

OPEN

Structures of GapR reveal a central channel which could accommodate B-DNA

Michael J. Tarry¹, Christoph Harmel¹, James A. Taylor^{2,3}, Gregory T. Marczyński^{2*} & T. Martin Schmeing^{1*}

GapR is a nucleoid-associated protein required for the cell cycle of *Caulobacter crescentus*. We have determined new crystal structures of GapR to high resolution. As in a recently published structure, a GapR monomer folds into one long N-terminal α helix and two shorter α helices, and assembles into a tetrameric ring with a closed, positively charged, central channel. In contrast to the conclusions drawn from the published structures, we observe that the central channel of the tetramer presented here could freely accommodate B-DNA. Mutation of six conserved lysine residues lining the cavity and electrophoretic mobility gel shift experiments confirmed their role in DNA binding and the channel as the site of DNA binding. Although present in our crystals, DNA could not be observed in the electron density maps, suggesting that DNA binding is non-specific, which could be important for tetramer-ring translocation along the chromosome. In conjunction with previous GapR structures we propose a model for DNA binding and translocation that explains key published observations on GapR and its biological functions.

In prokaryotes, control of DNA replication and related processes for cell division are aided by a family of DNA-binding proteins known as nucleoid-associated proteins (NAPs)¹. GapR is an essential NAP in *Caulobacter crescentus* involved in DNA replication, chromosome segregation and cell division^{2–4}. It has been shown to bind DNA both *in vivo* and *in vitro*³. A preference for AT-rich DNA has been reported, though the strength of this preference differs between reports^{2,4}. GapR has dynamic chromosome-binding patterns and changing GapR distributions during the cell cycle. In *C. crescentus* “swarmer” cells before chromosome replication, fluorescently tagged GapR shows a concentration gradient across the whole chromosome from high GapR at the origin of replication region to low GapR towards the terminus region^{3,4}. Upon starting chromosome replication, a GapR-depleted region of the chromosome is created that expands while tracking the mitotic-like movements of the newly duplicated DNA³ and the moving replication forks⁴. Chromatin immunoprecipitation (ChIP) experiments showed that GapR preferentially binds to the 3' ends of actively transcribed genes, and that blocking transcription with antibiotics quickly (within minutes) dissipates GapR peaks and redistributes GapR towards the 5' ends of these genes^{3,5}. Arias-Cartin *et al.* proposed that replication forks encounter and dislodge tightly bound (presumed static) GapR molecules and that GapR molecules released into the cytoplasm rebind the chromosome at random places⁴.

GapR does not resemble other DNA binding proteins. It is a small protein of ~90 amino acids, with mainly α -helical secondary structure, and until very recently its mode of DNA binding was unknown. Last year, while this work was underway, a highly informative study which included crystal structures of GapR in the presence and absence of DNA was published⁵. In the reported structure of GapR determined in absence of DNA, GapR is seen to be a dimer, with each monomer folded into two extended α helices. In the structure of GapR determined in presence of DNA, GapR is seen to be a tetramer, with the C-terminal α helix of each monomer reorganized into two shorter helices, which allows GapR to encircle the DNA. The DNA in the published structure is 100% AT, over-twisted and slightly narrower than B-form DNA. In the accompanying experiments, GapR is proposed to associate with positively supercoiled chromosomal DNA⁵, which is often found ahead of the replication fork and RNA polymerase⁶. GapR was found to stimulate *in vitro* the topoisomerases gyrase and topo IV, and removing supercoils is required for DNA replication to proceed *in vivo*⁵.

¹Department of Biochemistry, McGill University, 3649 Promenade Sir William Osler, Montreal, QC, H3G 0B1, Canada.

²Department of Microbiology and Immunology, McGill University, 3775 University Street, Montreal, QC, H3A 2B4, Canada. ³Present address: Department of Cellular and Molecular Medicine, University of Ottawa, 451 Smyth 15 Road, Ottawa, ON, K1H 8L1, Canada. *email: gregory.marczynski@mcgill.ca; martin.schmeing@mcgill.ca

Name	Sequence (5' to 3')	Source
GapR_Fwd	AAAACCATGGCCGACGACGCCATTCC	Integrated DNA Technologies
GapR_Rev	AAAAGAATTCAACGCTCGACCATACGTCTC	Integrated DNA Technologies
GapR_NtDel_Fwd	ACCTGTATTTTCAGGGCCTGAACTCGACCGC	BioCorp DNA
GapR_NtDel_Rev	GCGGTCGAGTTCAGGCCCTGAAAATACAGGT	BioCorp DNA
19mer_Fwd	GTTAAGCAACCGTTAACGG	BioCorp DNA
19mer_Rev	CCGTAAACGGTTGCTTAAC	BioCorp DNA
10mer	CCGTAAACGG	Integrated DNA Technologies
9mer_Fwd	CCGTAAACG	Integrated DNA Technologies
9mer_Rev	CGTAAACGG	Integrated DNA Technologies
8mer	CGTAAACG	Integrated DNA Technologies

Table 1. DNA oligonucleotides used in this study.

