The replication initiator of the cholera pathogen's second chromosome shows structural similarity to plasmid initiators

Natalia Orlova[1,2,†], Matthew Gerding[3,4,5,†], Olha Ivashkiv[1], Paul Dominic B. Olinares[6], Brian T. Chait[6], Matthew K. Waldor[3,4,5,*] and David Jeruzalmi[1,2,7,8,*]

[1]Department of Chemistry and Biochemistry, City College of New York, New York, NY 10031, USA, [2]Ph.D. Program in Biochemistry, The Graduate Center of the City University of New York, NY 10016, USA, [3]Department of Microbiology and Immunobiology, Harvard Medical School, Boston, MA, USA, [4]Division of Infectious Diseases, Brigham and Women's Hospital, Boston, MA, USA, [5]Howard Hughes Medical Institute, Boston, MA 02115, USA, [6]Laboratory for Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, NY 10021, USA, [7]Ph.D. Program in Biology, The Graduate Center of the City University of New York, NY 10016, USA and [8]Ph.D. Program in Chemistry, The Graduate Center of the City University of New York, NY 10016, USA

ABSTRACT

The conserved DnaA-oriC system is used to initiate replication of primary chromosomes throughout the bacterial kingdom; however, bacteria with multipartite genomes evolved distinct systems to initiate replication of secondary chromosomes. In the cholera pathogen, *Vibrio cholerae*, and in related species, secondary chromosome replication requires the RctB initiator protein. Here, we show that RctB consists of four domains. The structure of its central two domains resembles that of several plasmid replication initiators. RctB contains at least three DNA binding winged-helix-turn-helix motifs, and mutations within any of these severely compromise biological activity. In the structure, RctB adopts a head-to-head dimeric configuration that likely reflects the arrangement in solution. Therefore, major structural reorganization likely accompanies complex formation on the head-to-tail array of binding sites in *oriCII*. Our findings support the hypothesis that the second *Vibrionaceae* chromosome arose from an ancestral plasmid, and that RctB may have evolved additional regulatory features.

INTRODUCTION

Regulated initiation is a common feature of DNA replication systems of chromosomes. Although incompletely understood at the atomic level, studies in *Escherichia coli* and other bacterial model systems have yielded great insight into this critical process (1–8). Two molecular players play a central role in the current model: (i) the origin of DNA replication, a site on the chromosome where DNA synthesis begins, and (ii) the initiator protein that recognizes segments of double-stranded DNA and single-stranded DNA within the origin (1). In *E. coli*, where initiation of bacterial chromosomal replication has been extensively studied, binding of multiple DnaA initiator proteins ($\sim$53 kD) to sites within the 245 bp *oriC* DNA sequence leads to assembly of a large multi-protein DNA complex (1,9,10). The DnaA–OriC complex mediates initial melting of origin DNA within an A-T rich segment of the origin (1,11,12); the resulting single-stranded DNA is bound and stabilized by an oligomeric form of DnaA (4,13). Melted DNA at the origin serves as the entry point for the replicative helicase, and additional events lead to establishment of the replisome (14). Notably, DnaA is conserved in all bacteria (15), suggesting that the *E. coli* paradigm for initiation of chromosome replication applies throughout the bacterial kingdom. Furthermore, many elements of the bacterial paradigm can be discerned in the more elaborate replication systems found in eukaryotes (16–18).

However, organization of the bacterial genome into the paradigmatic single circular chromosome found in *E. coli* is by no means universal. For example, the genomes of several bacterial families, including the *Vibrionaceae* and the related *Photobacteriacea*, are distributed across more than one chromosome (19). Relatively little is known about the factors and mechanisms that govern replication initiation of secondary chromosomes in bacteria with multipartite

*To whom correspondence should be addressed. Tel: +1 212 650-6062; Fax: +1 212 650-6107; Email: dj@ccny.cuny.edu
Correspondence may also be addressed to Matthew K. Waldor. Tel: +1 617 525 4646; Fax: +1 617 525-4660; Email: mwaldor@research.bwh.harvard.edu
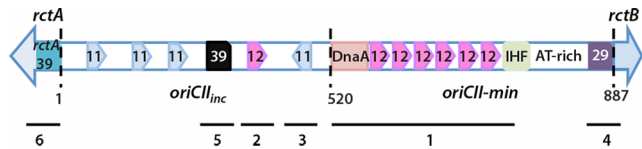† These authors contributed equally to this work as the first authors.

Figure 1. Architecture of the origin of DNA replication of the secondary chromosome (*oriCII*) from *Vibrio cholerae*. *oriCII* contains five distinct binding sites for the RctB initiator, named for their length: 12-mer (pink), 11-mer (cyan), 39-mer (black), 29-mer (purple) and the rctA-39-mer (teal). The *oriCII-min* segment contains six direct repeats of the 12-mer site, and is sufficient to direct replication initiation in the presence of RctB. The other RctB binding sites are proposed to serve regulatory purposes. The origin also contains binding sites for DnaA (light orange) and integration host factor (IHF, light green), and an A-T rich region (white) that is the locus of the initial melting at the origin. The numbered black bars represent the probes used for DNA binding electrophoretic mobility shift assay.

genomes. In *Vibrio cholerae*, the causative agent of cholera, replication of the larger primary chromosome (chrI) is managed by a DnaA-*oriC* system that closely resembles that of *E. coli* (20–22). In contrast, replication of the smaller secondary chromosome (chrII) is managed by a parallel system that contains unique components (20,21). Neither the chrII origin (*oriCII*), nor RctB, its cognate replication initiator protein, bear any sequence similarity to functional analogs utilized by characterized chromosome or plasmid replication systems (19).

RctB is a highly conserved 75.3 kD protein (658 residues), which is unique to the *Vibrionaceae*, and shows no detectible relationship to any other protein in the sequence database. The first ∼500 amino acids of RctB are sufficient to mediate *oriCII*-based replication (19,23,24) and its C-terminal 165 residues may mediate regulatory processes (19,23–25). The restriction of RctB to the *Vibrionaceae*, a large family of organisms that includes several important human and fish pathogens, suggests it as a potential target for discovery and design of novel selective antibacterial agents (26).

The *V. cholerae oriCII* DNA element spans 887 base-pairs, and is organized into two functional domains (Figure 1) (25). These are: (i) a 367 bp segment (*oriCII-min*) that supports RctB-based replication of plasmids containing this sequence in *V. cholerae* and *E. coli* (21), and (ii) an adjacent 520 bp segment (*oriCII*$_{inc}$), which exerts a negative regulatory role on *oriCII*-based replication (21,25). Both *oriCII*$_{inc}$ and *oriCII-min* harbor a variety of sites, referred to as 12-mers, 11-mers, 39-mers and 29-mers based on their lengths, which are known to bind RctB (19,21,27–29) (Supplementary Table S1). *OriCII-min* contains a 167 bp region that harbors six 12-mer sites, arranged with a regular spacing (10 or 11 bps apart) in a head-to-tail manner. Thus, six copies of RctB (or a multiple thereof) are expected to bind to *oriCII-min*, and associate into an oligomeric entity that should retain the head to tail configuration of the 12-mer sites. *OriCII-min* also contains a single binding site for DnaA, which is required for chrII replication (21). The remaining 190 bp of the *oriCII-min* element contains an A-T rich segment, which is likely melted to initiate replisome assembly (25), and a 29-mer RctB binding site that overlaps with the *rctB* promoter (29). Thorough mutational analysis of *oriCII-min* revealed high sensitivity to introduced

changes (e.g. changes in the spacing between 12-mer binding sites impaired *oriCII*-based replication (25)).

With a mass of 75.3 kDa, RctB is larger than other initiators for bacterial (DnaA: ∼53 kDa) or plasmid DNA replication (RepE: 29 kDa, π: 35 kDa), implying that the second *V. cholerae* initiator (DnaA is the first) may encode additional functions not found in other initiators. To gain insights into mechanisms implemented by RctB at *oriCII*, we describe biochemical and structural analyses of RctB. Our findings suggest that RctB is comprised of four structural domains. The two central domains of RctB are structurally related to the plasmid replication initiators RepE and π. However, RctB contains two additional domains not found in the plasmid initiators, and we found that one of these domains is also critical for the initiator to bind to *oriCII* and mediate replication. The finding that the DNA binding surface of RctB is comprised of domains 1, 2 and 3 provokes reexamination of models of binding to the array of 12-mers in *oriCII-min*. The head-to-head dimeric configuration seen in the RctB structure is incompatible with binding to the head-to-tail arrangement of binding sites in *oriCII*; this suggests that dissociation and/or conformational switching in RctB dimers must accompany the initiator's binding to *oriCII*.

MATERIALS AND METHODS

Plasmids

The wild-type RctB expression construct has been described (19). Other RctB expression constructs were generated using conventional PCR-based cloning. Point mutants were generated using the QuikChange® II XL site-directed mutagenesis kit (Agilent).

The *oriCII-min* transformation plasmid, as well as plasmids from which the electrophoretic mobility shift assay (EMSA) probes corresponding to (i) the array of six 12-mer sites, (ii) the inc11 site and (iii) the inc12 site were generated have been described (25). Plasmids containing the EMSA probes corresponding to the inc39 site, the rctA39 site and the PrctB sites were generated by inserting the relevant double-stranded DNA (Supplementary Table S4) into the SmaI restriction of pBlueScript II KS+.

A complete list of plasmids, primers and EMSA probes appears in Supplementary Tables S3–S5, respectively. The sequence of the insert in each plasmid was verified by DNA sequencing (Genewiz).

