j.jmb.2008.01.067 (2008).

24 Caruthers, J. M. & McKay, D. B. Helicase structure and mechanism. *Current opinion in structural biology* **12**, 123–133 (2002).

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

26 Mott, M. L., Erzberger, J. P., Coons, M. M. & Berger, J. M. Structural synergy and molecular crosstalk between bacterial helicase loaders and replication initiators. *Cell* **135**, 623–634, doi:10.1016/j.cell.2008.09.058 (2008).

27 Duderstadt, K. E., Chuang, K. & Berger, J. M. DNA stretching by bacterial initiators promotes replication origin opening. *Nature* **478**, 209–213, doi:10.1038/nature10455 (2011).

28 Erzberger, J. P. & Berger, J. M. Evolutionary relationships and structural mechanisms of AAA+ proteins. *Annual review of biophysics and biomolecular structure* **35**, 93–114, doi:10.1146/annurev.biophys.35.040405.101933 (2006).

29 Singleton, M. R., Dillingham, M. S. & Wigley, D. B. Structure and mechanism of helicases and nucleic acid translocases. *Annual review of biochemistry* **76**, 23–50, doi:10.1146/annurev.biochem.76.052305.115300 (2007).

30 Snider, J., Thibault, G. & Houry, W. A. The AAA+ superfamily of functionally diverse proteins. *Genome biology* **9**, 216, doi:10.1186/gb-2008-9-4-216 (2008).

31 Berger, J. M. SnapShot: nucleic acid helicases and translocases. *Cell* **134**, 888–888.e881, doi:10.1016/j.cell.2008.08.027 (2008).

32 Itsathitphaisarn, O., Wing, R. A., Eliason, W. K., Wang, J. & Steitz, T. A. The hexameric helicase DnaB adopts a nonplanar conformation during translocation. *Cell* **151**, 267–277, doi:10.1016/j.cell.2012.09.014 (2012).

33 Ilyina, T. V., Gorbalenya, A. E. & Koonin, E. V. Organization and evolution of bacterial and bacteriophage primase-helicase systems. *Journal of molecular evolution* **34**, 351–357 (1992).

34 Leipe, D. D., Aravind, L., Grishin, N. V. & Koonin, E. V. The bacterial replicative helicase DnaB evolved from a RecA duplication. *Genome research* **10**, 5–16 (2000).

35 Rupnik, M., Wilcox, M. H. & Gerding, D. N. Clostridium difficile infection: new developments in epidemiology and pathogenesis. *Nature reviews. Microbiology* **7**, 526–536, doi:10.1038/nrmicro2164 (2009).

36 Smits, W. K., Lyras, D., Lacy, D. B., Wilcox, M. H. & Kuijper, E. J. Clostridium difficile infection. *Nature reviews. Disease primers* **2**, 16020, doi:10.1038/nrdp.2016.20 (2016).

37 Bruand, C., Ehrlich, S. D. & Janniere, L. Primosome assembly site in Bacillus subtilis. *The EMBO journal* **14**, 2642–2650 (1995).

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

39 Rokop, M. E., Auchtung, J. M. & Grossman, A. D. Control of DNA replication initiation by recruitment of an essential initiation protein to the membrane of Bacillus subtilis. *Molecular microbiology* **52**, 1757–1767, doi:10.1111/j.1365-2958.2004.04091.x (2004).

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

41 Brouwer, M. S., Warburton, P. J., Roberts, A. P., Mullany, P. & Allan, E. 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 (2011).

42 Karna, S. L. *et al.* A bacterial two-hybrid system that utilizes Gateway cloning for rapid screening of protein-protein interactions. *BioTechniques* **49**, 831–833, doi:10.2144/000113539 (2010).

43 Miller, J. H. in *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory Press, 1972).

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

45 Biswas, E. E., Chen, P. H. & Biswas, S. B. Modulation of enzymatic activities of Escherichia coli DnaB helicase by single-stranded DNA-binding proteins. *Nucleic acids research* **30**, 2809–2816 (2002).

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

47 Grainger, W. H., Machon, C., Scott, D. J. & Soultanas, P. DnaB proteolysis in vivo regulates oligomerization and its localization at oriC in Bacillus subtilis. *Nucleic acids research* **38**, 2851–2864, doi:10.1093/nar/gkp1236 (2010).

48 Bujalowski, W., Klonowska, M. M. & Jezewska, M. J. Oligomeric structure of Escherichia coli primary replicative helicase DnaB protein. *The Journal of biological chemistry* **269**, 31350–31358 (1994).

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

50 Strycharska, M. S. *et al.* Nucleotide and partner-protein control of bacterial replicative helicase structure and function. *Molecular cell* **52**, 844–854, doi:10.1016/j.molcel.2013.11.016 (2013).

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

52 Chodavarapu, S. & Kaguni, J. M. Replication Initiation in Bacteria. *The Enzymes* **39**, 1–30, doi:10.1016/bs.enz.2016.03.001 (2016).

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

54 Gao, D. & McHenry, C. S. 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. *The Journal of biological chemistry* **276**, 4441–4446, doi:10.1074/jbc.M009830200 (2001).

55 Afonso, J. P. *et al.* Insights into the structure and assembly of the Bacillus subtilis clamp–loader complex and its interaction with the replicative helicase. *Nucleic acids research* **41**, 5115–5126, doi:10.1093/nar/gkt173 (2013).

56 Haroniti, A. *et al.* The clamp–loader–helicase interaction in Bacillus. Atomic force microscopy reveals the structural organisation of the DnaB–tau complex in Bacillus. *Journal of molecular biology* **336**, 381–393 (2004).

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

58 Chang, P. & Marians, K. J. Identification of a region of Escherichia coli DnaB required for functional interaction with DnaG at the replication fork. *The Journal of biological chemistry* **275**, 26187–26195, doi:10.1074/jbc.M001800200 (2000).

59 Stordal, L. & Maurer, R. Defect in general priming conferred by linker region mutants of Escherichia coli dnaB. *Journal of bacteriology* **178**, 4620–4627 (1996).

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

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

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

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

64 Mizushima, T. *et al.* Site–directed mutational analysis for the ATP binding of DnaA protein. Functions of two conserved amino acids (Lys–178 and Asp–235) located in the ATP–binding domain of DnaA protein in vitro and in vivo. *The Journal of biological chemistry* **273**, 20847–20851 (1998).

65 Vallet–Gely, I., Donovan, K. E., Fang, R., Joung, J. K. & Dove, S. L. Repression of phase–variable cup gene expression by H–NS–like proteins in Pseudomonas aeruginosa. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 11082–11087, doi:10.1073/pnas.0502663102 (2005).

Chapter 5

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

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

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

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

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

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

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

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

Abstract

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

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

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

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

Background

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

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

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

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

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

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

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

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

Results

Identification of primase

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

C. difficile helicase is activated by primase

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

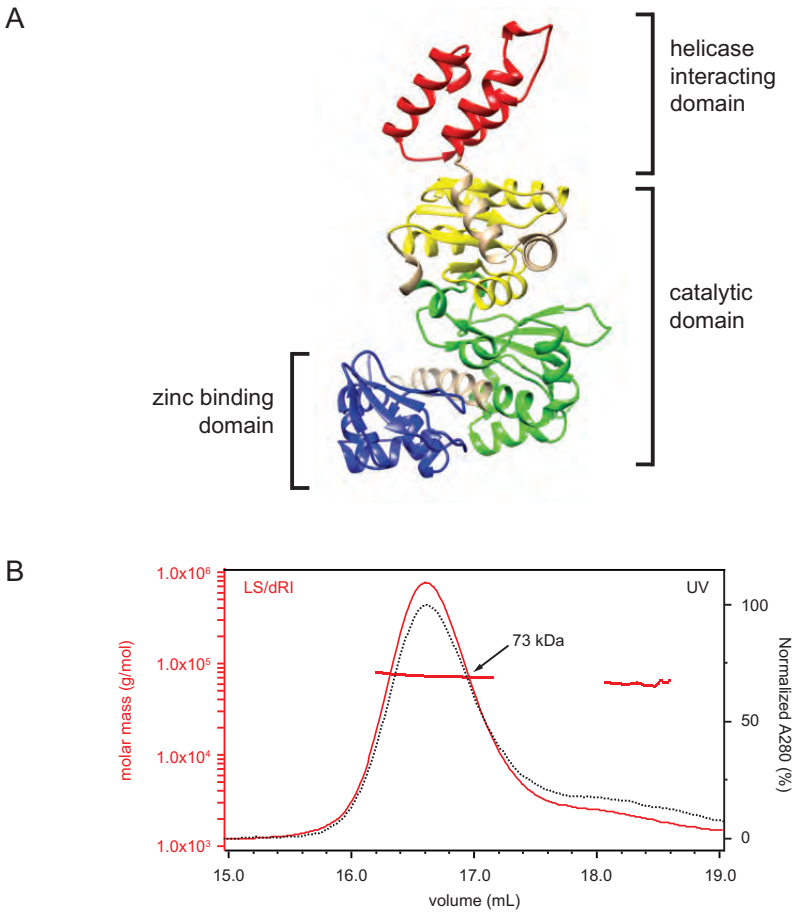


Figure 1. CD1454 is the *C. difficile* primase.

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

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

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

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

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

The putative loader protein CD3654 'locks' the CD3657 helicase

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

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

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

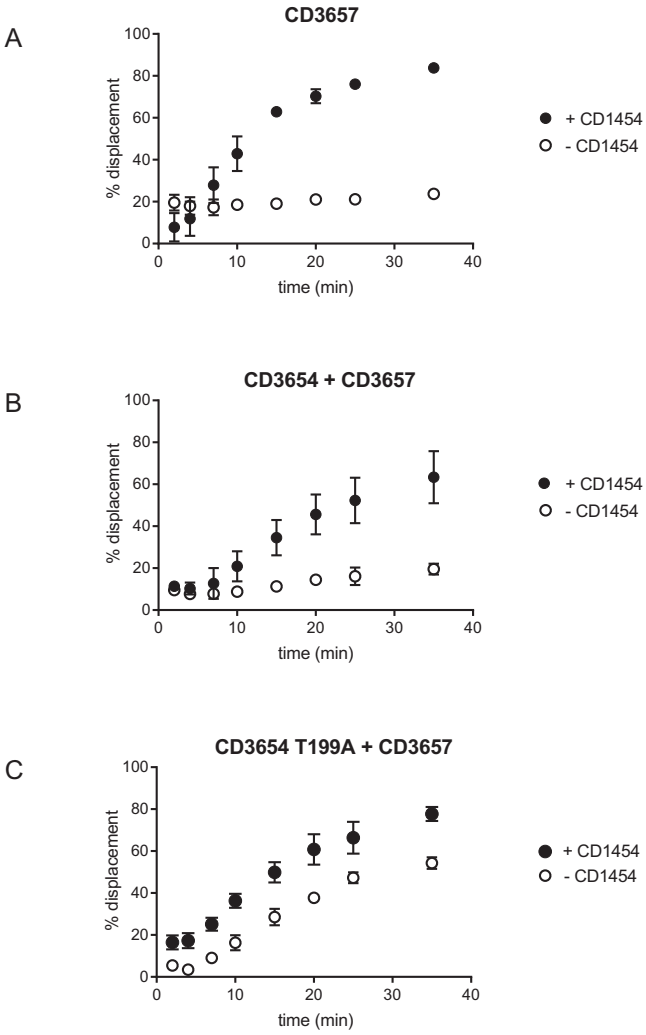


Figure 2. Helicase activity is stimulated by primase.

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

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

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

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

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

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

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

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

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

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

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

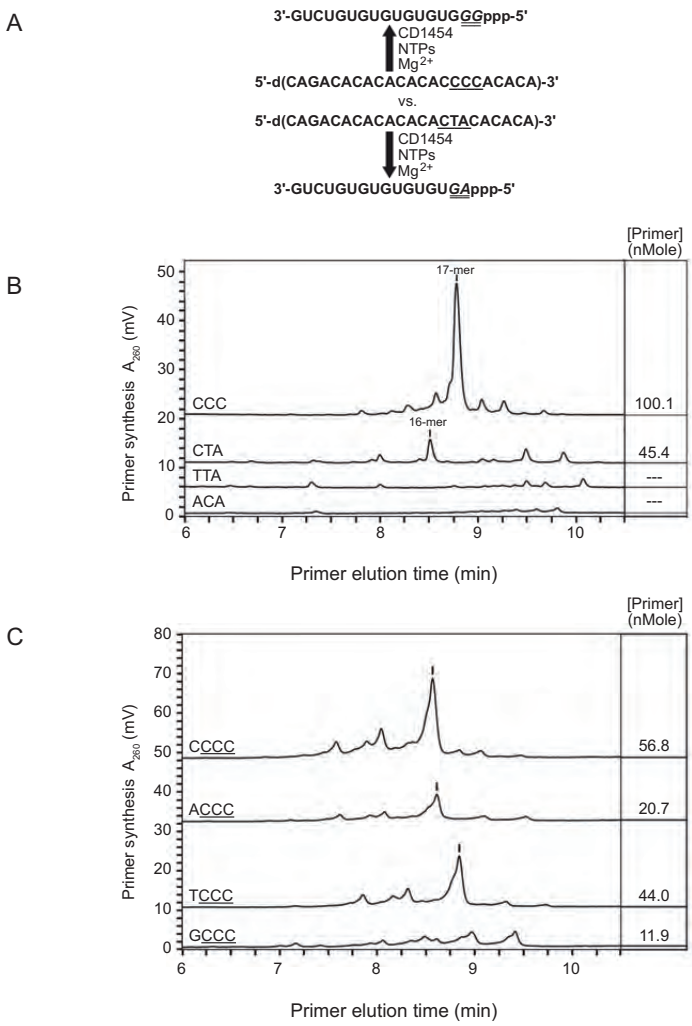


Figure 3. Trinucleotide specificity of CD1454 primase.

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

B. Analysis of CD1454 priming activity at different trinucleotides.

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

Helicase influences trinucleotide specificity of primase

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

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

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

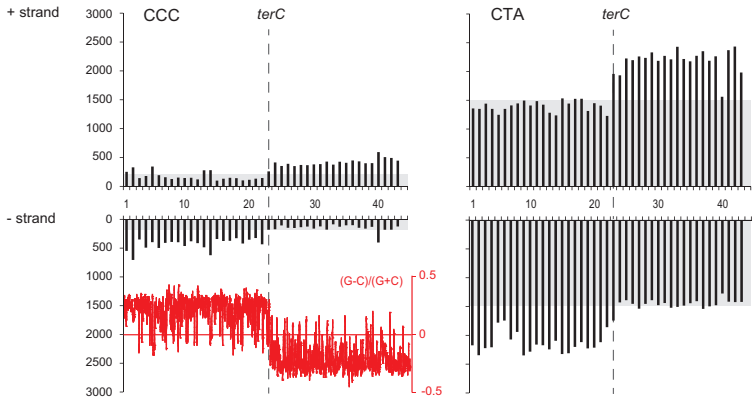


Figure 4. Trinucleotide frequency in the genome of *C. difficile*.

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

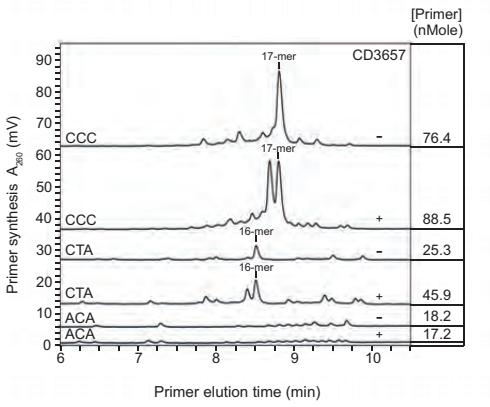


Figure 5. CD3657 helicase affects CD1454 primase activity.

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

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

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

A lysine residue contributes to trinucleotide specificity of primase

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

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

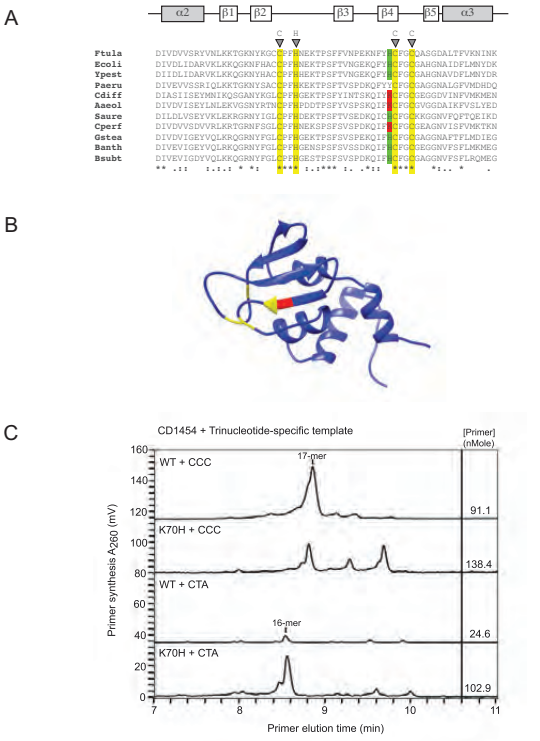


Figure 6. A lysine residue affects trinucleotide specificity.

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

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

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

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

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

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

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

Discussion

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

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

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

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

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

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

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

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

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

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

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

Materials and Methods

Plasmid construction

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

Site-directed mutagenesis

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

Protein purifications

Purification of helicase CD3657 and putative loader CD3654

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

Purification of primase CD1454 and CD1454 K70H

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

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

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

Size exclusion chromatography coupled with multiangle laser light scattering analysis

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

RNA primer synthesis assay and thermally denaturing HPLC analysis

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

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

Helicase assays

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

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

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

Acknowledgements

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

Supplemental information

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

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

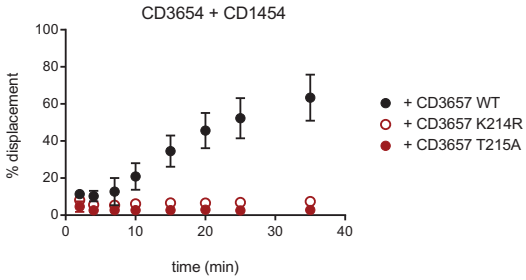
1 Department of Medical Microbiology, Leiden University Medical Center, Leiden, the Netherlands;

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

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

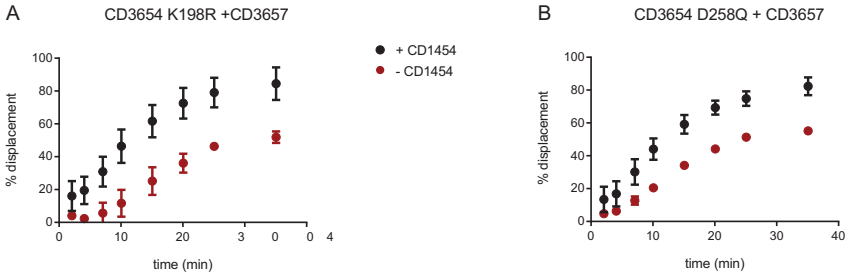
4 University of Nebraska Center for Biosecurity, Omaha, NE 68198-4080,

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



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

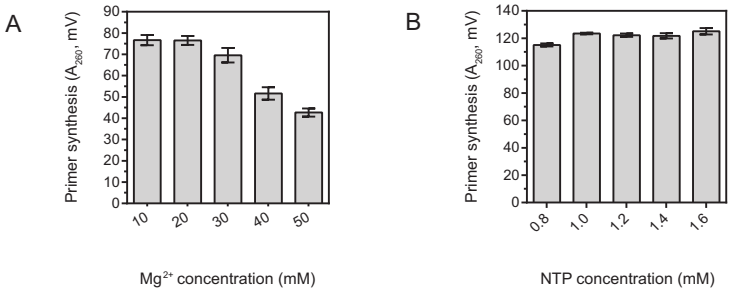
Helicase activity was assayed by quantifying the displacement of a radiolabelled (γ^{32}P–ATP) oligonucleotide partially annealed to the single stranded circular DNA m13mp18. Percent displaced signal from the helicase assays in time. Helicase activity of the wild type (WT; data series depicted in black)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



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

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

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



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

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

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

Supplemental Table 1. Oligonucleotides used in this study

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

Supplemental Table 2. Plasmids used in this study

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

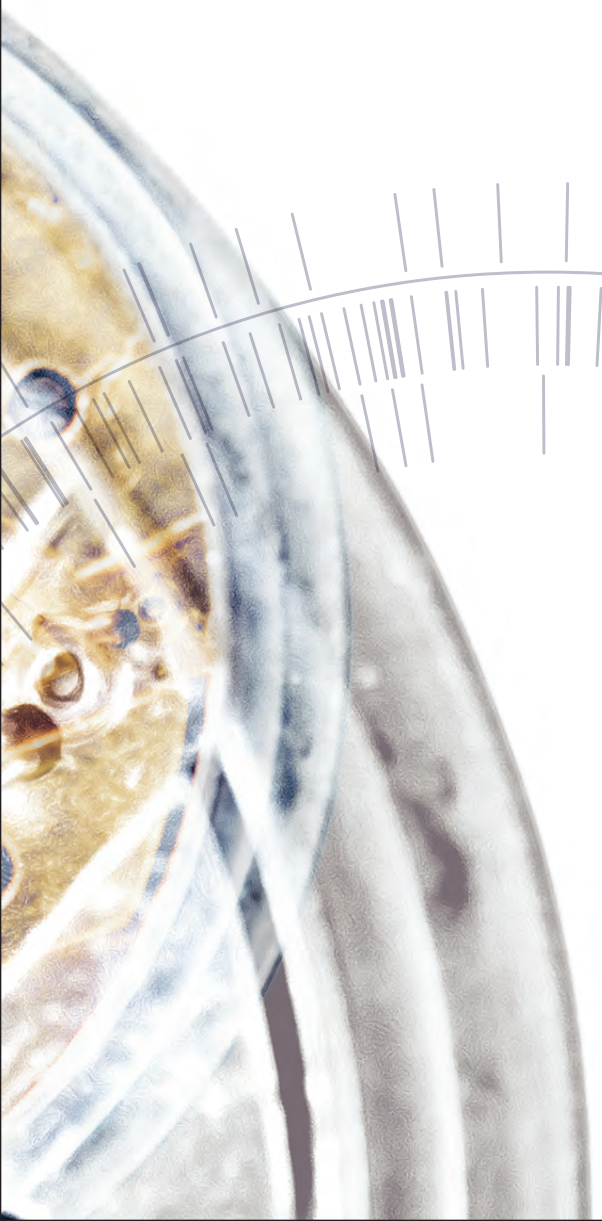
58 Hussain, H. A., Roberts, A. P. & Mullany, P. Generation of an erythromycin-sensitive derivative of Clostridium difficile strain 630 (630Deltaerm) and demonstration that the conjugative transposon Tn916DeltaE enters the genome of this strain at multiple sites. *Journal of medical microbiology* **54**,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

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

Chapter 6

The transcriptional effects of sub–inhibitory concentrations of PolC–inhibitors suggest a gene–dosage dependent response to replication inhibition in *Clostridium difficile*

Erika van Eijk [1]
Ed J. Kuijper [1]
Ingrid Bos–Sanders [1]
George Wright [2]
Wiep Klaas Smits [1,3]



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

2 GL Synthesis Inc., Worcester, Massachusetts, The United States of America;

3 Centre for Microbial Cell Biology, Leiden, The Netherlands.

This chapter is published as part of van Eijk *et al.* Antimicrobial Agents and Chemotherapy, 2019

Abstract

Clostridium difficile **is a potentially lethal gut pathogen that causes nosocomial and community-acquired infections. Limited treatment options and reports of reduced susceptibility to current treatment emphasize the necessity for novel antimicrobials. The DNA-polymerase of Gram-positive organisms is an attractive target for the development of antimicrobials. 362E (*N*^{2}-(3,4-dichlorobenzyl)-7-(2-[1-morpholinyl] ethyl) guanine; MorE-DCBG) is a DNA polymerase inhibitor in preclinical development as a novel therapeutic against *C. difficile* infection. This synthetic purine shows preferential activity against *C. difficile Pol* IIIC (PolC) over those of other organisms *in vitro* and is effective in an animal model of *C. difficile* infection. Its specificity may limit the negative effects on the colonic microbiota. In this study we have determined its efficacy against a large collection of clinical isolates. At concentrations below the minimal inhibitory concentration (MIC), the presumed slowing (or stalling) of replication forks due to 362E leads to a growth defect. We have determined the transcriptional response of *C. difficile* to replication inhibition and observed an overrepresentation of upregulated genes near the origin of replication in the presence of PolC-inhibitors, but not when cells were subjected to sub-inhibitory concentrations of other antibiotics. This phenomenon can be explained by a gene dosage shift, as we observed a concomitant increase in the ratio between origin-proximal and terminus-proximal gene copy number upon exposure to PolC-inhibitors.**

Background

Clostridium difficile (*Clostridioides difficile*) [1] is a Gram-positive, anaerobic bacterium that can asymptomatically colonize the intestine of humans and other mammal [2-4]. However, when the normal flora is disturbed, *C. difficile* can overgrow and cause fatal disease, as has been dramatically demonstrated in the Stoke Mandeville Hospital outbreaks in 2004 and 2005 [5]. The ability to form highly resistant endospores coupled to its extensive antibiotic resistance have contributed to its success as a nosocomial and community-acquired pathogen [2-4]. Recent years have seen an increase in the incidence and severity of *C. difficile* infections (CDI), due to the emergence of certain PCR ribotypes [4,6]. Antibiotic use is a well-established risk factor for CDI [7], and the emergence of the epidemic PCR ribotype 027 has been linked to fluoroquinolone resistance [8]. At present, two antibiotics, metronidazole and vancomycin, are commonly used to treat CDI, and a third, fidaxomicin, is indicated for the treatment of relapsing CDI [9,10]. Clearly, limited treatment options and reports of reduced susceptibility to current treatment [11-13] emphasise the necessity for the development of novel antimicrobials and a better understanding of tolerance and resistance towards existing therapeutics.