In this paper we present three crystal structures of GapR from *C. crescentus*, crystallised in the presence of double-stranded DNA with a sequence from the origin of chromosome replication. In all three structures, GapR adopts the same overall tetramer ring form as seen by Guo *et al.*, albeit with small but important increases in the diameter of the central channel. DNA could not be visualized in any of our three structures, but DNA could be detected in gel electrophoresis, indicating that DNA is present in the crystals. Furthermore, these crystals do not form without DNA in the crystallization protocol. Notably, double stranded B-form DNA can be docked into the channel of our structure without any clashes. We identified residues that our structural analyses suggested should be involved in DNA binding and showed that mutation of these residues abolishes DNA binding while retaining tetramer oligomerization. We propose that our crystals capture a physiologically-relevant GapR structure distinct from and complementary to the physiologically-relevant structures captured by Guo *et al.* Finally, we combine these structural data and present an updated model for DNA binding and translocation which more fully accounts for the observed dynamics of GapR binding.

Results

Structure determination of GapR. At the time of undertaking these experiments, no structural information on GapR was known. We set out to crystallize GapR alone and in the presence of DNA to structurally characterize the protein and protein-DNA complex. Initial attempts to crystallize GapR in the absence of added DNA yielded abundant microcrystals. Despite extensive efforts, conditions producing these microcrystals could not be optimized to yield diffraction quality crystals. We next tried crystallizing GapR in the presence of double stranded DNA (dsDNA). For this, we used both full length GapR and GapR $_{\Delta 1-11}$ (a construct in which we removed the first eleven residues, that were predicted to be disordered) and a 19-bp dsDNA oligomer with a sequence from the *C. crescentus* origin of replication (Table 1), which GapR binds with a nanomolar dissociation constant³. GapR-DNA complexes were subjected to sparse matrix crystallization, and multiple crystallization conditions were identified. A single condition was optimized to yield three sets of morphologically distinct diffraction quality crystals, which were used to determine three independent structures of GapR (Table 2).

The best diffracting crystals were of full-length GapR plus DNA. A diffraction dataset was collected which included reflections to 1.85 Å resolution and was of the space group I4₁32. At the time, there were no available structures of GapR, and the protein shared no substantial identity with any other DNA-binding protein structures. Attempts to use DNA as a search model in molecular replacement were unsuccessful. GapR was categorized as containing the domain of unknown function 2312^{2,4} which had previously been subjected to *ab initio* modelling and theoretical screening for nucleic acid binding potential⁷. We submitted the sequence of GapR to the R α betta server for protein structure prediction⁸, and the top five models produced were used as search models in molecular replacement. One of these models gave a possible solution, with residues 35–65 well placed in the electron density maps generated by Phaser⁹. A subsequent search model using only these residues gave a definite solution and the remaining residues other than the first 11 could be manually built into the electron density to give a final model at 1.85 Å with an R_{free} of 0.223 (Fig. 1).

Overall structure of GapR. In this structure, there is a single GapR monomer in the asymmetric unit. It has an extended N-terminal α helix followed by two shorter α helices (Fig. 1a). The three helices are relatively open and not bundled together, suggesting that this conformation would not be stabilized in the absence of a binding partner. Indeed, applying crystallographic symmetry assembles the monomers into repeating tetramer units (Fig. 1c) which have a large central channel. The tetramer is very similar, but not identical to the published tetramer⁵.

The same crystallization condition also produced crystals with a second morphology, which were of the space group P22₁. The diffraction from these crystals displayed severe anisotropy and had resolution limits between 2.0 and 2.7 Å. Following ellipsoidal truncation and anisotropic correction using the Diffraction Anisotropy Server^{10,11}, the structure was solved using our first GapR structure as a molecular replacement search model and refined to a final R_{free} of 0.279. In this crystal, GapR is a dimer in the asymmetric unit, with the tetramer again being formed by crystallographic symmetry. The tetramer could be overlaid onto the I4₁32 tetramer with a root-mean-square deviation (rmsd) of 1.5 Å (Fig. 1d).