Protein biochemistry

Proteins in this study were produced using standard methods for preparing recombinant proteins in bacteria (30). Expression plasmids for full length (19) and designed variants (Supplementary Table S2) included C-terminal hexahistidine affinity tags; the full-length RctB constructs (wild type and the mutants), as well as the RctB-1-499 constructs, contained an additional alanine in position 2 following the first methionine, and additional leucine and glutamic acid residues at the C-terminus preceding the hexahistidine affinity tag; RctB-2-124 and RctB-155-483 constructs contained the hexahistidine affinity tag only. All the expression plasmids were grown in *E. coli* BL21. Small-scale growths were performed in LB media supplemented

with 50 μg/ml kanamycin. Cultures were started by addition of an overnight 'starter' culture prepared from a fresh transformation at a 5% volume ratio of starter to culture volume. Cells were cultured at 37°C until their density reached at $OD_{600} \approx 0.6$–0.7, the culture was cooled to ∼20°C, and protein expression was induced by addition of 1 mM isopropyl-β-D-thiogalactopyranoside to the culture medium. Protein expression was allowed to continue at 16°C for 14–18 h. Cells were then harvested by centrifugation and resuspended in buffer A (500 mM NaCl, 50 mM sodium phosphate (pH 8.0), 5% glycerol) at a ratio of 5 ml of buffer A per gram of cells. Large scale growths were carried out in a fermenter as above, except that SuperBroth (12 g/l tryptone, 24 g/l yeast extract, 2.3 g/l KH2PO4, 12.5 g/l K2HPO4, 3.2% glycerol), supplemented with 1 mM $MgCl_2$ and 0.1 mM $CaCl_2$ was used in place of LB, 100% oxygen was bubbled through the media at 0.5–1 l/min, and the agitation rate was set to 450 RPM. The culture was grown at 37°C until its density reached $OD_{600} \approx 2$. Protein expression was induced by addition of 1 mM isopropyl-β-D-thiogalactopyranoside to the media. Protein expression was allowed to proceed for 14–18 h at 27°C. RctB proteins substituted with selenomethionine were prepared as described (31,32).

RctB proteins used for EMSAs were expressed and purified as described previously (21,33).

RctB proteins that were used for crystallization were purified using different purification protocols with a number of chromatography steps. The first purification step was common for all the purification protocols. The proteins were initially purified by thawing frozen biomass cells expressing the appropriate construct into Ni buffer A (500 mM NaCl, 50 mM potassium phosphate, pH 8.0) such that a 3-fold dilution was achieved. Cell lysis was achieved by sonication. The soluble fraction was then isolated by centrifugation and incubated with nickel-nitrilotriacetic acid (Ni-NTA) agarose (QIAGEN) for 40 min at 4°C.

For RctB-2-124, the Ni-NTA agarose beads were washed with a set of buffer solutions with increasing imidazole concentration containing up to 40 mM imidazole; RctB-2-142 was then eluted by washing with 500 mM imidazole. The Ni-NTA purified material was diluted 5-fold (to achieve a final NaCl concentration of 100 mM) with Q-SP buffer A (20 mM Tris 7.4, 5% glycerol, 5 mM beta-mercaptoethanol), and applied to a Q column (Q Sepharose Fast Flow, GE Healthcare) arranged inline with an SP-column (SP Sepharose Fast Flow, GE Healthcare), both equilibrated with Q-SP buffer A. RctB was eluted from the SP column to which it bound using a gradient from 0.1 to 2 M sodium chloride. Fractions containing pure protein were dialyzed into the following buffer: 50 mM sodium chloride, 20 mM Tris pH 7.4, 5% glycerol, 5 mM beta-mercaptoethanol, concentrated and either were used fresh or were flash-frozen in liquid nitrogen, and stored until use. The yield was ∼18 mg/l of culture (3 mg of protein per 1 gram of cells).

For RctB-2-124-L48M (selenomethionine labeled), the Ni-NTA agarose beads were washed with a set of buffer solutions with increasing imidazole concentration containing up to 40 mM imidazole; RctB-2-124-L48M was then eluted by washing with 500 mM imidazole. The resulting

Ni-NTA purified protein was concentrated, and applied to a size-exclusion column (Superdex 200 beads, GE healthcare). Chromatography was carried out in SEC buffer: 500 mM NaCl, 20 mM Tris pH 7.4, 5% glycerol, 5 mM beta-mercaptoethanol. Fractions containing pure protein were concentrated, and either were used fresh or were flash-frozen in liquid nitrogen, and stored until use. The yield was ∼18 mg/l of culture (3 mg of protein per 1 g of cells).

For RctB-155-483 (selenomethionine labeled), the Ni-NTA agarose beads were washed with a set of buffer solutions with increasing imidazole concentration containing up to 60 mM imidazole, RctB-2-155-483 was then eluted with 500 mM imidazole. Fractions containing pure protein were pooled and brought to 1.4 M ammonium sulfate by addition of powder. Then protein was loaded onto butyl column (Macro-Prep® t-Butyl HIC Support, BIO-RAD), and eluted with reverse gradient of ammonium sulfate (gradient from 1.4 to 0.07 M ammonium sulfate). The fractions containing the pure protein were pooled, concentrated and further purified using size-exclusion chromatography (Superdex 200 media, GE healthcare). The final buffer (SEC buffer) contained 500 mM NaCl, 10 mM Tris pH 7.4, 5% glycerol, 5 mM beta-mercaptoethanol. Fractions containing pure protein were concentrated, and either were used fresh or were flash-frozen in liquid nitrogen and stored until use. The yield was ∼15 mg/l of culture (2.5 mg of protein per 1 g of cells).

One measure of the integrity of point mutants of RctB, in comparison to wild-type, was to assess solution properties by SEC (Supplementary Figure S6). SEC was performed using a 21.6 ml column packed with Superdex 200 prep grade (GE Healthcare) in the following buffer: 500 mM NaCl, 10 mM Tris-HCl pH 7.4, 5% glycerol, 5 mM 2-merceptoethanol. Estimates of the masses of various RctB proteins were obtained by comparing elution volumes against those by a set of molecular weight standards (GE Healthcare).

Electrophoretic mobility shift assay

Due to the requirement for methylated DNA (methylation at the N6 position of the adenine residues in the sequence GATC) for RctB binding (25), it was necessary to produce EMSA probes by excising them from methylated plasmid DNA. Probe fragments were cloned into the pBlueScript II SK(+) vector and the constructs were prepped from Dam(+) *E. coli*. The constructs were then digested with XbaI and XhoI and treated with CIP (NEB) for 2 h at 37°C. The digests were separated on 1% agarose gels and the probes were excised and extracted from the gel. The DNA was then desalted using Illustra MicroSpin G-50 Columns (GE) and the concentration was quantitated with a NanoDrop (ThermoFisher Scientific). The probe ends were then labeled with T4 PNK (NEB) and a slight excess of gamma P32 ATP. The probes were separated from the nucleotide using Illustra MicroSpin G-50 Columns (GE). The labeled probes were then phenol–chloroform extracted and subjected to ethanol precipitation.

Binding reactions were conducted in 20 μl of 1x EMSA reaction buffer: 20 mM TrisCl pH 7.5, 1 mM ethylenediaminetetraacetic acid, 150 mM NaCl, 1 mM $MgCl_2$, 100

μg/ml bovine serum albumin,12.5 μg/ml poly (dI-dC). Radiolabeled probes were added to a final concentration of 0.1 nM. The reactions were incubated for 10 min at room temperature. Five microliters of 1x EMSA reaction buffer with 50% glycerol was added to the reactions, which were then loaded onto 6% DNA retardation gels (ThermoFisher Scientific) and run in 0.5% TBE buffer. The probes contained binding sites embedded in a larger DNA sequence, and for the complete sequence of the probes, please refer to Supplementary Table S5. The gels were then dried onto filter paper and then exposed to a Phosphor Screen and imaged with Fuji FLA-5000 imager. Band intensities were quantified using Image Studio™ Lite software (LI-COR, Inc). The data were fit to appropriate binding equations using KaleidaGraph 4.5.

Transformation efficiency assay

The transformation assay was performed as previously described (19,25). To place the results of this assay on a quantitative basis, we noted that, when cells that harbored a plasmid expressing wild-type RctB were transformed with an *oriCII-min* containing plasmid, 200–300 colonies were obtained. Under these conditions, this number of colonies was set as the maximum in our quantitative scale, thus, this assay could be used to analyze mutant RctB proteins that have ∼1% replication competence. The standard deviation for all RctB proteins tested was 0.09 or less across the replicates.

Matrix-assisted laser desorption time-of-flight mass spectrometry (MALDI-TOF MS) analysis

To probe for RctB domain organization, trypsin proteolysis of full-length RctB was performed at room temperature with a trypsin:RctB ratio of 1:500 followed by MALDI-TOF MS analysis. At various time points during digestion, 0.5 μl of the sample was collected and mixed with 9.5 μl of matrix consisting of a saturated solution of α-cyano-4-hydroxycinnamic acid in a 1:3:2 (v/v/v) mixture of formic acid/water/isopropanol. An aliquot of 0.5 μl of this protein-matrix solution was spotted onto a MALDI plate precoated with an ultrathin layer of matrix (34,35). The sample spots were washed for a few seconds with 2 μl of cold 0.1% aqueous trifluoroacetic acid solution. MALDI spectra were acquired in linear, delayed extraction mode using a Spiral TOF JMS-S3000 (JEOL, Tokyo, Japan). The instrument is equipped with a Nd:YLF laser, delivering 10-Hz pulses at 349 nm. Delayed extraction time was set at 1 μs and acquisition was performed with a sampling rate of 2 ns. Each MALDI spectrum corresponded to an average of 500 scans. Mass calibration was performed using a technique of pseudo-internal calibration wherein a few shots on a nearby calibrant spot are collected and averaged with the sample shots into a single spectrum. The spectra were processed and analyzed using MoverZ (Proteometrics, LLC).

For characterizing protein degradation in the crystallization drop, 1–2 protein crystals covered in residual mother liquor were removed from a crystallization drop, and dissolved in the matrix solution (same as above). A 0.5 μl aliquot of the resulting protein-matrix solution was spotted onto a MALDI plate precoated with an ultrathin layer of matrix (34,35). The sample spots were then washed for a few seconds with 2 μl of cold 0.1% aqueous trifluoroacetic acid solution. MALDI spectra were acquired and processed as detailed above.

Native mass spectrometry

RctB proteins samples were diluted to 10–20 μM with 10 mM Tris pH 7.4, 500 mM NaCl and subsequently buffer-exchanged into the native MS buffer (500 mM ammonium acetate, 0.01% Tween-20) using the Zeba microspin desalting columns (Thermo Scientific) with a 40-kDa molecular weight cut-off. The buffer-exchanged samples were then further diluted with the native MS buffer into the desired concentrations ranging from 0.1 to 5 μM. An aliquot (2–3 μl) of the sample was loaded into an in-house fabricated gold-coated quartz capillary and sprayed using a static nanospray source into the Exactive Plus EMR instrument (Thermo Fisher Scientific). The EMR was calibrated using cesium iodide. Typical native MS parameters included spray voltage, 0.9–1.5 kV; capillary temperature, 100°C; S-lens RF level, 200; resolving power, 8,750 or 17,500 at m/z of 200 corresponding to 32 or 64 ms analyzer transient duration, respectively; AGC target, 5×105; number of microscans, 5; maximum injection time, 200 ms; injection flatapole, 8 V; interflatapole, 7 V; bent flatapole, 6 V; ultrahigh vacuum pressure, $3–5 \times 10{-}10$ mbar; total number of scans, 100. The in-source dissociation and high energy collision dissociation parameters were varied accordingly. RAW files were processed manually using Thermo Xcalibur Qual Browser (version 3.0.63).

