Staphylococcal self-loading helicases couple the staircase mechanism with inter domain high flexibility

Cuncun Qiao[1,2], Gianluca Debiasi-Anders[1,2] and Ignacio Mir-Sanchis [1,2,*]

[1]Department of Medical Biochemistry and Biophysics, Umeå University, Umeå, Sweden and [2]Wallenberg Centre for Molecular Medicine, Umeå, Sweden

ABSTRACT

Replication is a crucial cellular process. Replicative helicases unwind DNA providing the template strand to the polymerase and promoting replication fork progression. Helicases are multi-domain proteins which use an ATPase domain to couple ATP hydrolysis with translocation, however the role that the other domains might have during translocation remains elusive. Here, we studied the unexplored self-loading helicases called Reps, present in *Staphylococcus aureus* pathogenicity islands (SaPIs). Our cryoEM structures of the PriRep5 from SaPI5 (3.3 Å), the Rep1 from SaPI1 (3.9 Å) and Rep1–DNA complex (3.1Å) showed that in both Reps, the C-terminal domain (CTD) undergoes two distinct movements respect the ATPase domain. We experimentally demonstrate both *in vitro* and *in vivo* that SaPI-encoded Reps need key amino acids involved in the staircase mechanism of translocation. Additionally, we demonstrate that the CTD´s presence is necessary for the maintenance of full ATPase and helicase activities. We speculate that this high interdomain flexibility couples Rep´s activities as initiators and as helicases.

INTRODUCTION

The initiation of DNA replication is a tightly controlled process in cellular organisms, and it is initiated at a discrete region of the genome called the origin of replication (*ori*C in bacteria) (1). At *ori*C, the initiator of replication protein binds and distorts the double helix of DNA generating a replication bubble and marking the location where other components will assemble to form the replisome. One of these components, the replicative helicase, forms a ring-shaped homo hexamer in bacteria called DnaB. In archaea the replicative helicase is a homo hexameric ring called the mini-chromosome maintenance or MCM, while in eukaryotes the MCM is a hetero hexamer complex that interacts further with other components to form the final replicative helicase called CMG (2,3). Replicative helicases are loaded onto the origin of replication by the helicase loaders in both prokaryotes and eukaryotes. Depending on the system, helicase loaders may act as helicase ring-makers or ring-breakers (4). Ring-breakers induce conformational changes in the stable helicase ring and thereby break the ring. Ring-makers aid the helicase subunit assembly around the DNA so that the ring is formed while being loaded. In these rink-maker systems the hexameric helicase is not sufficiently stable to maintain the ring *per se*. Although in different systems replicative helicases might be loaded onto single-stranded DNA (ssDNA) or double-stranded DNA (dsDNA), in both cases the final product is a hexameric ring-shaped helicase encircling one strand of DNA and excluding the complementary strand.

Once the helicase is loaded onto ssDNA, it couples the nucleotide turnover with ssDNA translocation. In hexameric translocases, three translocation mechanisms have been proposed to link nucleotide hydrolysis (firing) with substrate translocation: stochastic, concerted and rotary firing (5). Rotary firing coupled with a staircase disposition of DNA binding elements has been shown to be conserved in prokaryotic as well as eukaryotic systems (6–10), however, other pieces of evidence indicated that eukaryotic systems may couple non-rotary ATP hydrolysis to translocation (11–14). Ultimately, and whichever is the translocation mechanism, as the helicase unwinds the DNA provides the DNA template strand to the DNA polymerase and allows replication fork progression. Having helicase loaders adds an extra layer of regulation in the system because the loader may be bound to the replicative helicase keeping it inactive until the appropriate cellular signal triggers the replication process. There are also simpler biological systems with fewer layers of regulation where the helicase is loaded through a loader-independent mechanism. These helicases are known as self-loading helicases and quite often they are present in subcellular entities such as viruses and mobile genetic elements (MGEs) (15).

The MGEs used as the model system in this work are the *Staphylococcus aureus* pathogenicity islands (SaPIs) (16). SaPIs are a subfamily of MGEs of about ~15 Kbp in size that are quiescently integrated into the staphylococcal chromosome and they belong to a broader family called

---

*To whom correspondence should be addressed. Tel: +46 766974363; Email: ignacio.mir-sanchis@umu.se

phage-inducible chromosomal islands. SaPIs harbor virulence factors such as antibiotic resistance proteins and the Toxic Shock Syndrome Toxin rendering them medically relevant. Throughout the manuscript we refer to these MGEs as SaPIs or islands interchangeably. SaPIs are highly efficient molecular parasites of staphylococcal phages where the phage is referred to as helper phage [16,17]. Once a helper phage induces a SaPI, it is likely that the bacterial host where both are located will perish, thus, all molecular mechanisms inherently present in the island will be triggered to escape the cell. The induction of the island is highly specific, where only one phage-encoded protein acts as anti-repressor and, by protein-protein interaction, it triggers SaPI excision-replication-packaging cycle [18–20]. After induction, SaPIs unfold a plethora of mechanisms to exquisitely affect and interfere with the phage but not in excess so that its own high transfer is assured. This detailed parasitic strategy ranges from inhibiting phage-encoded anti-repressor protein functions and phage DNA packaging mechanism to exploit bacterial SOS response where SaPI-encoded packaging related proteins are expressed in a timely fashion [21–24]. SaPIs DNA replication, however, has not been investigated structurally to fully understand mechanistically how they work. While in cellular organisms the replication of DNA should be carefully linked to the appropriate cell cycle, in temperate mobile elements such as prophages and SaPIs the cycle is an on/off binary system. After induction, SaPI´s replication is supported by its own replication module, thus being DNA replication one of the mechanisms that supports SaPI transfer and ultimately assures SaPI survival in nature. In this work, we studied the initiators–helicases encoded by SaPIs, called Reps. For simplicity we will refer to them as helicases. These islands contain two more components, in addition to Rep, that allow SaPI's own phage-independent DNA replication: one gene upstream of *rep* coding for a primase (Pri) and the SaPI origin of replication (*ori*), which is located immediately downstream of the *rep* gene [25]. There are three prototypes of Reps: Rep1, Rep2 and PriRep5, encoded by SaPI1, SaPI2 and SaPI5 respectively (Supplementary Figure S1). SaPI-encoded Rep proteins are unrelated at the sequence level showing low sequence similarities among them (Supplementary Figure S1G). Based on homology predictions and published literature [26], the ATPase domain of Rep1 and PriRep5 belong to the superfamily 3 (SF3) of helicases whereas Rep2's ATPase domain belongs to SF6 where archaeal and eukaryotic MCM helicases belong too (Supplementary Figure S1B and C). Structural classification of ATPases associated with diverse cellular activities (AAA + proteins) place the SF3 group in the superfamily III helicase clade, whereas the MCM group of helicases belong to a clade called pre-sensor II insert clade [27]. SaPI5 has the *pri* and *rep* genes fused, so that the putative primase and Rep are part of the same polypeptide chain. The Reps, in addition to having helicase activity also act as the initiator of replication by means of a winged helix domain (WHD) present in their C-terminal domain (CTD), which specifically recognize short penta- or hexanucleotide repeats called iterons [28]. Depending on the SaPI, between six and twelve iterons are located immediately downstream the *rep*'s open reading frame generating the *ori* [28]. Approximately

at the center of the *ori* there is an AT rich region so that half of the iterons are disposed at each side [28,29]. Despite the variability of the number of iterons in different SaPIs, what seems conserved is the presence of at least two iterons at each side of the AT region pointing in opposite directions [28]. No additional components have been identified to aid Reps during loading onto DNA, indicating that Reps are self-loading helicases [25]. Reps do not need ATP for their *ori* binding and melting activities *in vitro* [28,29], which is common among the initiators present in viruses (including phages), but uncommon in bacteria, archaea, and eukaryotes [15,30–32]. The function of the Rep´s N-terminal domain (NTD) is unknown. Recently, distant homologs of SaPI-encoded Rep initiators–helicases have been identified in different and medically relevant staphylococcal genomic islands called the staphylococcal cassette chromosome or SCC*mec* [26,33,34]. Two different variants of Rep homologs have been studied in SCC*mec*, Cch (similar to Rep2) and Cch2 (similar to PriRep5) [26]. The Cch X-ray crystal structure showed that when in helicase mode as a closed-ring hexamer, the WHD responsible for dsDNA binding remains covered by the neighboring subunit. This suggests that for dsDNA recognition by the CTD there should be some rearrangement or conformational change in the hexamer so that the correct area is exposed and available for dsDNA binding. Additionally, the molecular mechanism that these group of helicases employ for translocation along the DNA remains unexplored. We study here Reps from SaPI1 and SaPI5 and show that they have the canonical AAA + fold, thus showing 3′ to 5′ polarity, which is a hallmark of eukaryotic helicases. Our structures complemented with biochemical and *in vivo* analysis showed that Reps employ key residues involved in the staircase mechanism of helicase translocation. We identified a high flexibility of the CTD with respect to the ATPase domain which is conserved among Reps and showed that the presence of the CTD is unexpectedly essential for ATPase and helicase activities. We speculate that this flexibility efficiently couples the functions as initiator and as helicase of these multi-domain proteins.

MATERIALS AND METHODS

DNA manipulations, cloning, protein expression and purification

The primers, plasmids and strains used are listed in Supplementary Tables S1, S2 and S3, respectively. DNA encoding wild-type Rep1 and PriRep5 were amplified by PCR from the genomes of *Staphylococcus aureus* strains U93688 and USA300 genomes, respectively. The PCR products of *rep1* and *prirep5* were cloned into pET28a using the restriction enzyme pairs NdeI/Xhol and NcoI/XhoI, respectively. For the *prirep*5 CTD truncation, the oligos Rep_SaPI5HisNcoI_F and Rep_SaPI5_S675_XhoI_R were used to amplify the corresponding region, which was cloned into pET28a vector. A new construct containing both *pri* and *rep* genes from SaPI1 was generated to solve the structure of the Rep1–DNA complex (pIMS503). Note that the nomenclature Rep1–DNA complex is used throughout the manuscript because the primase