It is increasingly realized that off-target effects, that occur when cells are exposed to antimicrobials [14], can contribute to its efficacy but also facilitate the emergence of tolerance and/or resistance. Antimicrobials may act as signalling molecules which modulate gene expression [14]. Additionally, in particular, those targeting DNA replication (such as polymerase inhibitors) can cause transcriptional effects as a result of differences in gene dosage [15].

The polymerase of Gram-positive organisms is an attractive target for the development of novel antimicrobials [16]. First, these PolC-type polymerases are absent from Gram-negative organisms and humans [17,18]. HPUra (6(*p*-Hydroxyphenylazo)-uracil), one of the first such compounds, is therefore highly active against a wide range of Gram-positive bacteria but does not affect Gram-negative bacteria [17,18]. Template-directed elongation is blocked by the inhibitor through simultaneous binding to the cytosine of the DNA strand and near the active site of PolC. Second, compounds can be derived that have an increased specificity towards specific microorganisms. 362E (**Figure 1**) is a compound in pre-clinical

development as a novel therapeutic against *C. difficile*, as it is shows preferential activity against *C. difficile* PolC over those of other organisms *in vitro* [19,20].

PolC-inhibitors can cause a stress response and cell death after prolonged exposure. In *Bacillus subtilis*, this stress is characterized by a combination of DNA damage (SOS) response, and an SOS-independent pathway dependent on the DNA replication initiator DnaA [21,22]. In *Streptococcus pneumoniae* cells, devoid of an SOS-response, competence for genetic transformation is induced upon replication stress [23]. The response of *C. difficile* to this particular class of compounds is unknown.

In this study, we characterized aspects of the action of PolC-inhibitors towards *C. difficile*. Minimal inhibitory concentrations for HPUra and 362E were determined using agar dilution for a large collection of clinical isolates. Next, we investigated the effects of sub-inhibitory levels of PolC-inhibitors on growth of *C. difficile* in liquid medium and performed RNA sequencing (RNA-Seq) analyses to determine the transcriptional response to PolC-inhibitors in our laboratory strain 630Δ*erm*. Finally, marker frequency analysis was used to provide a mechanistic explanation for the observed up-regulated of origin-proximal genes under conditions of replication inhibition.

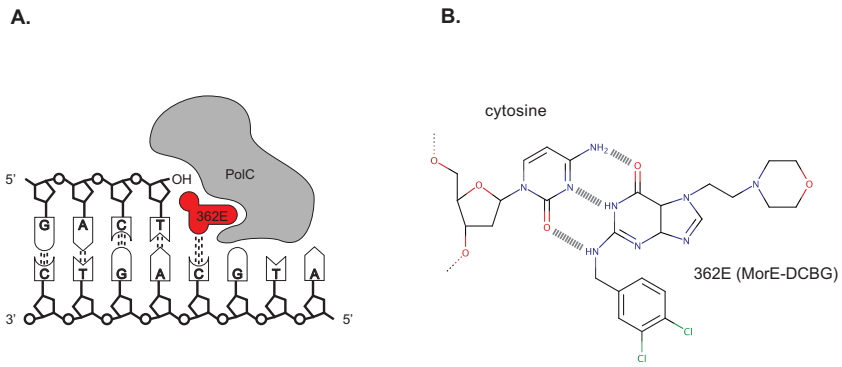


Figure 1. Mechanism of action of the PolC-inhibitor 362E.

A. Ternary complex of inhibitor 362E, DNA, and PolC.
B. H-bonding between inhibitor molecule 362E and a cytosine residue of DNA.

Results

362E is a potent inhibitor of diverse clinical isolates of *C. difficile*

To date, reports on activity of PolC-inhibitors towards *C. difficile* are limited. For only 4 [19] and 23 [20] *C. difficile* strains minimal inhibitory concentrations were published, and no analysis was performed on possible differences in efficacy between various phylogenetic groups [8,24]. Therefore, we assessed the sensitivities of a diverse collection of *C. difficile* clinical isolates towards PolC-inhibitors and determined if 362E was indeed superior to the general PolC-inhibitor HPUra.

HPUra and 362E were tested by the agar dilution method, according to Clinical and Laboratory Standards Institute (CLSI) guidelines for testing of antimicrobial susceptibility of anaerobes [25,26], against 363 *C. difficile* clinical isolates collected earlier in the framework of a pan-European study [6,27].

We found that 362E (MIC_{50}: 2 µg/ml; MIC_{90}: 4 µg/ml) demonstrates lower inhibitory concentrations than the general Gram-positive PolC-inhibitor HPUra (MIC_{50}: 16 µg/ml; MIC_{90}: 32 µg/ml) (**Figure 2**), consistent with previous *in vitro* activities observed against purified PolC [19].

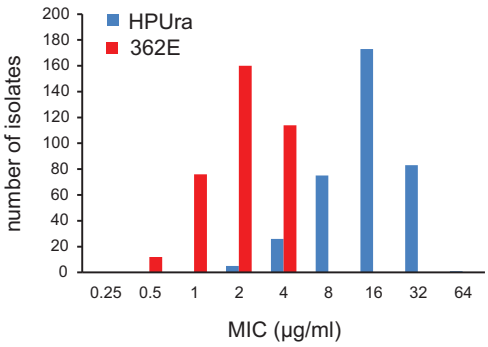


Figure 2. Minimal inhibitory concentrations of PolC-inhibitors.

MIC was determined by agar dilution according to CLSI standards [25] and is expressed in µg/mL. The distribution in the MIC for the collection of clinical isolates (*n*=363) is given for the PolC-inhibitors HPUra (blue) and 362E (red).

We observed no significant difference in 362E susceptibilities between clades (**Table 1**) and the different PCR ribotypes demonstrated a similar distribution in MIC values (*data not shown*).

Table 1. Minimal inhibitory concentrations expressed in µg/ml of PolC-inhibitors stratified by clade

	MIC_{50}	MIC_{90}	Range	No. of isolates	PCR Ribotypes
Clade 1					
HPUra	16	32	2-64	230	001, 002, 003, 005, 009, 010, 011, 012, 014, 015, 018, 025, 026, 029, 031, 037, 050, 051, 053, 056, 057, 064, 070, 081, 084, 087, 106 and 118
362E	2	4	0,25-4		
Clade 2					
HPUra	16	32	2-32	24	016, 019, 027, 075 and 208.
362E	2	4	0,5-4		
Clade 3					
HPUra	16	32	4-32	7	023
362E	4	4	1-4		
Clade 4					
HPUra	8	16	2-16	9	017
362E	2	4	1-4		
Clade 5					
HPUra	16	32	4-32	43	033, 045, 078 and 126
362E	1	2	0,5-4		
Clade 6					
HPUra	4	4	4	1	131
362E	4	4	4		
Clade unknown					
HPUra	16	32	4-32	49	013, 024, 039, 046, 063, 090, 093, 097, 099, 101, 107, 110, 137, 139, 150, 154, 159, 161, 176, 202, 205, 207, 228, 229, 230, 231, 232, and 234
362E	2	4	0,5-4		

No growth at the highest concentration of compounds tested for either one of the PolC-inhibitors was observed among the collection of clinical isolates (n=363) (**Supplementary table 1**).

Notably, we observed only a 2-fold difference in MIC_{50} and MIC_{90}, indicating that the compounds have similar activities against nearly all strains. In contrast, the Gram-negative obligate anaerobe *Bacteroides fragilis*, was resistant to both polymerase inhibitors under the conditions tested (MIC >265 µg/ml), as expected for an organism lacking PolC. The Gram-positive bacterium *Staphylococcus aureus*, which was included as a control for the activity of HPUra against this group of bacteria [16,28,29], was sensitive to both HPUra and 362E, with MIC values of 2 µg/mL and 1 µg/mL, respectively.

We conclude that 362E is highly active against diverse clinical isolates of *C. difficile*, and resistance is not a concern in currently circulating strains.

Treatment with 362E leads to a pleiotropic transcriptional response

In order to determine the transcriptional response of PolC-inhibitors, we established the optimal concentration of both inhibitors which affected growth (sub-MIC levels) of *C. difficile* in liquid medium. The laboratory strain *C. difficile* 630Δ*erm* (PCR ribotype 012, multi-locus sequence type [MLST] Clade 1) [30,31] was grown in liquid medium with various amounts of HPUra (10-40 µg/mL) or 362E (0.25-8 µg/mL). We note that concentrations up to the MIC_{90} (as determined by agar dilution) did not lead to a complete growth arrest in liquid medium in the time course of the experiment (growth was abolished at >2-fold MIC in liquid culture; *data not shown*). A difference in the MIC values between agar dilution and (micro)broth has been observed before [32]. The growth kinetics of *C. difficile* under the influence of varied concentrations of HPUra was marginally affected when using concentrations from 10-40 µg/ml, at approximately 80 percent of the non-treated culture. Growth kinetics of cultures containing PolC-inhibitor 362E at 1 to 8 µg/ml were similar and resulted in 30 to 40 percent reduced growth compared to that of the non-treated culture (**Figure 3**). For subsequent experiments, we used concentrations of PolC-inhibitors that result in a maximum reduction in growth of 30 percent compared to that of a non-treated culture (HPUra, 35 µg/ml; 362E, 4 µg/ml).

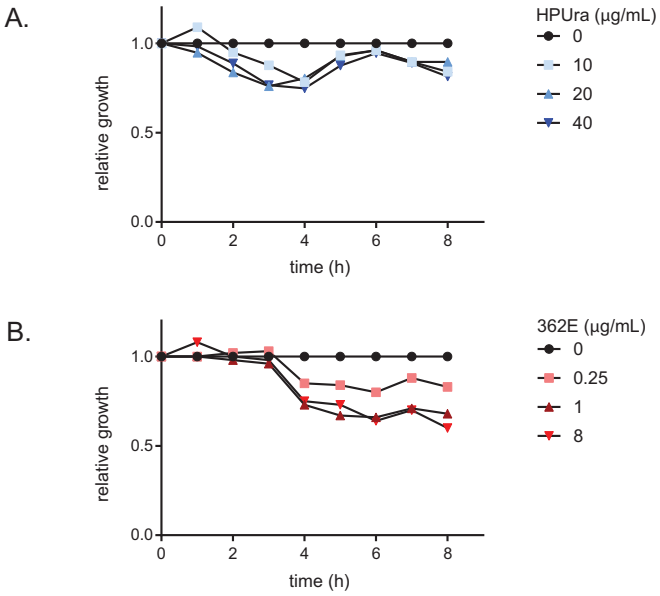


Figure 3. Relative inhibition of growth by varying concentrations of PolC-inhibitors.

A. Growth kinetics of a culture grown in the presence of the indicated amount of HPUra, relative to a non-treated culture.

B. Growth kinetics of a culture grown in the presence of the indicated amount of 362E, relative to a non-treated culture. Growth in liquid media was abolished at >2x the MIC (*data not shown*).

As described above, we established that growth of *C. difficile* is partially inhibited at certain concentrations of PolC-inhibitors. Slowing down or stalling of replication forks might lead to a stressed state, as observed for other organisms [21,23]. As nothing is known about the effect of replication inhibition on the physiology of *C. difficile*, we determined the transcriptional response to replication inhibition by sub-MIC levels of PolC inhibitors through strand-specific RNA sequencing (RNA-Seq).

C. difficile 630Δ*erm* was grown for 5h in medium with HPUra (35 µg/mL) or 362E (4 µg/mL) starting from an optical density at 600 nm (OD_{600}) of 0.05 after which cells were harvested for RNA isolation. Purified RNA was converted to cDNA and used for RNA-Seq as described in the Materials and Methods. For 362E, 722 genes were differentially expressed, of which 438 genes were up-regulated and 284

genes were down-regulated. The number of differentially expressed genes in HPUra-treated samples was ~2-fold lower, at 360, of which 124 genes were up-regulated and 236 genes were down-regulated. The full list of differentially regulated genes is available as **Supplementary Table 2** and the top 10 of up- and down-regulated genes are shown in **Table 2** (for HPUra) and **Table 3** (for 362E).

Here, we highlight three aspects of the results. First, we performed a Gene Set Enrichment Analysis (GSEA) [33] via the Genome2D web server [34] using the locus tags of the differentially regulated genes (**Supplementary Table 2**) as input. Among the genes up-regulated by 362E, there is a strong overrepresentation of those involved in translation, ribosome structure and ribosome biogenesis. Not unexpectedly, replication, recombination and repair processes are also affected. This suggests that genes from these pathways show a coordinated response in the presence of 362E. Among the genes down-regulated in the presence of 362E, the levels of significance for specific processes are generally much lower, suggesting that there is a more heterogeneous response among genes from the same pathway. Nevertheless, metabolic pathways (especially carbon metabolism and coenzyme A transfer) and tellurite resistance were found to be significantly affected. Strikingly, a GSEA analysis on lists of genes that are differentially expressed in the presence of HPUra revealed similar processes to be affected.

The findings from the GSEA analysis prompted us to evaluate the overlap in the lists of differentially regulated genes between the 362E and HPUra datasets in more detail. If the two compounds act via a similar mechanism, we expect a conserved response. Indeed, we observe that >90% of the genes that are up-regulated in the presence of HPUra compared to the non-treated condition, are also identified as up-regulated in the presence of 362E (**Figure 4A**). Though the overlap is not as strong for the down-regulated genes, we find that >30% of the genes affected by HPUra are also identified as affected by 362E (**Figure 4B**). Notably, the directionality of the response is conserved, as no genes were found to be up-regulated under one condition but down-regulated under the other condition. Based on these observations, we believe that the differentially expressed genes identified in this study are representative for a typical response to inhibition of PolC in *C. difficile*.

Table 2. Top 10 up-and down regulated genes in HPUra treated samples

Top 10 significantly changed genes HPUra UP

RefSeq locus tag [1]	logFC	log CPM	LR	pvalue	adj_ pvalue	Fold	min FDR	Old locus tag [2]	Gene name [3]	Product	Protein ID
CD630DERM_RS12480	11.16	3.84	9.8	1.8e-03	2.4e-02	2291.8	5.37	CD630DERM_23052	CD2305B	Hypothetical protein	WP_004454601.1
CD630DERM_RS15730	6.02	12.24	39.2	3.9e-10	1.9e-07	64.9	22.30	–	–	Hypothetical protein	WP_032506906.1
CD630DERM_RS17185	5.12	8.69	48.8	2.9e-12	3.8e-09	34.8	27.97	CD630DERM_32060	CD3206	Hypothetical protein	WP_003434302.1
CD630DERM_RS17015	4.64	8.61	45.7	1.3e-11	1.0e-08	25.0	26.55	CD630DERM_31740	CD3174	Type I glyceraldehyde-3-phosphate dehydrogenase	WP_003421962.1
CD630DERM_RS17020	4.62	8.68	54.7	1.4e-13	2.7e-10	24.6	31.77	CD630DERM_31750	CD3175	Transcriptional regulator	WP_003429564.1
CD630DERM_RS00700	4.37	10.08	33.9	5.8e-09	2.1e-06	20.7	18.85	CD630DERM_00801	CD0080A	50S ribosomal protein L29	WP_003421158.1
CD630DERM_RS00655	4.26	8.55	22.7	1.9e-06	1.4e-04	19.1	12.77	CD630DERM_00720	CD0072	30S ribosomal protein S10	WP_011860633.1
CD630DERM_RS16270	3.96	8.16	34.9	3.5e-09	1.4e-06	15.6	19.44	CD630DERM_30310	CD3031	Transcriptional antiterminator	WP_003432034.1
CD630DERM_RS19860	3.93	10.72	31.8	1.7e-08	4.1e-06	15.2	17.89	CD630DERM_36630	CD3663	30S ribosomal protein S6	WP_003420522.1
CD630DERM_RS00775	3.83	6.09	15.1	1.0e-04	2.7e-03	14.2	8.53	CD630DERM_00930	CD0093	50S ribosomal protein L14	WP_009895197.1

Top 10 significantly changed genes HPUra DOWN

RefSeq locus tag [1]	logFC	log CPM	LR	pvalue	adj_ pvalue	Fold	min FDR	Old locus tag [2]	Gene name [3]	Product	Protein ID
CD630DERM_RS09945	-14.85	4.52	17.0	3.7e-05	1.2e-03	-29514.0	9.72	CD630DERM_17940	CD1794	Hypothetical protein	WP_003430297.1
CD630DERM_RS12730	-14.52	4.33	17.3	3.1e-05	1.1e-03	-23428.2	9.87	CD630DERM_23550	CD2355	Thiol reductase thioredoxin	WP_003416870.1
CD630DERM_RS00870	-14.39	4.91	21.8	3.0e-06	1.9e-04	-21483.7	12.35	–	–	tRNA-Asn	–
CD630DERM_RS00255	-14.35	5.48	45.8	1.3e-11	1.0e-08	-20874.2	26.55	–	–	tRNA-His	–
CD630DERM_RS00925	-14.18	4.91	23.1	1.5e-06	1.3e-04	-18514.8	12.89	–	–	tRNA-Ser	–
CD630DERM_RS17905	-14.02	5.11	26.3	3.0e-07	4.1e-05	-16646.0	14.57	–		Hypothetical protein	WP_042741280.1
CD630DERM_RS05300	-14.01	4.27	12.7	3.7e-04	7.3e-03	-16475.8	7.11	CD630DERM_09230	CD0923	Hypothetical protein, putative phage protein	WP_021362052.1
CD630DERM_RS00170	-13.98	4.82	13.8	2.0e-04	4.5e-03	-16156.5	7.78	–	–	tRNA-Asn	–
CD630DERM_RS03115	-13.96	5.24	56.8	4.8e-14	1.9e-10	-15908.4	32.28	CD630DERM_04981	CD0498A	Hypothetical protein	WP_011860839.1
CD630DERM_RS05160	-13.92	3.89	15.3	9.1e-05	2.5e-03	-15523.8	8.65	CD630DERM_08960	CD0896	Hypothetical protein	WP_003418667.1

1 At time of analysis https://www.ncbi.nlm.nih.gov/nuccore/NZ_LN614756.1
2 https://www.ncbi.nlm.nih.gov/nuccore/LN614756,1
3 https://www.ncbi.nlm.nih.gov/nuccore/AM180355.1

Table 3. Top 10 up- and down regulated genes in 362E treated samples

362E **top 10 significant changed genes UP**

RefSeq locus tag [1]	logFC	log CPM	LR	pvalue	adj_ pvalue	Fold	min FDR	Old locus tag [2]	Gene name [3]	Product	Protein ID
CD630DERM_RS12480	11.31	3.84	9.9	1.6e-03	1.3e-02	2544.2	6.25	CD630DERM_23052	CD2305B	Hypothetical protein	WP_004454601.1
CD630DERM_RS17185	7.73	8.69	87.6	8.1e-21	3.3e-17	211.9	54.77	CD630DERM_32060	CD3206	Hypothetical protein	WP_003434302.1
CD630DERM_RS17955	6.88	7.78	28.5	9.5e-08	5.0e-06	117.8	17.62	CD630DERM_33520	CD3352	AraC family transcriptional regulator	WP_009891821.1
CD630DERM_RS17950	6.87	9.82	31.9	1.6e-08	1.0e-06	117.3	19.88	CD630DERM_33510	CD3351	ATP-dependent Clp protease proteolytic subunit	WP_003436212.1
CD630DERM_RS15730	6.67	12.24	45.3	1.7e-11	4.0e-09	101.9	27.89	-	-	Hypothetical protein	WP_032506906.1
CD630DERM_RS00990	5.92	8.75	50.1	1.4e-12	5.7e-10	60.4	30.70	CD630DERM_01090	CD0109	Anaerobic ribonucleoside-triphosphate reductase activating protein	WP_003432598.1
CD630DERM_RS00655	5.65	8.55	35.0	3.2e-09	2.7e-07	50.3	21.80	CD630DERM_00720	CD0072	30S ribosomal protein S10	WP_011860633.1
CD630DERM_RS00985	5.50	8.94	66.3	3.8e-16	7.6e-13	45.2	40.27	CD630DERM_01080	CD0108	Anaerobic ribonucleoside triphosphate reductase	WP_003436320.1
CD630DERM_RS05405	5.49	10.12	43.5	4.2e-11	8.0e-09	45.0	26.89	CD630DERM_09410	CD0941	Hypothetical protein	WP_009894743.1
CD630DERM_RS15695	5.24	10.08	41.5	1.2e-10	2.0e-08	37.9	25.61	CD630DERM_29250	CD2925	Hypothetical protein	WP_009894743.1