Crystallization of RctB

All aspects of the crystallization of full length and shorter variants of RctB were carried out using an automated crystallization and analysis instrument available in house. Crystals of RctB domain 1 (1 - 124) were prepared using the sitting drop vapor diffusion method by mixing 0.1, or 0.2 or 0.4 μL of the protein solution (22.4 mg/ml RctB-2-124 in 20 mM Tris pH 7.4, 50 mM sodium chloride, 5% glycerol, 5 mM 2-mercaptoethanol) and 0.2 μL of reservoir solution (0.24 M Sodium malonate pH 7.0, 20% w/v PEG 3350). Crystals grew within 7 days. Crystals were flash-frozen in liquid nitrogen without additional cryoprotection for X-ray diffraction.

Crystals of selenomethionine-substituted RctB domain 1 (1–124, L48M) were grown by mixing 0.1, or 0.2 or 0.4 μl of the protein solution (20.1 mg/ml RctB-2-124-L48M in 20 mM Tris pH 7.4, 500 mM sodium chloride, 5% glycerol, 5 mM 2-mercaptoethanol) and 0.2 μl of reservoir solution (0.1 M Sodium HEPES pH 7.5, 20% w/v PEG10000). Crystals grew within 7 days. In preparation of cryogenic X-ray diffraction, crystals were transferred sequentially, over a period of 10 min, into a set of drops that contained 5%, 10%, 15 and 20% glycerol.

Crystals of selenomethionine-substituted RctB domains 2–3 (155–483) were prepared by mixing 0.1, or 0.2 or 0.4 μl of the protein solution (22.7 mg/ml RctB-155-483 in 20 mM Tris pH 7.4, 500 mM sodium chloride, 5% glycerol, 5 mM 2-mercaptoethanol) and 0.2 μl of reservoir solution (0.1 M

Bis-Tris propane pH 6.5, 0.2 M Magnesium chloride, 2% w/v PEG 8000). Crystals grew 10–14 days. In preparation of cryogenic X-ray diffraction, crystals were transferred sequentially into a set of drops that contained 5%, 10%, 15 and 20% glycerol over a period of 10 min.

X-ray data collection

Diffraction data for the RctB-2-124 crystal was recorded at the X-25 beam line at Brookhaven National Laboratory using a wavelength of 0.979 Å. The data extended to Bragg spacings of 2.0 Å. RctB-2-124 crystallized in space group P2$_1$, with the following cell parameters: a = 45.84 Å, b = 52.15 Å and c = 63.53 Å, α = 90°, β = 101.5°, γ = 90°. Matthews analysis indicated that the crystal had two molecules in the asymmetric unit (V$_m$ = 2.51 A^3/Da). Diffraction data for crystals of RctB-2-124-L48M were measured at the Stanford Synchrotron Radiation Lightsource at SLAC National Accelerator Laboratory using a wavelength 0.9791 Å. The data extended to Bragg spacings of 2.0 Å. RctB-2-124-L48M crystallized in space group P1, with the following cell parameters a = 32.45 Å, b = 38.17 Å and c = 63.04 Å, α = 97.46°, β = 91.49°, γ = 98.43°. Matthews analysis revealed that two molecules in the crystallographic asymmetric unit (V$_m$ = 2.58 A^3/Da). Data for the RctB-155-483 crystal were recorded at the Northeastern Collaborative Access Team facility at the Advanced Photon Source at Argonne National Laboratory using a wavelength 0.9792 Å. The data extended to Bragg spacings of 2.6 Å. RctB-155-483 crystallized in space group R3, cell dimensions are a = 128.54 Å, b = 128.54 Å and c = 127.8 Å, α = 90°, β = 90°, γ = 120°. Matthews analysis suggested that the two molecules resided in the asymmetric unit (V$_m$ = 2.58 A^3/Da).

Structure determination and refinement

Diffraction data were processed using HKL2000 software (36). Phenix (37) was used to solve the structures of RctB-2-124-L48M and RctB-155-483 using the single wavelength anomalous dispersion method and crystals with selenomethionine substituted protein. The final model of RctB-2-124-L48M consists of residues 7–122 with a crystallographic *R* factor of 21.46% and *R*free of 25.09%. The final model of RctB-155-483 consists of residues 182–472 (with a 14-residue gap 242–255) with a crystallographic *R* factor of 24.00% and *R*free of 28.38%.

The RctB-AA-2-124 structure was solved using molecular replacement (with the RctB-2-124-L48M structure as a search model) in Phenix (37). The final model of RctB-2-124 consists of residues 7–122 with a crystallographic *R* factor of 24.03% and *R*free of 28.04%. In all cases, initial models were improved using several rounds of model building and refinement as implemented in Phenix (37), Coot (38) and Phenix.refine (37). Structural models were visualized with Coot (38) and PyMol (MacPyMOL: The PyMOL Molecular Graphics System, v1.7.0.5 Schrödinger, LLC).

Crystallographic computing and structural analyses

Global structural alignments were performed using DALI on-line server (39) and PDBefold online-server (40). Struc-
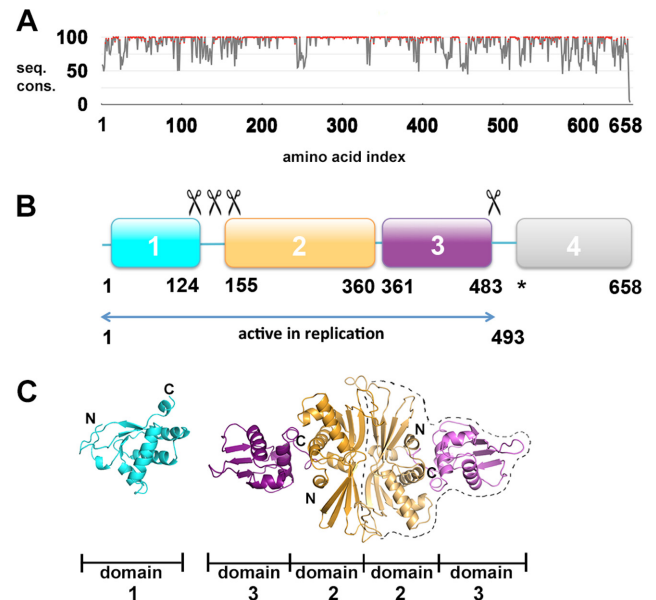

Figure 2. Sequence conservation, domain architecture and structures of the RctB initiator protein. (**A**) A BLAST alignment (63) consisting of 99 RctB orthologues was converted to a numerical conservation score where equivalence of amino acid at each position was established using a normalized BLOSUM62 matrix (64). Conservation score is plotted against the primary sequence (gray lines). Reds dots represent positions with greater 90% sequence conservation. (**B**) Domain architecture of RctB as deduced from mass spectrometric analysis of proteolytic digestion products. A precise boundary for the fourth domain of RctB could not be obtained owing to this region's sensitivity to limited proteolysis. The cleavage sites revealed by our analysis are depicted with scissors. (**C**) The crystal structures of RctB domain 1 and domains 2–3 are shown in a cartoon representation. The coloring scheme employed corresponds to that in Figure 2B. The dotted line represents one protomer of the RctB dimer. Full-length RctB forms a dimer, and the dimerization interface localizes to domain 2.

tural alignments using particular regions of a structure were performed in Pymol (MacPyMOL: The PyMOL Molecular Graphics System, v1.7.0.5 Schrödinger, LLC). Other calculations were carried out in the CCP4 (41) and the Uppsala Software Factory (42,43) software suites.

RESULTS

RctB folds into a four-domain structure organized around two central domains related to plasmid initiators

Our efforts to crystallize full-length RctB were thwarted by spontaneous proteolysis in the crystallization drop. Consequently, we used limited proteolysis and MS to identify stable fragments more amenable for structure determination by X-ray crystallography. Limited proteolysis of RctB resulted in rapid release of a ∼14 kDa N-terminal segment (residues 1–124, referred to as domain 1 below) and a ∼38 kDa segment (residues 155–483, referred to as domains 2–3 below) (Figure 2, Supplementary Figures S1, and S2). Although our analysis did not yet identify a stable fragment corresponding to the C-terminus (residues 484–658), RctB mutants deleted for these C-terminal residues exhibit defects in binding to the *inc* and *rctA* 39-mer sequences in *oriCII* (19,24), suggesting that this segment constitutes a

fourth domain. Thus, RctB appears to adopt an architecture that includes four structural domains (Figure 2).

The structures of both the 14 kDa N-terminal (domain 1) and the 38 kDa middle fragments (domains 2–3) were determined using X-ray crystallography (Supplementary Table S2). Two crystal forms of domain 1 (one with the wild-type sequence and native sulfomethionine, and a second with an L48M substitution containing selenomethionine) were used to decipher its structure; both forms contain two copies in the asymmetric unit, but with different crystal packing arrangements. The four structures of the N-terminal domain of RctB (molecules A and B from each of the two crystal forms) were virtually identical (RMSD over C-alpha atoms varies from 0.23 to 0.543 Å, Supplementary Figure S3). We focused our analyses on molecule A of crystal form I (residues 2–124, L48M), as this was the best-defined structure. RctB domain 1 consists of an array of four helices packed against a four-stranded beta sheet (Figure 2C). Comparative structural analyses using the Dali (39) and PDBefold (40) tools revealed that RctB's domain 1 closest structural neighbors are a number of DNA binding proteins, including transcription factors and replication initiators. The top hit (1Q1H) was archaeal TFIIE (a component of the core transcriptional machinery). Closer analysis of hits with Z-scores of 5.0 or higher revealed a high degree of structural similarity between three alpha helices and two beta-sheets of RctB domain 1 (residues 42–57, 65–72, 78–91, 94–97, 111–114) and a family of winged-helix-turn-helix motif proteins (RMSD from 1.3 to 3.0 Å, Supplementary Figure S4). No function has yet been ascribed to RctB domain 1; however, these comparisons raise the possibility (supported by findings shown below) that domain 1 binds to DNA.