was not visualized in the map. For this construct, the oligos SaPI1-Pri-NdeHis-1F and SaPI1_Rep_XhoI_3R were used to amplify the *pri-rep*1-containing operon and cloned into pET21a. Although the *pri-rep*1 construct (pIMS503) was used to successfully solve the structure of Rep-1-DNA complex, the *rep*1 construct (pIMS502) was used for biochemical studies, for solving the Rep-apo structure and as template to generate the mutant variants. Site-directed mutagenesis was employed to generate the Rep1 Triple Mutant (residues R248, Y251 and K253 were substituted to alanine: R248A-Y251A-K253A) and PriRep5 K591A variants from the aforementioned *rep1* and *prirep5* genes cloned into pET28a using primer pairs PriRep1_248A_251A_253A_F/PriRep1_248A_251A_253A_R and Rep5_K591A-F/Rep5_K591A-R, respectively. All of the resulting proteins had a His-tag moiety at their N-terminus. The clones were transformed into BL21 (DE3) Rosetta cells which were grown in Luria Broth medium supplemented with 50 µg/ml kanamycin and 25 µg/ml chloramphenicol at 37°C until they reached an OD_{600} of 0.6–0.8. The *Escherichia coli* cultures carrying the PriRep1 construct were supplemented with 100 µg/ml ampicillin. Protein expression was induced by addition of 0.5 mM IPTG at 18°C for 16 h. Cell pellets were harvested and re-suspended in Buffer A containing 50 mM sodium phosphate (pH 7.0), 5% glycerol, 1 mM DTT and 1 M NaCl in the presence of a protease inhibitor cocktail (Roche) and then lysed with 0.5 mg/ml of lysozyme for 30 min at 37°C followed by sonication. After centrifugation, the supernatant was loaded onto a HisTrap™ HP column equilibrated with Buffer A, and the target protein was eluted with a gradient from 0 to 500 mM imidazole using an ÄKTA pure system. Rep1 and PriRep5 peak fractions were concentrated and injected into HiLoad® 26/600 Superdex® 200 pg columns (Cytiva) equilibrated with Buffer B (20 mM Tris pH 8.0, 200 mM NaCl, 0.5 mM EDTA and 2 mM DTT). The eluted fractions were analyzed by SDS-PAGE and those containing the target protein were pooled. After dialysis in stock buffer containing 20% glycerol, the proteins were concentrated, flash-frozen in liquid nitrogen, and stored at –80°C. For PriRep1 and PriRep5-ΔCTD, after the elution from the HisTrap column, the pooled fractions were dialyzed with 10 mM sodium phosphate pH 7.0, 0.4 M NaCl, 5% glycerol, 1 mM DTT, and injected into a HiTrap™ Heparin HP column (Cytiva) equilibrated in the same buffer. Heparin-bound protein was eluted with a gradient concentration of NaCl from 0.4 to 2 M. Positive fractions were pooled and injected into a Superdex 200 26/300 column (Cytiva) as a last step. Positive fractions were dialyzed, concentrated and stored in the same way as previous constructs. Rep1 Triple Mutant (Rep1-TM) was purified in a similar manner to PriRep1 and PriRep5-ΔCTD, except the dialysis prior to injection into the HiTrap™ Heparin HP column was done in 20 mM Tris pH 8.0, 0.2M NaCl, 5% glycerol, 0.5 mM EDTA, 1 mM DTT (final pH adjusted to 7.0). Purification of PriRep5 K591A was done in a similar manner to Rep1 Triple Mutant, except all buffers were adjusted to pH 8.0 and there was an additional purification step through a HiTrap™ Capto Q (Cytiva) anion exchange column after pooling and dialysis of peak fractions from the Heparin column. Protein concentration values throughout the text refer to hexamers.

Size exclusion chromatography

A superdex 200 26/300 column (Cytiva) connected to an ÄKTA Pure 25 M1 GoldSeal with F9-C fraction collector was used to perform gel filtration analysis at 4°C. The system was equilibrated with buffer containing 20 mM Tris pH 8.0, 200 mM NaCl, 0.5 mM EDTA and 2 mM DTT. The standard proteins were obtained from Sigma Aldrich (MWGF1000-1KT): carbonic anhydrase (29 kDa), albumin bovine serum (66 kDa), alcohol deshidrogenase (150 kDa), β-amylase (200 kDa), apoferritin (443 kDa), thyroglobulin (669 kDa).

Electrophoretic mobility shift assay (EMSA)

EMSAs were performed as described in ([26](#)) with minor modification. Protein stocks were diluted in protein dilution buffer (20 mM Tris pH 8.0, 200 mM NaCl, 0.5 mM EDTA, 20% glycerol, 2 mM DTT and 200 mM NaCl). In the case of Rep1 EMSAs, protein stocks were diluted in protein dilution buffer containing 5% glycerol instead of 20%. The origin regions of SaPI1 and SaPI5 were amplified with the primer pairs ori1_FAM-F/*Ori*1_SaPI1_BamHI_R and FAM-Ori_SaPI5_F and Ori_SaPI5_HindIII_R, generating PCR fragments of 367 and 318 bp, respectively. After PCR purification using a Qiagen PCR cleaning kit, 2 nM of FAM-labeled ori-containing PCR were used as substrate dsDNA and mixed with Rep protein in binding buffer (20 mM Tris, pH 8.0, 100 mM NaCl, 5% glycerol, 50 ng/µl BSA, 5 mM DTT, 10 mM $MgSO_4$ and 1 ng/µl salmon sperm DNA). The total volume of the reaction was 20 µl. After incubation at 37°C for 30 min, 4 µl loading buffer (20 mM Tris pH 8.0, 60% glycerol, 30 mM EDTA and 0.03% xylene cyanol/bromophenol blue) was added to the reaction and the entire volume was then loaded on native 4% polyacrylamide gels (acrylamide/bis = 29:1). Gels were run in 0.5× TBE at 4°C for 4 h. The gels were imaged using a Typhoon 9400 (Amersham Biosciences) or Amersham Typhoon (Cytiva). The *ori* DNA of SaPI5 was used as non-specific binding control for Rep1, whereas the ori DNA of SaPI2 was used as control for PriRep5 bindings. The control *ori*2 DNA comprising 347 bp was amplified with primer pairs *ori*2-FAM-F/*Ori*2_SaPI2_HindIII_R. For EMSAs with PriRep5-ΔCTD, the *ori* region of SaPI5 was amplified by PCR using the same oligonucleotides as for the wild-type version. The reaction was performed as described above for the other EMSAs. For the bindings of PriRep5-WT and PriRep5-ΔCTD to ssDNA, 2 nM of 27-mer poly(dT) oligonucleotides were used as substrate. The gels were quantified using ImageJ software and the data were fitted with the Hill equation using Prism 9.3.1 (GraphPad) to calculate the K_d and the Hill coefficient (*n*).

Helicase assays

Different combinations of DNA were prepared by annealing the 5′-FAM-labeled oligo Rep_helassFAM-R with the following complementary oligos: Rep_helass-F/

Rep_helassFAM-R (8 bp, mismatched); Rep_heliass-F2/ Rep_helassFAM-R (8 bp, 5′ ssDNA tail); Rep_heliass-F3/ Rep_helassFAM-R (8 bp, 3′ ssDNA tail). Annealing was performed by mixing equal volumes of equimolar oligonucleotides in a solution containing 200 mM NaCl. After being incubated at 90°C for 20 min, the microtube was allowed to cool down overnight to room temperature. The reaction buffer was composed of 20 mM Tris (pH 8.0), 100 mM NaCl, 10 mM $MgCl_2$, 5% glycerol, 5 ng/ml BSA, 5 mM DTT and 10 mM ATP or other nucleotide analogs. The protein and the 5 nM FAM-labeled duplex DNA were incubated at 37°C for 3 min, and 2 μl of 500 nM unlabeled Rep_helass-R was added into the reaction system. The reaction was terminated 27 min later by addition of stop buffer (75 mM EDTA, 1.25% SDS, 25% glycerol, 0.01% xylene cyanol/bromophenol blue). For the heating control (H), the sample was heated at 95°C for 5 min then transferred immediately to ice. Aliquots of 25 μl were run on 10% acrylamide native gels in 0.5× TBE. at room temperature for 3 h. Gels were visualized with Typhoon 9400 (Amersham Biosciences) or Amersham Typhoon (Cytiva). The bands were quantified with software ImageJ (35). To remove the background, values were normalized against the controls with the ecuation unwound DNA(%) = ($ssDNA_{sample}$ − $ssDNA_{0\_protein}$) / (($ssDNA_{sample}$ − $ssDNA_{0\_protein}$) + ($dsDNA_{sample}$ − $dsDNA_H$)). To test the unwinding activities of PriRep5-ΔCTD with forked DNA with longer overhang, Rep_helass-R was labeled by 32P-γ-ATP (PerkinElmer). Annealing was performed by incubating 32-P labeled Rep_helass-R with 10-fold excess of Rep_helass_R and Rep_helas_38-F in a solution containing 200 nM NaCl at 90°C for 20 min, followed by overnight cool down to room temperature. The reaction was the same as what was described above. After reaction, DNA was resolved by electrophoresis in 10% polyacrylamide gels and visualized by autoradiography.

ATPase activity assay

The ATP hydrolysis activity of PriRep5-WT and PriRep5-ΔCTD at 37°C was determined using BIOMOL® GREEN Reagent for phosphate detection (BML-AK111, Enzo Life Sciences). Briefly, 0.01 μM of PriRep5 and mutant proteins were incubated with 1 mM ATP in assay buffer (40 mM Tris, pH 7.5, 80 mM NaCl, 8 mM MgAc and 1 mM EDTA) at different time points. The reaction was stopped by adding 0.5 M EDTA at different time points and subsequently, 100 μl of BIOMOL® GREEN reagent was added for the quantification. The absorbance of the sample as well as the manufacturer provided standards was measured at 620 nm on a microtiter plate reader FLUOstar Omega (BGM Labtech). 27mer-poly(dT) was added at different concentrations (5, 10, 50 and 100 nM) to test whether ssDNA could stimulate the ATPase activities of PriRep5-WT and PriRep5-ΔCTD. The reaction was stopped after 30 min and measured as above.

Specimen preparation for Cryo-EM and data collection

For Rep1, an aliquot of the protein was passed through a Superdex 200 26/300 column to remove the glycerol. 0.4

mg/ml of the Rep1 sample was used for plunge freezing. 0.4 mg/ml of PriRep5 containing 0.5% glycerol was incubated with 5 mM $MgCl_2$ and 5 mM ADPNP for 20 min at room temperature, then used for specimen preparation. 0.22 mg/ml of PriRep1 containing 4% glycerol was incubated with 5 mM $MgCl_2$, 2mM ATPγS and 2.5 μM of duplex oligo (Rep-2in-FAM-F/Rep-2in_79bp-R) during 30 min. In all cases, 4 μl of the sample were applied to freshly glow discharged carbon grids, blotted for 5 s, and flash frozen in liquid ethane using a Vitrobot (Thermo Fisher Scientific) with a controlled humidity of 100% and temperature of 4°C. Quantifoil 2/2 was used for PriRep5, while Quantifoil 1.2/1.3 and 2 nm-carbon coated grid Quantifoil 2/2 were used for Rep1. Quantifoil 2/1 was used for PriRep1–DNA. Data were acquired by a Titan Krios Cryo transmission electron microscope operating at 300 kV using EPU as the automated data collection software. The images were collected using a Gatan K2 BioQuantum (energy filter) 4k × 4k direct electron detector camera at a normal magnification of 165 000× and a pixel size of 0.82 Å per pixel. For Rep1, PriRep5 and PriRep1–DNA data sets the following micrographs and defocus ranges (μm) were acquired: 3,345, 1.2–2.4; 3,762, 1.2–2.7; 3,121, 1.5–3. Complete data collection parameters are summarized in Supplementary Table S4.