362E **top 10 significant changed genes DOWN**

RefSeq locus tag [1]	logFC	log CPM	LR	pvalue	adj_ pvalue	Fold	min FDR	Old locus tag [2]	Gene name [3]	Product	Protein ID
CD630DERM_RS05350	-14.17	5.04	40.4	2.1e-10	3.1e-08	-18382.7	24.94	CD630DERM_09310	CD0931	Hypothetical protein	WP_011861014.1
CD630DERM_RS05160	-13.92	3.89	15.0	1.1e-04	1.6e-03	-15523.8	9.33	CD630DERM_08960	CD0896	Hypothetical protein	WP_003418667.1
CD630DERM_RS05375	-13.72	4.09	14.3	1.5e-04	2.0e-03	-13454.4	8.95	CD630DERM_09360	CD0936	Endonuclease	WP_011861020.1
CD630DERM_RS15535	-13.57	4.30	23.5	1.3e-06	4.4e-05	-12181.6	14.47	CD630DERM_28950	CD2895	Membrane protein	WP_011861046.1
CD630DERM_RS13040	-13.55	3.91	11.4	7.5e-04	7.1e-03	-11987.2	7.15	CD630DERM_24140	CD2414	PTS sorbose transporter subunit IIB	WP_003430855.1
CD630DERM_RS05345	-13.55	4.41	16.5	4.9e-05	8.2e-04	-11969.7	10.25	CD630DERM_09301	CD0930A	Hypothetical protein	WP_011861013.1
CD630DERM_RS11925	-13.53	4.10	14.0	1.9e-04	2.3e-03	-11799.9	8.76	CD630DERM_21991	CD2199A	4Fe-4S ferredoxin	WP_003423406.1
CD630DERM_RS01060	-13.52	4.51	14.9	1.1e-04	1.6e-03	-11718.7	9.33	-	-	tRNA-Glu	-
CD630DERM_RS15350	-13.44	3.89	15.3	9.1e-05	1.4e-03	-11076.8	9.53	-	-	Membrane protein	WP_003426729.1
CD630DERM_RS06360	-13.43	4.12	23.9	1.0e-06	3.7e-05	-11012.2	14.72	CD630DERM_11160	CD1116	Transposase	WP_011861118.1

1 At time of analysis https://www.ncbi.nlm.nih.gov/nuccore/NZ_LN614756.1
2 https://www.ncbi.nlm.nih.gov/nuccore/LN614756.1
3 https://www.ncbi.nlm.nih.gov/nuccore/AM180355.1

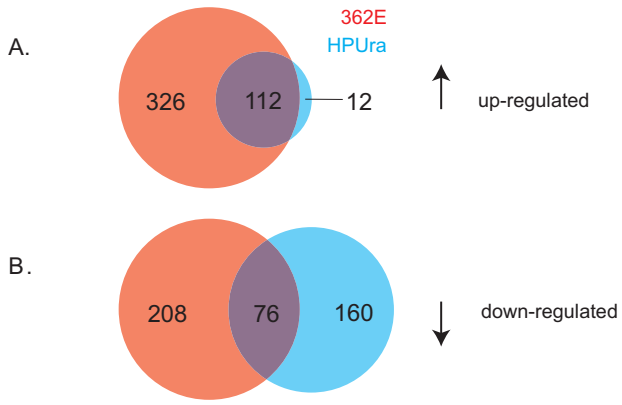


Figure 4. Overlap in the transcriptional response to different PolC-inhibitors.

A. Venn diagram of the number of genes up-regulated in the presence of 362E (red), in the presence of HPUra (blue) or under both conditions (overlapping region).

B. Venn diagram of the number of genes down-regulated in the presence of 362E (red), in the presence of HPUra (blue) or under both conditions (overlapping region). The size of the circles is proportional to the number of genes that showed differential expression.

Finally, we related the changes in transcription to genome location. *C. difficile* has a single circular chromosome and one origin of replication (*oriC*) from which the process of DNA replication occurs bi-directionally towards the terminus (*terC*) (**Figure 5A**). Though neither *oriC* nor *terC* has been definitively defined for *C. difficile*, it is assumed that *oriC* is located at or near *dnaA* (CD0001; CD630DERM_RS00005). The terminal region is generally located at the inflection point of a GC skew ([G - C]/[G + C]) plot. Such a plot places the *terC* region around 2.2Mb from CD0001, near the CD1933 (CD630DERM_RS10465) open reading frame (**Figure 5A**) [35].

We noted that the differential expression appeared to correlate with genome location (**Table 2, Table 3** and **Supplemental Table 2**), as many of the up-regulated genes have either low or high gene identifiers (CD numbers) indicative of an origin proximal location, conversely, many of the down-regulated genes appear to be located away from *oriC*. Though this correlation is not absolute, we observed a clear trend when plotting the mean fold change against genome location for all genes (**Figure 5B**).

Overall, our data show that inhibition of DNA replication by PolC-inhibitors causes a consistent and pleiotropic transcriptional response that is at least in part directly dictated by genome location.

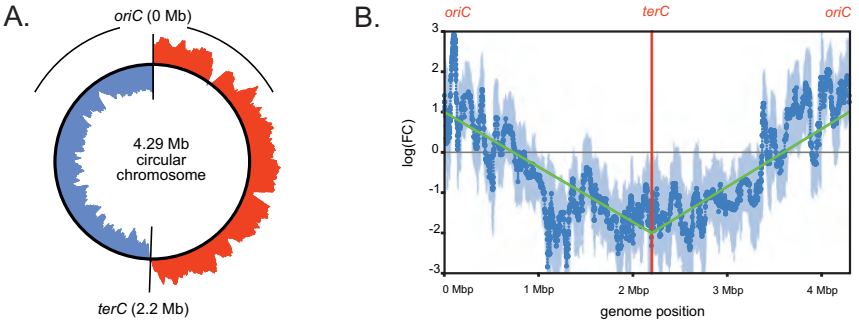


Figure 5. Genome location correlates with differential expression upon PolC inhibition.

A. Schematic representation of the chromosome of *C. difficile*. Higher than average GC skew ([G – C]/[G + C]) (red) and lower than average GC skew (blue) were calculated with DNAplotter [36]. Vertical lines indicate the position of the predicted origin (*oriC*) and terminus (*terC*) of replication. Arrows indicate the direction of replication.

B. Sliding window analysis (bins of 51 loci, step size 1) of the median log fold change (FC) projected on a linear genome map. The *oriC* of the circular chromosome is located on either size of the linear graph (0/4.29Mb), whereas *terC* is indicated with a vertical red line. The trend in log(FC) is highlighted using a green line. Light blue shading indicates the median absolute deviation of the mean [23].

6

Gene dosage shift occurs at sub-inhibitory concentration of 362E PolC-inhibitor

A possible explanation for the relative up-regulation of *oriC*-proximal genes and down-regulation of *terC*-proximal genes is a gene dosage shift, due to the fact that PolC inhibition slows down replication elongation but does not prevent re-initiation of DNA replication [23]. To determine if this in fact occurs in *C. difficile* when replication elongation is inhibited, we performed a marker frequency analysis (MFA) to determine the relative abundance of an origin proximal gene relative to terminus proximal gene on chromosomal DNA isolated from treated and non-treated cells.

We designed quantitative PCR (qPCR) probes against the CD0001 and CD1931 regions, representing *oriC* and *terC*, respectively [31,35]. Using these probes, we could show that *C. difficile* demonstrates multi-fork replication in exponential growth phase and that the MFA assay detects the expected decrease in *oriC*:*terC* ratio when cells enter stationary growth phase (*data not shown*). Next, we analysed the effects of PolC-inhibitors on the *oriC*:*terC* ratio. When cells were treated with HPUra (35 µg/mL), the MFA showed a modest increase in *oriC*:*terC* ratio of 2.6-fold compared to non-treated cells. However, when cells were treated with 362E (4 µg/ml), the MFA showed a >8-fold increase in the *oriC*:*terC* ratio compared to non-treated cells (**Figure 6**). By contrast, such an increase was not observed for cells treated with metronidazole (0.25 µg/mL; a DNA damaging agent), fidaxomicin (0.00125 µg/mL; an RNA polymerase inhibitor) or surotomycin (0.625 µg/mL; a cell-wall synthesis inhibitor) (**Figure 6**) or chloramphenicol (2 µg/mL; a protein synthesis inhibitor) (*data not shown*). We conclude that inhibition of PolC-activity, but not the actions of any of the other tested antimicrobials, lead to a gene dosage shift in *C. difficile*.

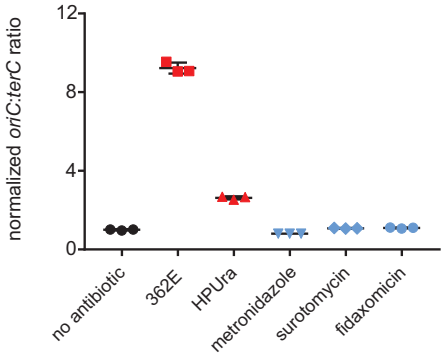


Figure 6. Polymerase-inhibitors lead to an increase in *oriC*:*terC* ratio.

A marker frequency analysis of the effects of sub-inhibitory amounts of polymerase inhibitors (red; HPUra: 35 µg/mL; 362E: 4 µg/mL) and three antibiotics with different modes of action (blue; metronidazole: 0.25 µg/mL; fidaxomicin: 0.00125 µg/mL; surotomycin: 0.625 µg/mL) compared to non-treated cells (black). Data points are averages of technical replicates (n=3). Black lines behind the data points indicate the average of the biological replicates (n=3) and whiskers indicate the standard deviation of the mean. Data have been normalized compared to the non-treated control. The mean of HPUra and 362E treated samples is statistically different from the other samples (p <0.0001).

Discussion

Activity and specificity of 362E

Limited treatment options and reports of reduced susceptibility to current treatment [11,12,37] emphasise the necessity for the development of novel antimicrobials. More broadly, reducing the widespread and prolonged use of vancomycin, first-line therapy of severe CDI [9,10], may also contribute in minimizing the risk of inducing and shedding of vancomycin-resistant enterococci and staphylococci [19,20]. As CDI can be induced by use of broad-spectrum antibiotics [7], new antimicrobials ideally should only target *C. difficile*, thereby maintaining integrity of the colonic microbiota. In this study, we have tested the inhibitors HPUra and 362E which specifically target the PolC enzyme, which is essential for DNA replication. The majority of PolC-inhibitors target Gram-positive bacteria with low G+C content, but 362E has been reported to demonstrate increased specificity towards *C. difficile* PolC *in vitro* and showed promising results for the efficacy *in vivo*, based on a limited set of *C. difficile* strains [19,20]. The compound will progress to clinical trials in the near future (Acurx Pharmaceuticals, *personal communication*). The present study is the largest survey of the efficacy of HPUra and 362E against a large collection of clinical isolates consisting of many relevant PCR ribotypes to date. We have established that 362E demonstrated lower inhibitory concentrations than the general Gram-positive PolC-inhibitor HPUra in agar dilution experiments. The MIC_{50} and MIC_{90} of 362E are similar to those of antimicrobials commonly used to treat *C. difficile* infection (for metronidazole: MIC_{50} = 2 µg/mL and MIC_{90} = 4 µg/mL; for vancomycin MIC_{50} = 2 µg/mL and MIC_{90} = 4 µg/mL [20]). We did not detect a significant difference in MICs between clades and ribotypes, demonstrating that PolC-inhibitors have the potential to be used as treatment for the majority of – if not all – circulating *C. difficile* strains. This includes the epidemic types of PCR ribotype 027 and 078 [8,38]. These results are in line with other work that demonstrated only 2- to 4-fold differences in antimicrobial susceptibility between different clades for metronidazole, fidaxomicin, and semi-synthetic thiopeptide LFF571 [27]. In the course of our experiments, we did not find any strains that grew at the highest concentrations of either HPUra or 362E tested, suggesting that resistance against these particular PolC-inhibitors is

rare or absent in *C. difficile*. PolC-inhibitors are competitive inhibitors of polymerase activity by binding in the active site. Mutations that abolish binding of HPUra or 362E are likely to affect the essential enzymatic activity of the polymerase and for that reason unlikely to occur *in vivo*. A single mutation (*azp-12*) has been described in *B. subtilis* that confers resistance to HPUra [39]. This T>G transversion results in the replacement of a serine with an alanine in the highly conserved PFAM07733 domain of the polymerase [40]. To our knowledge, it is unknown whether this mutation prevents binding of HPUra to PolC of *B. subtilis*. Few other mutations have been described that confer resistance against other PolC-inhibitors [41,42]. It will be of interest to see if similar mutations in *C. difficile* result in resistance to HPUra and/or 362E and what the effect is on binding of these compounds to *C. difficile* PolC.

In our experiments, we included *S. aureus* as a control for the activity of HPUra against Gram-positive bacteria. Surprisingly, we found that *S. aureus* was highly sensitive to both HPUra and 362E, and even more so than *C. difficile*. The reasons for this sensitivity are unknown. It is conceivable that 362E also targets PolC of *S. aureus* and if so, this information could be used to further characterize the interaction of 362E with PolC. Alternatively, 362E may also affect the activity of the other PolIII-type polymerase, DnaE. PolIII-inhibitors can affect PolC, DnaE or both [41], though *in vivo* activity appears to correlate with PolC-inhibition. Both *C. difficile* and *S. aureus* possess PolC and DnaE polymerase, but the DnaE enzymes belong to different families (DnaE1 in *C. difficile* and DnaE3 in *S. aureus*). The phylogenetic split coincides with the taxonomic division as DnaE3- polymerases are found in the class Bacilli (which includes *S. aureus*), whereas the DnaE1 polymerase are present in the classes Clostridia and Negativicutes [43]. To test whether a difference in activity of 362E towards different DnaE-type polymerases explain the increased sensitivity of *S. aureus* compared to *C. difficile*, the activity of 362E towards purified DnaE from both organisms should be determined. Finally, 362E may have off-target effects in *S. aureus* unrelated to its replication-inhibitory activity.

Though it is clear that 362E inhibits *C. difficile* efficiently and shows limited activity towards certain other anaerobes [19], these findings highlight that it is necessary to perform additional (microbiome) studies to more clearly define the antimicrobial spectrum of this compound. It also shows that 362E may have therapeutic potential outside treatment of CDI.

Regulators of the transcriptional response to *PolC inhibitors*

The present study is the first to describe the transcriptional response of *C. difficile* to inhibition of DNA replication. We find that ~200 genes show differential expression under conditions of PolC-inhibition by both HPUra and 362E, compared to non-treated cells (**Supplemental Table 2, Figure 4**). When considering only 362E, approximately 13 percent of all genes in the genome show statistically significant altered transcription. We demonstrate that this large reprogramming of transcription is likely to be caused directly by a gene dosage shift (**Figure 5** and **6**).

The main limitation of our study is the fact that we cannot conclusively identify one or more mechanisms that explains the transcriptional response. There are several reasons for this.

First, our list of differentially regulated genes includes several putative regulators; sigma factors (including *sigE*, *sigG* and *sigH*), transcription factors and anti-terminators. The relatively long induction time (5h) at sub-MIC levels of antimicrobials may have contributed to secondary effects, through one or more of these regulators. Though shorter induction times are thought to provoke more compound-specific responses [44], we did observe a highly consistent transcriptional signature with both HPUra and 362E (**Supplemental Table 2, Figure 4**).

Second, major stress response pathways are poorly characterized in *C. difficile*. On the basis of experiments in other organisms [21,23,45-47], we expect that inhibition of DNA replication inhibition might possibly induce an SOS response (LexA) [48], a DnaA-dependent transcriptional response [21], and possibly a heat shock response (HrcA/CtsR) [50] and/or a general stress response (SigB) [49]. Putative LexA-regulated genes of *C. difficile* were identified *in silico* [51] and some of these (such as the *uvr* excinuclease and 30S ribosomal protein S3) were differentially expressed in our dataset (**Supplemental Table 2**). However, pleiotropic phenotypes have been described for a *C. difficile lexA* mutant [52] and it is likely that other LexA targets exist. To date, no genes have been identified that are regulated by DnaA in *C. difficile* and direct regulation of genes through the other pathways has not been demonstrated. No mutants of *hrcA* or *ctsR* have been described for *C. difficile*, but transcriptome and proteome analyses have been performed with heat shocked cells 42°C [53], or 41°C [54-56]. Similarities between these datasets and our data include genes encoding

proline racemase (*prdF*), chaperones (*groEL*, *groES*), thioredoxin systems (*trxA*, *trxB*) and Clp-proteases (*clpC*, *clpP*). In contrast to most anaerobic Gram-positive organisms, *C. difficile* encodes a homologue of the general stress response sigma factor, σ^B [49,57]. A transcriptome analysis comparing a *sigB* mutant versus wild type cells was recently published [58]. Strikingly, we find ~35% of the genes (20/58) identified as involved in stress response under the control of σ^B to be differentially expressed in our 362E dataset (including many non-characterized genes), suggesting that the response to DNA replication inhibition is at least partially dependent on σ^B. It should be noted that the *sigB*-operon was not differentially expressed. This phenomenon has been observed before in other organisms and is attributed to a persistent upregulation of *sigB* followed by a persistent response of σ^B-dependent gene expression [50,53,55]. In our experiment, cells were grown for 5 hours at sub-MIC levels of inhibitors, making it unlikely that altered *sigB* transcription would still be detected. Alternatively, σ^B activity might be regulated post-translationally.

Many parameters (such as the medium used, cell density, concentration of antibiotics, and protocol used to arrest transcription between cell harvest and lysis) can influence overall transcription signatures and can also govern an incomplete overlap between our data and the stress regulons determined by others [44]. To conclusively demonstrate an involvement for any of these pathways in the response of *C. difficile* to replication inhibition, independent validation of the results from the RNA-Seq analysis is necessary. Validation of RNA-Seq data frequently using qPCR on the same RNA will not validate any biological conclusions but validates the technology [59]. For that reason, our current efforts are directed to the construction of luciferase reporter-fusions [60] that can be used to study promoter activity under different conditions and in different mutant backgrounds.

Genome location contributes to the transcriptional response to PolC-inhibition

Our analysis of differential regulation in relation to genome location revealed a striking pattern of relative up-regulation for *oriC*-proximal genes, and down-regulation for *terC*-proximal genes under conditions of PolC inhibition (**Figure 5**). Antimicrobials directed at DNA replication in bacteria have a profound negative effect on the processivity of replication forks, though initiation of DNA replication is

not or only marginally affected [23, 39]. As a consequence, the presence of multiple replication forks simultaneously increases the copy numbers of genes located in close proximity of the origin of replication, and such a gene dosage differences can result in functionally relevant transcriptional differences, either directly or indirectly [15]. We found an increase of *oriC*:*terC* ratio when performing MFA on chromosomal DNA of cells subjected to a sub-inhibitory concentration of 362E (and HPUra, albeit less pronounced) (**Figure 6**), consistent with findings in other organisms [23].

An example of direct regulation by gene dosage can be found in *Vibrio*, for instance, where the location of ribosomal protein clusters close to the origin is crucial for fast growth, because increased copy number under condition of multi-fork replication allows for higher expression levels [61]. We note that ribosomal gene clusters are up-regulated when DNA replication is inhibited in our experiments (**Table 2, Table 3, Supplemental Table 2**), suggesting that a similar mechanism may be active in *C. difficile*.

An example of indirect regulation as a result of gene dosage is the induction of competence genes in *S. pneumoniae* [23]. Competence is believed to be a stress response in this organism, that lacks a canonical (σ^B-dependent) stress response pathway. Key regulatory genes for competence development are located close to the origin, and replication inhibition therefore leads to the induction of origin distal competence genes [23]. In our experiments, the large overlap with the proposed σ^B regulon [58] and the origin proximal location of the *sigB* operon (8.5kb-10kb) suggest that part of the transcriptional response to PolC-inhibition can be explained by an indirect gene dosage effect. The positioning of stress-response regulators close to *oriC* may therefore be a conserved strategy in bacteria to respond to DNA replication insults, independent of the nature of the regulator.

Though it is likely that an increase in gene copy number leads to an increase in transcription of these genes, it is less clear whether this is the case for the observed down-regulation. Most methods of normalization for transcriptome analyses are based on the assumption that there is no overall change in transcription or that the number of transcripts per cells is the same for all conditions and this may not be the case when a global copy number shift occurs [15]. Absolute transcript levels for down-regulated genes might therefore be similar under both conditions (but lower, relative to *oriC*-proximal transcripts).

It is interesting that certain processes are highly enriched in the list of genes up-regulated under conditions of PolC-inhibition (most notably ribosome function and DNA-related functions), whereas this is less so for the down-regulated genes. This suggests that pathways susceptible to replication-dependent gene dosage effects demonstrate a functional clustering of genes near *oriC*, whereas clustering of genes belonging to specific pathways in the *terC*-proximal region is less pronounced. Indeed, most ribosomal gene clusters in *C. difficile* are located close to the origin of replication [31,57] and many genes involved in DNA replication and repair are located in these regions. Consistent with this, positioning of genes involved in transcription and translation close to the origin appears to be under strong selection as such genomes tend to be more stable [65].

In conclusion, both direct and indirect effects of gene dosage shifts are likely to contribute to the transcriptional response of *C. difficile* to replication inhibition.

Materials and Methods

Agar dilution

HPUra and 362E were tested against a collection of *C. difficile* clinical isolates. 375 clinical isolates have been collected during the ECDIS study [6]. All strains were characterized by PCR ribotyping [62] and by PCR to confirm the presence of genes encoding toxins A, B and binary toxin [63-65]. Of the 375 clinical isolates, we excluded stocks that were found to contain more than one strain and isolates that could not be re-cultured. Testing was therefore performed on 363 isolates (**Supplemental Table 1**). *C. difficile* ATCC 700057, *B. fragilis* ATCC 25285 and *S. aureus* ATCC 29213 were used as controls [66].