	6OZX	6OZY	6OZZ		
Data collection					
Wavelength (Å)	0.979	0.979	0.979		
Resolution range (Å)	82.36–1.85 (1.89–1.85)	50.0–2.00 (2.05–2.00)	50.0–3.30 (3.38–3.30)		
Space group	I4 ₁ 32	P222 ₁	P4 ₃ 22		
Unit cell <i>a b c</i> (Å) $\alpha \beta \gamma$ (°)	116.5 116.5 116.5 90 90 90	39.7 59.3 93.9 90 90 90	106.3 106.3 139.1 90 90 90		
		Pre-correction	Post-correction	Pre-correction	Post-correction
Total reflections	821532 (38807)	201701 (13944)	130424 (589)	323588 (22821)	242400 (1083)
Unique reflections	11872 (709)	15659 (1134)	10203 (50)	12618 (901)	9555 (49)
Multiplicity	69.2 (54.7)	12.9 (12.2)	8.5 (0.5)	25.6 (25.3)	19.3 (1.2)
Completeness (%)	99.9 (98.2)	99.9 (99.3)	66.5 (4.5)	99.9 (99.8)	76.1 (5.4)
Mean <i>I</i> / σ (<i>I</i>)	30.2 (2.6)	10.9 (0.5)	16.5 (3.1)	13.8 (0.5)	18.0 (3.1)
<i>R</i> _{merge}	0.113 (2.45)	0.137 (4.74)	0.095 (0.808)	0.156 (7.88)	0.116 (1.25)
<i>CC</i> _{1/2}	0.999 (0.532)	0.999 (0.349)	0.998 (0.809)	0.998 (0.367)	0.999 (0.880)
Refinement					
Resolution range (Å)	47.55–1.85 (1.917–1.85)	46.97–2.01 (2.09–2.01)	49.7–3.30 (3.42–3.30)		
<i>R</i> _{work}	0.193 (0.255)	0.2373 (0.356)	0.2852 (0.409)		
<i>R</i> _{free}	0.223 (0.282)	0.279 (0.357)	0.322 (0.461)		
Non-hydrogen atoms	673	1342	3661		
Macromolecules	627	1251	3661		
ligands	—	1	—		
water	46	90	0		
Protein residues	78	156	455		
RMS bonds (Å)	0.011	0.002	0.003		
RMS angles (°)	1.21	0.42	0.58		
Ramachandran favored (%)	100	100	100		
Clashscore	0.00	0.00	2.39		
Average <i>B</i> -factor	41.6	49.6	146		
macromolecules	41.2	49.7	146		
ligands	—	45.7	—		
solvent	46.6	48.0	—		

Table 2. Data collection and refinement statistics. Statistics for the highest-resolution shell are shown in parentheses.

We were also able to solve the structure of GapR _{Δ 1-11} from crystals of space group P4₃22. Diffraction from these crystals also showed high anisotropy, with resolution limits between 3.3 Å and 3.8 Å. Following ellipsoidal truncation and anisotropic correction using the Diffraction Anisotropy Server¹⁰ the structure was refined to an *R*_{free} of 0.322. There are six monomers in the asymmetric unit, arranged as a tetramer and a dimer. The tetramer could be overlaid onto the I4₁32 tetramer with an rmsd of 1.9 Å (Fig. 1d). Crystallographic symmetry again showed that the dimer of the asymmetric unit paired with symmetry-related molecules to form tetramers. The main difference between GapR in the three crystal forms arises from slight changes in the orientation of helix three, which results in minor variations in the size of the central channel (Fig. 1d).

The central channel seen in the GapR tetramers is an obvious binding site for DNA. Interestingly, the channel is quite short, at around ~30 Å long. This is much shorter than the length of the 19-bp oligonucleotide, which would be ~65 Å long. In each crystal, GapR forms a complete and tight lattice (Fig. 1e–g), suggesting that although the presence of the DNA was required for crystal growth, it did not participate in crystal packing, and that there is not a 1:1 stoichiometry between 19 bp DNA and GapR tetramer. Furthermore, in the I4₁32 and P4₃22 structures, a 19 bp DNA oligo would have to adopt substantial bends to thread through the central channel of adjacent tetramers, because channels formed by adjacent GapR tetramers do not line up.

An electrostatic surface representation of the GapR tetramer showed that the lining of the channel is highly positively charged (Fig. 2a). We manually placed model B-DNA (PDB ID: 1BNA)¹² into the channel and found that it could be fit without clashing with the GapR tetramer (Fig. 2b). Closer analysis of the channel identified six highly conserved lysine residues per monomer (Fig. 2c,d) that seemed likely to contribute to DNA binding by tetrameric GapR.