Image processing and 3D reconstruction

The workflows for 3D reconstruction of PriRep5, Rep1-apo and PriRep1–DNA are illustrated in Supplementary Figures S2A, S6A and S7A respectively. Image processing was performed with RELION 3.1 (36). The movie frames were first aligned and superimposed with MotionCorr2.0 (37) and CTF parameters were calculated with GCTF (38). Around 2,000 particles were picked based on Laplacian-of-Gaussian auto picking and then used to generate several 2D averages that were subsequently used as a template for automatic picking. Automatic particle picking was then performed for the whole dataset. 724,786 particles of PriRep5 and 510,024 of Rep1 were initially picked. Several rounds of 2D classifications were performed, and particles in bad classes were removed. Ab-initio 3D model implemented in RELION was used to generate the initial map as the reference for 3D auto-refinement. After 3D classification, the final sets of good particles were 91,171 of Rep1-apo and 185,358 of PriRep5. Subsequently, three rounds of CTF refinements were performed: first refining magnification anisotropy; then refining fourth order aberrations; and finally refining per-particle defocus and per-micrograph astigmatism. A final round of auto-refinement was carried out after Bayesian polishing to optimize per-particle beam-induced motion tracks, resulting in a 4.5 Å map of Rep1-apo and 3.28 Å map of PriRep5 in C1 symmetry. Another round of 3D refinement with C6 symmetry was performed, resulting in a 3.9 Å map of Rep1-apo and 3.1 Å map of PriRep5. The resolution estimation was based on the FSC 0.143 threshold. The 3D map was corrected for the modulation transfer function of the detector and was sharpened by applying a negative *B*-factor. Local resolution was estimated using RELION. The final C1 map of Rep1-apo and C6 map of PriRep5 were subsequently multibody refined (39) where three bodies encompassing the NTD, ATPase

domain, and CTD were defined as illustrated in Supplementary Figure S5A.

For the 3D-reconstruction of PriRep1–DNA, 3121 movies were first aligned with MotionCorr2 and the defocus values were determined with GCTF. Particle picking was performed as above. Total 420,745 particles were picked initially and extracted with a box size of 360×360 pixels, followed by four rounds of 2D classification resulting in a set of 101,309 particles. No symmetry was imposed for subsequent processing. Initial 3D auto-refinement gave an initial map with an overall resolution of 4.1 Å. The resolution reached 3.51 Å after Bayesian polishing and three cycles of CTF refinement as described above. The map was then used as a reference for a new autopicking, giving a total of 538 681 particles. A set of 258,945 particles was selected after four rounds of 2D classification. After two rounds of 3D classification a total of 141,445 particles were extracted. An initial 3.74 Å map was CTF refined and polished giving a resolution of 3.14 Å. The 3D classification and local resolution was subjected. The 3DFSC curves shown in Supplementary Figure S8 were calculated using Remote 3DFSC Processing Server (40).

Structural modeling, refinement, and validation of PriRep5

Buccaneer (41) was used for initial automatic model building in maps not subjected to multibody refinements. All models were further built manually in Coot (42). Density for the CTD of PriRep5 was incomplete, and we built this part based on a homology model generated by the I-TASSER server (43). The entire PriRep5 model was further fitted and energy minimized using Namdinator (44) followed by several rounds of rigid-body refinement using Phenix (45) and subsequent manual adjustment in Coot. The density visualized in the center of Rep1–DNA map was fitted with the arbitrary sequence 5′-TAAAA-3′ followed by real-space refinement in Phenix. The final models were validated using MolProbity in Phenix. Structural figures were prepared using ChimeraX (46,47), Pymol (https://pymol.org/2/) and Biorender.com.

Software resources were provided by SBGrid (48).

In vivo assays, western and Southern blots

The *ori*1 and *ori*5 were cloned into separate pMAD (49) constructs, generating ori1-pMAD and ori5-pMAD, respectively. These were then transformed via electroporation into RN4220 *S. aureus*. All culturing protocols involving RN4220 were conducted either with Tryptic Soy Broth (TSB) as liquid media or Tryptic Soy Agar (TSA) as solid media. Genes *pri* and *rep* from SaPI1 (*pri-rep*1) as well as *prirep* from SaPI5 (*prirep*5) were cloned into a modified version of pCN51 (pCN51_cat194), where the original *erm*C cassette had been replaced with the *cat*194 cassette from pCN50 (50), generating a chloramphenicol-selective version of the pCN51 construct. From these, mutants were generated through site-directed mutagenesis. Oligonucleotide pairs PriRep1_deltaCTD_1–346_F/PriRep1_deltaCTD_1–346_R and PriRep5_deltaCTD_1–681_F/PriRep5_deltaCTD_1–681_R were used to introduce stop codons covering all three reading frames in the CTD of both Rep1 and PriRep5 respectively so that Rep1 protein was truncated after residue K346 and so was PriRep5 from residue V681 in these constructs for the *in vivo* assays.

All construct variants were then electroporated into RN4220 strain positive for either *ori*1-pMAD or *ori*5-pMAD. Cultures carrying both *ori*-containing pMAD and gene-variants-containing pCN51_cat194 were grown overnight at 30°C, 200 RPM in liquid media containing chloramphenicol at 10 μg/ml and erythromycin at 2.5 μg/ml. The overnight cultures were diluted to a starting OD_{600} of 0.1 and grown at 30°C, 200 RPM to an OD_{600} of 0.3. Gene expression in pCN51_cat194 was induced with the addition of 7.5 μM $CdCl_2$ to each culture and incubated at 30°C, 200 RPM for an additional 2 h. 1 ml aliquots were taken at T + 0 and T + 2 h post-induction. Processing of the aliquots as well as the general Southern blot procedure were done in accordance with reference (19). To confirm protein expression, we generated FLAG-tagged versions of all Pri-Rep1 and PriRep5 variants. The addition of a FLAG tag to the N-terminus of PriRep5 was done via site-directed mutagenesis directly on the previosiuly generated pCN51_cat194 constructs. For Pri-Rep1, since direct addition of a FLAG-tag sequence to the 5′ end of *rep*1 via site-directed mutagenesis proved difficult, both *pri*1 and 5′ tagged FLAG-*rep*1 were sequentially introduced into the multiple cloning site of pCN51_cat194 as separate cassettes. FLAG-*rep*1 was generated with regular amplification PCR, which was followed by another round of PCR both introducing a restriction site (BamHI, allowing for ligation to the region directly flanking the stop codon of *pri*1) and reintroducing the intergenic region upstream of FLAG-*rep*1. Having generated the FLAG-tagged versions of both Pri-Rep1 and PriRep5, the aforementioned *in vivo* experiments were repeated and colorimetric Western blots were conducted in the post-induction aliquots (T + 2 h). Pellets were resuspended in 25 μl PBS and lysed with lysostaphin at 0.25 mg/ml for 30 min at 37°C, after which SDS-PAGE loading buffer was added to the samples which were then incubated at 98°C for 10 min and further centrifuged at 14 000 RPM for another 10 min. 10 μl of the lysed samples were then separated by SDS-PAGE and transferred onto Amersham Hybond P 0.45 PVDF blotting membranes (Cytiva) overnight at 20V. 3 μl of PageRuler™ Plus Prestained Protein Ladder, 10 to 250 kDa (Thermo Scientific) was used as a marker for the Western blot experiments. Membrane staining was carried out with mouse ANTI-FLAG® M2 primary antibody (Sigma-Aldrich) and alkaline phosphatase-conjugated anti-mouse goat secondary antibody (Sigma-Aldrich). The membranes were then briefly incubated with a chromogen (SIGMAFAST™ BCIP®/NBT, Sigma-Aldrich) according to the manufacturer's specifications and the resulting bands were recorded with a ChemiDoc™ Touch Imaging System (Bio-Rad).

RESULTS

Rep proteins bind their cognate origins and have 3′ to 5′ polarity

Rep proteins from SaPI1 and SaPI5 islands were cloned, expressed in *E. coli*, and purified (Supplementary Figure S1).

We first verified that Rep proteins are stable hexamers in solution, thus suggesting that Rep helicases might self-load through the ring-breaker mechanism. The dsDNA binding activity of SaPIbov-Rep (the homolog of Rep2 in Supplementary Figure S1) and Rep1 was previously characterized *in vitro* (28). PriRep5 (SaPI5) is not yet characterized. To characterize PriRep5 and fully characterize Rep1, we performed electromobility shift assays using FAM-labeled *ori*-containing PCR as substrate. Figure 1A shows that both Rep proteins could bind to their corresponding iterons-containing substrates in the absence of ATP, thus indicating that they were active and exhibited the ATP-independent binding activity to their origins, a conserved hallmark of these types of initiators (28,29). To confirm binding specificity, we used FAM-labeled *ori*-containing PCRs as substrates from other SaPI origins as controls. We used *ori*5-containing PCR as control for Rep1 and *ori*2 as control for PriRep5. As shown in Figure 1A, PriRep5 bound their cognate origin with higher affinity than the control DNA, whereas Rep1 showed a low discrimination factor in our experimental setup indicating that either we have not found proper *in vitro* conditions or that some other unidentified specificity factor is present in physiological conditions. In both cases, fitting the binding curves with the Hill equation suggested positive cooperativity with two potential binding sites, as expected for one hexamer binding at one side of the *ori*'s AT rich region and a second hexamer binding the other side (Supplementary Figure S1F). Although Cch encoded by SCC*mec* has been proven to be a functional homolog of MCM replicative helicases (26) in terms of polarity, SaPI-encoded initiators have not been characterized in this regard. To do so, we performed the helicase assay as in (26) but using FAM labeled oligonucleotides. First, we used 38-mer oligonucleotides duplexes with an 8-nucleotides mismatch on one side to generate both 3′ and 5′ ssDNA. As shown in Figure 1B, all Rep proteins tested were active helicases, and single-stranded DNA oligonucleotides could be resolved in the presence of ATP but not in the absence of ATP or in the presence of ADP or non-hydrolysable ATP analogs ADPNP and ATPγS. We then used a set of three different substrates to investigate the polarity of these helicases, including 3′ overhang and 5′ overhang double-stranded oligos (Figure 1C). In all the cases, the Rep proteins showed the highest activity when a forked substrate was used followed by a substrate with a 3′ overhang and a smaller amount of activity was seen for the 5′ overhang in both cases. Our findings show that in addition to the helicases present in SCC*mec* elements, namely Cch and Cch2 (26,33), the helicases present in SaPI1 and SaPI5 are also functional homologs in terms of polarity to eukaryotic helicases such as E1 and MCMs translocating along the leading strand (5,14,51). The fact that these Reps showed the highest activity with forked substrate suggests that they might use a similar mechanism for strand separation than in other eukaryotic and bacterial systems (52–54).