The 38 kDa fragment of RctB (residues 155–483) crystallized as a dimer in the asymmetric unit; the two RctB monomers are configured in a head-to-head arrangement. The dimerization interface of RctB domains 2–3 localizes exclusively to domain 2, and is comprised of two seven-stranded beta sheets arranged in a domain swapped configuration whereby one monomer contributes four of the seven strands to one sheet, and the remaining three come from the second monomer; this arrangement is reversed in the second sheet (Figure 2, Supplementary Figure S5). This configuration represents the most extensive protein-protein interface in the crystal ($\sim$4300 Å^2), and is likely to be functionally significant (44). We pursue this question further below. Each 38 kDa monomer is composed of two domains – residues 182 to 360 (henceforth domain 2) and residues 361 to 472 (henceforth domain 3). Superposition analysis revealed a small ($\sim$9°) difference between the relative orientation of domains 2 and 3 in the two copies present in the dimer seen in the asymmetric unit, suggesting flexibility between the two domains (Supplementary Figure S6).

Structural comparisons against the PDB (39,40) revealed that the 38 kDa fragment of RctB bears significant similarity to several replication initiator proteins from plasmid DNA replication systems, including π (2NRA, (45)), RepE (2Z90, (46), 1REP, (47)) and RepA (1HKQ, (48)), (Z scores of between 7.7 and 9 using the DALI server). Nearly every secondary structure element of π or RepE can be mapped on to a corresponding element of RctB domain 2 or 3 (Figure 3, Supplementary Figures S7 and S8). For RepA, the structure of only one domain of the two is available, and its secondary structure elements correspond to RctB domain 2 (Figure 3); the structure of the second domain of RepA is not known, but it likely resembles the corresponding domain of RepE based on primary sequence considerations (47). However, domains 2 and 3 of RctB also include some unique structural elements (Figure 3). Both RepE and RepA crystallized as dimers (46,48), with a beta sheet arranged in a domain swapped configuration as in RctB. However, unlike in RctB, the interfacial beta sheet for the RepE, and RepA initiators contains five strands instead of seven; however, all—RepE, RepA and RctB—are arranged as head-to-head dimers. The structures of these plasmid initiators, like that of the corresponding RctB fragment, consist of two domains; both have been shown to bind DNA (45,47,48), suggesting that RctB domains 2 and 3 might likewise both bind to DNA (as is confirmed below). Additional database searches using the RctB domain 2 and 3 structures individually revealed similarities with a variety of winged-helix-turn-helix DNA binding domains, including the archaeal and eukaryotic replication initiator proteins Cdc6 and Orc2 proteins (Supplementary Figures S9 and S10). Collectively, these analyses suggest that RctB is a four-domain protein with a core region (domains 2 and 3) structurally homologous to plasmid initiators, and two unique peripheral domains (domains 1 and 4), not present in plasmid initiators.

RctB harbors at least three DNA binding surfaces

Mutational analyses were used to explore the possibility that the winged-helix-turn-helix motifs in RctB domains 1, 2 and 3 mediate DNA binding. Comparisons against close structural homologs bound to DNA were used to predict RctB residues likely to contact DNA. Mutations at the selected sites were introduced into full-length RctB, and the DNA binding capacity of mutant proteins was subsequently assessed using EMSA. Additionally, the biological activity of mutant proteins was assessed using a transformation assay in which the capacity of RctB variants to support *oriCII-min*-based plasmid replication was determined (19,25). Solution properties of mutant RctBs were tested as well; unless otherwise noted, solution properties of the mutant proteins determined by size-exclusion chromatography did not differ from those of the wild-type protein, indicating that substitutions did not cause aggregation or degradation of mutant proteins (Supplementary Figure S11). Notably, our analyses below do not represent a complete census of DNA binding contacts by RctB.

For RctB domain 1, comparative analyses using five distinct protein–DNA complexes (Figure 4A) suggested that Gln 83 on helix αD might be important for DNA binding. Additionally, given their 100% conservation in RctB amino acid sequences from diverse *Vibrio* species (Supplementary Figure S12), we hypothesized that the neighboring positively charged residues Arg 84 and Arg 86 might also have roles in DNA binding. To evaluate these predictions, we mutated all three positions to alanine, and measured the affinity of the resulting triple mutant (RctB Q83A-R84A-R86A, referred to as domain 1 triple mutant below) to six probes
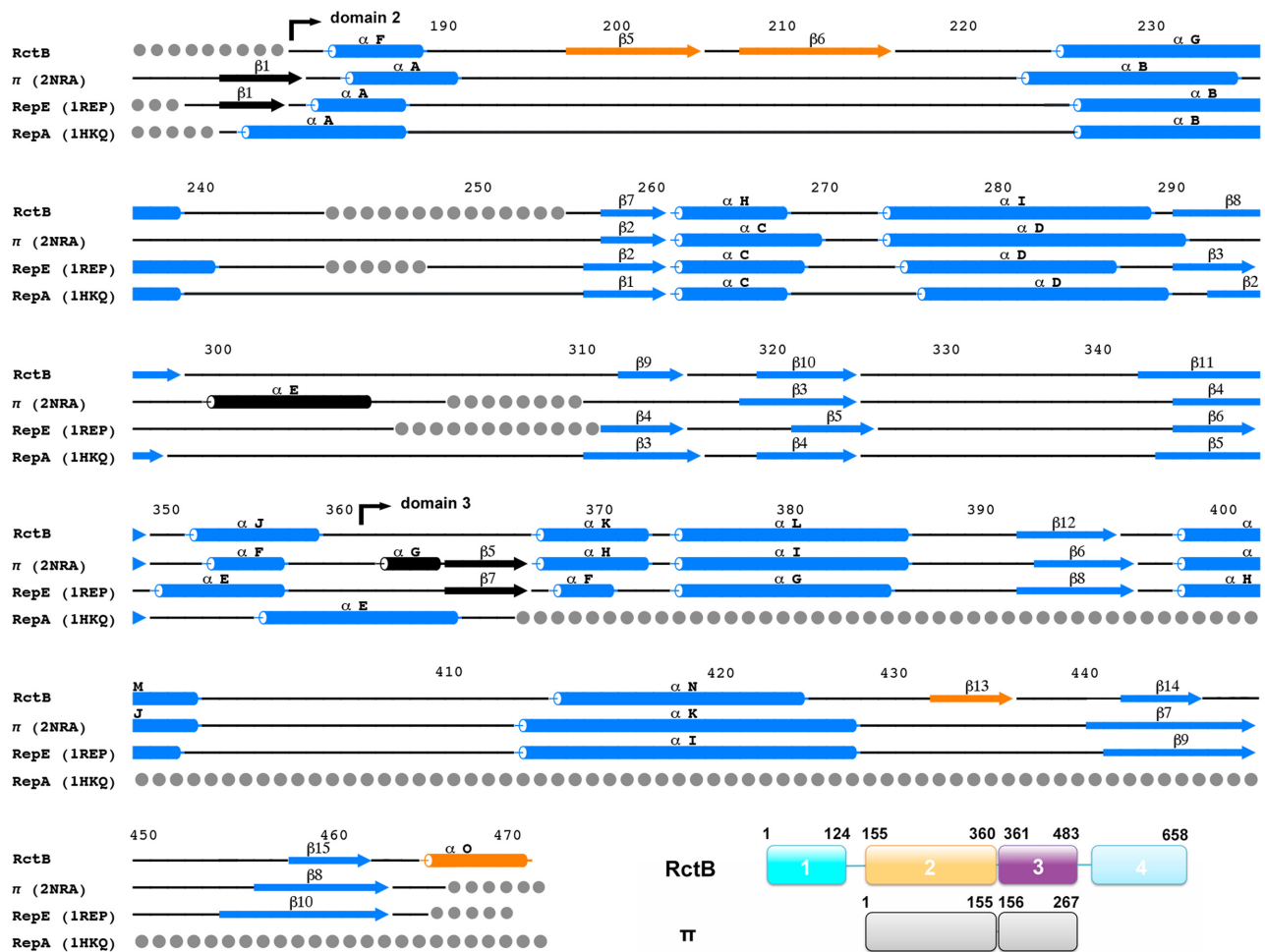
Figure 3. The structures of the central domains of RctB resemble the structures of plasmid initiators π, RepE and RepA. The structure of the central domains (domains 2–3) of RctB was aligned to the π (Z-score = 9.0, RMSD = 4.9 Å), RepE (Z-score = 8.4, RMSD = 6.1 Å) and RepA (Z-score = 7.7, RMSD = 2.9 Å) plasmid replication initiators. Depicted here is an alignment of secondary structure elements extracted from this alignment. Elements shared by each protein are colored in cyan. Elements unique to RctB are colored in orange, while elements present in the plasmid initiators, but not in RctB, are colored in black. Grey circles represent regions of the various structures that were not modeled. Depiction of RepA is limited to the one available domain. A schematic in the bottom right corner shows that RctB middle portion (domains 2 and 3) aligns with the entire structure of π, therefore, the secondary structure alignment is shown only for domains 2–3 of RctB and structures of the plasmid initiators.

containing nucleotide sequences derived from *oriCII*: (i) the array of six 12-mers in *oriCII-min*, (ii) a single 12-mer sequence, (iii) a single 11-mer sequence from the *inc* region, (iv) the 29-mer sequence (corresponding to the RctB promoter), (v) the inc39-mer sequence and (vi) the *rctA*39-mer sequence (Figure 1). The domain 1 triple mutant bound to the 6 × 12-mer array EMSA probe with an apparent Kd (K_{dapp}) > 23 000x higher than that of wild-type RctB (Figure 4, Supplementary Figure S13). Determination of precise K_{dapp} values from EMSAs using other probes was challenging owing to complex binding curves; nevertheless, the trend we observed with probe #1 was recapitulated with probes #2 and #3 (Supplementary Figures S14 and S15). However, domain 1 triple mutant binding to the 29-mer, 39-mer and rctA sequences was similar to that of wild-type RctB (Supplementary Figures S16–S18). The near wild-type binding of the RctB-Q83A-R84A-R86A mutant to a subset of the probes examined supports the idea that its structural integrity is intact. Thus, the role of domain 1 binding to DNA

appears to vary depending upon the target sequence, and domain 1 does not appear to play a critical role in binding to most regulatory sequences outside of *oriCII-min*. Consistent with its severe deficiency in binding to the *oriCII-min* probe, RctB domain 1 DNA-binding mutant failed to support *oriCII-min*-based replication (Figure 4). Taken together, these observations strongly suggest that RctB domain 1 binds *oriCII* DNA, and that this function is critical for the capacity of RctB to mediate *oriCII*-based replication.