The strains were tested for the different concentrations of antimicrobial using the agar dilution method according to Clinical & Laboratory Standards Institute guidelines [25]. In short, the antimicrobials were diluted into Brucella Blood Agar (BBA) supplemented with hemin and vitamin K1. Bacterial isolates were cultured on blood agar plates and after 24 hours re-suspended to a turbidity of 0.5 McFarland in phosphate buffered saline (PBS). The strains were inoculated onto BBA solid media containing the PolC-inhibitors using multipoint inoculators to a final con-

centration of 10^4 CFU per spot. Each series of antimicrobial agents was tested from the lowest concentration to the highest concentration. Two control plates without antibiotics were inoculated to control for aerobic contamination and purity of anaerobic growth. At the end of the final series, two control plates were inoculated to verify the final organism viability and purity. Plates were incubated anaerobically in a variable atmosphere cabinet (VA1000, Don Whitley Scientific) and the MICs were determined after 24 and 48 hours.

Sub-MIC determination

C. difficile 630Δ*erm* [30,31] was grown in 20 mL Brain Heart Infusion (Oxoid) supplemented with 0.5% yeast extract (Sigma-Aldrich) (BHI/YE) starting from an optical density measured at 600 nm (OD_{600}) of 0.05 using an exponentially growing starter culture (3 biological replicates per concentration). To determine the effects on growth of sub-inhibitory concentration of 362E, cells were cultured in the presence of the following concentrations; 0.25, 0.5, 1, 2, 4, and 8 µg/mL 362E and compared to an untreated culture. To determine the effects on growth of sub-inhibitory concentration of HPUra, cells were cultured in the presence of the following concentrations; 10, 20, 40, 80 µg/mL HPUra and compared to an untreated culture. The OD_{600} was monitored every hour until stationary phase was reached.

Marker Frequency analysis

C. difficile 630Δ*erm* [30,31], was grown in 20 mL BHI/YE with sub-MIC amounts of antimicrobials (HPUra: 35 µg/mL; 362E: 4 µg/mL; metronidazole: 0.25 µg/mL; fidaxomicin: 0.00125 µg/mL, surotomycin: 0.625 µg/mL), starting from an OD_{600} of 0.05 using an exponentially growing starter culture. We confirmed that these concentrations did not lead to a >30% reduction in growth compared to non-treated cultures (**Figure 3** and *data not shown*). In parallel, cultures were grown without inhibitors from the same starter culture. All conditions were performed in biological triplicates. After 5 hours, 1mL cells was harvested (OD_{600} ~ 0.5), and stored at -20°C. Isolation of chromosomal DNA was performed the next day with the QIAamp DNA Blood Mini kit (Qiagen) according to the instructions of the manufacturer. Marker frequency analysis (MFA) was performed to assess the relative abundance of origin proximal

genes relative to terminus proximal genes. As a proxy for *oriC*, a probe was designed that targets the CD0001 region (CD0001-probe-FAM). By using plots of the GC skew ([G - C]/[G + C]) generated using DNAPlotter [36], the approximate location of the terminus region for the *C. difficile* chromosome was determined and a probe targeting this region (CD1931) was designed (CD-1931-probe-TXR). Probe design was performed with Beacon Designer™ (Premier Biosoft, Palo Alto CA, USA). Real-time PCR reactions were performed on a Biorad CFX96™ real-time PCR detection system (95°C 15 m, 39 × (94°C 30 s + 55°C 30 s + 72°C 30 s). Sequences for primers and probes are listed in **Table 4**.

Table 4. Oligonucleotides and probes used in this study

Name	Sequence (5' – 3')	Description
CD-0001- F	GAGACAAGAATTGCTATACTTA	Forward primer CD0001 MFA (oriC)
CD-0001- R	CAACCACTCTAGTTAATGC	Reverse primer CD0001 MFA (oriC)
CD-0001-probe-FAM	CTCAACTAGAACGTATAGATGTGCCAA	Probe CD0001 MFA (oriC)
CD-1931- F	GCAGGAATTTTAGATGAAGA	Forward primer CD1931 MFA (terC)
CD-1931- R	GGCTGAAGTCTTATTAAATTTC	Reverse primer CD1931 MFA (terC)
CD-1931-probe-TXR	CCTCTTAACTGTAGCAGATTCACCAT	Probe CD1931 MFA (terC)

For each biological replicate, three technical replicates were performed. Amplification efficiency was determined using standard curves obtained from DNA late stationary phase cells of strain 630Δ*erm*, for which an *oriC*:*terC* ratio of 1 was assumed. RT-PCR results from antibiotic treated cells were normalized to the *oriC*:*terC* ratio of DNA samples (3 biological replicates) from non-treated cells. Calculations were performed in Microsoft Office Excel 2010, plotted using Prism 7 (GraphPad) and prepared for publication in Corel Draw Suite X8. Significance was determined using a One-way ANOVA and a Tukey's test for multiple comparisons (GraphPad).

Growth and RNA isolation for RNA-Seq

For RNA-Seq analysis, *C. difficile* 630Δ*erm* was grown for 5h in BHI medium with HPUra (35 µg/mL) or 362E (4 µg/mL) starting from an OD_{600} of 0.05 using an

exponentially growing starter culture, after which cells (3mL) were harvested for RNA isolation. RNA isolation was performed with NucleoSpin© RNA kit (Macherey-Nagel). Although the kit includes on column rDNAse digestion, a second treatment was performed in solution and RNA was precipitated and recovered by NaAc precipitation to remove residual DNA. Concentration determination and quality control (16S/23S ratio and RNA integrity number [RIN]) was performed with a fragment analyser (Agilent bio-analyzer), according to the instructions of the manufacturer. Samples with a RIN>9 and 16S/23S ratio >1.4 were submitted for analysis by RNA-Seq.

RNA-Seq

RNA-Seq was performed at a commercial provider (GenomeScan, Leiden, The Netherlands). In short, the NEBNext Ultra Directional RNA Library Prep Kit for Illumina was used to process the samples. Sample preparation was performed according to the protocol "NEBNext Ultra Directional RNA Library Prep Kit for Illumina" (NEB #E7420S/L). Briefly, after selective removal of rRNA (Ribo-Zero rRNA Removal Kit for Gram-Positive Bacteria) and fragmentation of the mRNA, cDNA synthesis was performed. cDNA was ligated to the sequencing adapters and the resulting product was PCR amplified. Clustering and DNA sequencing using the Illumina NextSeq 500 platform was performed according manufacturer's protocols. A concentration of 1.5 pM of DNA was used. Image analysis, base calling, and quality check was performed with the Illumina data analysis pipeline RTA vl.18.64 and Bcl2fastq v2.17. Per sample, four technical replicates were included in the RNA-Seq experiment. In case of insufficient reads, the sample was re-run on another flow cell to reach satisfactory quantities (≥ 20 M).

Analysis of RNA-Seq data

Analysis of the data was performed using T-REx, a user-friendly webserver which has been optimized for the analysis of prokaryotic RNAseq-derived expression data [67]. The pipeline requires raw RNA expression level data as an input for RNA-Seq data analysis. For data normalisation and determination of the genes, the factorial design statistical method of the RNA-Seq analysis R-package EdgeR is implemented in the T-Rex pipeline. Some samples displayed incomplete rRNA depletion and rRNA mapping reads had to be removed manually prior to analysis.

To analyse the genome-wide pattern in differential gene expression a sliding window analysis was performed essentially as described[23]. In short, genome locations (start of the locus tag) were coupled to the locus tags in the T-Rex output. Next, the median log(FC) was calculated for bins of 51 loci with a step size of 1. For each bin of $[X_1, X_2...X_{51}]$ the median absolute deviation of the median (MAD=median($|X_i$-median(X)|) was calculated as a robust indication of the distribution around calculated median values. Calculations were performed and three curves (median, median-MAD and median+MAD) were plotted in Microsoft Office Excel 2010 and the graph was prepared for publication using Adobe Photoshop CC and Corel Draw Suite X8.

A GSEA analysis [33] was performed via the Genome2D webserver [34], using our reference genome sequence for *C. difficile* 630Δ*erm*, GenBank identifier LN614756.1 (listed in Genome2D as "Clostridioides_difficile_630Derm") [31]. As input a single list of locus tags was used of either up- or down regulated genes. The output was copied to Microsoft Excel 2010. The single list column was split, and a column was inserted to calculate the significance of the overrepresentation using the formula "(# hits in list/ClassSize)*-log(p-value; 2)" to allow for sorting of the output of the GSEA analysis by significance.

Acknowledgements

This work was supported, in part, by a Veni fellowship from the Netherlands Organization for Scientific Research and a Gisela Thier Fellowship from the Leiden University Medical Center to WKS. Anne de Jong is acknowledged for expert assistance with the RNA-Seq analysis, Els Wessels for assistance in designing the MFA real time PCR and Jelle Slager for helpful suggestions for the sliding window analysis.

Supplemental Information

Supplemental tables associated with this chapter can be found online at bioRxiv (pre-print server for Biology) via the following link: https://bit.ly/2mqcoLw

Supplemental Table 1. Characteristics of the clinical isolates used in the agar dilution experiments.

Supplemental Table 1_ agar dilution (as XLSX)

Supplemental Table 2. Lists of the genes that are differentially expressed in the presence of PolC inhibitors compared to non-treated cells.

Supplemental Table 2_ Significant (as XLSX)

Supplemental Table 3A. SigB-dependent stress response genes differentially regulated in the 362E dataset.

Supplemental Table 3_ SigB (as XLSX)

References

1 Lawson, P. A., Citron, D. M., Tyrrell, K. L. & Finegold, S. M. Reclassification of Clostridium difficile as Clostridioides difficile (Hall and O'Toole 1935) Prevot 1938. *Anaerobe* **40**, 95–99, doi:10.1016/j.anaerobe.2016.06.008 (2016).

2 Abt, M. C., McKenney, P. T. & Pamer, E. G. Clostridium difficile colitis: pathogenesis and host defence. *Nature reviews. Microbiology* **14**, 609–620, doi:10.1038/nrmicro.2016.108 (2016).

3 Leffler, D. A. & Lamont, J. T. Clostridium difficile Infection. *The New England journal of medicine* **373**, 287–288, doi:10.1056/NEJMc1506004 (2015).

4 Smits, W. K., Lyras, D., Lacy, D. B., Wilcox, M. H. & Kuijper, E. J. Clostridium difficile infection. *Nature reviews. Disease primers* **2**, 16020, doi:10.1038/nrdp.2016.20 (2016).

5 Investigation into outbreaks of *Clostridium difficile* at Stoke Mandeville Hospital, Buckinghamshire Hospitals NHS Trust. (Health Care Commission 2006).

6 Bauer, M. P. *et al.* Clostridium difficile infection in Europe: a hospital-based survey. *Lancet (London, England)* **377**, 63–73, doi:10.1016/s0140-6736(10)61266-4 (2011).

7 Hensgens, M. P., Goorhuis, A., Dekkers, O. M. & Kuijper, E. J. Time interval of increased risk for Clostridium difficile infection after exposure to antibiotics. *The Journal of antimicrobial chemotherapy* **67**, 742–748, doi:10.1093/jac/dkr508 (2012).

8 He, M. *et al.* Emergence and global spread of epidemic healthcare-associated Clostridium difficile. *Nature genetics* **45**, 109–113, doi:10.1038/ng.2478 (2013).

9 Debast, S. B., Bauer, M. P., Kuijper, E. J., European Society of Clinical, M. & Infectious, D. European Society of Clinical Microbiology and Infectious Diseases: update of the treatment guidance document for Clostridium difficile infection. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases* **20 Suppl 2**, 1–26, doi:10.1111/1469-0691.12418 (2014).

10 McDonald, L. C. *et al.* Clinical Practice Guidelines for Clostridium difficile Infection in Adults and Children: 2017 Update by the Infectious Diseases Society of America (IDSA) and Society for Healthcare Epidemiology of America (SHEA). *Clin Infect Dis* **66**, 987–994, doi:10.1093/cid/ciy149 (2018).

11 Baines, S. D. *et al.* Emergence of reduced susceptibility to metronidazole in Clostridium difficile. *The Journal of antimicrobial chemotherapy* **62**, 1046–1052, doi:10.1093/jac/dkn313 (2008).

12 Chong, P. M. *et al.* Proteomic analysis of a NAP1 Clostridium difficile clinical isolate resistant to metronidazole. *PloS one* **9**, e82622, doi:10.1371/journal.pone.0082622 (2014).

13 Moura, I., Spigaglia, P., Barbanti, F. & Mastrantonio, P. Analysis of metronidazole susceptibility in different Clostridium difficile PCR ribotypes. *The Journal of antimicrobial chemotherapy* **68**, 362–365, doi:10.1093/jac/dks420 (2013).

14 Davies, J., Spiegelman, G. B. & Yim, G. The world of subinhibitory antibiotic concentrations. *Current opinion in microbiology* **9**, 445–453, doi:10.1016/j.mib.2006.08.006 (2006).

15 Slager, J. & Veening, J. W. Hard-Wired Control of Bacterial Processes by Chromosomal Gene Location. *Trends in microbiology* **24**, 788–800, doi:10.1016/j.tim.2016.06.003 (2016).

16 Tarantino, P. M., Jr., Zhi, C., Wright, G. E. & Brown, N. C. Inhibitors of DNA polymerase III as novel antimicrobial agents against gram-positive eubacteria. *Antimicrobial agents and chemotherapy* **43**, 1982-1987 (1999).

17 Brown, N. C. 6-(p-hydroxyphenylazo)-uracil: a selective inhibitor of host DNA replication in phage-infected Bacillus subtilis. *Proceedings of the National Academy of Sciences of the United States of America* **67**, 1454-1461 (1970).

18 Butler, M. M., Dudycz, L. W., Khan, N. N., Wright, G. E. & Brown, N. C. Development of novel inhibitor probes of DNA polymerase III based on dGTP analogs of the HPUra type: base, nucleoside and nucleotide derivatives of N2-(3,4-dichlorobenzyl)guanine. *Nucleic acids research* **18**, 7381-7387 (1990).

19 Torti, A. *et al.* Clostridium difficile DNA polymerase IIIC: basis for activity of antibacterial compounds. *Current enzyme inhibition* **7**, 147-153 (2011).

20 Dvoskin, S. *et al.* A novel agent effective against Clostridium difficile infection. *Antimicrobial agents and chemotherapy* **56**, 1624-1626, doi:10.1128/aac.06097-11 (2012).

21 Goranov, A. I., Katz, L., Breier, A. M., Burge, C. B. & Grossman, A. D. A transcriptional response to replication status mediated by the conserved bacterial replication protein DnaA. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 12932-12937, doi:10.1073/pnas.0506174102 (2005).

22 Goranov, A. I., Kuester-Schoeck, E., Wang, J. D. & Grossman, A. D. Characterization of the global transcriptional responses to different types of DNA damage and disruption of replication in Bacillus subtilis. *Journal of bacteriology* **188**, 5595-5605, doi:10.1128/jb.00342-06 (2006).

23 Slager, J., Kjos, M., Attaiech, L. & Veening, J. W. Antibiotic-induced replication stress triggers bacterial competence by increasing gene dosage near the origin. *Cell* **157**, 395-406, doi:10.1016/j.cell.2014.01.068 (2014).

24 Knetsch, C. W. *et al.* Comparative analysis of an expanded Clostridium difficile reference strain collection reveals genetic diversity and evolution through six lineages. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* **12**, 1577-1585, doi:10.1016/j.meegid.2012.06.003 (2012).

25 CLSI 2012 Methods for antimicrobial susceptibility testing of anaerobic bacteria; approved standard, 8th ed. CLSI document M11-A8. Clinical and Laboratory Standards Institute, Wayne, PA, USA.

26 Freeman, J., Stott, J., Baines, S. D., Fawley, W. N. & Wilcox, M. H. Surveillance for resistance to metronidazole and vancomycin in genotypically distinct and UK epidemic Clostridium difficile isolates in a large teaching hospital. *The Journal of antimicrobial chemotherapy* **56**, 988-989, doi:10.1093/jac/dki357 (2005).

27 Debast, S. B., Bauer, M. P., Sanders, I. M., Wilcox, M. H. & Kuijper, E. J. Antimicrobial activity of LFF571 and three treatment agents against Clostridium difficile isolates collected for a pan-European survey in 2008: clinical and therapeutic implications. *The Journal of antimicrobial chemotherapy* **68**, 1305-1311, doi:10.1093/jac/dkt013 (2013).

28 Barnes, M. H. *et al.* DNA polymerase III of Mycoplasma pulmonis: isolation and characterization of the enzyme and its structural gene, polC. *Molecular microbiology* **13**, 843-854 (1994).

29 Pacitti, D. F., Barnes, M. H., Li, D. H. & Brown, N. C. Characterization and overexpression of the gene encoding Staphylococcus aureus DNA polymerase III. *Gene* **165**, 51-56 (1995).

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

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

32 Hastey, C. J. *et al.* Comparison of Clostridium difficile minimum inhibitory concentrations obtained using agar dilution vs broth microdilution methods. *Anaerobe* **44**, 73-77, doi:10.1016/j.anaerobe.2017.02.006 (2017).

33 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).

34 de Jong, A. Genome2D webserver for analysis and visualization of bacterial genomes and transcriptome data, <http://genome2d.molgenrug.nl>

35 van Eijk, E. *et al.* Primase is required for helicase activity and helicase alters the specificity of primase in the enteropathogen Clostridium difficile. *Open biology* **6**, doi:10.1098/rsob.160272 (2016).

36 Carver, T., Thomson, N., Bleasby, A., Berriman, M. & Parkhill, J. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* **25**, 119-120, doi:10.1093/bioinformatics/btn578 (2009).

37 Lynch, T. *et al.* Characterization of a stable, metronidazole-resistant Clostridium difficile clinical isolate. *PloS one* **8**, e53757, doi:10.1371/journal.pone.0053757 (2013).

38 Knetsch, C. W. *et al.* Zoonotic Transfer of Clostridium difficile Harboring Antimicrobial Resistance between Farm Animals and Humans. *Journal of clinical microbiology* **56**, doi:10.1128/jcm.01384-17 (2018).

39 Ott, R. W., Barnes, M. H., Brown, N. C. & Ganesan, A. T. Cloning and characterization of the polC region of Bacillus subtilis. *Journal of bacteriology* **165**, 951-957 (1986).

40 Barnes, M. H., Hammond, R. A., Foster, K. A., Mitchener, J. A. & Brown, N. C. The cloned polC gene of Bacillus subtilis: characterization of the azp12 mutation and controlled in vitro synthesis of active DNA polymerase III. *Gene* **85**, 177-186 (1989).

41 Barnes, M. H., Butler, M. M., Wright, G. E. & Brown, N. C. Antimicrobials targeted to the replication-specific DNA polymerases of gram-positive bacteria: target potential of dnaE. *Infectious disorders drug targets* **12**, 327-331 (2012).

42 Barnes, M. H., Hammond, R. A., Kennedy, C. C., Mack, S. L. & Brown, N. C. Localization of the exonuclease and polymerase domains of Bacillus subtilis DNA polymerase III. *Gene* **111**, 43-49 (1992).

43 Timinskas, K., Balvociute, M., Timinskas, A. & Venclovas, C. Comprehensive analysis of DNA polymerase III alpha subunits and their homologs in bacterial genomes. *Nucleic acids research* **42**, 1393-1413, doi:10.1093/nar/gkt900 (2014).

44 Wecke, T. & Mascher, T. Antibiotic research in the age of omics: from expression profiles to interspecies communication. *The Journal of antimicrobial chemotherapy* **66**, 2689-2704, doi:10.1093/jac/dkr373 (2011).

45 Kazmierczak, M. J., Mithoe, S. C., Boor, K. J. & Wiedmann, M. Listeria monocytogenes sigma B regulates stress response and virulence functions. *Journal of bacteriology* **185**, 5722-5734 (2003).

46 van Schaik, W. *et al.* Identification of the sigmaB regulon of Bacillus cereus and conservation of sigmaB-regulated genes in low-GC-content gram-positive bacteria. *Journal of bacteriology* **189**, 4384-4390, doi:10.1128/jb.00313-07 (2007).

47 Weber, H., Polen, T., Heuveling, J., Wendisch, V. F. & Hengge, R. Genome-wide analysis of the general stress response network in Escherichia coli: sigmaS-dependent genes, promoters, and sigma factor selectivity. *Journal of bacteriology* **187**, 1591-1603, doi:10.1128/jb.187.5.1591-1603.2005 (2005).

48 Kreuzer, K. N. DNA damage responses in prokaryotes: regulating gene expression, modulating growth patterns, and manipulating replication forks. *Cold Spring Harbor perspectives in biology* **5**, a012674, doi:10.1101/cshperspect.a012674 (2013).

49 Hecker, M., Pane-Farre, J. & Volker, U. SigB-dependent general stress response in Bacillus subtilis and related gram-positive bacteria. *Annual review of microbiology* **61**, 215-236, doi:10.1146/annurev.micro.61.080706.093445 (2007).

50 Helmann, J. D. *et al.* Global transcriptional response of Bacillus subtilis to heat shock. *Journal of bacteriology* **183**, 7318-7328, doi:10.1128/jb.183.24.7318-7328.2001 (2001).

51 Walter, B. M. *et al.* The LexA regulated genes of the Clostridium difficile. *BMC microbiology* **14**, 88, doi:10.1186/1471-2180-14-88 (2014).

52 Walter, B. M., Cartman, S. T., Minton, N. P., Butala, M. & Rupnik, M. The SOS Response Master Regulator LexA Is Associated with Sporulation, Motility and Biofilm Formation in Clostridium difficile. *PloS one* **10**, e0144763, doi:10.1371/journal.pone.0144763 (2015).

53 Emerson, J. E., Stabler, R. A., Wren, B. W. & Fairweather, N. F. Microarray analysis of the transcriptional responses of Clostridium difficile to environmental and antibiotic stress. *Journal of medical microbiology* **57**, 757-764, doi:10.1099/jmm.0.47657-0 (2008).

54 Jain, S., Graham, C., Graham, R. L., McMullan, G. & Ternan, N. G. Quantitative proteomic analysis of the heat stress response in Clostridium difficile strain 630. *Journal of proteome research* **10**, 3880-3890, doi:10.1021/pr200327t (2011).

55 Ternan, N. G., Jain, S., Srivastava, M. & McMullan, G. Comparative transcriptional analysis of clinically relevant heat stress response in Clostridium difficile strain 630. *PloS one* **7**, e42410, doi:10.1371/journal.pone.0042410 (2012).

56 Ternan, N. G., Jain, S., Graham, R. L. & McMullan, G. Semiquantitative analysis of clinical heat stress in Clostridium difficile strain 630 using a GeLC/MS workflow with emPAI quantitation. *PloS one* **9**, e88960, doi:10.1371/journal.pone.0088960 (2014).