CryoEM structure of PriRep5

To mechanistically understand these initiators–helicases, we solved the PriRep5 structure at an overall resolution of 3.2 Å (Figure 2 and Supplementary Figure S2). PriRep5 is a 790 amino acid residues protein and based on our sizing column analysis it forms stable homo hexamers in solution (Supplementary Figure S1). The PriRep5 cryoEM map showed a three-tiers barrel composed of six subunits in a pseudo six-fold symmetry disposition. Density for residues 1–322 could not be seen due to its high degree of flexibility. Figure 2A and B shows the PriRep5 cryoEM map and model. The NTD that could be resolved in our structure (residues 317–460) was composed of alpha helices 1–4 and beta sheets 1–8 grouped independently (Supplementary Figure S3). The middle tier of the PriRep5 barrel is the ATPase domain (Figure 2B and Supplementary Figure S3). The closest structural homologs pointed out by the Dali server (55) were the superfamily 3 helicases (SF3). The helicase domain of bacteriophage Nrs-1 polymerase was the closest homolog (56). As for all other replicative helicases, the nucleotide binding site was formed at the interface of adjacent subunits. Among the six nucleotide binding sites present within the hexamer, four showed density that corresponded to ADPNP (Supplementary Figure S4). All conserved catalytic residues were present in our structure in agreement with the extensive literature of AAA + proteins (1,27,31,57,58). Replicative helicases couple translocation with ATP hydrolysis (firing). In the rotary cycling model (5), nucleotide firing occurs sequentially from one subunit to the next and correlates with the height of the beta hairpins located inside the channel that are responsible for binding to ssDNA. The beta hairpins resemble a right-handed staircase. In our case, we visualized the nucleotides in the ATPase active sites at subunit interfaces A–F, B–A, C–B and D–C, while the PriRep5 beta hairpins form a staircase disposition as seen previously in other helicases (Figure 2B). Here, subunit F beta hairpin is at the very bottom closest to the CTD and following the right-handed staircase the E beta hairpin is reached at the very top pointing towards the NTD and with some degree of disorder. This suggests the staircase mechanism of translocation.

Previous studies showed that the CTD of SaPI-encoded and SCC-encoded helicases is responsible for *ori* recognition and iteron binding (26,28,33). This domain appears highly blurred in the 2D Classes (Supplementary Figure S2) and in the C1-processed map is disordered in three subunits although it appears with some order in the other three (Figure 2A) indicating high degree of motion. Structural studies using SCC-encoded Cch (26) showed that part of the WHD potentially responsible for dsDNA binding was occluded by the adjacent subunit in the closed ring conformation (helicase mode). To mimic Cch's conformation, we refined our PriRep5 structure imposing C6 symmetry and identified the same occlusion in the CTD region as shown in the electrostatic representation (Figure 2C). However, our C1 refined map and model indicate that the CTD is flexible enough to expose the *ori* binding motif. Additionally, patches suggesting a potential involvement in DNA interactions were seen between subunits, as well as surface-exposed positively charged residues pointing towards the lumen of the barrel (Figure 2C). The diameter of the lumen within the NTD and CTD regions expands 16.8 and 30 Å respectively, whereas the diameter in the ATPase domain is 10.9 Å (Supplementary Figure S9) suggesting that dsDNA might
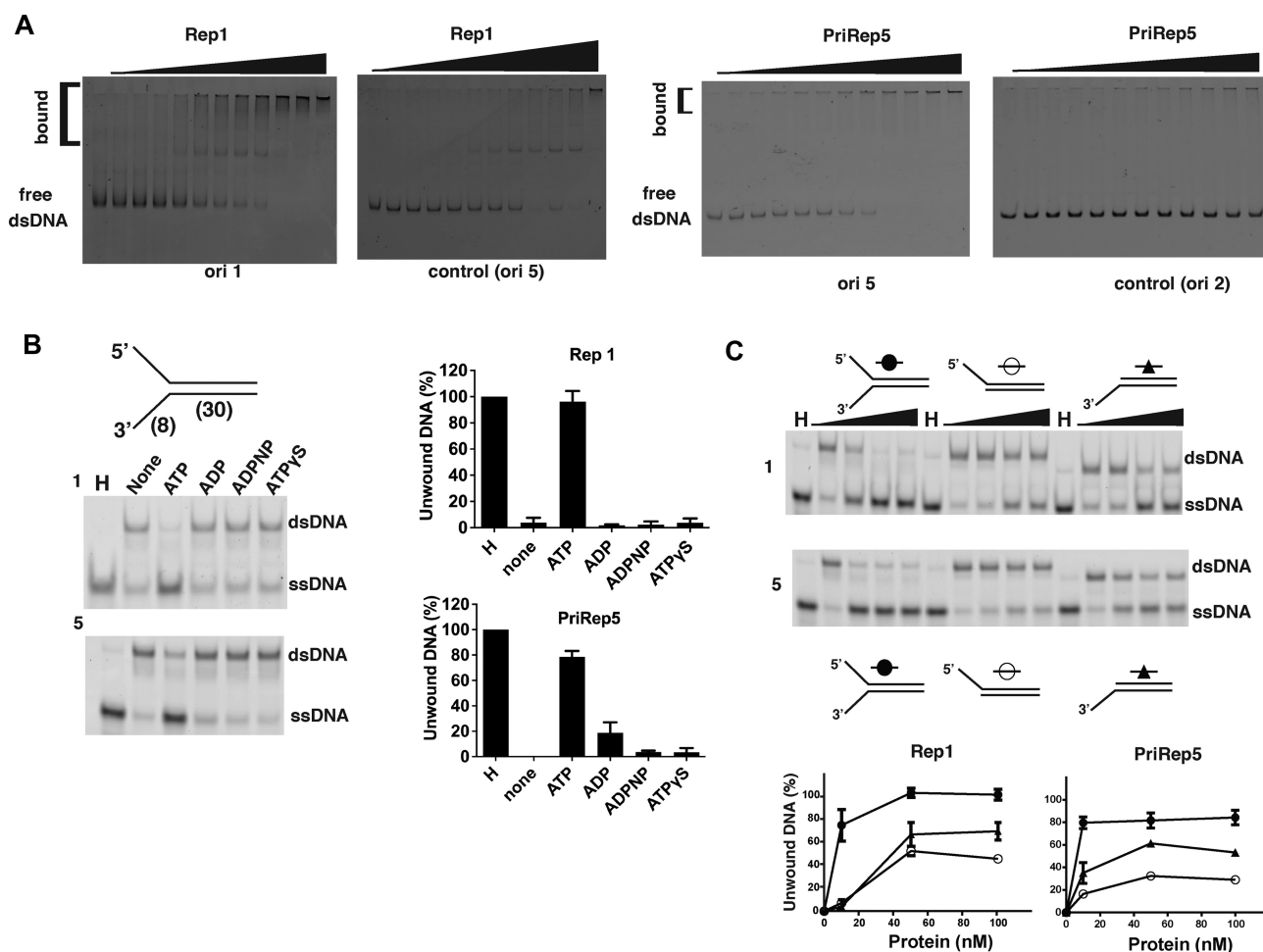
Figure 1. SaPI-encoded Reps bind their cognate *oris* and have 3′ to 5′ polarity. (**A**) Electromobility shift assay of Rep1 and PriRep5. *ori*-containing FAM-labelled PCRs were amplified from SaPI1, SaPI2 and SaPI5 *oris* and 2 nM were used as substrates. For Rep1, protein concentrations are 0, 2, 5, 10, 20, 30, 40, 50, 60, 70, 80 and 100 nM. For PriRep5, protein concentrations are 0.2, 0.4, 0.6, 0.8, 1, 1.5, 2, 4, 6, 8 and 10 nM. (**B**) Cartoon of the one-strand-FAM-labelled oligonucleotide duplex (mimicking forked DNA) used as the substrate is depicted on top of the gel, where numbers in parenthesis represent the nucleotides in each portion: 30 nt as duplex DNA and 8 nt as forked DNA. A total of 50 nM of protein was incubated without nucleotide or with ATP, with ADP, or with the non-hydrolyzable analogs ADPNP or ATPγS. H is the heated substrate used as the positive control. Numbers 1 and 5 refer to Rep1 and PriRep5 respectively. Representative gels of experiments repeated at least three times. The right side includes the graphical representations of the quantification of the gels' bands. Averages with standard deviations are shown. (**C**) The helicase assay as in B but with the different substrate duplex oligos drawn on the top. H, heated, protein concentrations are 0, 10, 50, 100 nM. Numbers 1 and 5 refer to Rep1 and PriRep5 respectively. Representative gels of experiments repeated at least three times. The bottom part includes the graphical representations of the quantification of the gels' bands. In all graphs, the vertical axes correspond to unwound DNA (%). Averages with standard deviations are shown. In those values where the standard deviation is smaller than the symbol, the error bars are not drawn.

be accommodated in the CTD and maybe in the NTD if the protein breathes a bit.

The PriRep5 CTD moves regarding the ATPase domain via two motions

The CTD of SaPI-encoded Reps is responsible for origin recognition and initiation of replication activity (29). The PriRep5 CTD in our C6 map shows that the dsDNA binding motif is occluded by the neighboring subunit when it is in helicase mode (closed ring conformation). The CTD appeared blurred in the 2D classes and partially ordered in the final C1 map suggesting motion (Figure 2A and Supplementary Figure S2). To better understand the potential movements of the PriRep5 CTD we performed multi-body

refinement in RELION3.1 (36). We segmented three different bodies corresponding to the NTD, the ATPase domain, and the CTD (Supplementary Figure S5). No movement was detected between the NTD and the ATPase domain. However, between ATPase and CTD domains a rotation motion of ∼30° respect to each other was identified (Figure 3A and supp video 1). Rotation movements between domains have been proposed to be involved in *ori* recognition and melting in eukaryotic systems (12,59,60), which agrees with the role assigned to Rep's CTD domain by *in vivo* and *in vitro* experiments (28,29). Unexpectedly, a tilting movement was also identified between ATPase and CTD domains. From a coplanar disposition of domains, the CTD tilts itself both to the left and to the right so that the planes of the two domains are no longer parallel (Figure 3B), in-
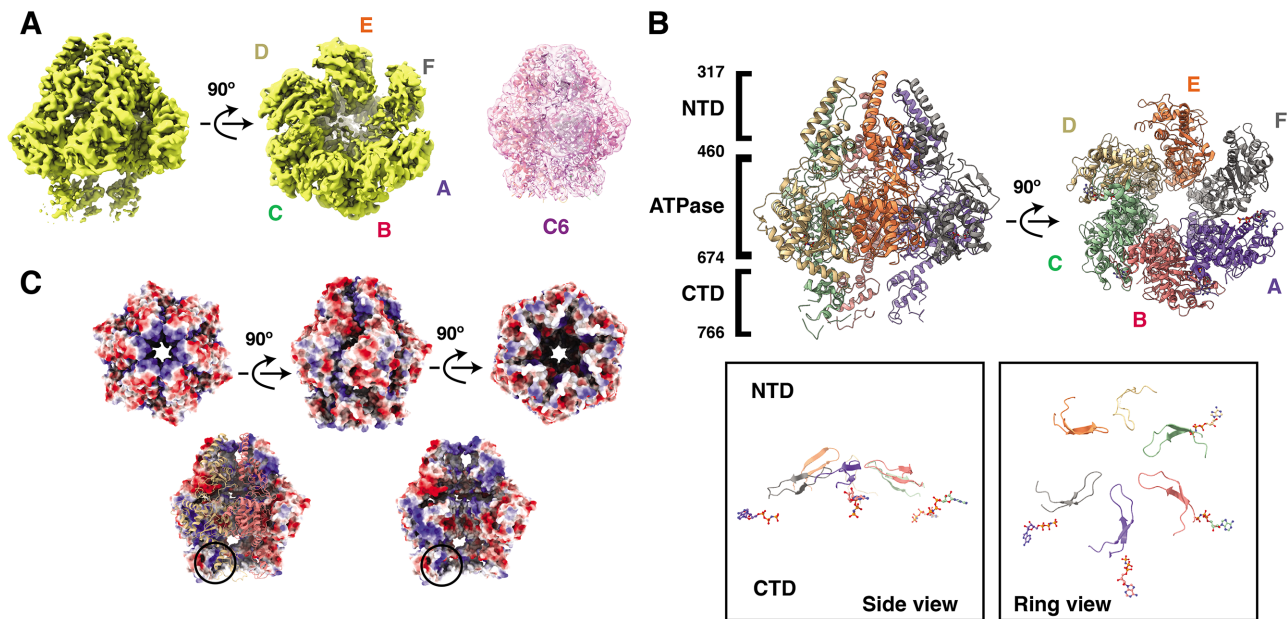
Figure 2. The PriRep5 cryoEM structure. (**A**) The PriRep5 map processed in C1. Only the CTDs of subunits A, B and C were visible in the C1 map. Letters match the color code of subunits in (**B**). On the right, a transparent pink PriRep5 map processed in C6 and docked with the C6 model is shown. (**B**) Top. The PriRep5 C1 model. NTD, N-terminal domain, CTD, C-terminal domain. Only the CTDs of subunits A, B and C were built. The ADPNP nucleotide is depicted as sticks between subunits F–A, A–B, B–C and C–D. No density corresponding to ADPNP was observed between D–E and E–F. Bottom. Beta hairpin loops in a staircase disposition and nucleotides are shown from the side view (left box), and from the ring view (right box). (**C**) Electrostatics representation of C6 model. The upper row shows the N-terminal pore, the side view, and the C-terminal pore of the closed barrel. The lower row shows four subunits as electrostatics and two frontal subunits as cartoons on the left. On the right, the two cartons have been removed to show the lumen of the barrel. The black circles in the lower row indicate the positively charged region that is responsible for dsDNA binding activity, which is occluded by the yellow subunit, as seen in (26).