A similar experimental approach was used to assess candidate DNA-binding residues in RctB domains 2 and 3. The structure of RctB domain 2 was compared to those of plasmid initiators (RepE and π) in complex with DNA, as well as to a variety of winged-helix-turn-helix containing protein–DNA complexes. As such, domain 2 residues Lys 271, Lys 272, Ser 274, Arg 278, Asp 279 and Arg 282 were selected for analysis. The residues at these positions were absolutely conserved in all RctB sequences examined (Sup-
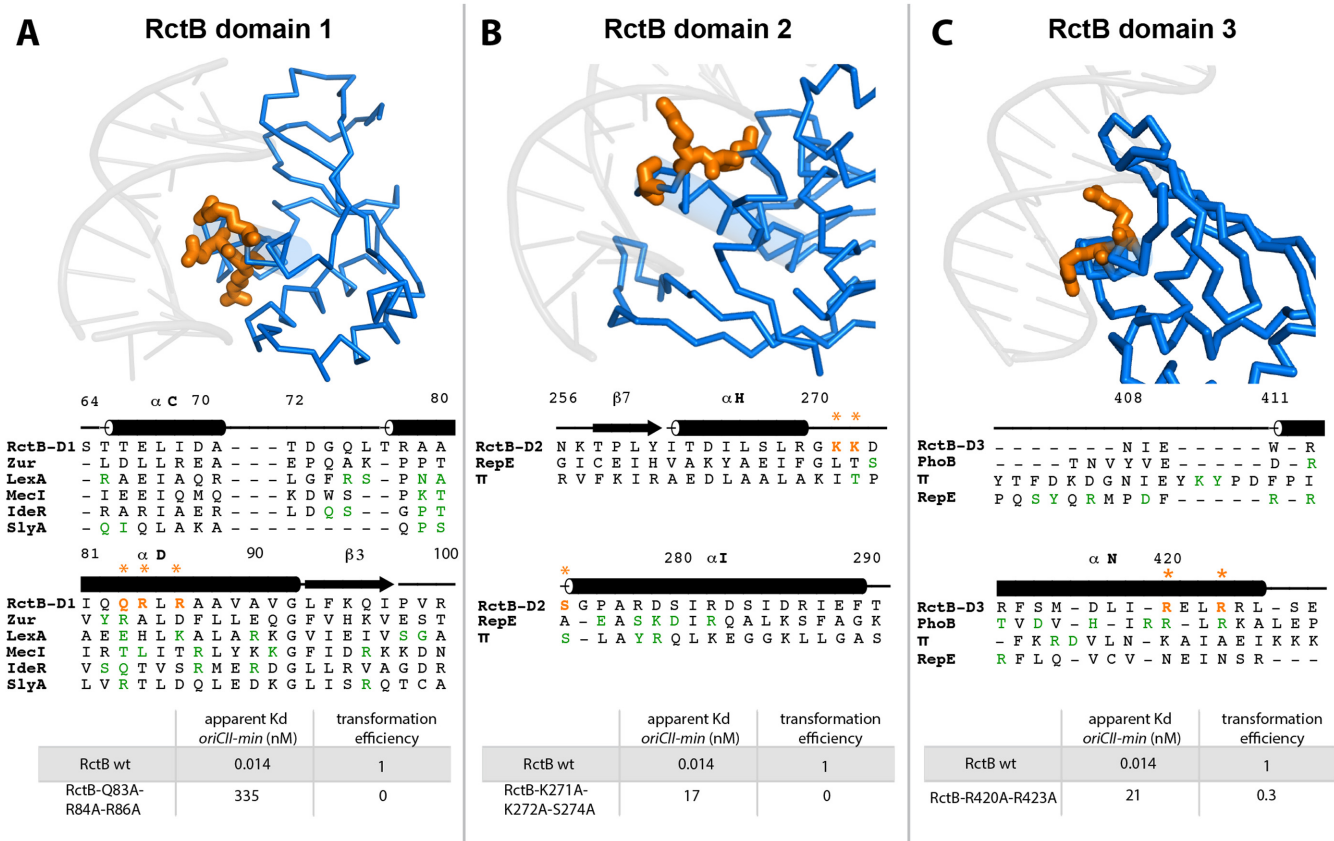
Figure 4. RctB has three DNA binding domains. Structural comparisons predicted the presence of at least three distinct DNA binding domains in RctB. To evaluate these predictions, three mutant forms of RctB were studied: (**A**) RctB Q83A-R84A-R86A (domain 1), (**B**) RctB K271A-K272A-S274A (domain 2) and (**C**) RctB R420A-R423A (domain 3). The top of each panel depicts a ribbon representation (colored blue) of the predicted DNA binding domain of RctB modeled onto a DNA molecule taken from a structural homolog bound to its DNA target. The residues selected for analysis are shown in a ball and stick representation (colored orange). The middle portion of each panel summarizes, in a sequence alignment format, the structural alignment of the DNA binding domains that emerged from database searches. The residues tested in this study are shown in orange, and labeled with orange stars. Shown in green are positions implicated in DNA binding by other studies (45,47,65–70). The lower portion of each panel shows the binding affinity for *oriCII-min* and performance in the transformation assay by wild-type and the mutant RctB proteins.

plementary Figure S12). Two distinct triple mutant proteins, RctB-K271A-K272A-S274A (referred to as the first domain 2 triple-mutant below) and RctB-R278A-D279A-R282A, were prepared to test the importance of the substituted residues in RctB binding to *oriCII* and in replication. The first domain 2 triple-mutant (RctB-K271A-K272A-S274A) exhibited reduced binding affinity to all six *oriCII* derived DNA probes examined (Figure 4, Supplementary Figures S13–S18); e.g. its binding affinity (apparent Kd) for the 6 × 12-mer array probe was reduced by ∼1000-fold. A similar trend was observed with the remaining probes tested (Supplementary Figures S14–S18). Concordant with its markedly defective binding to *oriCII* DNA sequences, the first domain 2 triple-mutant (RctB-K271A-K272A-S274A) was also unable to support *oriCII*-based replication (Figure 4). The RctB-R278A-D279A-R282A mutant could not be produced in soluble form and was not analyzed.

Candidate DNA-binding residues in domain 3 were identified through structural alignment of RctB to PhoB bound to its target DNA (PDB entry 2Z33); based on this analysis, we anticipated that a number of residues, including Arg

420 and Arg 423, would be required for DNA binding and generated the RctB-R420A-R423A, referred to as domain 3 double-mutant. Similar to the domain 1 triple-mutant, domain 3 double-mutant bound to three of the six DNA probes tested differently than wild-type RctB. The K_{dapp} of RctB R420A-R423A binding to the 6 × 12-mer array probe was ∼1500-fold lower than wild-type RctB (Figure 4, Supplementary Figure S13), and similar marked reductions in binding to individual 12-mer and 11-mer containing probes were observed (Supplementary Figures S14 and S15). However, domain 3 double-mutant binding to the 29-mer, 39-mer and rctA sequences was similar to that of wild-type RctB (Supplementary Figures S16–S18). We note that the near wild-type binding of the domain 3 double-mutant to a subset of the probes examined supports the idea that its structural integrity is intact. Moreover, in contrast to the domain 1 and domain 2 mutants (RctB Q83A-R84A-R86A and RctB-K271A-K272A-S274A, respectively) domain 3 double-mutant could support *oriCII-min*-based replication, albeit at reduced efficiency compared to wild-type RctB.

Collectively, these experiments strongly suggest that at least three of the four RctB domains are involved in contact-

ing DNA, and, thus, that the protein contains a much more extensive DNA binding surface than was previously appreciated (49). Moreover, the observation that the R420A-R423A mutation only disrupts binding to a subset of *oriCII*-derived sequences raises the possibility that RctB forms structurally distinct complexes on its varied DNA targets within *oriCII*, and that these complexes rely on different RctB domains to contact DNA (a summary of the DNA-binding phenotypes of all the mutants appears in Supplementary Table S6). However, elucidation of the precise division of labor between the three RctB DNA binding domains will require future structural and functional analyses.

RctB forms a head-to-head dimer not compatible with origin binding

RctB crystallized in a head-to-head dimeric configuration. However, the head-to-tail array of six 12-mer sites at *oriCII* implies that the complex on DNA will feature an RctB oligomer with a matched configuration. Also, RctB is known to be a dimer in solution, but its configuration has not been described (49). To better understand RctB oligomer dynamics, we performed mass measurements in solution, examined crystal packing for clues on the nature of potentially distinct oligomers (dimers and higher order oligomers), and measured the effects of disrupting the dimer seen in the crystal. First, we analyzed the oligomeric state of full-length RctB and a panel of single and multidomain RctB fragments, using native MS (Figure 5A, Supplementary Figure S19 and S20 shows SEC data and native mass spectrometry data as well). Our findings indicate that full-length RctB is a dimer in solution, consistent with previous reports (24,49). In addition, only the segments containing the wild-type domains 2–3 form dimers in solution, while all others are monomeric under the conditions tested (Figure 5A, Supplementary Figure S19). This finding implies that, in solution, the dimer interface is mediated by the core plasmid initiator homology domains (domains 2–3) of RctB.

Second, we examined the packing environments associated with the two crystal forms of domain 1 and the single crystal form of domains 2–3 for potential physiologically relevant interfaces. Both RctB domain 1 and domains 2–3 crystallized as dimers in the asymmetric unit. The surface area buried by the various interfaces made by domain 1 in the crystal ranged from 30 to 1340 Å^2, values at the low end for a physiologically relevant interface (44). Thus, we conclude that the likelihood of physiologic relevance for one of the interfaces made by domain 1 in the crystal is low. This finding is in concert with results from native MS of wild-type domain 1 (bottom spectrum in Figure 5A).