57 Sebaihia, M. *et al.* Genome sequence of a proteolytic (Group I) Clostridium botulinum strain Hall A and comparative analysis of the clostridial genomes. *Genome research* **17**, 1082-1092, doi:10.1101/gr.6282807 (2007).

58 Kint, N. *et al.* The alternative sigma factor sigmaB plays a crucial role in adaptive strategies of Clostridium difficile during gut infection. *Environmental microbiology*, doi:10.1111/1462-2920.13696 (2017).

59 Fang, Z. & Cui, X. Design and validation issues in RNA-seq experiments. *Briefings in bioinformatics* **12**, 280-287, doi:10.1093/bib/bbr004 (2011).

60 Oliveira Paiva, A. M., Friggen, A. H., Hossein-Javaheri, S. & Smits, W. K. The Signal Sequence of the Abundant Extracellular Metalloprotease PPEP-1 Can Be Used to Secrete Synthetic Reporter Proteins in Clostridium difficile. *ACS synthetic biology* **5**, 1376-1382, doi:10.1021/acssynbio.6b00104 (2016).

61 Soler-Bistue, A. *et al.* Genomic location of the major ribosomal protein gene locus determines Vibrio cholerae global growth and infectivity. *PLoS genetics* **11**, e1005156, doi:10.1371/journal.pgen.1005156 (2015).

62 Bidet, P. *et al.* Comparison of PCR-ribotyping, arbitrarily primed PCR, and pulsed-field gel electrophoresis for typing Clostridium difficile. *Journal of clinical microbiology* **38**, 2484-2487 (2000).

63 Kato, H. *et al.* Identification of toxin A-negative, toxin B-positive Clostridium difficile by PCR. *Journal of clinical microbiology* **36**, 2178-2182 (1998).

64 Paltansing, S. *et al.* Characteristics and incidence of Clostridium difficile-associated disease in The Netherlands, 2005. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases* **13**, 1058-1064, doi:10.1111/j.1469-0691.2007.01793.x (2007).

65 Stubbs, S. *et al.* Production of actin-specific ADP-ribosyltransferase (binary toxin) by strains of Clostridium difficile. *FEMS microbiology letters* **186**, 307-312 (2000).

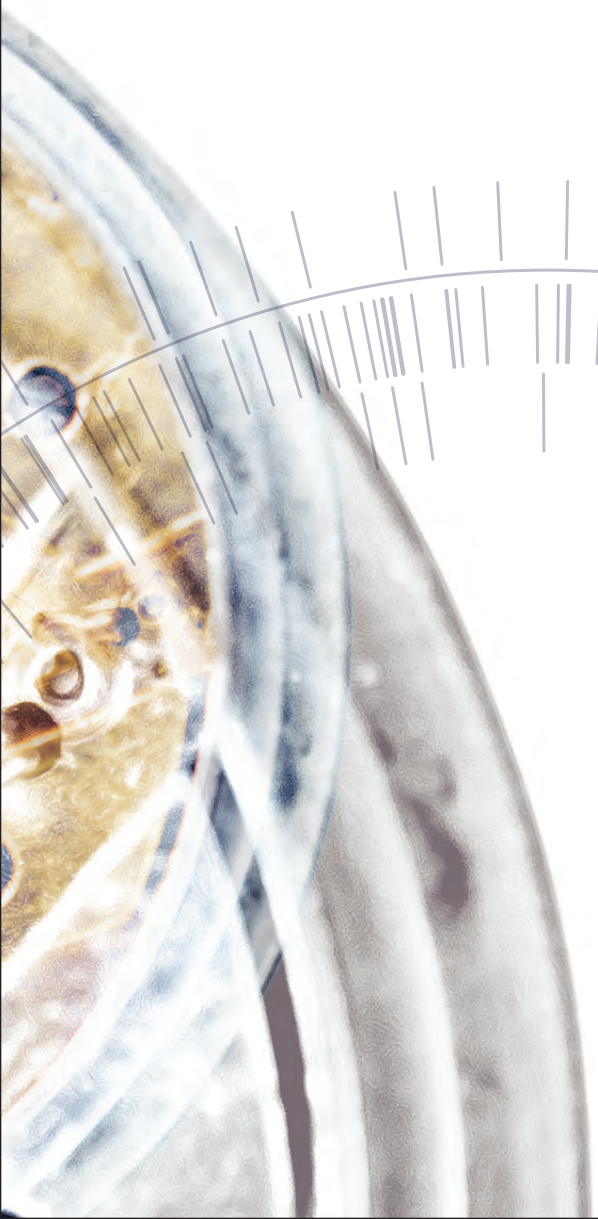
66 American Type Culture Collection (ATCC), <https://www.lgcstandards-atcc.org/en/Products/Cells_and_Microorganisms/Bacteria.aspx>

67 de Jong, A., van der Meulen, S., Kuipers, O. P. & Kok, J. T-REx: Transcriptome analysis webserver for RNA-seq Expression data. *BMC genomics* **16**, 663, doi:10.1186/s12864-015-1834-4 (2015).

Chapter 7

General Discussion

Erika van Eijk



General discussion

Phenotypic consequences of genetic diversity

In 2006, the complete genome sequence of a multiple drug resistant clinical isolate of *C. difficile* was published [1], and this particular strain (strain 630) became the reference strain for laboratories interested in the characterization of the organism itself and/or studying the genetic basis of the pathogenicity of this bacterium. A relatively large proportion of the genome of this historical PCR ribotype 012 isolate consists of mobile genetic elements (11 percent) [1]. Most of these are present in the chromosome as conjugative transposons, non-conjugative transposons, or integrated bacteriophages [1-3], thereby increasing genetic diversity in this species. The availability of the genome sequence spurred the development of tools for genetic manipulation of *C. difficile*, which is essential for gaining insight in pathogenesis of *C. difficile* infection [4]. Whereas the historical 630 strain demonstrated transferable erythromycin resistance [1], this resistance was lost due to serial culturing on non-selective media, to allow for the use of the *erm*B gene as antibiotic-selective marker in genetic studies in *C. difficile* [5]. The generation of these erythromycin sensitive strains, commonly referred to as 630Δ*erm* and 630E/JIR8094, occurred independently in laboratories based in Europe and Australia, respectively [4-7]. Both erythromycin-sensitive strains were distributed for research-purposes to other laboratories around the world. Additionally, the 630 and 630Δ*erm* strains were deposited and can be obtained from culture collections based in the US (ATCC), UK (NCTC), and Germany (DSMZ) [5], further contributing to the global distribution of these strains for laboratory-based studies [5,8]. However, when subjecting these different erythromycin-sensitive strains to similar experiments, differences in phenotypes - and thus primary outcomes - were observed, for instance in studies investigating the role of toxins in *C. difficile* infection [9,10]. Indeed, side-by-side comparison of 630E and 630Δ*erm* strains used for exploring the relative role of toxin A and toxin B confirmed their phenotypic- and genotypic distinction [6]. Similarly, significant phenotypic differences with respect to sporulation were observed within a collection of epidemic strains of PCR ribotype 027, refuting the assumption that the rate of sporulation was PCR ribotype-specific [11-13]. Taken together, the phenotypic differences demonstrated between the 630E and 630Δ*erm* strain as well as the parental

strain 630 underline the importance of documenting provenance, and hint at an underappreciate source of inter-laboratory experimental variation [5,6,8]. Moreover, the diversification of genetically distinct sub-strains caused by different passaging conditions in the different laboratories has been demonstrated for the attenuated strain of *Mycobacterium bovis*, Bacillus Calmette-Guérin (BCG), which is used as a prophylactic measure against tuberculosis [14]. Thus far, the molecular factors that impact the properties of the of BCG strains used for the vaccines are not known, but studies delineating these strains into phenotypic and genetic categories may provide insight into vaccine efficiency [14]. Such an approach may also be feasible for the laboratory strains of *C. difficile* that are currently used and determination of the complete genome sequence of our reference strain *C. difficile* 630Δ*erm* and comparison to its parental strain may be a starting point (**Chapter 3**).

In hindsight, the world-wide distribution of the *C. difficile* reference strain 630 and its erythromycin-sensitive derivatives and their propagation within numerous laboratories can be viewed as a long-term evolution experiment, akin to the ongoing long-term evolution experiment of *E. coli* (LTEE). This project started in the late eighties in order to study genetic changes as a result of repeated subculturing of initially identical populations of *Escherichia coli*, in a controlled laboratory environment [8,15,16]. In contrast to the LTEE, the *C. difficile* strains, however, encountered many different conditions in the individual laboratories, for instance in culturing and storage conditions [8]. Adaptation to these different conditions may have altered the genetic background these strains, most probably causing phenotypic variations [8]. For *C. difficile*, it is conceivable that growth conditions in a laboratory environment requires adaptation to nutrients that are readily available in the culture medium, but absent in the host environment of the bacteria. This may be reflected in an altered metabolic pathway or other adaptations. Such an adaptation to altered growth conditions has been observed in the LTEE, in which *E. coli* acquired a novel trait, aerobic citrate utilization, after being propagated for thousands of generations in a glucose-limited medium that also contains citrate [15,17]. This carbon source is normally not utilized under oxic conditions [15,17]. This observation in the long-term evolutionary experiment was a starting point to decipher the molecular basis of this specific adaptation [17]. The similarity between LTEE and the dissemination of *C. difficile* laboratory strains is that evolution of the species can be studied over time. However, in *C. difficile* strains, different selective pressures due to variability in laboratory conditions may have may accelerated the

accumulation of single nucleotide polymorphisms (SNPs) as opposed to a controlled environment such as the LTEE. Overall, though, in both species adaptation occurred over a vast period of time. Studying the (experimental) evolution of these bacteria may provide insight into the molecular basis of adaptation [18].

Even a single SNP can be responsible for phenotypic differences between genetically closely related strains. For instance, a single mutation in the gyrase gene is the underlying genetic mechanism for fluoroquinolone resistance in *C. difficile* [2,19–21]. When comparing the reference genome sequence of strain 630 to the sequence of our laboratory strain 630Δ*erm* (LUMC), we uncovered numerous insertions, deletions, single nucleotide polymorphisms (SNPs), and large scale chromosomal rearrangements such as the transposition of a mobile genetic element and the acquisition of an additional ~5 kb rRNA/tRNA cluster (**Chapter 3, Table 1, Figure 4**). It is not inconceivable that these genetic changes influence the phenotype of 630-derived strains. Moreover, when such phenotypic differences are already evident for genetically closely related strains, differences may be even more pronounced in genetically distant strains. For instance, studies in *Streptococcus pneumoniae* have shown that genomic content between strains can vary to such an extent, that an important part of experiencing and processing antibiotic stresses of a bacteria may be strain-dependent [22]. Indirectly, an antimicrobial can trigger a complex, multi-factorial process that resonates through the bacterium, involving numerous and diverse pathways such as regulation, metabolism and/or energy generation [22–26]. Due to the genetic diversity, or more specific, the absence and presence of genes between strains, the genome has to display remarkable plasticity to maintain function [22]. This implies that newly acquired genetic elements have to be integrated into regulatory networks, leading to changes within these networks (i.e. establishing new connections and altering existing connections). As a result, genomes may function differently, which may cause phenotypic variation ranging from adaptation of antimicrobials to virulence [22]. This may also apply to *C. difficile*, considering that the a large portion (11 percent) of the genome of *C. difficile* 630 consists of horizontally acquired elements [1], the average proportion of core genome (genes shared between strains) may be as little as 40 percent [27] and strains may or may not contain plasmids [28]. We observed already considerable differences between closely related strains belonging to PCR ribotype 012 (**Figure 1**), described in **Chapter 3** of this thesis.

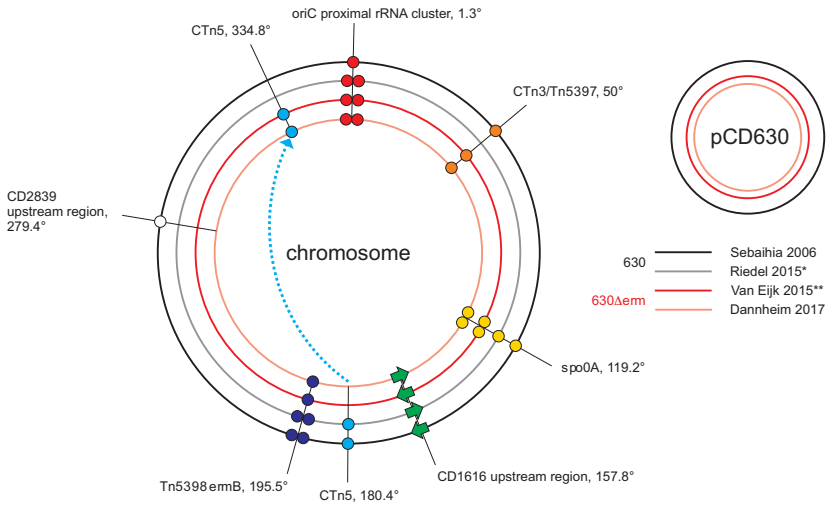


Figure 1. Whole genome comparison of *C. difficile* strain 630 and derivatives

Genome sequences (black/grey) and 630 *erm* sequences (red/orange) are represented as concentric circles. Structural variants are highlighted using symbols. Dots are used to indicate presence (dot) /absence (no dot) or duplications (two dots) of a sequence. Arrows indicate inversions. The dashed arrow between positions 180.4° and 334.8° indicates transposition of the CTn5 element. Note that this global overview does not allow discrimination of single nucleotide polymorphisms, but highlights presence/absence and location of certain larger structural variants (Figure reproduced from Roberts *et al.* (2018)[8] with permission from the publisher)

This PCR ribotype, however, is not the most prevalent circulating ribotype based on surveys conducted in Europe and the US [29,30] and large differences exist between ribotypes [1,3,31,32]. For RT 078, it even has been suggested that it forms a separate sub-species [33]. This indicates that there is a large variable component between genomes (accessory genome), that may be relevant for coping with stress, for instance exposure to antimicrobials [27,34]. The genetic diversity discussed above was one of the reasons to test a large collection of clinical isolates of *C. difficile* against the antimicrobial properties of new drugs such as 362E (**Chapter 6**). By testing a diverse collection, it can be established if a new antimicrobial is potent against all isolates despite the potential differences in genetic background, or only a subset. Our results show that there was no significant difference in 362E susceptibility

between clades (**Chapter 6, Table 1**) and that the different PCR ribotypes demonstrated a similar distribution in MIC values (range MIC: 0.5-4.0 µg/ml; MIC_{50}: 2 µg/ml; MIC_{90}: 4 µg/ml) (**Chapter 6, Figure 2**).

Genome organisation as regulator

When subjecting *C. difficile* vegetative cells to a sub-MIC level of the DNA polymerase inhibitor 362E, we observed an increase in the *oriC*:*terC* ratio, indicating the presence of an increased number of replication forks (**Chapter 6, Figure 6**). Although the replisome was (presumably) slowed or stalled by the inhibitor, initiation of replication seemed to continue or was possibly even stimulated [35]. Similarly, a coupling between replication inhibition and replication initiation was reported for *B. subtilis* where inhibition of gyrase by novobiocin leads to increased over-initiation [36]. As a direct consequence of this multi-fork replication, the copy-number of genes which are located in close proximity to the origin increase compared to genes located distally from the origin. A correlation between gene copy number and gene expression implies that the transcriptome of a bacterial cell is influenced by gene dosage [37-40] (**Chapter 6, Figure 5**). It should be noted that fluctuation in gene copy number due to genomic location is also affected by growth rate. In our experiment, cells were grown in nutrient-rich liquid medium which promotes fast growth and leads to multi-fork replication (i.e. more than one round of DNA replication occurs at the same time). In fast growing *E. coli*, which has a circular chromosome comparable in size to *C. difficile*, 8 copies of the origin of replication can be observed (doubling time ~ 20 minutes). As a result, the copy number of origin-proximal genes in comparison to genes located near the terminus may increase by 4-fold [41]. However, in our study, the *oriC*:*terC* ratio of the bacterial cells treated with 362E was compared to untreated cells and results of the Marker Frequency Analysis showed a relative increase up to 16-fold (**Chapter 6, Figure 6**). This result is independent of the increase of *oriC*:*terC* ratio caused by growth rate. The effect of 362E is concordant with experiments performed in *S. pneumoniae*, *B. cereus* and *S. aureus*, in which other antimicrobials affecting DNA replication significantly increased the *oriC*:*terC* ratio [42]. When cells of these organism were subjected to DNA-polymerase inhibitor HPUra, the *oriC*:*terC* ratio increased three- to tenfold [42]. Ciprofloxacin-treated cells (*S. pneumoniae*, *B. cereus*) showed a 6-fold increase in *oriC*:*terC* ratio compared to untreated cells [42]. In several organisms, such as *S.*

pneumoniae and *Vibrio cholerae*, it was shown that genes which are directly or indirectly important for dealing with (replication) stress and cell differentiation are located close to the origin of replication and that the functional genome organization is important for these bacteria to respond adequately to specific conditions [42-44]. Moreover, in *E. coli*, in which the origin of replication was relocated to an ectopic chromosome location, a significant loss of fitness was demonstrated when exposed to sub-inhibitory concentrations of ciprofloxacin [45]. Like 362E, ciprofloxacin generates replication fork stalling, resulting in increased number of copies of genes near the origin in wild-type *E. coli*. Cells with an ectopic origin display a different gene dosage effect, which affects replication fork repair and causes chromosome instability [45]. This experiment shows that the gene distribution in the origin region is pivotal for dealing with specific stresses, such as ciprofloxacin [45]. Nevertheless, the importance of chromosomal gene location has not received much attention yet compared to transcriptional regulation [39].

When analysing the sequence of our laboratory strain 630Δ*erm*, we found that it had acquired an additional copy of an origin-proximal rRNA/tRNA cluster, resulting into a total of twelve ribosomal RNA (*rrn)* operons [1]. Higher copy numbers of *rrn* operons seem to occur more often in fast growing organisms and those with a relatively large genome size [46,47]. It is plausible that strain 630Δ*erm* adapted to fast growth in the laboratory to cope with the increased demand for protein imposed by growth in nutrient-rich media [48]. Our assumption regarding adaptation to fast growth is further supported by the specific location of this extra copy of *rrn* operon, near the origin, which effectively amplifies rRNA gene dosage during rapid growth [49]. Comparative analysis of genomes of more than a thousand bacterial species showed that *rrn* copy number predicts traits associated with resource availability [50]. From an ecological perspective, a nutrient-rich environment enables the bacterium to acquire an additional *rrn* operon, as the substantial energetic cost of ribosome synthesis may be compensated by the excess of nutrients [47], whereas in nutrient-poor conditions an excess of ribosomes may result in suboptimal growth [51,52]. In *E. coli*, it was experimentally demonstrated that the optimal number of *rrn* operons is determined by the advantage of both rapid adjustment to influx of nutrients and the capacity for fast growth [47]. In conditions with fluctuating nutrient levels isogenic *E. coli* strains with 7 (wild-type copy number) and 8 operons were the most effective competitors, whereas in stable, relatively nutrient-poor conditions it was suggested that the number of operons would be lower

over time [47]. In our laboratory, our laboratory strain is subjected to fluctuating conditions when performing growth curves in which overnight starter cultures are diluted to a certain optic density into a fresh medium. To our knowledge, it is unknown if, or to what extent, nutrient levels fluctuate in the gut. The conditions, in terms of temperature and nutrients, between the laboratory and the environment to which a bacterium is subjected, vary considerably. For instance, growth (and doubling-time) of a soil-dwelling organism such as *B. subtilis* will be much slower in its natural habitat where (ambient) temperature will be far below 37 degrees Celsius. Moreover, growth at extreme temperatures causes stress for *B. subtilis* (laboratory conditions) and may even lead to activation of the alternative sigma B factor, and thereby the general stress response [53-56]. Perhaps, growth under laboratory conditions should be considered a stress for *C. difficile* as well.

Interestingly, the gene encoding the alternative sigma factor B homologue (*sigB*) is directly adjacent to a rRNA/tRNA cluster, that includes the additional copy found in our laboratory strain 630Δ*erm*. The linkage between the protein synthesis machinery and the protein σ^B, and the genomic location close to the origin was also noticed by others, and it was suggested that as a result, *sigB* might transiently be expressed (at an early stage of DNA replication) [1]. In most Firmicutes, σ^B governs the so-called general stress response that is important to cope with environmental stresses and starvation [53]. However, the conservation of genes that are σ^B-dependent in one organism cannot be taken as an indication that their σ^B-dependency is equally conserved, as genes regulated by σ^B can differ per species [57]. In *Bacillus subtilis*, many of the factors involved in regulating σ^B activity are encoded in the *sigB* operon (Rsb proteins) [58-60]. When analysing the gene-clusters which involve the genes responsible for regulating *sigB* from other bacteria, it is clear that large differences in operon-structure between species exist; *Bacillus subtilis* and *Listeria monocytogenes* both possess an 8-gene cluster, in contrast to the 4-gene cluster found in *Bacillus cereus* and *Bacillus anthracis* [58]. *S. aureus* has also a 4-gene *sigB* operon, that encodes three homologs of the *sigB* regulatory genes identified in *B. subtilis* (*rsbU*, *rsbV* and *rsbW*). *C. difficile* only has 2 homologues of the *sigB* regulatory genes (*rsbV*, *rsbW*) in common with *B. subtilis* in its 3-gene cluster [61]. The additional factors that influence the regulatory proteins encoded by the operon in *B. subtilis* (*rsbR*, *rsbS*, *rsbT*, *rsbX*) are absent from *S. aureus* [60] and *C. difficile* [1,61], strongly suggesting that *sigB* activation in these organisms is different.

In *C. difficile*, σ^B seems to control a wide array of genes involved in sporulation, metabolism, cell surface biogenesis and the management of stresses [61]. Inactivation of *sigB* reduced survival in the stationary growth phase and increased sensitivity to for instance, acidification, cationic antimicrobial peptides, nitric oxide and reactive oxygen species [61]. Mutants of *sigB* of other bacterial species demonstrated increased susceptibility to a diverse range of environmental and physiological stresses, including ethanol, acidic or alkaline pH, high osmolarity, bile acids, certain antibiotics, temperature variations or oxidative stress [62-66].