dicating a pumpjack motion (Figure 3 and Supplemental Video 1). RELION's multi body refinement algorithm indicated that these are not two conformational states, but instead a continuous movement. All 10 maps generated by the algorithm are overlapped and showed in Figure 3. This flexibility in the CTD therefore explains how these helicases might expose the WHD present in the CTD and responsible for *ori* recognition (Figure 3 C). However, the tilting movement rose the question whether the pumpjack movement was related to helicase translocation or not.

Intact PriRep5 CTD is required for its helicase activity

We then decided to investigate if the CTD was playing a role in the ATPase domain activity (motor). We purified a truncated version of PriRep5 where residues 676–790 were removed, generating PriRep5-ΔCTD (Figure 3D and Supplementary Figure S3). This mutant has intact NTD and ATPase domains, plus a small portion of the helix connecting the ATPase domain and the CTD. See methods. First, we verified by size exclusion chromatography that PriRep5-ΔCTD retained its hexameric state (Supplementary Figure S5). We then performed EMSA with *ori*5-containing ds-DNA as substrates (Figure 3E). The dsDNA substrate was the same as in Figure 1A. As expected, wild-type PriRep5 bound its origin while the PriRep5-ΔCTD did not. To investigate whether the presence of the CTD was important for the ATPase domain's activity we performed the helicase assay using the forked substrate. In this case and to rule out the possibility that the forked substrate used in our previ-

ous helicase assays had too short tails (8nt), we performed this helicase assay with a P32-labeled oligonucleotide resembling a forked structure with longer tails (20 nt). Unexpectedly, PriRep5-ΔCTD lost the unwinding (helicase) activity, indicating that the presence of CTD is needed for the helicase activity (Figure 3F). We then investigated the ssDNA binding capabilities of PriRep5-ΔCTD to verify if the lack of helicase activity was due to an impaired ssDNA binding function. As shown in Figure 3G, PriRep5-ΔCTD bound 27-mer poly(dT) with slightly less efficiency than the wild type. After fitting the binding curve with the Hill equation, we estimated the apparent dissociation equilibrium of PriRep5-WT and PriRep5-ΔCTD as 46.42 ± 1.589 and 60.85 ± 6.688 respectively (nM). These data suggest that the CTD might have a role in ssDNA binding, potentially during *ori* melting, during ssDNA threading or during helicase translocation. Although the ssDNA binding capabilities of PriRep5-ΔCTD were slightly affected and although the absence of the CTD could impair its ability to properly load onto the forked DNA in the helicase assays, we were surprised that the CTD mutant did not show any helicase activity whatsoever. We therefore hypothesized that there could be something else affected in PriRep5-ΔCTD and decided to verify the ATPase functionality of PriRep5-ΔCTD by performing an ATPase assay. We measured the ability of PriRep5-WT and PriRep5-ΔCTD to hydrolyze ATP at several time points. As shown in Figure 3H, only PriRep5-WT showed ATPase activity while the CTD mutant did not. Since the presence of ssDNA might promote the ATPase activity in some replicative helicases, we measured the ATPase
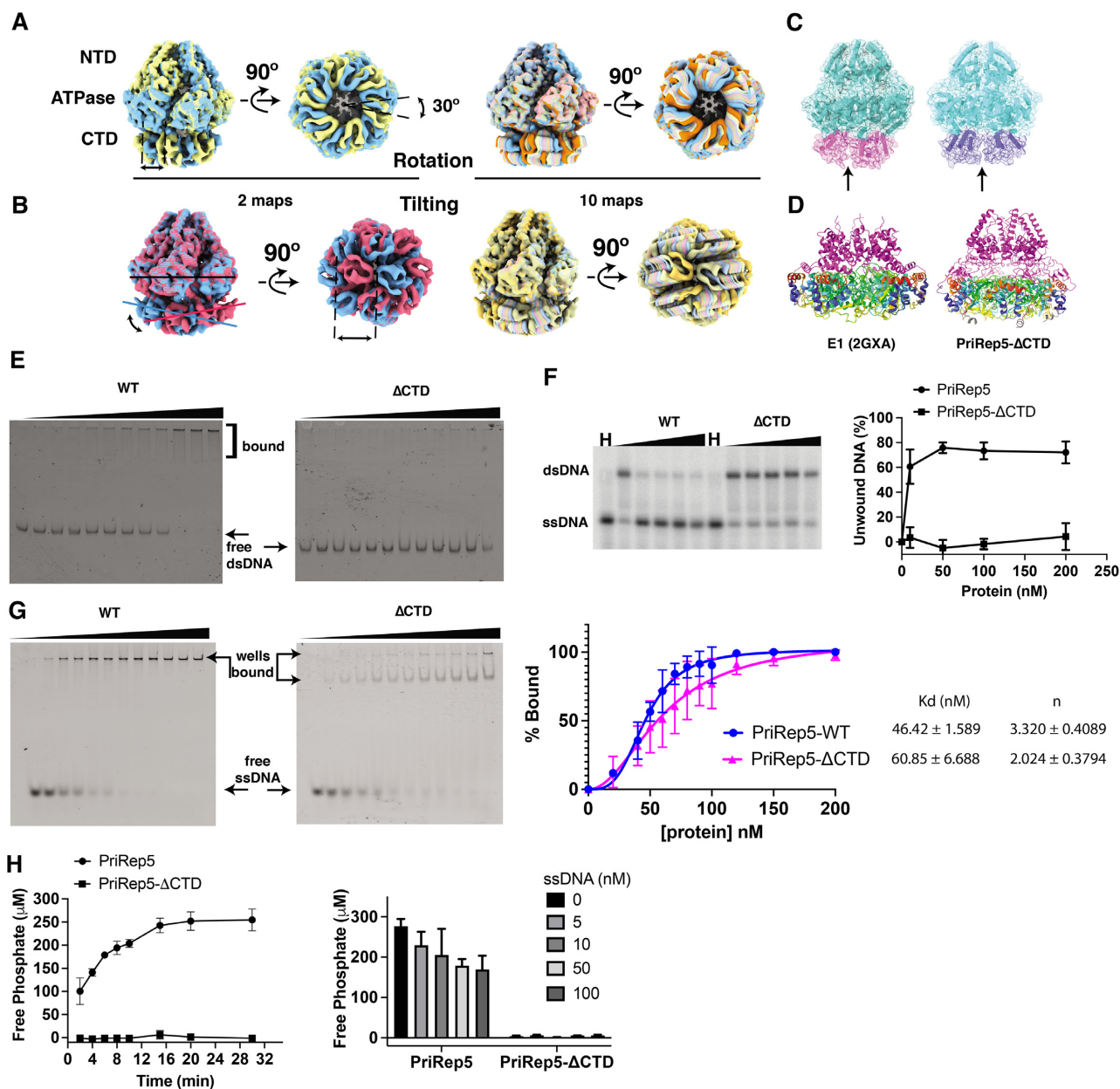
Figure 3. PriRep5 CTD motions and characterization. (**A**) The rotation movement of the CTD with respect to the ATPase domain. Overlapped multibody-refined maps are shown with random colors. On the left, the two maps depicted are those whose CTDs are the most distant from each other. Right, all 10 maps are overlapped. See Supplemental Video 1. (**B**) The tilting movement is illustrated. The black line indicates the plane of the ATPase domain. The blue and red lines indicate the non-planar (~30°) disposition of the CTD. Left, the two maps depicted are those whose CTDs are the most distant from each other. Right, all 10 maps are overlapped. (**C**) Models corresponding to the first map of tilting series (left) and rotation series (right). Transparent surface and cartoon representation of the models are shown in light green for NTD and ATPase domains, and magenta (left) and blue (right) for the CTD domains. Arrows indicate a potential gate where the *ori* recognizing residues get solvent exposed. (**D**) Model to illustrate the truncated mutant PriRep5-ΔCTD, where residues 676 to 790 have been removed. For comparison, E1 helicase (2GXA) is shown on the left. N tiers are colored in pink: E1 304 to 393; PriRep5-ΔCTD from 307 to 467 ATPase domains rainbow: E1 from 394 to 577; PriRep5-ΔCTD 468 to 659. PriRep5 residues from 660 to 675 are colored in grey. (**E**) PriRep5 WT and PriRep5-ΔCTD differ in their abilities to bind *ori*-containing DNA. EMSA with FAM labeled PCR as substrate where the entire *ori5* has been amplified as in figure 1A, see methods. WT, wild type PriRep5, ΔCTD, truncated mutant with CTD deleted. Protein concentrations are 0.2, 0.4, 0.6, 0.8, 1, 1.5, 2, 4, 6, 8 and 10 nM. (**F**) Left, helicase Assay using substrate as in Figure 1B but with 20 nt overhang and 18nt duplex DNA. H, heated control. Protein concentrations are 0, 10, 50, 100, 200 nM. Right, quantification of the gel´s bands. (**G**) EMSAs with FAM labeled 27-mer poly(dT) oligonucleotide as ssDNA substrate. Protein concentrations are 0, 20, 40, 50, 60, 70, 80, 90, 100, 120, 150 and 200 nM. Right panel, quantification of the bindings and fitting with the Hill equation. The values for the K_d (in nM) and Hill coefficient (n) are shown on the right side. (**H**) ATPase assay where the amount of phosphate generated by the ATPase activity of the protein is calculated. WT and PriRep5-ΔCTD are active and inactive respectively for ATP hydrolysis. The assay was performed at the specified time points (left) or (right) in the presence of different amounts of 27-mer poly(dT) as ssDNA. E and F are representative gels from two experiments, G, five experiments, H has been performed twice with three experimental replicas each time. In F, G and H, average and standard deviations are shown.

activity of wild-type and PriRep5-ΔCTD variants without and with different amounts of ssDNA. The ATPase activity of PriRep5 was not ssDNA dependent whereas the ATPase activity of PriRep5-ΔCTD was lost in all cases (Figure 3H), indicating that an intact CTD is necessary for nucleotide hydrolysis and hence for the helicase activity (unwinding). Our interpretation of these data is that the CTD is involved not only in *ori* recognition and melting in an ATP-independent way (28,29), but also plays a critical role in ATP hydrolysis and helicase translocation. In our model system, therefore, the ATPase domain is not sufficient to provide the staircase mechanism of translocation and needs the involvement of the CTD to fire ATP and ultimately unwind DNA.