In contrast, the non-crystallographic dimer of RctB domains 2–3 buries an extremely large amount of surface area (~4300 Å^2), a value consistent with physiologic relevance (44). To further explore the biological role of the RctB dimer seen in the crystal, we substituted a proline residue (D314P) in the beta strand closest to the dimer interface to disrupt the dimerization process and produce a monomeric form; such a strategy was used with the RepE plasmid initiator (47). This D314P substitution was intro-
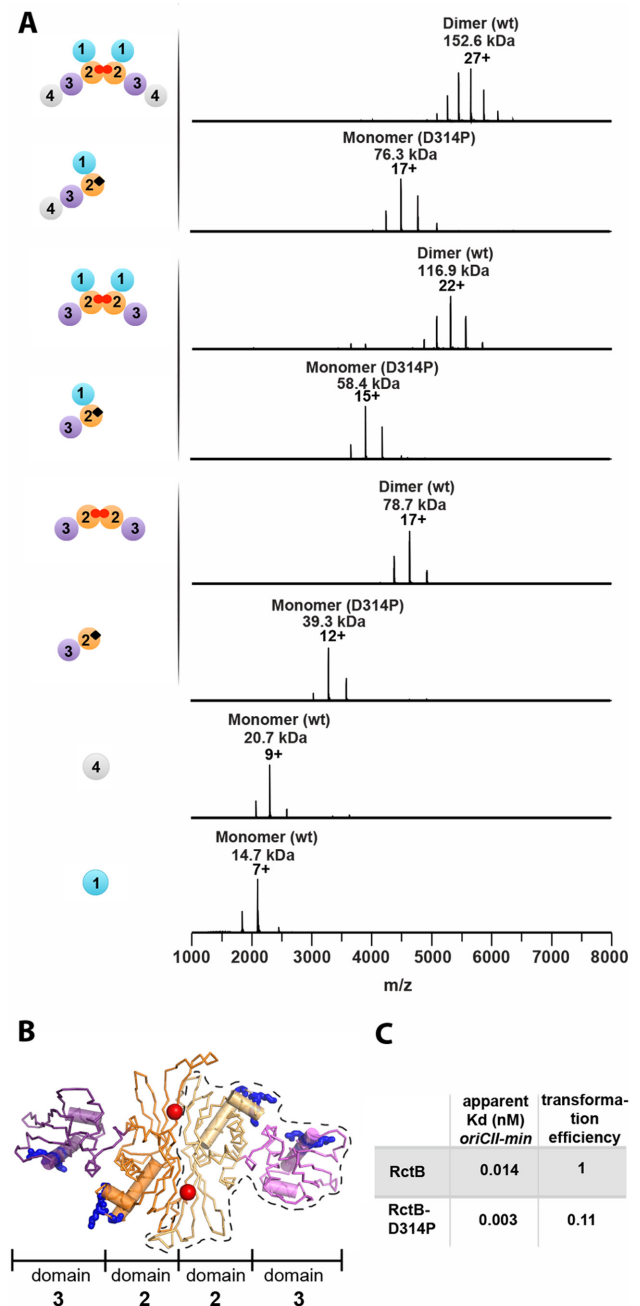

Figure 5. RctB adopts a head-to-head dimeric configuration in solution. (**A**) Native mass spectrometric analysis of the oligomeric state of full length and truncated constructs of RctB, including variants that harbored the D314P mutation (wild-type: red, mutant: black). Spectra for the wild-type and D314P entities are grouped together. To the left of each spectrum appears a schematic, colored as in Figure 2B, of the configuration revealed by the analysis. (**B**) The head-to-head dimer of RctB seen in the crystal. The Asp314 residue on each protomer is depicted as a red sphere. Residues shown to be involved in contacts to DNA are depicted in the ball-and-stick representation and colored dark blue. The two monomers of RctB are colored in varying shades (domain 2: orange, domain 3: purple). One of the RctB monomers is outlined with a dashed line. (**C**) Binding affinity for *oriCII-min* and performance in the transformation assay for wild-type and RctB-D314P.

duced into three RctB constructs: (i) full-length (residues 1–658), (ii) the smallest fragment that is active in replication initiation (residues 1–499) and (iii) the domains 2–3 construct (residues 155–483). Native MS analyses of these mutant proteins revealed that they were all monomers under the conditions tested (Figure 5A, Supplementary Figure S19). Furthermore, DNA binding assays indicated that monomeric RctB-D314P bound to all six probes with near wild-type affinity (K_{dapp} for the binding of the D314P mutant to *oriII-min* was 0.003 ± 0.006 nM versus 0.014 ± 0.01 nM for the wild-type) (Figure 5C). This stands in contrast to results from the transformation assay where the capacity of the RctB-D314P mutant to support replication was reduced (efficiency of 0.11 versus 1 for wild-type) (Figure 5C). These findings suggest that the head-to-head dimer of RctB observed in the crystal corresponds to the dimer revealed by native MS in solution. Additionally, the solution configuration of the RctB dimer implies incompatibility with binding to the head-to-tail array of 12-mer binding sites seen in *oriCII*. It is likely that a substantial rearrangement will accompany formation of the RctB – origin DNA complex that mediates replication initiation. The incompatibility of RctB head-to-head dimer with the head-to-tail array of the 12-mer binding sites is not entirely surprising, since the same is true for plasmid initiator systems. Plasmid initiators, which are structurally related to RctB, also exist as head-to-head dimers in solution, and the current model suggests monomerization takes place prior to binding the head-to-tail sites on the replication origin (46).

Taken together, our data show that RctB adopts a head-to-head dimeric configuration in solution; this arrangement resembles similarly configured dimers of the RepA and RepE plasmid initiators (46,48). Moreover, the dimerization interface is localized to RctB domain 2. Notably, our findings do not exclude the possibility that other segments of RctB may play significant roles in oligomeric forms of RctB, indeed the symmetry mismatch between the 2-fold rotational symmetry of the head-to-head dimer and the translational symmetry of the RctB binding sites at the origin make this very likely.

DISCUSSION

In contrast to the well-studied DnaA-OriC ensemble that operates in all bacteria, little is known about molecular mechanisms that mediate replication of secondary chromosomes in bacteria with multipartite genomes. RctB, the conserved initiator of chrII replication among the *Vibrionaceae*, lacks homologs outside of this large family of organisms whose genomes are divided between two chromosomes. Although RctB bears no significant sequence similarity to other proteins, we demonstrate here that the structure of the two central domains of RctB (RctB 2–3) bears significant structural similarity to several well-characterized plasmid initiators including RepE (from the F-plasmid), RepA (from the pPS10 plasmid) and π (from the R6K plasmid). However, RctB is considerably larger, and contains at least 2 additional domains. Three RctB domains contain winged-helix-turn-helix DNA binding motifs, all of which were implicated in binding to *oriCII*, and in the initiator's capacity to mediate *oriCII*-based replication. In the crystal and
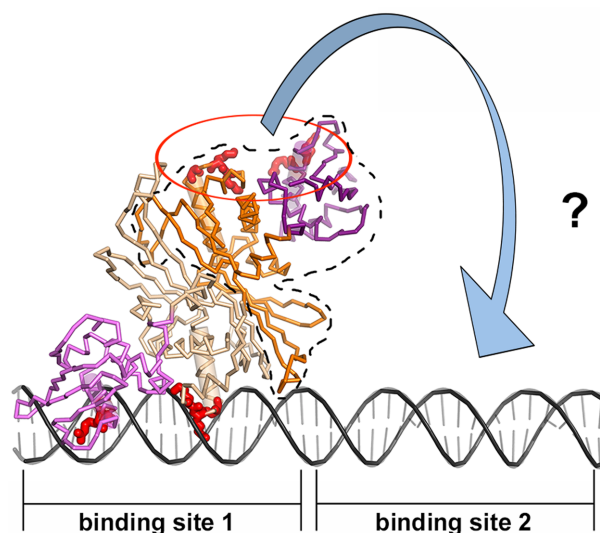

Figure 6. Incompatibility of a head-to-head dimer structure with origin binding. RctB middle fragment structure is shown as a ribbon representation. Domain 2 is colored in different shades of orange, domain 3 is colored in different shades of purple. The top RctB monomer is shown by dotted line. The bottom monomer is modeled to be bound to its site on the origin, according to our findings about the residues involved in the DNA binding and structural alignments; the two winged-helix-turn-helix domains contact two adjacent major grooves of the DNA, the DNA binding residues are shown as red sticks. When one of the monomers is bound to DNA, the DNA binding residues of the second monomer are located very far away from the DNA, and they can not interact with the following binding site on the DNA. Therefore a head-to-head RctB dimer is incompatible with origin binding not only because of the binding site orientation (direct repeats), but also because of the molecule geometry that does not allow the second monomer to contact the same DNA molecule.

in solution, RctB adopts a head-to-head dimeric configuration mediated by interactions between residues in domain 2. However, this arrangement is not structurally compatible with binding to the head-to-tail array of 12-mer RctB binding sites in *oriCII* (Figure 6). Additionally, we found that dimerization-deficient RctB retained affinity to *oriCII*, but exhibited a greatly reduced ability to support replication.

A segment of RctB between domains 3 and 4 has also been proposed to mediate RctB dimerization and DNA binding (49). Our data do not support these results. Rather, our structural, mutational and native MS analyses provide strong evidence that DNA binding and dimerization are instead dependent upon other regions of RctB. However, we cannot exclude the involvement of this or other segments in weak contacts in the expected oligomer formed on origin DNA.

The oligomeric state of plasmid initiators, which like RctB, are dimers in solution, is thought to regulate their activity. It has been proposed that plasmid initiator dimers dissociate into monomers prior to binding their respective replication origins, whose arrangement of binding sites resembles that *oriCII* (45,47,48). It is tempting to propose that formation of the RctB – *oriCII* replication initiation complex may involve dissociation of the RctB dimer into monomers, which then seed formation of a new RctB oligomer in the complex on origin DNA; such a complex is also predicted to form on plasmid origins (47,48). How-

ever, our observation that disruption of the RctB dimer diminishes, rather than enhances, biological activity as the above model predicts and as observed with plasmid initiators (50,51) does not, at present, allow us to rule out more complicated protein–DNA complexes. Alternate models, such as, e.g. where an array of RctB dimers, not monomers, bind to origin DNA are possible; however, in such models, steric constraints make it unlikely that both members of the head-to-head dimer contact DNA. This observation has a precedent with the bacteriophage lambda cII protein, where two protein dimers, each with two DNA-binding domains, however, only one DNA-binding domain within each dimer binds to the major groove of the DNA molecule (52). In addition, match in symmetry between the array of binding sites at the origin and the proteins that will populate these sites requires clarification. The question of symmetry between protein configuration and DNA target sites has also been considered with the steroid hormone receptors (53–57). Typically, these proteins bind to a pair of target sites that exhibit head-to-head or head-to-tail configurations. With rare exception (58), the symmetry of the DNA target matches that of the protein (53–57) (i.e. a head-to-tail array of DNA sites is bound by proteins that are arranged in head-to-tail manner, etc), and we anticipate this to be true in the RctB–DNA complex. Comparisons between RctB - DNA complexes and those made by hormone receptors to their target sites are limited, though, because RctB binds to an array of six sites and the receptors are limited to two sites. Indeed, it is likely that a series of novel contacts, not seen in our head-to-head dimer structure, will further stabilize the RctB oligomer. A more precise definition of the RctB origin DNA complex must await future studies.