Despite that differential expression of *sigB* in response to 362E inhibition was not demonstrated in our experiment, a substantial number of genes that are presumed to be under the control of σ^B were differentially expressed (**Chapter 6, Table 3**). Transient upregulation of *sigB* may have occurred before harvesting the cells in our experiment; such kinetics are known from other organisms and believed to minimize the fitness costs associated with *sigB* expression [53,61,67,68]. Our hypothesis is that the origin-proximal position of *sigB* and the inhibitor induced gene dosage increases transcription of *sigB* and σ^B-dependent genes in order to cope with this particular stress. To test this, *sigB* could be translocated from the origin-proximal location to an ectopic (origin-distal) location on the chromosome and repeat our experiment and transcriptome analysis (**Chapter 6**). The sensitivity to 362E of such a strain in comparison with a wild-type and a *sigB*-mutant strain will reveal the importance of the origin proximal location of *sigB* to cope with inhibition by 362E. Of note, Kint and co-workers reported an increased sensitivity of a *sigB* Clostron mutant towards several antibiotics, supporting the notion that σ^B can be important for dealing with growth inhibition [61]. It is most likely that the transcriptional response to 362E inhibition is driven by a combination of the direct effect of gene dosage based on genomic location, and an indirect effect exerted through the induction of genes dependent on an origin-proximal transcriptional regulator, such as *sigB*.

As mentioned earlier, expression and function of the general stress response deviates between species [53]. Within the group of pathogenic *Clostridia*, strikingly, a *sigB* operon was only identified in *C. difficile* [1] and *C. sordellii* [61], suggesting that the other species may have other stress-responsive transcriptional regulators. The fact that in *C. difficile sigB* is origin-proximal and that in other organisms stress-regulators are functionally clustered at the origin of replication [42,44], raises the intriguing possibility that these regulators in other Clostridia are also located near the origin.

When analysing the genome sequences of *Clostridium tetani* (E88 GenBank: AE015927.1), *Clostridium perfringens* (ATCC 13124; GenBank:CP000246) and *Clostridium botulinum* (ATCC 19397; GenBank:CP000726.1), several transcriptional regulators could be identified. An origin-proximal putative transcriptional regulator (CLB_0016; 14910-15431 bp) was found in *C. botulinum*. In *C. perfringens* and *C. tetani*, genes encoding TetR-family transcriptional regulators are located near the origin (CPF_0034; 42133-42702 bp, and CTC00079; 26462-27028, respectively). This family of regulators is a vast and diverse family of one component signal transduction systems. Most known for its role as regulator of antibiotic efflux, they are known to interact with an exceptionally diverse set of small molecules, including metabolites and cell-cell signalling molecules[69] and regulate a variety of important functions, including osmotic stress, catabolic pathways, homeostasis, biosynthesis of antibiotics, efflux pumps, multidrug resistance, and virulence of pathogenic bacteria[70,71]. TetR-family regulators are more abundant and diverse in microbial species exposed to environmental changes, suggesting important regulatory roles in microbial adaptation[70–72]. Although the function of both regulators in *C. perfringens* and *C. tetani* are not described in literature, it is interesting that they are located origin-proximal.

Antimicrobials targeting DNA replication proteins

In our view, compound 362E has the potential to become a new therapeutic against *C. difficile* infections, as it has a unique mechanism of action, with a low potential for resistance and a lack of cross-resistance with existing agents [73–77]. Furthermore, it has low bio-availability which ensures high local concentration at the site of infection in the gut[77]. Though specificity to *C. difficile* was claimed, activity to this compound was only tested against a small panel of intestinal bacteria, including *Lactobacillus casei*, *Lactobacillus acidophilus*, *Bifidobacterium breve*, *Eubacterium lentum*, and *Bacteroides fragilis* [74]. The conclusion may therefore be premature; indeed, we demonstrated in our study that *S. aureus* is highly sensitive to 362E (**Table 1**).

Table 1. MIC values of control organisms used for testing PolC inhibitors with agar dilution method (expressed in µg/ml)

	HPUra		362E	
Strain	**24h**	**48h**	**24h**	**48h**
Bacteroides fragilis ATCC 25285	32	>64	>16	>16
Bacteroides thetaiotaomicron	32	>64	>16	>16
Clostridioides difficile ATCC 43603 #	2	16	2	4
Clostridioides difficile ATCC 700057 $	2	16	1	2
Clostridium glycolicum	2	16	1	2
Clostridium perfringens	4	16	8	8
Staphylococcus aureus ATCC 29213	0,5	1	1	1

This strain is PCR ribotype 085, tcdA– tcdB–, cdtB not amplified by PCR, serogroup X
$ This strain is PCR ribotype 038, tcdA– tcdB–, cdtB not amplified by PCR, non-toxigenic

Admittedly, *S. aureus* does not reside in the gut, but this finding underlines the importance of assessing how a microbial community, such as the gut microbiome, changes upon exposure to 362E. It would be very interesting to compare the effect on the microbiome of 362E with the one of the standard treatment options of CDI, vancomycin. Considering that broad-spectrum antibiotic treatment is one of the most important risk factors for CDI [78,79], it is desirable that a new antimicrobial against *C. difficile* has a minimal impact on the microbiome. Another approach to improve the specificity of an antimicrobial is to combine antimicrobials with different spectra, or antimicrobials and non-antibiotic compounds, that show synergistic activity towards a specific pathogen, such as *C. difficile* [80,81]. Such combinations may also shorten the duration treatment and/or decrease dosage compared to mono-therapy [82], thereby minimizing the time of exposure and facilitating a decrease in antibiotic concentrations. Improving the specificity of treatment in this fashion may also have a profound effect on the prevention of development of resistance. Combination therapy does not have to imply the use of two separate antibiotics but may be engineered into a single molecule. These hybrid antibacterial agents may be the ultimate combination of two antibiotics, and prevention or reduction of *de novo* generation of complete resistance is based on the assumption that the targets of both moieties are inhibited [83]. However, true dual targeting may not be achieved for certain hybrid agents [84]. This includes the anti-*C. difficile* agent cadazolid, an oxazolidinone-fluoroquinolone hybrid. Despite promising results in Phase I and Phase II clinical trials [85,86], Phase III clinical studies demonstrated it was equivalent or inferior

to vancomycin [87]. Mechanistically, cadazolid is a potent inhibitor of bacterial protein synthesis, but barely inhibits the fluoroquinolone target [88]. Nonetheless, the fluoroquinolone pharmacophore contributes to favourable physicochemical properties of the hybrid molecule, which probably increases the activity of the hybrid [84]. For some molecules, such as MBX-500, there is evidence that they do act as dual-target hybrids. MBX-500 consists of a fluoroquinolone moiety covalently linked to a PolC (PolIIIC)- inhibitor pharmacophore (HB-EMAU) (**Chapter 2**). It was developed as an agent to treat antimicrobial-resistant Gram-positive aerobic pathogens, but is also active against certain *Bacillus* species, *C. difficile* and *S. aureus* strains [89,90]. To this date, no clinical studies in humans have been announced for MBX-500 in relation to treatment of *C. difficile* [91]. The evidence for its true dual targeting is based on a demonstration that full resistance to MBX-500 in *S. aureus* required mutations that individually confer resistance to fluoroquinolones as well as PolC-inhibitors [89]. However, also in this study, the hybrid molecule proved to be strong inhibitor compared to the weak antimicrobial activity of an equimolar combination of parent molecules. This suggests that other antibacterial properties than the effect of dual targeting may be involved. It is possible that the fluoroquinolone-moiety of MBX-500 also influences bacterial cell permeation, as suggested for cadazolid. Unfortunately, no *in vitro* biochemical assays with purified *C. difficile* PolC or DNA topoisomerase have been reported. Performing DNA polymerase activity assay for *C. difficile* PolC [74,92] and DNA supercoiling-and decatenation assays, as described by Locher *et al.* 2014 [88] could shed light on the activity of the compound towards the individual targets. To gain further insight into the mode of action of MBX-500 in *C. difficile* it would be interesting to construct a PolC-fluoroquinolone resistant (double) mutant and to assess if resistance will occur, as seen for *S. aureus* [89]. Elucidation of pathways and target identification of MBX-500 or other antimicrobials in *C. difficile* may be further explored using targeted transcriptional repression using CRISPR interference (CRISPRi) [93]. Recently, progress has been made by using CRISPR-based systems in *C. difficile* [93–97].

If the activity of one of the pharmacophores is dominant over the other, an elegant solution to preserve the physico-chemical properties of the combined molecule could be to link both pharmacophores with a peptide bond which can only be cleaved by an enzyme of *C. difficile*, such as endopeptidase PPEP-1 (ZMP1/CD2830) [98], if possible. The advantage of this approach over combination therapy of the individual parent molecules is that the pharmacokinetics of a hybrid

molecule are more predictable. Furthermore, such a cleavable peptide bond-linker may be also be a new approach for designing new hybrid antibiotics which are inactive (prodrug) [83] until they encounter the pathogen of interest, and therefore display high specificity.

Although the spectrum of activity of MBX-500 is reasonably limited, it still targets a fair number of other low-GC Gram-positive organisms and some Gram-negatives [89,90]. Depending on the organism of interest, the niche of that specific micro-organism and the bio-availability of the antimicrobial compound, this may or may not pose a problem. If MBX-500 is aimed at the treatment of *S. aureus*, the PolC-inhibitor pharmacophore could be refined to improve its specificity to this species, as was done for *C. difficile*-specific compound 362E [74]. Surprisingly, we observed that *S. aureus*, which was one of the control organisms in our experiments (**Chapter 6**), displayed a higher sensitivity to 362E than most of the *C. difficile* strains, suggesting this may be a good starting point for designing such hybrids (**Table 1**). A topical formulation of compound 362E could be considered for treatment of various skin infections caused by *S. aureus*, as its bioavailability is low and *S. aureus*' niche is clearly separated from *C. difficile*'s in the human gut. If the therapeutic potential of compound 362E in relation to *S. aureus* would be further explored, a number of experiments should be performed. Apart from the assessment of antimicrobial activity with the agar dilution method (**Chapter 6**), we do not know whether the exquisite sensitivity of *S. aureus* to 362E is due to its PolC-inhibiting activity or if the compound demonstrates an alternative mode of action. A first step in elucidating the mode of action of compound 362E in *S. aureus*, is to perform a marker frequency analysis (MFA) [42]. If *S. aureus* PolC is the primary target of compound 362E, we expect a similar increase in *oriC:terC* ratio as determined for *C. difficile* (**Chapter 6, Figure 6**). If an increase in *oriC:terC* ratio is confirmed, purifying PolC of *S. aureus* [99] would be the next logical step in order to assess the *in vitro* activity of 362E against this essential replication protein. Identification of the target of 362E in *S. aureus in vivo*, can be performed by induced strain sensitivity assays where components of specific pathways are depleted by for instance antisense interference, leading to increased susceptibility towards a particular compound [100-102].

Biochemistry of replication proteins in drug development

Above we mentioned that *in vitro* experiments using purified proteins (gyrase and DNA-polymerase) are important for drug design and target validation. This has been an important consideration for the work discussed in **chapters 4** and **5** of this thesis. Apart from DNA- polymerase and gyrase, no information on replication proteins or the mechanism of replication in *C. difficile* was available at the beginning of our research. In order to facilitate development of antimicrobials against replication proteins, basic research to identify and characterize these proteins, along with the mechanistic details of their activity and interactions is required. The focus of our attention were replication proteins that together form an essential component of the replisome: the primosome [103-105]. In *B. subtilis*, the involvement of the primosomal proteins helicase loader, helicase and primase in the initiation of chromosomal replication has been recognized for a long time and many details about its functioning have been elucidated [106-110]. We hypothesize that helicase and primase form an interesting target for the development of antimicrobials against *C. difficile*. First, the interaction between helicase and primase is essential for bacterial survival and interference with this interaction should therefore lead to cell death [111] (**Chapter 2**). Helicase-primase interactions can be disrupted with peptides [112], but due to unfavourable pharmacological properties these peptides are not antibacterial. It does indicate however, that the interaction may be druggable. Second, the variation in stability of the interaction between species [113-117] and species-specific cross-stimulation by helicase and primase [118,119] suggest that the molecular details of the interaction may be species-specific and this may therefore also apply for the molecule that inhibits the interaction.

In our study, we successfully characterized aspects of the replicative helicase and primase of *C. difficile* (**Chapters 4** and **5**). We established that the helicase and the helicase loader protein interact directly (**Chapter 4**) and provided indirect evidence for an interaction of helicase and primase *in vitro*, through assays measuring helicase and primase activity in the presence and absence of the partner protein (**Chapter 5**). However, attempts to demonstrate a direct interaction between helicase and primase using a bacterial two-hybrid system have been unsuccessful so far. This situation is reminiscent of the transient and/or weak interactions observed for *E. coli* helicase and primase [114-116], which can only be detected by sensitive techniques [117]. One potentially more sensitive technique to demonstrate

an interaction between helicase and primase is a luciferase protein fragment complementation assay, in which both proteins are genetically fused with two inactive fragments of luciferase [120]. Association of these fragments reconstitutes the enzyme activity, allowing for a luminescence read out. The next step to study this interaction is to apply site-directed mutagenesis, to identify the residues important for the interaction as we have performed for the (self-) interaction of helicase and helicase-loader protein (**Chapter 4, Figure 6**). Interaction between helicase and primase may also be demonstrated with cross-linking experiments.

Although the direct evidence of an interaction between helicase and primase was not established, the activation of helicase and/or the helicase-helicase loader complex by primase as determined in the helicase assay (**Chapter 5**) may be sufficient to design a high-throughput small molecule screen and provide new leads for anti-*C. difficile* agents. Potentially, a molecular beacon-based helicase assay (MBHA) can be employed to measure the activity of helicase (**Figure 2**). In such a Förster resonance energy transfer (FRET)-based assay, one strand of a helicase substrate is made of a molecular beacon of which one end is attached to a fluorescent molecule (donor) and the other a quencher (acceptor) [121–123]. The unique feature of a molecular beacon is that is can form a stem-loop structure (hairpin) upon strand separation by the activity of helicase, thereby quenching the fluorescent signal as they are in close proximity [121] (**Figure 2A**).

A major advantage of a molecular beacon-based assay over other FRET-bases assays is that non-specific binding of a compound can be detected by a decrease in fluorescence in absence of helicase (or ATP). Non-specific hits due to DNA-binding compounds are common in helicase assays and their identification is critical for efficient inhibitor development [123]. The MBHA has to be adjusted and optimized in order to meet the unique criteria for inducing *C. difficile* helicase activity and to safeguard that the inhibition of activity observed is a product of disruption of the helicase-primase interaction. This includes the order of addition of the proteins or small molecules, the effect on the activity of non-related (primase-independent) AAA+ helicase (for instance from *B. subtilis* [125,126]), and appropriate negative controls (e.g. Walker-motif mutants of helicase). For instance, we have learned through our experiments that in the presence of ATP and primase, helicase activity is observed (**Chapter 5, Figure 2A**). Therefore, 96-well plates with the small molecule library should be set up under conditions of low helicase activity (in the absence of ATP, and/or

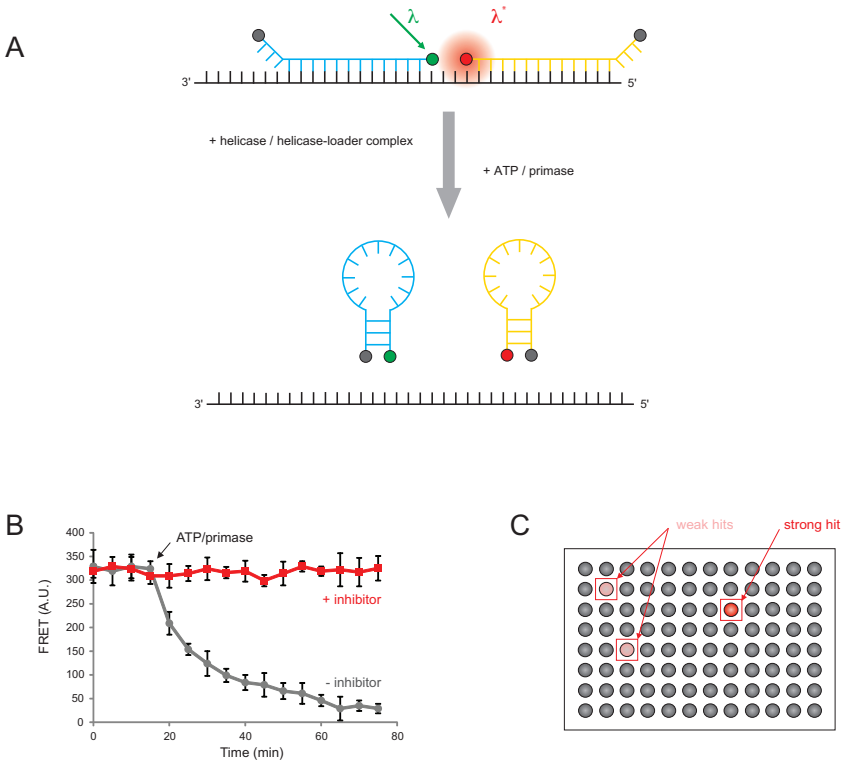


Figure 2. High throughput screen of helicase–primase inhibitors based a molecular beacon-based helicase assay for *C. difficile*.

A. Schematic depiction of the molecular beacon–based assay of helicase activity (based on Belon and Frick (2018)[121], Santangelo (2004)[124]). In the absence of helicase activity, Förster resonance energy transfer between the two fluorescent groups of the molecular beacons (red and green spheres) results in an emission of light (λ*) upon excitation (λ). When the beacons are displaced by helicase, the fluorescent signal is quenched by the proximity of the quencher group (grey sphere). Helicase activity of *C. difficile* helicase requires the presence of ATP and primase.

B. Inhibition of the helicase–primase interaction leads to persistent fluorescence (in arbitrary units, A.U.), whereas helicase activity results in a decrease of fluorescence over time [121]. Data is hypothetical and does not reflect actual experimental data.

C. In a high–throughput screen, inhibitory compounds are readily identified by the persistent fluorescent signal at experimental endpoint. Depicted is a 96–well plate in which to each well a small molecule compound has been added, in addition to helicase, primase and ATP.

primase) and should be monitored for fluorescence upon addition of this compound/protein. Background noise in this assay may be reduced by the addition of the loader protein simultaneously with helicase in the initial plate set-up, as this protein has a negative effect on helicase-activity in a two-hybrid system (**Chapter 5, Figure 2B**).

In conclusion, the work presented in this thesis has provided new insights into the genome of *C. difficile*, and the mechanisms that ensure its faithful replication. We have made important steps in understanding how *C. difficile* responds to antimicrobial compounds targeting DNA replication. Finally, in this discussion we have outlined how our findings can be related to other DNA replication and stress resistance in organisms and how the results of our research can be used in the development of novel antimicrobials.

References

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

2 He, M. *et al.* Emergence and global spread of epidemic healthcare-associated Clostridium difficile. *Nat Genet* **45**, 109–113, doi:10.1038/ng.2478 (2013).

3 He, M. *et al.* Evolutionary dynamics of Clostridium difficile over short and long time scales. *Proc Natl Acad Sci U S A* **107**, 7527–7532, doi:10.1073/pnas.0914322107 (2010).

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

5 Smits, W. K. SNP-ing out the differences: Investigating differences between Clostridium difficile lab strains. *Virulence* **8**, 613–617, doi:10.1080/21505594.2016.1250998 (2017).

6 Collery, M. M. *et al.* What's a SNP between friends: The influence of single nucleotide polymorphisms on virulence and phenotypes of Clostridium difficile strain 630 and derivatives. *Virulence* **8**, 767–781, doi:10.1080/21505594.2016.1237333 (2017).

7 O'Connor, J. R. *et al.* Construction and analysis of chromosomal Clostridium difficile mutants. *Mol Microbiol* **61**, 1335–1351, doi:10.1111/j.1365-2958.2006.05315.x (2006).

8 Roberts, A. P. & Smits, W. K. The evolving epidemic of Clostridium difficile 630. *Anaerobe*, doi:10.1016/j.anaerobe.2018.04.015 (2018).

9 Kuehne, S. A. *et al.* The role of toxin A and toxin B in Clostridium difficile infection. *Nature* **467**, 711–713, doi:10.1038/nature09397 (2010).

10 Lyras, D. *et al.* Toxin B is essential for virulence of Clostridium difficile. *Nature* **458**, 1176–1179, doi:10.1038/nature07822 (2009).

11 Burns, D. A., Heap, J. T. & Minton, N. P. The diverse sporulation characteristics of Clostridium difficile clinical isolates are not associated with type. *Anaerobe* **16**, 618–622, doi:10.1016/j.anaerobe.2010.10.001 (2010).

12 Burns, D. A., Heeg, D., Cartman, S. T. & Minton, N. P. Reconsidering the sporulation characteristics of hypervirulent Clostridium difficile BI/NAP1/027. *PLoS One* **6**, e24894, doi:10.1371/journal.pone.0024894 (2011).

13 Heeg, D., Burns, D. A., Cartman, S. T. & Minton, N. P. Spores of Clostridium difficile clinical isolates display a diverse germination response to bile salts. *PLoS One* **7**, e32381, doi:10.1371/journal.pone.0032381 (2012).

14 Abdallah, A. M. & Behr, M. A. Evolution and Strain Variation in BCG. *Adv Exp Med Biol* **1019**, 155–169, doi:10.1007/978-3-319-64371-7_8 (2017).

15 Blount, Z. D., Borland, C. Z. & Lenski, R. E. Historical contingency and the evolution of a key innovation in an experimental population of Escherichia coli. *Proc Natl Acad Sci U S A* **105**, 7899–7906, doi:10.1073/pnas.0803151105 (2008).

16 Lenski, R. E. Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. *ISME J* **11**, 2181–2194, doi:10.1038/ismej.2017.69 (2017).

17 Blount, Z. D., Barrick, J. E., Davidson, C. J. & Lenski, R. E. Genomic analysis of a key innovation in an experimental Escherichia coli population. *Nature* **489**, 513–518, doi:10.1038/nature11514 (2012).

18 Lang, G. I. & Desai, M. M. The spectrum of adaptive mutations in experimental evolution. *Genomics* **104**, 412–416, doi:10.1016/j.ygeno.2014.09.011 (2014).

19 Spigaglia, P., Barbanti, F., Dionisi, A. M. & Mastrantonio, P. Clostridium difficile isolates resistant to fluoroquinolones in Italy: emergence of PCR ribotype 018. *Journal of clinical microbiology* **48**, 2892–2896, doi:10.1128/JCM.02482-09 (2010).

20 Spigaglia, P., Barbanti, F., Louie, T., Barbut, F. & Mastrantonio, P. Molecular analysis of the gyrA and gyrB quinolone resistance-determining regions of fluoroquinolone-resistant Clostridium difficile mutants selected in vitro. *Antimicrob Agents Chemother* **53**, 2463–2468, doi:10.1128/AAC.01252-08 (2009).

21 Spigaglia, P. *et al.* Fluoroquinolone resistance in Clostridium difficile isolates from a prospective study of C. difficile infections in Europe. *J Med Microbiol* **57**, 784–789, doi:10.1099/jmm.0.47738-0 (2008).