We then turned to Rep1 to complement and verify if our findings in PriRep5 were specific to that island or otherwise conserved among SaPIs. In this case, we solved the cryoEM map of SaPI1 Rep in its apo state, without any ribonucleotide analog or magnesium ions, at 3.9 Å resolution (see methods and Supplementary Figure S6).

Rotation and tilting motions are ATP-independent and conserved in Reps

To generalize our findings in PriRep5, we performed the same motion analysis of the Rep1cryoEM map with RE-LION's multi-body algorithm (Supplementary Figure S5). Rep1 2D classes did not show blurriness in the CTD. As shown in Supplementary Figure S6 and Supplemental Video 1, the ATPase domains and CTDs moved with respect to each other by rotation and tilting, indicating that these movements are conserved among SaPI-encoded Reps. Additionally, because Rep was in apo state, the presence of these two movements indicates that they occur independent of nucleotide binding and/or hydrolysis. Moreover, we collected cryoEM data from a Rep1 Walker A mutant (K186A) and detected the same two movements (not shown). Instead of fitting a model into Rep1-apo map, which severely suffered preference particle orientation (Supplementary Figure S10), we screened several partners to complex with Rep1. The condition that gave better results and solved the preference orientation problem was co-purifying Rep1 with its cognate Primase (PriRep1) and subsequently mixing the pair of proteins with ATPγS and with 19 bp dsDNA bearing 2 iterons and 60 nucleotide long ssDNA tail. See methods for details.

CryoEM structure of Rep1–DNA complex

This complex yielded a cryoEM map of 3.1 Å resolution with good particle distribution (See Figure 4 and Supplementary Figures S7 and S10) and allowed us to structurally characterize the staircase mechanism. The primase could not be resolved and the only part of the DNA that could be visualized was a single stranded penta nucleotide segment making close contact with the beta hairpin ssDNA binding motif whose sequence was arbitrarily assigned to be 5′-TAAAA-3′. Rep1 forms also a three-tier ring shaped homo hexamer as depicted in Figure 4A and B. Rep1, as PriRep5, belongs to the SF3 group. Our cryoEM structure reveals basically the same features depicted in E1 helicase-
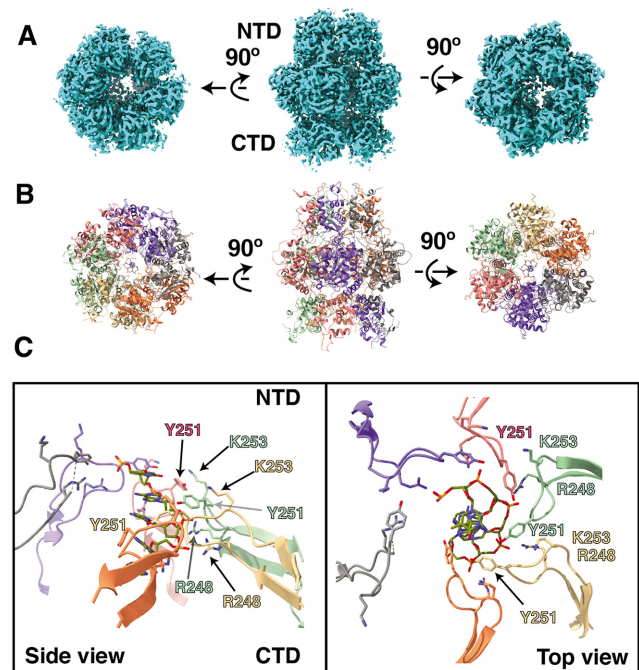
Figure 4. CryoEM map of Rep1–DNA complex shows the staircase mechanism. (**A**) CryoEM map of Rep1–DNA complex processed in C1. NTD, N-terminal domain, CTD, C-terminal domain. (**B**) Rep1–DNA model. (**C**) Zoomed vision of the beta hairpin-ssDNA backbone contacts. Subunits are colored grey, green, cyan, magenta, yellow. In the side view, NTD locates on top, whereas CTD is underneath the represented cartoon. Side chains of R248, Y251 and K253 of all subunits are shown as sticks. ssDNA is shown as sticks and carbon colored as olive green.

ssDNA complex crystal structure (6). ATPγS was visualized at every subunit interface where different degrees of nucleotide and inter subunits contacts correlates well with the beta hairpin height. ssDNA binding residues located in the beta hairpin of Rep1 are R248 and Y251 which interact with the phosphate and the sugar of the ssDNA backbone respectively, so that one subunit (one beta hairpin) interacts with one nucleotide suggesting a step size of one nucleotide (Figure 4C and Supplementary Figure S7). Additionally, Y251 is sandwiched between the DNA and K253 (Figure 4C and Supplementary Figure S7F). To experimentally verify our structure, we generated and purified a triple mutant where residues R248, Y251 and K253 were substituted to alanine (R248A-Y251A-K253A). We then performed a helicase assay using a 20nt forked substrate (see methods) and confirmed that the helicase activity was lost indicating that key residues in the staircase mechanism are necessary for the protein to maintain the helicase activity. To generalize our findings, we turned to PriRep5 helicase and generated a beta hairpin mutant PriRep5-K591A which also lost its helicase activity (Supplementary Figure S8). To rule out the possibility that the lack of activity was because these mutants had lost their ability to hexamerise, we performed size exclusion chromatography to analyze their oligomeric state in solution. As shown in Supplementary Figure S8C, these mutants eluted at volumes corresponding to their hexameric forms.
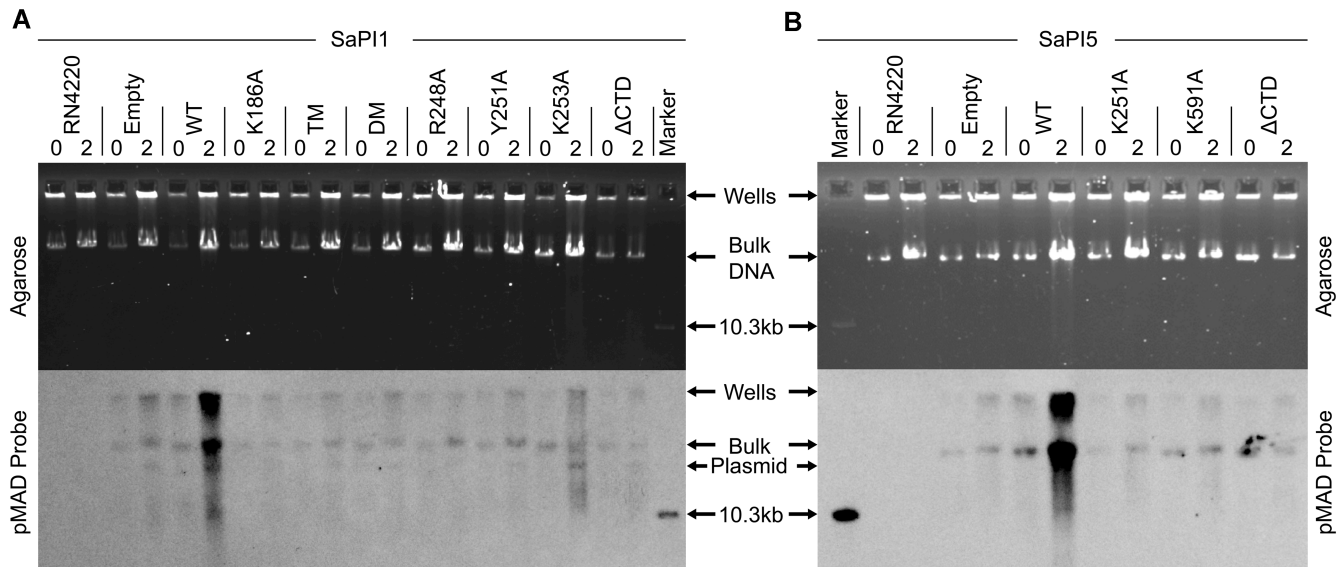
SaPI replication *in vivo*

To fully verify the cryoEM Rep1–DNA structure, we generated several constructs to analyze *in vivo* their replicative capabilities. We cloned SaPI1 *ori* in the staphylococcal plasmid pMAD and wild-type *pri-rep* genes in plasmid pCN51_*cat* (pCN51 derivative plasmid where the *erm*C cassette has been replace by a *cat*194 cassette selective for chloramphenicol) which harbors a cadmium inducible promoter, see methods for details. Staphylococcal strain RN4220 was co-transformed with both plasmids, and pMAD replication was monitored by Southern blot using a specific probe targeting the pMAD plasmid (Figure 5A). As expected, the wild-type construct produced a strong signal 2 h post induction, not only in the bulk of DNA but also close to the wells, which agrees with the fact that a concatemeric large molecular weight DNA molecule is generated under the SaPI's replicon. We then engineered a set of mutations in the beta hairpin dissecting the previous triple mutant with single and double mutants. Three single mutants: R248A, Y251A, K253A. One double mutant: R248A–Y251A. As controls we included the triple mutant: R248A–Y251A–K253A; the mutant K186A (Walker A) and a truncated version where three frame-shifted stop codons were introduced downstream K346 to generate the CTD truncate (ΔCTD in Figure 5A). None of the aforementioned mutations equaled the wild type´s Southern blotting signal during plasmid replication *in vivo*, indicating that the amino acids involved in the staircase mechanism are necessary to support DNA replication. To verify that all proteins were properly expressed, a new set of constructs were generated including a FLAG so that the presence of protein could be monitored by Western blot (Supplementary Figure S11). The only mutant showing a faint activity was K253A which is not directly contacting the DNA but flanking Y251 of the adjacent subunit on the opposite side of ssDNA (Figure 4C and Supplementary Figure S7). We subsequently expanded our *in vivo* findings to PriRep5. We included Walker A (K521A) and CTD truncated mutants as controls. Figure 5B shows that a single mutation in PriRep5 beta hairpin (K591A) abolished replication capabilities, which in agreement with our structural and biochemical data confirms that these initiators-helicases couple the staircase mechanism with the the presence of the CTD to support highly efficient replication.