Although similarities between RctB and plasmid initiators were not recognized prior to our work, previous studies have commented on similarities between iteron plasmid and *oriCII*-based replication systems (21,59). Identification of the structural similarity between RctB and plasmid initiator proteins provides greater understanding of parallels between these systems. For example, the origins from chrII and plasmids share a number of elements, including directly repeated initiator binding sites. However, close examination reveals important differences, e.g. the 12 bp length of the RctB binding site is considerably shorter than the 19–22 bp length of iterons in plasmid origins. Structures of plasmid initiators bound to DNA provide insight on likely interactions between RctB and its binding sites on *oriCII-min* (Figure 7). Notably, these models suggest that only one of the three DNA binding domains on RctB can be accommodated on one face of the 12-mer sequence, and make sequence specific contacts in the major groove. Thus, it seems likely that some DNA binding sites within RctB recognize sequences other than the 12-mer, even in the context of the 6 × 12 array, since we have shown that all three domains contribute to interactions made by RctB with this probe. One possibility is that RctB also interacts with the adjacent major groove in the 10–11 bp spacer sequences between the 12-mers, so that the effective target size of RctB is actually closer to that found in plasmid origins. If so, then at least one of RctB's core DNA-binding domains is likely to lack sequence specificity in binding since the nucleotide sequence of the spacer segments is not conserved (25). Given spatial
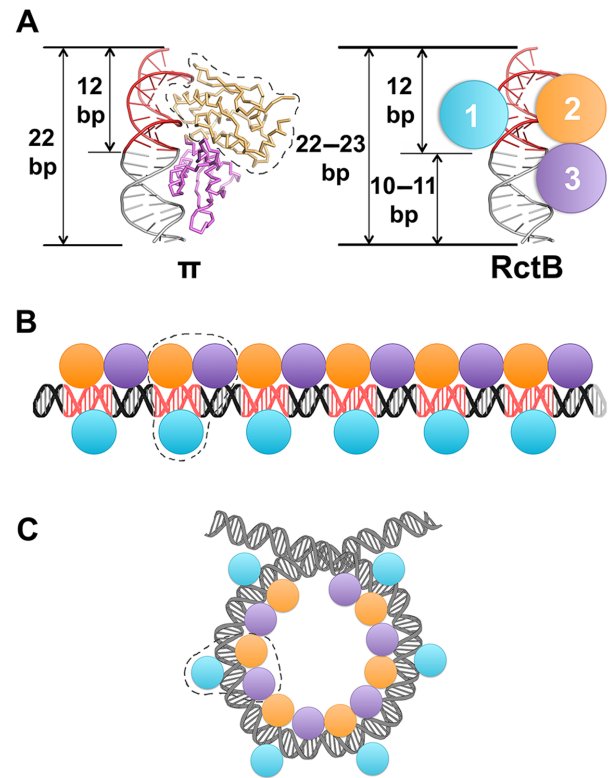

Figure 7. Model for the interaction of RctB with DNA. (**A**) The 12-mer binding site is of insufficient length to accommodate the three DNA binding domains of RctB. On the left is a ribbon representation of the experimental structure of the π initiator (domains of π are colored in light orange and purple) bound to its 22 bp iteron DNA target. Each domain of the π plasmid initiator binds to ∼10 bp of DNA. On the right is shown a schematic of the three DNA binding domains of RctB (domain1: cyan, domain 2: orange, domain 3: purple) and its 12-mer binding site (colored in red), drawn approximately to scale. We propose that domains of RctB (tentatively domains 1 and 2) will make contacts to positions in a 12-mer binding site on either side of the major groove, and that domain 3 (tentatively) will make contacts to positions in the 'spacer' sequence between the 12-mers. (**B**) Linear representation of a putative head-to-tail RctB oligomer formed on the array of 12-mer sites at *oriCII*min. One of the members of the RctB oligomer is outlined with a dashed line. (**C**) Schematic of the putative organization of the RctB oligomer on the 12-mer array in a DNA loop configuration to facilitate melting of the A-T-rich region of *oriCII*min. This model is constructed by analogy with that proposed for the plasmid initiator (47).

constraints, we postulate that domain 1 and domains 2–3 bind to opposite faces of the DNA target, where they are presumed to also interact with the major groove (Figure 7). This scheme is compatible with the expected head-to-tail arrangement of RctB on the direct repeats in *oriCII-min*, but does not rule out the possibility that there is a division of labor among the three RctB DNA binding domains, such that some specialize in contacts to a subset of its target sequences, as perhaps evidenced by mutational analysis of domains 1 and 3 (Figure 4, Supplementary Figures S13–S18). Future structural analyses of the nature of the oligomeric RctB initiator complex on *oriCII* DNA are required to address these issues.

The incompatibility of the head-to-head dimeric configuration of RctB with the directly repeated 12-mer binding

sites implies that a structural reorganization must take place prior to formation of the initiator complex on origin DNA. Indeed, this is known to be the case for the head-to-head dimeric plasmid initiators, which do not bind to the directly repeated binding sites within their cognate origins, unless a chaperone is provided to promote disruption of the dimer (60). RctB, however, appears to bind to *oriCII-min* without a chaperone (though it is impossible to exclude trace amounts in our preparations). Also, disruption of the RctB dimer into monomers does not promote DNA binding (Figure 5), as seen with the plasmid initiator RepA (60). This finding implies a potential role for the 12-mer RctB binding site itself in the necessary structural rearrangement. However, the precise mechanism that mediates rearrangement of the dimer remains to be clarified. It is possible that binding of the head-to-head RctB dimer to sites outside of the $6 \times$ 12mer array in *oriCII* is important for RctB-mediated regulation of initiation or of its own transcription.

Two basic scenarios for the evolution of multi-chromosomal bacteria have been put forward (61,62). A single large ancestral chromosome could have split into two chromosomes or alternatively, an ancestral strain could have acquired a plasmid, which subsequently acquired essential genes. In this context, our discovery that the structure of the core of RctB resembles plasmid initiator proteins lends strong support for the plasmid acquisition scheme. However, RctB and *oriCII* also contain features not found in plasmids. Notably RctB has two additional domains, one of which is critical for *oriCII* binding and replication. It seems plausible that these additional domains arose during the evolution of the *Vibrionaceae*, and allow for the more stringent regulatory requirements necessary for proper chromosome maintenance.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENT

REFERENCES

1. Mott,M.L. and Berger,J.M. (2007) DNA replication initiation: mechanisms and regulation in bacteria. *Nat. Rev. Microbiol.*, **5**, 343–354.
2. Wolanski,M., Donczew,R., Zawilak-Pawlik,A. and Zakrzewska-Czerwinska,J. (2014) oriC-encoded instructions for the initiation of bacterial chromosome replication. *Front. Microbiol.*, **5**, Article 735, 1–14.
3. Leonard,A.C. and Grimwade,J.E. (2011) Regulation of DnaA assembly and activity: taking directions from the genome. *Annu. Rev. Microbiol.*, **65**, 19–35.
4. Ozaki,S., Noguchi,Y., Hayashi,Y., Miyazaki,E. and Katayama,T. (2012) Differentiation of the DnaA-oriC subcomplex for DNA unwinding in a replication initiation complex. *J. Biol. Chem.*, **287**, 37458–37471.
5. Robinson,A. and van Oijen,A.M. (2013) Bacterial replication, transcription and translation: mechanistic insights from single-molecule biochemical studies. *Nat. Rev. Microbiol.*, **11**, 303–315.
6. O'Donnell,M., Langston,L. and Stillman,B. (2013) Principles and concepts of DNA replication in bacteria, archaea, and eukarya. *Cold Spring Harb. Perspect. Biol.*, **5**, a010108.
7. Leonard,A.C. and Grimwade,J.E. (2015) The orisome: structure and function. *Front. Microbiol.*, **6**, Article 545, 1–13.
8. Kaguni,J.M. (2011) Replication initiation at the Escherichia coli chromosomal origin. *Curr. Opin. Chem. Biol.*, **15**, 606–613.
9. Stepankiw,N., Kaidow,A., Boye,E. and Bates,D. (2009) The right half of the Escherichia coli replication origin is not essential for viability, but facilitates multi-forked replication. *Mol. Microbiol.*, **74**, 467–479.
10. Ozaki,S. and Katayama,T. (2009) DnaA structure, function, and dynamics in the initiation at the chromosomal origin. *Plasmid*, **62**, 71–82.
11. Flatten,I. and Skarstad,K. (2013) The Fis protein has a stimulating role in initiation of replication in Escherichia coli in vivo. *PloS One*, **8**, e83562.
12. Donczew,R., Zakrzewska-Czerwinska,J. and Zawilak-Pawlik,A. (2014) Beyond DnaA: the role of DNA topology and DNA methylation in bacterial replication initiation. *J. Mol. Biol.*, **426**, 2269–2282.
13. Duderstadt,K.E., Chuang,K. and Berger,J.M. (2011) DNA stretching by bacterial initiators promotes replication origin opening. *Nature*, **478**, 209–213.
14. Duderstadt,K.E., Mott,M.L., Crisona,N.J., Chuang,K., Yang,H. and Berger,J.M. (2010) Origin remodeling and opening in bacteria rely on distinct assembly states of the DnaA initiator. *J. Biol. Chem.*, **285**, 28229–28239.
15. Erzberger,J.P., Mott,M.L. and Berger,J.M. (2006) Structural basis for ATP-dependent DnaA assembly and replication-origin remodeling. *Nat. Struct. Mol. Biol.*, **13**, 676–683.
16. Remus,D. and Diffley,J.F. (2009) Eukaryotic DNA replication control: lock and load, then fire. *Curr. Opin. Cell Biol.*, **21**, 771–777.
17. Duncker,B.P., Chesnokov,I.N. and McConkey,B.J. (2009) The origin recognition complex protein family. *Genome Biol.*, **10**, Article 214, 1–8.

18. Li,H. and Stillman,B. (2012) The origin recognition complex: a biochemical and structural view. *Subcell. Biochem.*, **62**, 37–58.

19. Yamaichi,Y., Gerding,M.A., Davis,B.M. and Waldor,M.K. (2011) Regulatory cross-talk links Vibrio cholerae chromosome II replication and segregation. *PLoS Genet.*, **7**, e1002189.

20. Duigou,S., Knudsen,K.G., Skovgaard,O., Egan,E.S., Lobner-Olesen,A. and Waldor,M.K. (2006) Independent control of replication initiation of the two Vibrio cholerae chromosomes by DnaA and RctB. *J. Bacteriol.*, **188**, 6419–6424.