22 van Opijnen, T., Dedrick, S. & Bento, J. Strain Dependent Genetic Networks for Antibiotic-Sensitivity in a Bacterial Pathogen with a Large Pan-Genome. *PLoS Pathog* **12**, e1005869, doi:10.1371/journal.ppat.1005869 (2016).

23 Drlica, K., Malik, M., Kerns, R. J. & Zhao, X. Quinolone-mediated bacterial death. *Antimicrob Agents Chemother* **52**, 385–392, doi:10.1128/aac.01617-06 (2008).

24 Floss, H. G. & Yu, T. W. Rifamycin-mode of action, resistance, and biosynthesis. *Chemical reviews* **105**, 621–632, doi:10.1021/cr030112j (2005).

25 Tomasz, A. The mechanism of the irreversible antimicrobial effects of penicillins: how the beta-lactam antibiotics kill and lyse bacteria. *Annual review of microbiology* **33**, 113–137, doi:10.1146/annurev.mi.33.100179.000553 (1979).

26 Vakulenko, S. B. & Mobashery, S. Versatility of aminoglycosides and prospects for their future. *Clinical microbiology reviews* **16**, 430–450 (2003).

27 Lewis, B. B. *et al.* Pathogenicity Locus, Core Genome, and Accessory Gene Contributions to Clostridium difficile Virulence. *MBio* **8**, doi:10.1128/mBio.00885-17 (2017).

28 Smits, W. K., Weese, J. S., Roberts, A. P., Harmanus, C. & Hornung, B. A helicase-containing module defines a family of pCD630-like plasmids in Clostridium difficile. *Anaerobe* **49**, 78–84, doi:10.1016/j.anaerobe.2017.12.005 (2018).

29 Bauer, M. P. *et al.* Clostridium difficile infection in Europe: a hospital-based survey. *Lancet* **377**, 63–73, doi:10.1016/S0140-6736(10)61266-4 (2011).

30 Tickler, I. A. *et al.* Strain types and antimicrobial resistance patterns of Clostridium difficile isolates from the United States, 2011 to 2013. *Antimicrob Agents Chemother* **58**, 4214–4218, doi:10.1128/AAC.02775-13 (2014).

31 Knetsch, C. W. *et al.* Comparative analysis of an expanded Clostridium difficile reference strain collection reveals genetic diversity and evolution through six lineages. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* **12**, 1577–1585, doi:10.1016/j.meegid.2012.06.003 (2012).

32 Stabler, R. A. *et al.* Macro and micro diversity of Clostridium difficile isolates from diverse sources and geographical locations. *PLoS One* **7**, e31559, doi:10.1371/journal.pone.0031559 (2012).

33 Dingle, K. E. *et al.* Evolutionary History of the Clostridium difficile Pathogenicity Locus. *Genome Biology and Evolution* **6**, 36–52, doi:10.1093/gbe/evt204 (2014).

34 Jackson, R. W., Vinatzer, B., Arnold, D. L., Dorus, S. & Murillo, J. The influence of the accessory genome on bacterial pathogen evolution. *Mobile genetic elements* **1**, 55–65, doi:10.4161/mge.1.1.16432 (2011).

35 Goranov, A. I., Kuester-Schoeck, E., Wang, J. D. & Grossman, A. D. Characterization of the global transcriptional responses to different types of DNA damage and disruption of replication in Bacillus subtilis. *J Bacteriol* **188**, 5595-5605, doi:10.1128/jb.00342-06 (2006).

36 Samadpour, A. N. & Merrikh, H. DNA gyrase activity regulates DnaA-dependent replication initiation in Bacillus subtilis. *Mol Microbiol* **108**, 115-127, doi:10.1111/mmi.13920 (2018).

37 Block, D. H., Hussein, R., Liang, L. W. & Lim, H. N. Regulatory consequences of gene translocation in bacteria. *Nucleic Acids Res* **40**, 8979-8992, doi:10.1093/nar/gks694 (2012).

38 Sauer, C. *et al.* Effect of Genome Position on Heterologous Gene Expression in Bacillus subtilis: An Unbiased Analysis. *ACS synthetic biology* **5**, 942-947, doi:10.1021/acssynbio.6b00065 (2016).

39 Slager, J. & Veening, J. W. Hard-Wired Control of Bacterial Processes by Chromosomal Gene Location. *Trends Microbiol* **24**, 788-800, doi:10.1016/j.tim.2016.06.003 (2016).

40 Sousa, C., de Lorenzo, V. & Cebolla, A. Modulation of gene expression through chromosomal positioning in Escherichia coli. *Microbiology* **143 (Pt 6)**, 2071-2078, doi:10.1099/00221287-143-6-2071 (1997).

41 Jaruszewicz-Blonska, J. & Lipniacki, T. Genetic toggle switch controlled by bacterial growth rate. *BMC Syst Biol* **11**, 117, doi:10.1186/s12918-017-0483-4 (2017).

42 Slager, J., Kjos, M., Attaiech, L. & Veening, J. W. Antibiotic-induced replication stress triggers bacterial competence by increasing gene dosage near the origin. *Cell* **157**, 395-406, doi:10.1016/j.cell.2014.01.068 (2014).

43 Couturier, E. & Rocha, E. P. Replication-associated gene dosage effects shape the genomes of fast-growing bacteria but only for transcription and translation genes. *Mol Microbiol* **59**, 1506-1518, doi:10.1111/j.1365-2958.2006.05046.x (2006).

44 Soler-Bistue, A. *et al.* Genomic location of the major ribosomal protein gene locus determines Vibrio cholerae global growth and infectivity. *PLoS genetics* **11**, e1005156, doi:10.1371/journal.pgen.1005156 (2015).

45 Bos, J. *et al.* in *American Physical Society (APS) March Meeting 2017.*

46 Klappenbach, J. A., Dunbar, J. M. & Schmidt, T. M. rRNA operon copy number reflects ecological strategies of bacteria. *Appl Environ Microbiol* **66**, 1328-1333 (2000).

47 Gyorfy, Z. *et al.* Engineered ribosomal RNA operon copy-number variants of E.coli reveal the evolutionary trade-offs shaping rRNA operon number. *Nucleic Acids Res* **43**, 1783-1794, doi:10.1093/nar/gkv040 (2015).

48 Nannapaneni, P. *et al.* Defining the structure of the general stress regulon of Bacillus subtilis using targeted microarray analysis and random forest classification. *Microbiology* **158**, 696-707, doi:10.1099/mic.0.055434-0 (2012).

49 Vieira-Silva, S. & Rocha, E. P. The systemic imprint of growth and its uses in ecological (meta)genomics. *PLoS genetics* **6**, e1000808, doi:10.1371/journal.pgen.1000808 (2010).

50 Roller, B. R. K., Stoddard, S. F. & Schmidt, T. M. Exploiting rRNA Operon Copy Number to Investigate Bacterial Reproductive Strategies. *Nat Microbiol* **1**, 16160, doi:10.1038/nmicrobiol.2016.160 (2016).

51 Baracchini, E. & Bremer, H. Control of rRNA synthesis in Escherichia coli at increased rrn gene dosage. Role of guanosine tetraphosphate and ribosome feedback. *The Journal of biological chemistry* **266**, 11753-11760 (1991).

52 Stevenson, B. S. & Schmidt, T. M. Life history implications of rRNA gene copy number in Escherichia coli. *Appl Environ Microbiol* **70**, 6670–6677, doi:10.1128/aem.70.11.6670-6677.2004 (2004).

53 Hecker, M., Pane-Farre, J. & Volker, U. SigB-dependent general stress response in Bacillus subtilis and related gram-positive bacteria. *Annual review of microbiology* **61**, 215–236, doi:10.1146/annurev.micro.61.080706.093445 (2007).

54 Price, C. W. in *Bacterial Stress Responses* (eds G. Storz & R. Hengge-Aronis) 179–197 (American Society for Microbiology, 2000).

55 Price, C. W. in *Bacillus subtilis and its Closest Relatives: From Genes to Cells* (eds A.L. Sonenshein, J.A. Hoch, & R. Losick) 369–384 (American Society for Microbiology, 2002).

56 Price, C. W. in *Bacterial Stress Responses* (eds G. Storz & R. Hengge) 301–318 (American Society for Microbiology, 2011).

57 van Schaik, W. *et al.* Identification of the sigmaB regulon of Bacillus cereus and conservation of sigmaB-regulated genes in low-GC-content gram-positive bacteria. *J Bacteriol* **189**, 4384–4390, doi:10.1128/JB.00313-07 (2007).

58 Guldimann, C., Boor, K. J., Wiedmann, M. & Guariglia-Oropeza, V. Resilience in the Face of Uncertainty: Sigma Factor B Fine-Tunes Gene Expression To Support Homeostasis in Gram-Positive Bacteria. *Appl Environ Microbiol* **82**, 4456–4469, doi:10.1128/AEM.00714-16 (2016).

59 Hecker, M. & Volker, U. General stress response of Bacillus subtilis and other bacteria. *Advances in microbial physiology* **44**, 35–91 (2001).

60 Senn, M. M. *et al.* Molecular Analysis and Organization of the (B) Operon in Staphylococcus aureus. *J Bacteriol* **187**, 8006–8019, doi:10.1128/jb.187.23.8006-8019.2005 (2005).

61 Kint, N. *et al.* The alternative sigma factor sigma(B) plays a crucial role in adaptive strategies of Clostridium difficile during gut infection. *Environmental microbiology* **19**, 1933–1958, doi:10.1111/1462-2920.13696 (2017).

62 Bandow, J. E., Brotz, H. & Hecker, M. Bacillus subtilis tolerance of moderate concentrations of rifampin involves the sigma(B)-dependent general and multiple stress response. *J Bacteriol* **184**, 459–467 (2002).

63 Ferreira, A., O'Byrne, C. P. & Boor, K. J. Role of sigma(B) in heat, ethanol, acid, and oxidative stress resistance and during carbon starvation in Listeria monocytogenes. *Appl Environ Microbiol* **67**, 4454–4457 (2001).

64 Giachino, P., Engelmann, S. & Bischoff, M. Sigma(B) activity depends on RsbU in Staphylococcus aureus. *J Bacteriol* **183**, 1843–1852, doi:10.1128/jb.183.6.1843-1852.2001 (2001).

65 Reder, A., Hoper, D., Gerth, U. & Hecker, M. Contributions of individual sigmaB-dependent general stress genes to oxidative stress resistance of Bacillus subtilis. *J Bacteriol* **194**, 3601–3610, doi:10.1128/jb.00528-12 (2012).

66 Zhang, Q. *et al.* SigB plays a major role in Listeria monocytogenes tolerance to bile stress. *International journal of food microbiology* **145**, 238–243, doi:10.1016/j.ijfoodmicro.2010.12.028 (2011).

67 Emerson, J. E., Stabler, R. A., Wren, B. W. & Fairweather, N. F. Microarray analysis of the transcriptional responses of Clostridium difficile to environmental and antibiotic stress. *J Med Microbiol* **57**, 757–764, doi:10.1099/jmm.0.47657-0 (2008).

68 Ternan, N. G., Jain, S., Srivastava, M. & McMullan, G. Comparative transcriptional analysis of clinically relevant heat stress response in Clostridium difficile strain 630. *PLoS One* **7**, e42410, doi:10.1371/journal.pone.0042410 (2012).

69 Cuthbertson, L. & Nodwell, J. R. The TetR family of regulators. *Microbiol Mol Biol Rev* **77**, 440–475, doi:10.1128/MMBR.00018-13 (2013).

70 Ramos, J. L. *et al.* The TetR family of transcriptional repressors. *Microbiol Mol Biol Rev* **69**, 326–356, doi:10.1128/MMBR.69.2.326-356.2005 (2005).

71 Krushkal, J. *et al.* Genome diversity of the TetR family of transcriptional regulators in a metal-reducing bacterial family Geobacteraceae and other microbial species. *OMICS* **15**, 495–506, doi:10.1089/omi.2010.0117 (2011).

72 Ma, D. *et al.* Genes acrA and acrB encode a stress-induced efflux system of Escherichia coli. *Mol Microbiol* **16**, 45–55 (1995).

73 Butler, M. M., Dudycz, L. W., Khan, N. N., Wright, G. E. & Brown, N. C. Development of novel inhibitor probes of DNA polymerase III based on dGTP analogs of the HPUra type: base, nucleoside and nucleotide derivatives of N2-(3,4-dichlorobenzyl)guanine. *Nucleic Acids Res* **18**, 7381–7387 (1990).

74 Torti, A. *et al.* Clostridium difficile DNA polymerase IIIC: basis for activity of antibacterial compounds. *Current enzyme inhibition* **7**, 147–153 (2011).

75 Wright, G. E. *et al.* Active site directed inhibitors of replication-specific bacterial DNA polymerases. *Bioorganic & medicinal chemistry letters* **15**, 729–732, doi:10.1016/j.bmcl.2004.11.016 (2005).

76 Xu, W. C. *et al.* 7-Alkyl-N(2)-substituted-3-deazaguanines. Synthesis, DNA polymerase III inhibition and antibacterial activity. *Bioorganic & medicinal chemistry letters* **21**, 4197–4202, doi:10.1016/j.bmcl.2011.05.093 (2011).

77 Dvoskin, S. *et al.* A novel agent effective against Clostridium difficile infection. *Antimicrob Agents Chemother* **56**, 1624–1626, doi:10.1128/aac.06097-11 (2012).

78 Brown, K. A., Khanafer, N., Daneman, N. & Fisman, D. N. Meta-analysis of antibiotics and the risk of community-associated Clostridium difficile infection. *Antimicrob Agents Chemother* **57**, 2326–2332, doi:10.1128/AAC.02176-12 (2013).

79 Stevens, V., Dumyati, G., Fine, L. S., Fisher, S. G. & van Wijngaarden, E. Cumulative antibiotic exposures over time and the risk of Clostridium difficile infection. *Clin Infect Dis* **53**, 42–48, doi:10.1093/cid/cir301 (2011).

80 Lofmark, S., Edlund, C. & Nord, C. E. Metronidazole is still the drug of choice for treatment of anaerobic infections. *Clin Infect Dis* **50 Suppl 1**, S16–23, doi:10.1086/647939 (2010).

81 Maier, L. *et al.* Extensive impact of non-antibiotic drugs on human gut bacteria. *Nature* **555**, 623–628, doi:10.1038/nature25979 (2018).

82 Soothill, G., Hu, Y. & Coates, A. Can we prevent antimicrobial resistance by using antimicrobials better? *Pathogens* **2**, 422–435, doi:10.3390/pathogens2020422 (2013).

83 Pokrovskaya, V. & Baasov, T. Dual-acting hybrid antibiotics: a promising strategy to combat bacterial resistance. *Expert opinion on drug discovery* **5**, 883–902, doi:10.1517/17460441.2010.508069 (2010).

84 Parkes, A. L. & Yule, I. A. Hybrid antibiotics – clinical progress and novel designs. *Expert opinion on drug discovery* **11**, 665–680, doi:10.1080/17460441.2016.1187597 (2016).

85 Ivarsson, M. E., Leroux, J. C. & Castagner, B. Investigational new treatments for Clostridium difficile infection. *Drug Discov Today* **20**, 602–608, doi:10.1016/j.drudis.2014.12.003 (2015).

86 Roshan, N., Hammer, K. A. & Riley, T. V. Non-conventional antimicrobial and alternative therapies for the treatment of Clostridium difficile infection. *Anaerobe* **49**, 103–111, doi:10.1016/j.anaerobe.2018.01.003 (2018).

87 Wilcox, M. H. *et al.* in *European Congress of Clinical Microbiology and Infectious Diseases (ECCMID)*.

88 Locher, H. H. *et al.* Investigations of the mode of action and resistance development of cadazolid, a new antibiotic for treatment of Clostridium difficile infections. *Antimicrob Agents Chemother* **58**, 901–908, doi:10.1128/AAC.01831-13 (2014).

89 Butler, M. M. *et al.* Antibacterial activity and mechanism of action of a novel anilinouracil-fluoroquinolone hybrid compound. *Antimicrob Agents Chemother* **51**, 119–127, doi:10.1128/AAC.01311-05 (2007).

90 Butler, M. M. *et al.* MBX-500, a hybrid antibiotic with in vitro and in vivo efficacy against toxigenic Clostridium difficile. *Antimicrob Agents Chemother* **56**, 4786–4792, doi:10.1128/AAC.00508-12 (2012).

91 Feher, C., Soriano, A. & Mensa, J. A Review of Experimental and Off-Label Therapies for Clostridium difficile Infection. *Infect Dis Ther* **6**, 1–35, doi:10.1007/s40121-016-0140-z (2017).

92 Wright, G. E. & Brown, N. C. Inhibition of Bacillus subtilis DNA polymerase III by arylhydrazinopyrimidines. Novel properties of 2-thiouracil derivatives. *Biochimica et biophysica acta* **432**, 37–48 (1976).

93 Hawkins, J. S., Wong, S., Peters, J. M., Almeida, R. & Qi, L. S. Targeted Transcriptional Repression in Bacteria Using CRISPR Interference (CRISPRi). *Methods Mol Biol* **1311**, 349–362, doi:10.1007/978-1-4939-2687-9_23 (2015).

94 Hong, W., Zhang, J., Cui, G., Wang, L. & Wang, Y. Multiplexed CRISPR-Cpf1-Mediated Genome Editing in Clostridium difficile toward the Understanding of Pathogenesis of C. difficile Infection. *ACS synthetic biology*, doi:10.1021/acssynbio.8b00087 (2018).

95 McAllister, K. N., Bouillaut, L., Kahn, J. N., Self, W. T. & Sorg, J. A. Using CRISPR-Cas9-mediated genome editing to generate C. difficile mutants defective in selenoproteins synthesis. *Sci Rep* **7**, 14672, doi:10.1038/s41598-017-15236-5 (2017).

96 Neggers, J. E. *et al.* Target identification of small molecules using large-scale CRISPR-Cas mutagenesis scanning of essential genes. *Nature communications* **9**, 502, doi:10.1038/s41467-017-02349-8 (2018).

97 Wang, S. *et al.* Genome engineering of Clostridium difficile using the CRISPR-Cas9 system. *Clin Microbiol Infect*, doi:10.1016/j.cmi.2018.03.026 (2018).

98 Hensbergen, P. J. *et al.* Clostridium difficile secreted Pro-Pro endopeptidase PPEP-1 (ZMP1/CD2830) modulates adhesion through cleavage of the collagen binding protein CD2831. *FEBS letters* **589**, 3952–3958, doi:10.1016/j.febslet.2015.10.027 (2015).

99 Pacitti, D. F., Barnes, M. H., Li, D. H. & Brown, N. C. Characterization and overexpression of the gene encoding Staphylococcus aureus DNA polymerase III. *Gene* **165**, 51–56 (1995).

100 Donald, R. G. *et al.* A Staphylococcus aureus fitness test platform for mechanism-based profiling of antibacterial compounds. *Chem Biol* **16**, 826–836, doi:10.1016/j.chembiol.2009.07.004 (2009).

101 Painter, R. E. *et al.* Elucidation of DnaE as the Antibacterial Target of the Natural Product, Nargenicin. *Chem Biol* **22**, 1362–1373, doi:10.1016/j.chembiol.2015.08.015 (2015).

102 Phillips, J. W. *et al.* Discovery of kibdelomycin, a potent new class of bacterial type II topoisomerase inhibitor by chemical-genetic profiling in Staphylococcus aureus. *Chem Biol* **18**, 955–965, doi:10.1016/j.chembiol.2011.06.011 (2011).

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

104 Chintakayala, K., Larson, M. A., Griep, M. A., Hinrichs, S. H. & Soultanas, P. Conserved residues of the C-terminal p16 domain of primase are involved in modulating the activity of the bacterial primosome. *Mol Microbiol* **68**, 360–371, doi:10.1111/j.1365-2958.2008.06155.x (2008).

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

106 Gross, J. D., Karamata, D. & Hempstead, P. G. Temperature-sensitive mutants of B. subtilis defective in DNA synthesis. *Cold Spring Harbor symposia on quantitative biology* **33**, 307–312 (1968).

107 Imai, Y. *et al.* Subcellular localization of Dna-initiation proteins of Bacillus subtilis: evidence that chromosome replication begins at either edge of the nucleoids. *Mol Microbiol* **36**, 1037–1048 (2000).

108 Karamata, D. & Gross, J. D. Isolation and genetic analysis of temperature-sensitive mutants of B. subtilis defective in DNA synthesis. *Molecular & general genetics : MGG* **108**, 277–287 (1970).

109 Moriya, S., Imai, Y., Hassan, A. K. & Ogasawara, N. Regulation of initiation of Bacillus subtilis chromosome replication. *Plasmid* **41**, 17–29, doi:10.1006/plas.1998.1381 (1999).

110 Sueoka, N. Cell membrane and chromosome replication in Bacillus subtilis. *Progress in nucleic acid research and molecular biology* **59**, 35–53 (1998).

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

112 Gardiner, L., Coyle, B. J., Chan, W. C. & Soultanas, P. Discovery of antagonist peptides against bacterial helicase-primase interaction in B. stearothermophilus by reverse yeast three-hybrid. *Chem Biol* **12**, 595–604, doi:10.1016/j.chembiol.2005.04.007 (2005).

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

114 Lu, Y. B., Ratnakar, P. V., Mohanty, B. K. & Bastia, D. Direct physical interaction between DnaG primase and DnaB helicase of Escherichia coli is necessary for optimal synthesis of primer RNA. *Proc Natl Acad Sci U S A* **93**, 12902–12907 (1996).

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

116 Oakley, A. J. *et al.* Crystal and solution structures of the helicase-binding domain of Escherichia coli primase. *The Journal of biological chemistry* **280**, 11495–11504, doi:10.1074/jbc.M412645200 (2005).

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

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

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

120 Hatzios, S. K., Ringgaard, S., Davis, B. M. & Waldor, M. K. Studies of dynamic protein-protein interactions in bacteria using Renilla luciferase complementation are undermined by nonspecific enzyme inhibition. *PLoS One* **7**, e43175, doi:10.1371/journal.pone.0043175 (2012).

121 Belon, C. A. & Frick, D. N. Monitoring helicase activity with molecular beacons. *Biotechniques* **45**, 433-440, 442, doi:10.2144/000112834 (2008).

122 Hanson, A. M., Hernandez, J. J., Shadrick, W. R. & Frick, D. N. Identification and analysis of inhibitors targeting the hepatitis C virus NS3 helicase. *Methods Enzymol* **511**, 463-483, doi:10.1016/B978-0-12-396546-2.00021-8 (2012).

123 Shadrick, W. R. *et al.* Discovering new medicines targeting helicases: challenges and recent progress. *J Biomol Screen* **18**, 761-781, doi:10.1177/1087057113482586 (2013).

124 Santangelo, P. J., Nix, B., Tsourkas, A. & Bao, G. Dual FRET molecular beacons for mRNA detection in living cells. *Nucleic Acids Res* **32**, e57, doi:10.1093/nar/gnh062 (2004).