DISCUSSION

SaPIs were the first Gram positive pathogenicity islands whose replication was demonstrated *in vivo*. They replicate extensively and effectively so that, in conjunction with other parasitic mechanisms, they can be transferred as high or at even higher rates than the phages they parasite, which assures their successful spread in nature. The initiators-helicases, Reps, are critical to support SaPI's replication. Among the three types of Reps, Rep1 and PriRep5 belong to the SF3 group, whereas Rep2 (SaPIbov1 Rep homolog) belongs to SF6. SaPI-encoded Reps also show functional similarities with SF3 and SF6 helicases (both superfamilies bear eukaryotic helicases, whereas prokaryotic relatives are SF4) by showing preference for translocation along

the leading strand (Figure 1C). Eukaryotic MCM replicative helicases are loaded onto the *ori* as double hexamers in a head-to-head conformation (61). Previous studies also indicate that Reps most likely initiate replication by self-loading two hexamers, one on each side of the AT-rich region at their origins (28,29). If they do so in a head-to-head, tail-to-tail, or other conformation needs to be addressed. Reps do not need ATP binding and hydrolysis activity to mediate *ori* recognition and melting *in vitro* (29). CTD disposition in Reps, where there is a domain-swap with the ATPase body of the adjacent protomers seems a conserved feature in origin recognition complexes in eukaryotes (26,59).

In our speculative working model (Figure 6) for the initiation of replication process in SaPIs, the high flexibility in Reps CTD exposes the WHD to bind the iterons. The diameter of the CTD is wide enough to accommodate dsDNA. The flexibility between CTD and the motor (ATPase) domains allows them to rock respect each other, so that the same and opposite directions of movement between them are allowed. Hexamer-hexamer interactions might facilitate opposite rotation movements so that the DNA is distorted as it has been proposed in eukaryotic systems (12). As soon as Rep is self-loaded, it starts translocating along the leading strand coupling the hydrolysis of ATP with translocation. Our cryoEM structure suggests that the translocation step is one nucleotide per ATP hydrolyzed. Our above experiments with PriRep5-ΔCTD demonstrate that the CTD domain is essential for nucleotide firing and therefor unwinding activity. Whether the whole CTD or part of it is essential remains to be studied. It should be noted that we arbitrarily assumed that Rep1 translocates with the NTD tier ahead of the motor domain (ATPase domain) in a similar fashion as E1 helicase does, and accordingly positioned the ssDNA pentanucleotide with its 5′ end pointing to the NTD in our Rep1–DNA structure. Further experiments are needed to confirm this assumption. We have shown that in PriRep5, the staircase mechanism presumed for the motor domain is not sufficient for its helicase activity, since the presence of the CTD is necessary for its helicase activity *in vitro*. We speculate that this observed high flexibility between domains allows the complex to efficiently couple *ori* binding and melting with translocation. In eukaryotic GMC helicases, a tilting movement between domains suggested a pumpjack motion of translocation, although there was not DNA present in that structure (13). Recently, a cryoEM structure of papillomavirus E1 helicase in complex with a replication fork suggested a tilting motion of the collar domain respect the ATPase domain (62). Our work here points in the same direction where the tilting motion between domains seems to be a conserved feature among evolutionary unrelated initiators-helicases. The closest structural homolog of PriRep5 and Rep1 is deep-sea vent phage NrS-1 polymerase-helicase. In agreement with our model, NrS-1 CTD deletion also abolished its helicase activity (56). However, NrS-1 phage biology has been studied to much less extent than SaPIs, and yet it has not been addressed if NrS-1 helicase uses CTD for *ori* specific recognition. Deep-sea phages, as well as staphylococcal phages and SaPIs, contain evolutive traits that will help us to understand important molecular processes.

Figure 5. Rep-supported replication *in vivo*. Mini-lysates of RN4220 strain containing a pair of vectors, *ori*-containing pMAD plus gene-variants-containing pCN51_*cat*, were run in agarose gels (upper panel) and then blotted against a pMAD probe (lower panel) to investigate Rep-supported replication. (**A**) Constructs corresponding to Rep1-supported replication. RN4220, strain with no plasmids; empty, *ori*1-pMAD plus empty pCN51_*cat*; Rep1-WT, wildtype *pri* and *rep* genes from SaPI1; K186A, Walker A mutant, TM, triple mutant in Rep1 beta hairpin R248A–Y251A–K253A; DM, double mutant R248A–Y251A; single Rep1 beta hairpin mutants R248A, Y251A, K253A; ΔCTD, three frame-shifted stop codons were introduced downstream K346 in Rep1; marker is *ori*1-pMAD linearized with *Sal*I whose size is 10.3 kb; 0, timepoint when cells were induced with 7.5 μM CdCl$_2$; 2, two hours post induction, (**B**) *ori*5-containing pMAD plasmid was used in combination with SaPI5 gene variants in pCN51_*cat*. RN4220, empty and marker as in (A); WT, *prirep* gene from SaPI5; K521A, Walker A mutant; K591A, beta hairpin single mutant; ΔCTD, three frame-shifted stop codons were introduced downstream V681. Experiments repeated at least three times.
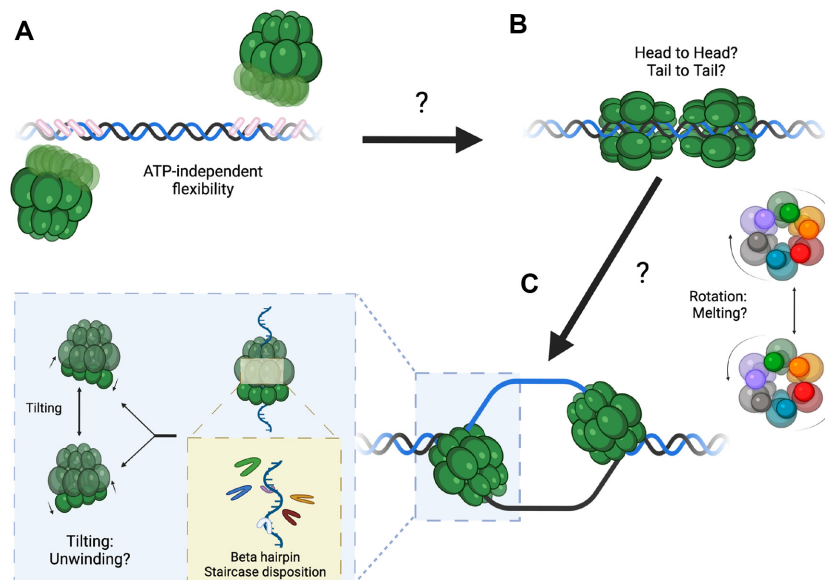
Figure 6. Model of Rep-mediated SaPI DNA initiation of replication. (**A**) The ATP-independent flexibility of CTD exposes the residues involved in iteron recognition. The diameter of the barrel in its N and C terminal sides might accommodate dsDNA. In (**B**), two hexamers have been speculatively depicted embracing dsDNA. Potential hexamer-hexamer interactions might induce opposite rotation movements that favor DNA distortion. Maybe electrostatically wrapping the DNA with the surface of the barrel (not depicted) might be involved in *ori* melting. (**C**) The tilting movement of CTD aids the ATPase domain's staircase mechanism to unwind the DNA and promote replication fork progression.

In summary, in the present work we have characterized Reps from SaPI1 and SaPI5 as 3′ to 5′ helicases. Rep1 and PriRep5 have a conserved flexibility between their CTD and ATPase domains. The presence of the PriRep5′s CTD is necessary for ATPase and helicase activities. Unveiling hidden biological aspects of these mobile elements and their helper phages will help us to understand DNA replication in particular and DNA transactions in general, not only in subcellular entities but also in cellular organisms from all domains of life.

DATA AVAILABILITY

The cryo-EM map and atomic coordinates of PriRep5 processed in C1 have been deposited in the Electron Microscopy Data Bank (EMDB) with the code (EMD-12975) and PDB code 7OLA. The C6 processing map and coordinates have been deposited in the EMDB and PDB under accession codes EMD-12982 and 7OM0 respectively. The cryo-EM map for Rep1 processed in C1 has been deposited in EMDB under accession code EMD-12987. CryoEM map and coordinates of Rep1–DNA have been deposited in EMDB code EMD-13342, PDB (7PDS). The raw images of the Rep1–DNA complex and PriRep5 data sets have been deposited in the EMPIAR Data Base under the codes EMPIAR-10880 and EMPIAR-10881 respectively. The raw images of the Rep1-apo data set have been deposited with the code EMPIAR-10945.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

Conflict of interest statement. None declared.

REFERENCES

1. Duderstadt,K.E. and Berger,J.M. (2008) AAA+ ATPases in the initiation of DNA replication. *Crit. Rev. Biochem. Mol. Biol.*, **43**, 163–187.
2. Bochman,M.L. and Schwacha,A. (2009) The mcm complex: unwinding the mechanism of a replicative helicase. *Microbiol. Mol. Biol. Rev.*, **73**, 652–683.
3. Costa,A. and Onesti,S. (2009) Structural biology of MCM helicases. *Crit. Rev. Biochem. Mol. Biol.*, **44**, 326–342.
4. Davey,M.J. and O'Donnell,M. (2003) Replicative helicase loaders: ring breakers and ring makers. *Curr. Biol.*, **13**, R594–R596.
5. Lyubimov,A.Y., Strycharska,M. and Berger,J.M. (2011) The nuts and bolts of ring-translocase structure and mechanism. *Curr. Opin. Struct. Biol.*, **21**, 240–248.
6. Enemark,E.J. and Joshua-Tor,L. (2006) Mechanism of DNA translocation in a replicative hexameric helicase. *Nature*, **442**, 270–275.
7. Thomsen,N.D. and Berger,J.M. (2009) Running in reverse: the structural basis for translocation polarity in hexameric helicases. *Cell*, **139**, 523–534.
8. Itsathitphaisarn,O., Wing,R.A., Eliason,W.K., Wang,J. and Steitz,T.A. (2012) The hexameric helicase DnaB adopts a nonplanar conformation during translocation. *Cell*, **151**, 267–277.
9. Gao,Y., Cui,Y., Fox,T., Lin,S., Wang,H., Val,N.de, Zhou,Z.H. and Yang,W. (2019) Structures and operating principles of the replisome. *Science*, **363**, eaav7003.
10. Meagher,M., Epling,L.B. and Enemark,E.J. (2019) DNA translocation mechanism of the MCM complex and implications for replication initiation. *Nat. Commun.*, **10**, 3117.
11. Ilves,I., Petojevic,T., Pesavento,J.J. and Botchan,M.R. (2010) Activation of the MCM2-7 helicase by association with cdc45 and GINS proteins. *Mol. Cell*, **37**, 247–258.
12. Abid Ali,F., Renault,L., Gannon,J., Gahlon,H.L., Kotecha,A., Zhou,J.C., Rueda,D. and Costa,A. (2016) Cryo-EM structures of the eukaryotic replicative helicase bound to a translocation substrate. *Nat. Commun.*, **7**, 10708.
13. Yuan,Z., Bai,L., Sun,J., Georgescu,R., Liu,J., O'Donnell,M.E. and Li,H. (2016) Structure of the eukaryotic replicative CMG helicase suggests a pumpjack motion for translocation. *Nat. Struct. Mol. Biol.*, **23**, 217–224.
14. O'Donnell,M.E. and Li,H. (2018) The ring-shaped hexameric helicases that function at DNA replication forks. *Nat. Struct. Mol. Biol.*, **25**, 122–130.
15. Weigel,C. and Seitz,H. (2006) Bacteriophage replication modules. *FEMS Microbiol. Rev.*, **30**, 321–381.
16. Novick,R.P., Christie,G.E. and Penadés,J.R. (2010) The phage-related chromosomal islands of Gram-positive bacteria. *Nat. Rev. Microbiol.*, **8**, 541–551.
17. Penadés,J.R. and Christie,G.E. (2015) The phage-inducible chromosomal islands: a family of highly evolved molecular parasites. *Annu Rev Virol*, **2**, 181–201.
18. Tormo-Más,M.A., Mir,I., Shrestha,A., Tallent,S.M., Campoy,S., Lasa,I., Barbé,J., Novick,R.P., Christie,G.E. and Penadés,J.R. (2010) Moonlighting bacteriophage proteins derepress staphylococcal pathogenicity islands. *Nature*, **465**, 779–782.
19. Mir-Sanchis,I., Martínez-Rubio,R., Martí,M., Chen,J., Lasa,Í., Novick,R.P., Tormo-Más,M.Á. and Penadés,J.R. (2012) Control of staphylococcus aureus pathogenicity island excision. *Mol. Microbiol.*, **85**, 833–845.
20. Bowring,J., Neamah,M.M., Donderis,J., Mir-Sanchis,I., Alite,C., Ciges-Tomas,J.R., Maiques,E., Medmedov,I., Marina,A. and Penadés,J.R. (2017) Pirating conserved phage mechanisms promotes promiscuous staphylococcal pathogenicity island transfer. *Elife*, **6**, 33.
21. Ubeda,C., Maiques,E., Tormo,M.A., Campoy,S., Lasa,I., Barbé,J., Novick,R.P. and Penadés,J.R. (2007) SaPI operon i is required for SaPI packaging and is controlled by lexA. *Mol. Microbiol.*, **65**, 41–50.
22. Damle,P.K., Wall,E.A., Spilman,M.S., Dearborn,A.D., Ram,G., Novick,R.P., Dokland,T. and Christie,G.E. (2012) The roles of sapi1 proteins gp7 (CpmA) and gp6 (CpmB) in capsid size determination and helper phage interference. *Virology*, **432**, 277–282.