21. Egan,E.S. and Waldor,M.K. (2003) Distinct replication requirements for the two Vibrio cholerae chromosomes. *Cell*, **114**, 521–530.

22. Heidelberg,J.F., Eisen,J.A., Nelson,W.C., Clayton,R.A., Gwinn,M.L., Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L. *et al.* (2000) DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae. *Nature*, **406**, 477–483.

23. Koch,B., Ma,X. and Lobner-Olesen,A. (2012) rctB mutations that increase copy number of Vibrio cholerae oriCII in Escherichia coli. *Plasmid*, **68**, 159–169.

24. Jha,J.K., Demarre,G., Venkova-Canova,T. and Chattoraj,D.K. (2012) Replication regulation of Vibrio cholerae chromosome II involves initiator binding to the origin both as monomer and as dimer. *Nucleic Acids Res.*, **40**, 6026–6038.

25. Gerding,M.A., Chao,M.C., Davis,B.M. and Waldor,M.K. (2015) Molecular dissection of the essential features of the origin of replication of the second Vibrio cholerae chromosome. *mBio*, **6**, e00973.

26. Yamaichi,Y., Duigou,S., Shakhnovich,E.A. and Waldor,M.K. (2009) Targeting the replication initiator of the second Vibrio chromosome: towards generation of vibrionaceae-specific antimicrobial agents. *PLoS Pathog.*, **5**, e1000663.

27. Demarre,G. and Chattoraj,D.K. (2010) DNA adenine methylation is required to replicate both Vibrio cholerae chromosomes once per cell cycle. *PLoS Genet.*, **6**, e1000939.

28. Venkova-Canova,T. and Chattoraj,D.K. (2011) Transition from a plasmid to a chromosomal mode of replication entails additional regulators. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 6199–6204.

29. Venkova-Canova,T., Saha,A. and Chattoraj,D.K. (2012) A 29-mer site regulates transcription of the initiator gene as well as function of the replication origin of Vibrio cholerae chromosome II. *Plasmid*, **67**, 102–110.

30. Sorensen,H.P. and Mortensen,K.K. (2005) Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli. *Microb. Cell Factories*, **4**, 1.

31. Van Duyne,G.D., Standaert,R.F., Karplus,P.A., Schreiber,S.L. and Clardy,J. (1993) Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J. Mol. Biol.*, **229**, 105–124.

32. Pakotiprapha,D., Inuzuka,Y., Bowman,B.R., Moolenaar,G.F., Goosen,N., Jeruzalmi,D. and Verdine,G.L. (2008) Crystal structure of Bacillus stearothermophilus UvrA provides insight into ATP-modulated dimerization, UvrB interaction, and DNA binding. *Mol. Cell*, **29**, 122–133.

33. Yamaichi,Y., Fogel,M.A., McLeod,S.M., Hui,M.P. and Waldor,M.K. (2007) Distinct centromere-like parS sites on the two chromosomes of Vibrio spp. *J. Bacteriol.*, **189**, 5314–5324.

34. Cadene,M. and Chait,B.T. (2000) A robust, detergent-friendly method for mass spectrometric analysis of integral membrane proteins. *Anal. Chem.*, **72**, 5655–5658.

35. Fenyo,D., Wang,Q., DeGrasse,J.A., Padovan,J.C., Cadene,M. and Chait,B.T. (2007) MALDI sample preparation: the ultra thin layer method. *J. Visualized Exp.*, **3**, 192.

36. Otwinowski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol.*, **276**, 307–326.

37. Adams,P.D., Afonine,P.V., Bunkoczi,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.

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

39. Holm,L. and Rosenstrom,P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res.*, **38**, W545–W549.

40. Squizzato,S., Park,Y.M., Buso,N., Gur,T., Cowley,A., Li,W., Uludag,M., Pundir,S., Cham,J.A., McWilliam,H. *et al.* (2015) The EBI Search engine: providing search and retrieval functionality for biological data from EMBL-EBI. *Nucleic Acids Res.*, **43**, W585–W588.

41. Winn,M.D., Ballard,C.C., Cowtan,K.D., Dodson,E.J., Emsley,P., Evans,P.R., Keegan,R.M., Krissinel,E.B., Leslie,A.G., McCoy,A. *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 235–242.

42. Kleywegt,G.J. (1997) Validation of protein models from Calpha coordinates alone. *J. Mol. Biol.*, **273**, 371–376.

43. Kleywegt,G.J. (1996) Use of non-crystallographic symmetry in protein structure refinement. *Acta Crystallogr. D Biol. Crystallogr.*, **52**, 842–857.

44. Lo Conte,L., Chothia,C. and Janin,J. (1999) The atomic structure of protein-protein recognition sites. *J. Mol. Biol.*, **285**, 2177–2198.

45. Swan,M.K., Bastia,D. and Davies,C. (2006) Crystal structure of pi initiator protein-iteron complex of plasmid R6K: implications for initiation of plasmid DNA replication. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 18481–18486.

46. Nakamura,A., Wada,C. and Miki,K. (2007) Structural basis for regulation of bifunctional roles in replication initiator protein. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 18484–18489.

47. Komori,H., Matsunaga,F., Higuchi,Y., Ishiai,M., Wada,C. and Miki,K. (1999) Crystal structure of a prokaryotic replication initiator protein bound to DNA at 2.6 A resolution. *EMBO J.*, **18**, 4597–4607.

48. Giraldo,R., Fernandez-Tornero,C., Evans,P.R., Diaz-Orejas,R. and Romero,A. (2003) A conformational switch between transcriptional repression and replication initiation in the RepA dimerization domain. *Nat. Struct. Biol.*, **10**, 565–571.

49. Jha,J.K., Ghirlando,R. and Chattoraj,D.K. (2014) Initiator protein dimerization plays a key role in replication control of Vibrio cholerae chromosome 2. *Nucleic Acids Res.*, **42**, 10538–10549.

50. Ishiai,M., Wada,C., Kawasaki,Y. and Yura,T. (1992) Mini-F plasmid mutants able to replicate in Escherichia coli deficient in the DnaJ heat shock protein. *J. Bacteriol.*, **174**, 5597–5603.

51. Abhyankar,M.M., Zzaman,S. and Bastia,D. (2003) Reconstitution of R6K DNA replication in vitro using 22 purified proteins. *J. Biol. Chem.*, **278**, 45476–45484.

52. Jain,D., Kim,Y., Maxwell,K.L., Beasley,S., Zhang,R., Gussin,G.N., Edwards,A.M. and Darst,S.A. (2005) Crystal structure of bacteriophage lambda cII and its DNA complex. *Mol. Cell*, **19**, 259–269.

53. Khorasanizadeh,S. and Rastinejad,F. (2001) Nuclear-receptor interactions on DNA-response elements. *Trends Biochem. Sci.*, **26**, 384–390.

54. Rastinejad,F. (2001) Retinoid X receptor and its partners in the nuclear receptor family. *Curr. Opin. Struct. Biol.*, **11**, 33–38.

55. Rastinejad,F., Perlmann,T., Evans,R.M. and Sigler,P.B. (1995) Structural determinants of nuclear receptor assembly on DNA direct repeats. *Nature*, **375**, 203–211.

56. Rastinejad,F., Wagner,T., Zhao,Q. and Khorasanizadeh,S. (2000) Structure of the RXR-RAR DNA-binding complex on the retinoic acid response element DR1. *EMBO J.*, **19**, 1045–1054.

57. Shaffer,P.L. and Gewirth,D.T. (2002) Structural basis of VDR-DNA interactions on direct repeat response elements. *EMBO J.*, **21**, 2242–2252.

58. Shaffer,P.L., Jivan,A., Dollins,D.E., Claessens,F. and Gewirth,D.T. (2004) Structural basis of androgen receptor binding to selective androgen response elements. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 4758–4763.

59. Jha,J.K., Baek,J.H., Venkova-Canova,T. and Chattoraj,D.K. (2012) Chromosome dynamics in multichromosome bacteria. *Biochim. Biophys. Acta*, **1819**, 826–829.

60. Wickner,S., Hoskins,J. and McKenney,K. (1991) Monomerization of RepA dimers by heat shock proteins activates binding to DNA replication origin. *Proc. Natl. Acad. Sci. U.S.A.*, **88**, 7903–7907.

61. Egan,E.S., Fogel,M.A. and Waldor,M.K. (2005) Divided genomes: negotiating the cell cycle in prokaryotes with multiple chromosomes. *Mol. Microbiol.*, **56**, 1129–1138.

62. Val,M.E., Soler-Bistue,A., Bland,M.J. and Mazel,D. (2014) Management of multipartite genomes: the Vibrio cholerae model. *Curr. Opin. Microbiol.*, **22**, 120–126.

63. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.

64. Henikoff,S. and Henikoff,J.G. (1992) Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 10915–10919.

65. Gilston,B.A., Wang,S., Marcus,M.D., Canalizo-Hernandez,M.A., Swindell,E.P., Xue,Y., Mondragon,A. and O'Halloran,T.V. (2014) Structural and mechanistic basis of zinc regulation across the E. coli Zur regulon. *PLoS Biol.*, **12**, e1001987.

66. Zhang,A.P., Pigli,Y.Z. and Rice,P.A. (2010) Structure of the LexA-DNA complex and implications for SOS box measurement. *Nature*, **466**, 883–886.

67. Garcia-Castellanos,R., Mallorqui-Fernandez,G., Marrero,A., Potempa,J., Coll,M. and Gomis-Ruth,F.X. (2004) On the transcriptional regulation of methicillin resistance: MecI repressor in complex with its operator. *J. Biol. Chem.*, **279**, 17888–17896.

68. Wisedchaisri,G., Chou,C.J., Wu,M., Roach,C., Rice,A.E., Holmes,R.K., Beeson,C. and Hol,W.G. (2007) Crystal structures, metal activation, and DNA-binding properties of two-domain IdeR from Mycobacterium tuberculosis. *Biochemistry*, **46**, 436–447.

69. Dolan,K.T., Duguid,E.M. and He,C. (2011) Crystal structures of SlyA protein, a master virulence regulator of Salmonella, in free and DNA-bound states. *J. Biol. Chem.*, **286**, 22178–22185.

70. Yamane,T., Okamura,H., Ikeguchi,M., Nishimura,Y. and Kidera,A. (2008) Water-mediated interactions between DNA and PhoB DNA-binding/transactivation domain: NMR-restrained molecular dynamics in explicit water environment. *Proteins*, **71**, 1970–1983.