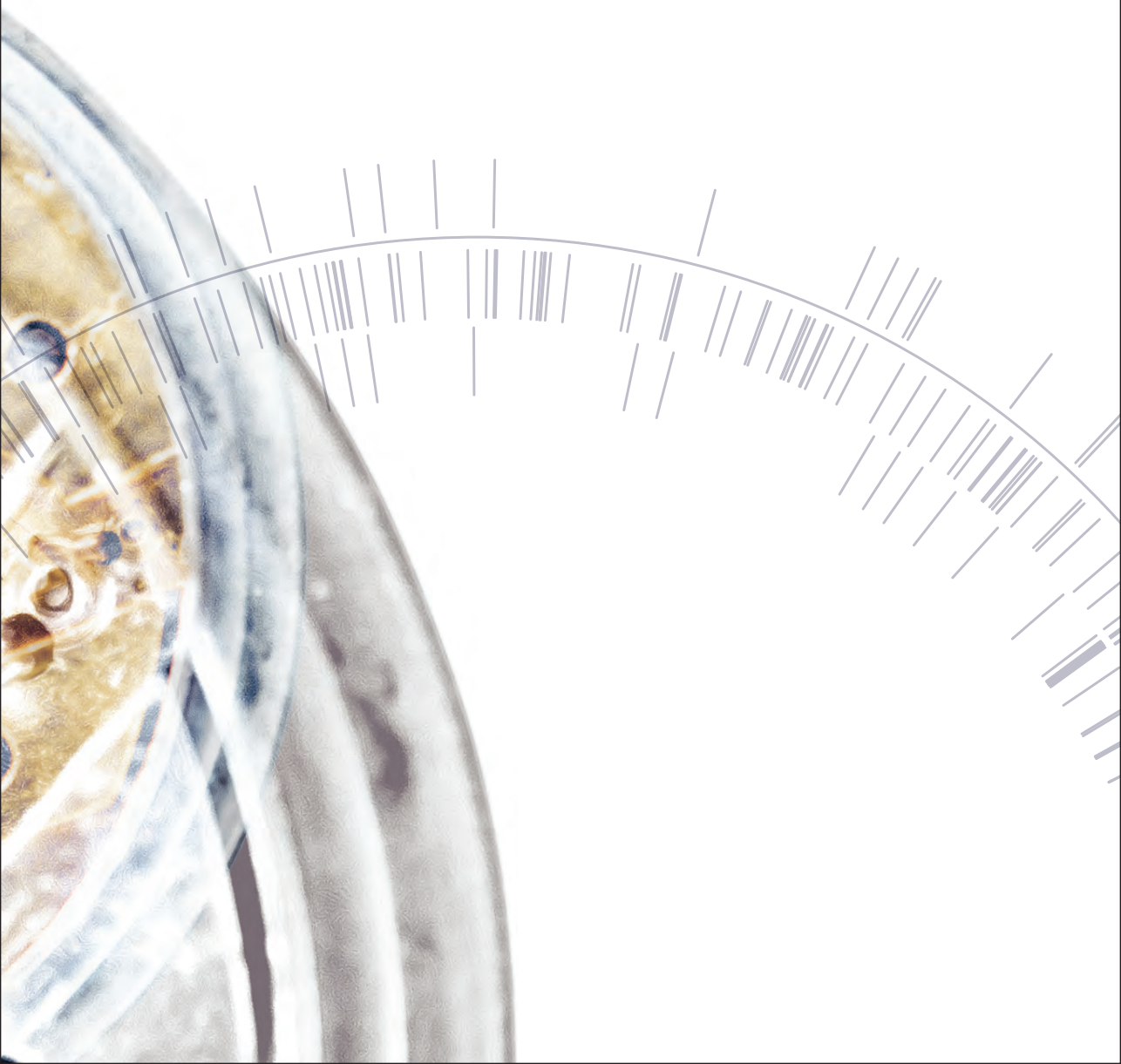
125 Ioannou, C., Schaeffer, P. M., Dixon, N. E. & Soultanas, P. Helicase binding to DnaI exposes a cryptic DNA-binding site during helicase loading in Bacillus subtilis. *Nucleic Acids Res* **34**, 5247-5258, doi:10.1093/nar/gkl690 (2006).

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

List of Publications

Summary in Dutch

Curriculum Vitae



List of Publications

(In chronological order)

van Eijk E, Anvar SY, Browne HP, Leung WY, Frank J, Schmitz AM, Roberts AP, Smits WK.
Complete genome sequence of the *Clostridium difficile* laboratory strain 630Δerm reveals differences from strain 630, including translocation of the mobile element CTn5. BMC Genomics. 2015 Jan 31;16:31. doi: 10.1186/s12864-015-1252-7.

van Eijk E, Paschalis V, Green M, Friggen AH, Larson MA, Spriggs K, Briggs GS, Soultanas P, Smits WK.
Primase is required for helicase activity and helicase alters the specificity of primase in the enteropathogen *Clostridium difficile*. Open Biol. 2016 Dec;6(12). pii: 160272. doi: 10.1098/rsob.160272.

van Eijk E, Wittekoek B, Kuijper EJ, Smits WK.
DNA replication proteins as potential targets for antimicrobials in drug-resistant bacterial pathogens. J Antimicrob Chemother. 2017 May 1;72(5):1275-1284. doi: 10.1093/jac/dkw548.

van Eijk E, Boekhoud IM, Kuijper EJ, Bos-Sanders IM , Wright G, Smits WK.
Genome location dictates the transcriptional response to sub-inhibitory concentrations of PolC-inhibitors in *Clostridium difficile*. Antimicrob Agents Chemother. 2019 Jan 29;63(2). pii: e01363-18. doi: 10.1128/AAC.01363-18.

Samenvatting / Summary in Dutch

Clostridium difficile is een bacterie die in het darmstelsel van gezonde mens en dier aanwezig kan zijn zonder symptomen van diarree; het zogenaamd asymptomatisch dragerschap. Als de normale darmflora verandert van samenstelling, bijvoorbeeld als gevolg van een antibiotische therapie, kan *C. difficile* uitgroeien en symptomen van enterocolitis veroorzaken; een *C. difficile* infectie (CDI). CDI treft ook personen met afweerstoornissen en ouderen met een afwijkende darmflora door andereoorzaken dan een voorafgaande antibiotische behandeling. De symptomen van CDI kunnen variëren van een milde vorm van diarree, welke spontaan kan genezen, tot een levensbedreigende ontsteking van de dikke darm (pseudomembraneuze colitis).

De bacterie is overal in het milieu aanwezig, met name in grond. Aangezien de bacterie strikt anaeroob is, heeft het een strategie nodig om buiten het darmstelsel in een zuurstofrijke omgeving te kunnen overleven. Deze strategie is het vormen van sporen en dit proces wordt gestimuleerd door blootstelling aan stress, zoals een tekort aan voedingsstoffen. De spore vertoont alle eigenschappen van een rusttoestand, waarin stofwisselingsprocessen stil liggen of tot een minimum beperkt zijn. De sporen zijn resistent tegen droogte, hitte, straling en chemicaliën. Via de fecaleorale route kunnen sporen in een nieuwe gastheer belanden en weer uitgroeien tot vegetatieve bacteriën in het darmstelsel. In zorginstellingen kan de verspreiding van sporen tussen patiënten via direct contact plaatsvinden of via een indirecte route zoals de besmette omgeving of het verzorgend personeel van de instelling. Overdracht van sporen speelt een grote rol bij het ontstaan van uitbraken van CDI in een zorginstelling. De bestrijding en preventie van CDI wordt bemoeilijkt door de intrinsieke resistentie van sporen tegen gangbare schoonmaakmiddelen en desinfectantia en het feit dat ze worden uitgescheiden door zowel symptomatische- als asymptomatische patiënten.

CDI kan optreden als de normale darmflora, die een beschermende functie heeft in de darm (zogenaamde kolonisatie resistentie), is aangetast. Het gebruik van

antibiotica wordt derhalve beschouwd als de grootste risicofactor voor het ontstaan van CDI, omdat het een groot deel van de darmbacteriën doodt, maar niet actief is tegen sporen van *C. difficile*. Het verstoorde evenwicht in de bacteriële microflora geeft *C. difficile* de gelegenheid om zich snel te vermenigvuldigen. De door de bacterie uitgescheiden toxinen veroorzaken vervolgens een ontstekingsreactie die de integriteit van de darmwand aantast en het klinische beeld van de infectie verklaart; diarree en colitis. Naast het gebruik van antibiotica, zijn er ook andere risicofactoren voor CDI, zoals de leeftijd en het onderliggend lijden van de patiënt. Deze risicofactoren komen samen in een ziekenhuisomgeving en dragen bij aan de incidentie van deze notoire nosocomiale infectie (ziekenhuisinfectie).

In het begin van dit millennium is er een sterke toename in het aantal en de ernst van *C. difficile* infecties waargenomen in zowel Noord-Amerika als Europa. Deze toegenomen incidentie en morbiditeit waren geassocieerd met uitbraken veroorzaakt door 'hyper-virulente' *C. difficile* stammen, overwegend van de PCR ribotypen 027 en 078. Beide typen veroorzaken ernstige CDI met een grote kans op mortaliteit, maar infecties met ribotype 078 vinden bovendien vaker plaats buiten het ziekenhuis en treffen jongere personen.

Los van de klinische implicaties, zijn er ook aanzienlijke kosten voor de gezondheidszorg gemoeid met CDI. In de Verenigde staten worden de jaarlijkse kosten geschat op 5.4 tot 6.3 miljard dollar. Onlangs heeft er nog een studie plaatsgevonden in meerdere ziekenhuizen in Groot-Brittannië waarin de mediane kosten voor de eerste episode CDI werd geschat op €7.100 en voor een eerste recidief op €8.500. Er is weinig data over de kosten gerelateerd aan CDI in Nederland. In een retrospectieve kostenanalyse uitgevoerd in een universitair medisch centrum, werden de kosten als gevolg uitbraak door ribotype 027 bij 72 patiënten, geschat op €1.222.376. Een groot deel van de kosten werd veroorzaakt door een langere opnameduur van patiënten met CDI, door maatregelen voor contactisolatie en door de extra activiteiten van de afdeling Infectiepreventie.

De huidige therapie voor CDI bestaat uit metronidazol, vancomycine of fidaxomicine. De keuze voor een van deze antibiotica wordt mede bepaald door de ernst van de infectie en de geschatte kans op een recidief. Het middel metronidazol staat ter discussie omdat het duidelijk minder effectief is dan vancomycine en fidaxomicine. Paradoxaal genoeg verstoren metronidazol en vancomycine de normale

beschermende darmflora als ze toegepast worden voor de behandeling van een infectie anders dan CDI, waardoor de kans op CDI toeneemt als het antibioticum wordt gestopt.

Hoewel er geen duidelijke microbiologisch gedocumenteerde resistentie tegen de standaardtherapie is waargenomen, is er wel een toename geconstateerd in het aantal gevallen waarbij de behandeling met metronidazol niet aanslaat (therapiefalen). Er zijn er sporadische meldingen van resistente *C. difficile* stammen die echter niet bevestigd konden worden in andere laboratoria. Toch zijn er voldoende argumenten die de noodzaak voor het ontwikkelen van nieuwe antimicrobiële middelen tegen *C. difficile* rechtvaardigen. Ten eerste, deze bacterie heeft reeds resistentie tegen meerdere antibiotica verworven en daardoor is het zeer aannemelijk dat er na verloop van tijd ook resistentie tegen de standaardtherapie ontstaat. Verder kan orale toediening van vancomycine leiden tot een verhoogde prevalentie van onder meer vancomycine-resistente entercocci (VRE) en/of overgroei van VRE in de darmen. Daarnaast bieden metronidazol en vancomycine geen bescherming tegen het ontstaan van recidieven. Dit wordt vermoedelijk veroorzaakt door de combinatie van een ongewenst neveneffect van de breed-spectrum antibiotica (de aantasting van de gewone darmflora) en het ontbreken van antimicrobiële activiteit tegen de sporevorm van *C. difficile*. Ondanks fidaxomicine een smal-spectrum antibioticum is en er aanwijzingen zijn dat het actief is tegen sporen van *C. difficile* wordt dit middel in Nederland niet gekozen als eerstelijns antibioticum tegen CDI vanwege de relatief hoge kosten. Verder biedt het middel in de groep patiënten met een recidiverende CDI in een kwart van de gevallen geen uitkomst.

Om deze redenen zijn er de afgelopen jaren zowel door de farmaceutische industrie als de academische wereld grote inspanningen geleverd om nieuwe behandelingen te ontwikkelen, gericht op de verschillende stadia van infectie. Dit varieert van vaccinatie, passieve immunisatie, fecale microbiota transplantatie tot nieuwe antibiotica voor preventie en behandeling van (recidiverende) infecties. Tot deze laatste groep behoren een paar veelbelovende kandidaten (ramoplanin, ridinazole, tygecycline) die in verschillende fasen van klinisch onderzoek zijn. Met uitzondering van ridinazole, zijn de “nieuwe” antibiotica gericht op klassieke processen zoals de bacteriële celwand- en eiwitsynthese. Een ander essentieel cellulair proces dat kan worden gebruikt als aangrijpingspunt voor antibiotica is DNA replicatie. Foutloze DNA replicatie is essentieel voor de bacterie en wordt bereikt door een nauwgezette

coördinatie van de actie van de betrokken eiwitten, zoals beschreven in hoofdstuk 1. Globaal genomen is het mechanisme van DNA-replicatie geconserveerd binnen de drie domeinen waarin het leven op aarde is ingedeeld, maar er is aanzienlijke variatie tussen organismen op het gebied van de eiwitten die bepaalde functies uitvoeren, de moleculaire mechanismen en de manier waarop de activiteit van de eiwitten is gereguleerd. Doordat deze eiwitten essentieel zijn voor het replicatieproces én specifiek voor bepaalde bacteriële species lijken deze uitermate geschikt om te fungeren als doelwit voor nieuw te ontwikkelen antibiotica.

Ondanks er veel verschillende eiwitten betrokken zijn bij dit proces, is het aantal middelen gericht tegen replicatie-eiwitten zeer gering. Tot op heden zijn topo-isomerase II-remmers, noodzakelijk voor ontwarring van DNA gedurende replicatie, de enige antibiotica klasse die uitgebreid klinisch worden toegepast. Voor de behandeling van CDI zijn er (nog) geen antibiotica die de DNA replicatie remmen beschikbaar als therapie. In hoofdstuk 2 wordt deze onderontwikkeling onder de aandacht gebracht en wordt er een overzicht gegeven van DNA replicatie eiwitten die mogelijk als doelwit kunnen dienen voor antibiotica in ziekteverwekkende bacteriën. Als eerste stap om inzicht te krijgen in het DNA replicatie proces van *C. difficile* hebben we de replicatie eiwitten van deze bacterie geïdentificeerd op basis van overeenkomsten met replicatie eiwitten van een verwante model-bacterie (*in silico* analyse). Aansluitend wordt er een inventarisatie gemaakt van de werkingsmechanismen van verschillende verbindingen gericht tegen replicatie eiwitten uiteengezet en worden de uitdagingen in de ontwikkeling deze nieuwe groep antimicrobiële middelen onder de aandacht gebracht.

In een aanzienlijk deel van het experimenteel onderzoek op het gebied van *C. difficile* worden erythromycine-gevoelige laboratoriumstammen gebruikt, die allemaal afstammen van de multiresistente stam 630. Deze stammen zijn mondiaal verspreid onder onderzoekers en worden gebruikt om genetische manipulatie te faciliteren. Aangezien het genoom van de originele stam een grote hoeveelheid mobiele genetische elementen bevat en de uitwisseling van deze elementen op frequente basis plaatsvindt, vonden wij het noodzakelijk om het genoom van onze eigen laboratoriumstam te bepalen en te vergelijken met de DNA-sequentie van 630 (om zo een referentie te verkrijgen voor ons eigen werk) (hoofdstuk 3). Variatie in genetische constitutie kan een oorzaak zijn van verschillen in fenotype (waarneembare kenmerken) en maakt het onderling vergelijken van resultaten van

experimenteel onderzoek gecompliceerd. Onafhankelijke replicatie, een van de pijlers van wetenschappelijke verificatie, is door gebruik van verschillende referentiestammen tussen laboratoria moeilijk haalbaar. Uit onze analyse bleek dat er talrijke mutaties in de DNA-sequentie van onze erythromycine- gevoelige laboratoriumstam (630Δ*erm*) waren ten opzichte van de genoomsequentie van de originele stam (630). Dit varieerde van inserties, deleties of veranderingen van de basen waaruit het DNA is opgebouwd, tot grote chromosomale herschikkingen zoals de verplaatsing van een mobiel genetisch element en het verwerven van een gencluster wat betrokken zou kunnen zijn bij eiwitsynthese. Tijdens dit promotieonderzoek werd het verschil tussen gerelateerde stammen (zoals groeisnelheid, spore-productie en beweeglijkheid) ook in onafhankelijk onderzoek nogmaals bevestigd. Onze resultaten pleiten voor het bepalen van de DNA-sequentie van iedere referentie laboratoriumstam voordat er experimenten mee uitgevoerd worden. Bovendien kan dit gedetailleerde overzicht van de genetische verschillen tussen de voorouder bacteriestam (630) en de erythromycine-gevoelige stam de basis vormen voor onderzoek naar de fenotypische verschillen tussen deze stammen.

Voortbordurend op de *in silico* identificatie van de replicatie eiwitten van *C. difficile* en het bepalen van de exacte DNA-sequentie van onze laboratoriumstam zijn er in hoofdstuk 4 en 5 experimenten verricht om de functionele rol van bepaalde replicatie eiwitten te karakteriseren en te valideren. De focus van dit experimentele werk is op helicase (hoofdstuk 4), het eiwit wat verantwoordelijk is voor het ontwinden van de dubbele streng van het DNA, en primase (hoofdstuk 5), wat de aanzet geeft voor de synthese van DNA. Onze experimenten hebben opgehelderd hoe helicase interacteert met een partner-eiwit, de “helicase lader” maar toonden ook aan dat het niet mogelijk was om helicase activiteit te meten met alleen deze twee eiwitten (hoofdstuk 4). Het primase eiwit, wat enkele bijzondere eigenschappen bleek te hebben, was noodzakelijk om helicase te activeren (hoofdstuk 5). De voornaamste conclusie van dit werk is dat de interacties tussen deze drie eiwitten die cruciaal zijn voor de eerste stappen van het kopiëren van het DNA in *C. difficile* verschillen van een verwante model-bacterie.

In hoofdstuk 6 hebben we onderzoek gedaan naar moleculen die DNA-polymerase kan remmen, zoals ook besproken in hoofdstuk 2. Het remmen van dit eiwit leidt tot een stop in DNA replicatie, en daarmee groei van bacteriële cellen en is dus een potentieel nieuw therapeuticum. Tot dusver zijn geen van deze stoffen

beschikbaar in de kliniek voor de behandeling van infecties die zich richten op dit eiwit. De gebruikte remmer is in een pre-klinische ontwikkelingsfase en is in ons onderzoek getest op een grote collectie klinische *C. difficile* stammen. De rationale hiervoor was dat er binnen de species genetische diversiteit bestaat (hoofdstuk 3), wat eventueel van invloed zou kunnen zijn op de antibiotica-gevoeligheid. Met behulp van deze collectie met verschillende (ribo-)typen *C. difficile* is er aangetoond dat het middel werkzaam is tegen alle geteste stammen en niet alleen tegen een specifieke subgroep. Ook is er een analyse uitgevoerd van de veranderingen in genexpressie van *C. difficile* onder invloed van DNA-polymerase remmers. Hieruit blijkt een verhoogde transcriptie (mRNA) van genen nabij het startpunt van DNA replicatie, een verschijnsel dat niet wordt waargenomen onder invloed van andere antibiotica. Wij hebben kunnen vaststellen dat dit samengaat met een toename van het aantal genen (DNA) nabij het startpunt van transcriptie. Afhankelijk van de functie van deze genen, zou dit invloed kunnen hebben op die manier waarop de bacterie reageert op blootstelling aan dit antimicrobiële middel.

Samenvattend, heeft het werk gepresenteerd in dit proefschrift geleidt tot nieuwe inzichten wat betreft het genoom van *C. difficile* en de mechanismen die betrokken zijn bij het nauwgezette proces van DNA replicatie. Ook hebben we belangrijke stappen gemaakt in ons begrip hoe *C. difficile* reageert op antimicrobiële middelen die gericht zijn op DNA-replicatie. Deze bevindingen kunnen bijdragen aan de screening en ontwikkeling van nieuwe therapeutica gericht op dit essentiële proces zowel *in C. difficile* als wellicht andere pathogene bacteriën.

Curriculum Vitae

Erika van Eijk (July 22th, 1976) was born in Aalten, The Netherlands. She is the daughter of Thijs van Eijk and Metteke van Eijk- Noteboom and the younger sister of Dirk. After finishing secondary school in the city of Leiden, she worked as a bartender for several years to fund her back-pack trips in South-East Asia. In 2000, she decided to resume her studies Nutrition and Dietetics at the Haagse Hogeschool in The Hague (higher professional education) and she received her Bachelor certificate in 2003. Subsequently, she finished a course for fitness trainer and worked in a fitness centre for a few years, combining her knowledge of sports and nutrition. In order to pursue a Master's degree, she enrolled in a pre-Master programme (Health Sciences) at the Vrije Universiteit Amsterdam, The Netherlands. In 2009, Erika completed the Master programme Biomedical Sciences at the Vrije Universiteit in Amsterdam (graduated with distinction), with a focus on epidemiology of infectious diseases. Internships of this research master programme were completed at Red Cross in Bangkok, Thailand (in collaboration with Mahidol University, Bangkok) and the Prince Leopold Institute for Tropical Medicine in Antwerp, Belgium. In 2009, Erika was awarded funding for specialization in clinical and molecular microbiology at the London School of Hygiene and Tropical Medicine (MSc Medical Microbiology) in London, United Kingdom (UK). During that Master programme, she completed an internship at the National Institute for Communicable Diseases in Johannesburg, South Africa. In 2011, she started her doctoral research at the Department of Medical Microbiology (section Experimental Bacteriology) of the Leiden Medical University Center (LUMC) under the supervision of Dr. W. K. Smits and Prof. dr. E.J. Kuijper. The work described in this thesis concerns the DNA replication process of the bacterium *Clostridium difficile*, which potentially may serve as target for the development of new antimicrobials, and is published in peer-reviewed international journals. Parts of this work are the result of close collaborations with other research groups, in particular the group of Dr. Panos Soultanas at School of Chemistry, University of Nottingham, UK. Since 2016, Erika works at the Clinical Microbiological Laboratory of the LUMC where she is involved in quality control and process innovation.