23. Dearborn,A.D., Wall,E.A., Kizziah,J.L., Klenow,L., Parker,L.K., Manning,K.A., Spilman,M.S., Spear,J.M., Christie,G.E. and Dokland,T. (2017) Competing scaffolding proteins determine capsid size during mobilization of staphylococcus aureus pathogenicity islands. *Elife*, **6**, e30822.

24. Kizziah,J.L., Manning,K.A., Dearborn,A.D., Wall,E.A., Klenow,L., Hill,R.L.L., Spilman,M.S., Stagg,S.M., Christie,G.E. and Dokland,T. (2017) Cleavage and structural transitions during maturation of staphylococcus aureus bacteriophage 80α and sapi1 capsids. *Viruses*, **9**, 384.

25. Úbeda,C., Maiques,E., Barry,P., Matthews,A., Tormo,M.Á., Lasa,Í., Novick,R.P. and Penadés,J.R. (2008) SaPI mutations affecting replication and transfer and enabling autonomous replication in the absence of helper phage. *Mol. Microbiol.*, **67**, 493–503.

26. Mir-Sanchis,I., Roman,C.A., Misiura,A., Pigli,Y.Z., Boyle-Vavra,S. and Rice,P.A. (2016) Staphylococcal SCCmec elements encode an active MCM-like helicase and thus may be replicative. *Nat. Struct. Mol. Biol.*, **23**, 891–898.

27. Erzberger,J.P. and Berger,J.M. (2006) Evolutionary relationships and structural mechanisms of AAA+ proteins. *Annu. Rev. Biophys. Biomol. Struct.*, **35**, 93–114.

28. Úbeda,C., Barry,P., Penadés,J.R. and Novick,R.P. (2007) A pathogenicity island replicon in staphylococcus aureus replicates as an unstable plasmid. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 14182–14188.

29. Ubeda,C., Tormo-Más,M.Á., Penadés,J.R. and Novick,R.P. (2012) Structure-function analysis of the sapibov1 replication origin in staphylococcus aureus. *Plasmid*, **67**, 183–190.

30. Mott,M.L. and Berger,J.M. (2007) DNA replication initiation: mechanisms and regulation in bacteria. *Nat. Rev. Microbiol.*, **5**, 343–354.

31. Duderstadt,K.E. and Berger,J.M. (2013) A structural framework for replication origin opening by AAA+ initiation factors. *Curr. Opin. Struct. Biol.*, **23**, 144–153.

32. Parker,M.W., Botchan,M.R. and Berger,J.M. (2017) Mechanisms and regulation of DNA replication initiation in eukaryotes. *Crit. Rev. Biochem. Mol. Biol.*, **52**, 107–144.

33. Bebel,A., Walsh,M.A., Mir-Sanchis,I. and Rice,P.A. (2020) A novel DNA primase-helicase pair encoded by SCCmec elements. *Elife*, **9**, e55478.

34. Mir-Sanchis,I., Pigli,Y.Z. and Rice,P.A. (2018) Crystal structure of an unusual single-stranded DNA-Binding protein encoded by staphylococcal cassette chromosome elements. *Structure*, **26**, 1144–1150.

35. Schneider,C.A., Rasband,W.S. and Eliceiri,K.W. (2012) NIH image to imagej: 25 years of image analysis. *Nat. Methods*, **9**, 671–675.

36. Zivanov,J., Nakane,T., Forsberg,B.O., Kimanius,D., Hagen,W.J., Lindahl,E. and Scheres,S.H. (2018) New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife*, **7**, e42166.

37. Zheng,S.Q., Palovcak,E., Armache,J.-P., Verba,K.A., Cheng,Y. and Agard,D.A. (2017) MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods*, **14**, 331–332.

38. Zhang,K. (2016) Gctf: Real-time CTF determination and correction. *J. Struct. Biol.*, **193**, 1–12.

39. Nakane,T., Kimanius,D., Lindahl,E. and Scheres,S.H. (2018) Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *Elife*, **7**, e36861.

40. Tan,Y.Z., Baldwin,P.R., Davis,J.H., Williamson,J.R., Potter,C.S., Carragher,B. and Lyumkis,D. (2017) Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat. Methods*, **14**, 793–796.

41. Cowtan,K. (2006) The *buccaneer* software for automated model building. 1. Tracing protein chains. *Acta. Crystallogr. D Biol. Crystallogr.*, **62**, 1002–1011.

42. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of coot. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 486–501.

43. Roy,A., Kucukural,A. and Zhang,Y. (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.*, **5**, 725–738.

44. Kidmose,R.T., Juhl,J., Nissen,P., Boesen,T., Karlsen,J.L. and Pedersen,B.P. (2019) Namdinator - automatic molecular dynamics flexible fitting of structural models into cryo-EM and crystallography experimental maps. *IUCrJ*, **6**, 526–531.

45. Adams,P.D., Afonine,P.V., Bunkóczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.-W., Kapral,G.J., Grosse-Kunstleve,R.W. *et al.* (2010) PHENIX a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 213–221.

46. Goddard,T.D., Huang,C.C., Meng,E.C., Pettersen,E.F., Couch,G.S., Morris,J.H. and Ferrin,T.E. (2018) UCSF chimerax: meeting modern challenges in visualization and analysis. *Protein Sci.*, **27**, 14–25.

47. Pettersen,E.F., Goddard,T.D., Huang,C.C., Meng,E.C., Couch,G.S., Croll,T.I., Morris,J.H. and Ferrin,T.E. (2021) UCSF chimerax: structure visualization for researchers, educators, and developers. *Protein Sci.*, **30**, 70–82.

48. Morin,A., Eisenbraun,B., Key,J., Sanschagrin,P.C., Timony,M.A., Ottaviano,M. and Sliz,P. (2013) Collaboration gets the most out of software. *Elife*, **2**, e01456.

49. Arnaud,M., Chastanet,A. and Débarbouillé,M. (2004) New vector for efficient allelic replacement in naturally nontransformable, low-gc-content, gram-positive bacteria. *Appl. Environ. Microbiol.*, **70**, 6887–6891.

50. Charpentier,E., Anton,A.I., Barry,P., Alfonso,B., Fang,Y. and Novick,R.P. (2004) Novel cassette-based shuttle vector system for gram-positive bacteria. *Appl. Environ. Microbiol.*, **70**, 6076–6085.

51. Gao,Y. and Yang,W. (2020) Different mechanisms for translocation by monomeric and hexameric helicases. *Curr. Opin. Struct. Biol.*, **61**, 25–32.

52. Rothenberg,E., Trakselis,M.A., Bell,S.D. and Ha,T. (2007) MCM forked substrate specificity involves dynamic interaction with the 5′-tail. *J. Biol. Chem.*, **282**, 34229–34234.

53. Graham,B.W., Schauer,G.D., Leuba,S.H. and Trakselis,M.A. (2011) Steric exclusion and wrapping of the excluded DNA strand occurs along discrete external binding paths during MCM helicase unwinding. *Nucleic Acids Res.*, **39**, 6585–6595.

54. Perera,H.M., Behrmann,M.S., Hoang,J.M., Griffin,W.C. and Trakselis,M.A. (2019) Contacts and context that regulate DNA helicase unwinding and replisome progression. *Enzymes*, **45**, 183–223.

55. Holm,L. and Sander,C. (1995) Dali: a network tool for protein structure comparison. *Trends Biochem. Sci.*, **20**, 478–480.

56. Chen,X., Su,S., Chen,Y., Gao,Y., Li,Y., Shao,Z., Zhang,Y., Shao,Q., Liu,H., Li,J. *et al.* (2020) Structural studies reveal a ring-shaped architecture of deep-sea vent phage nrs-1 polymerase. *Nucleic Acids Res.*, **48**, 3343–3355.

57. Dougan,D.A., Mogk,A., Zeth,K., Turgay,K. and Bukau,B. (2002) AAA+ proteins and substrate recognition, it all depends on their partner in crime. *FEBS Lett.*, **529**, 6–10.

58. Iyer,L.M., Leipe,D.D., Koonin,E.V. and Aravind,L. (2004) Evolutionary history and higher order classification of AAA+ ATPases. *J. Struct. Biol.*, **146**, 11–31.

59. Bleichert,F., Botchan,M.R. and Berger,J.M. (2015) Crystal structure of the eukaryotic origin recognition complex. *Nature*, **519**, 321–326.

60. Bleichert,F. (2019) Mechanisms of replication origin licensing: a structural perspective. *Curr. Opin. Struct. Biol.*, **59**, 195–204.

61. Miller,T.C.R., Locke,J., Greiwe,J.F., Diffley,J.F.X. and Costa,A. (2019) Mechanism of head-to-head MCM double-hexamer formation revealed by cryo-EM. *Nature*, **575**, 704–710.

62. Javed,A., Major,B., Stead,J.A., Sanders,C.M. and Orlova,E.V. (2021) Unwinding of a DNA replication fork by a hexameric viral helicase. *Nat. Commun.*, **12**, 5535.