GapR is a tetramer in solution that binds DNA through conserved lysine residues. We used size exclusion chromatogram – multiangle light scattering (SEC-MALS) to investigate the oligomeric state of GapR in solution (Fig. 3a). SEC-MALS of GapR in the absence of DNA provided an estimate of its solution molecular

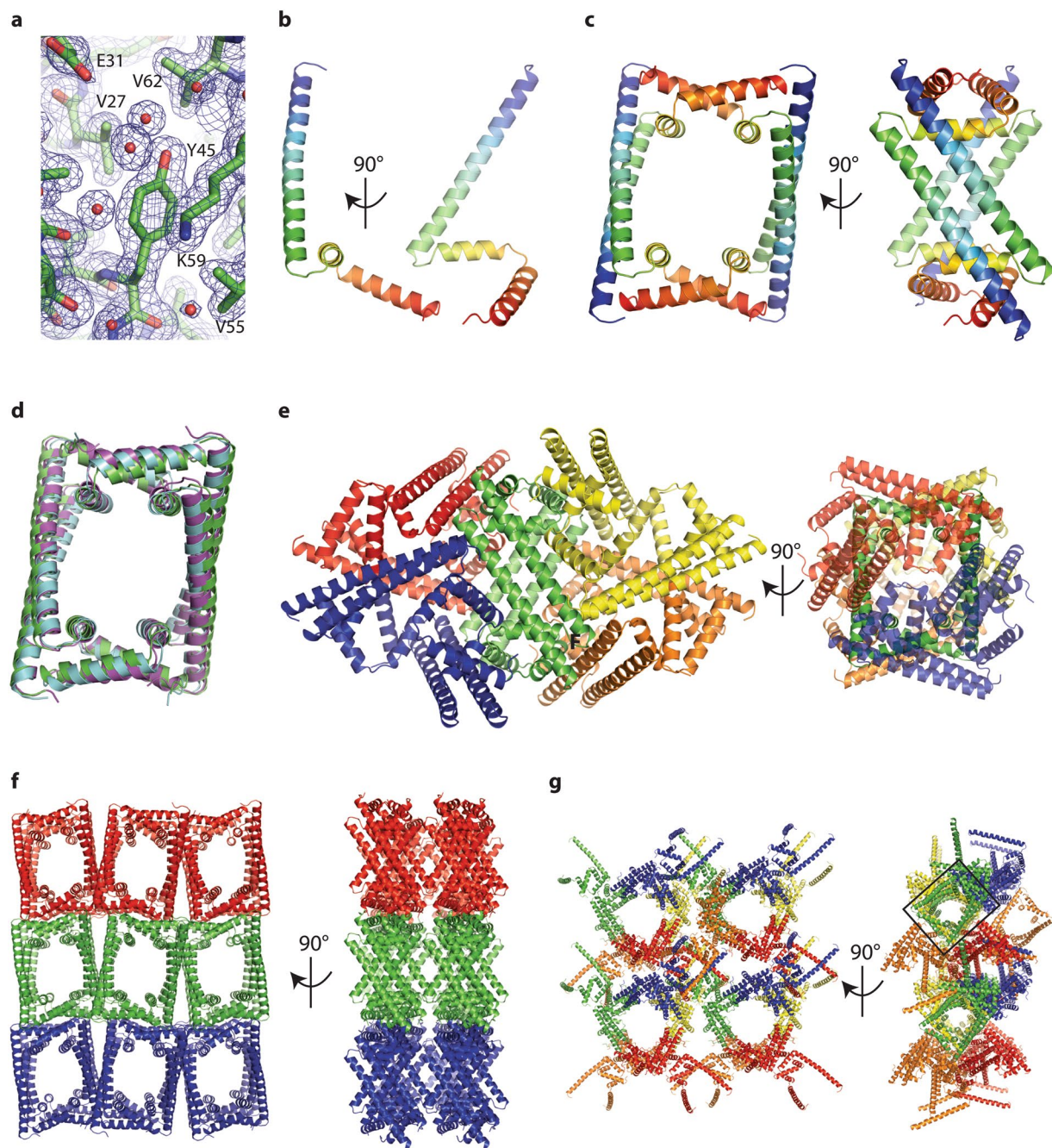


Figure 1. Crystal structures of GapR and packing in the crystal lattice. **(a)** GapR modelled in to a $2F_o - F_c$ electron density map, shown contoured at 1σ . **(b)** The GapR monomer from the asymmetric unit of the crystals of space group $I4_132$, colored blue through to red from the N to C termini. **(c)** Four symmetry related monomers assemble into the physiologically relevant tetramer. The left side of panel C shows the central channel. **(d)** Overlay of the biological tetramer of GapR from crystals of space group $I4_132$ (green), $P222_1$ (cyan) and $P4_322$ (purple) used in this study. **(e)** Packing in crystals of space group $I4_132$, with different tetramers shown in red, blue, green, yellow and orange. The right side of panel e orients the channel of the green tetramer toward to viewer. **(f)** Packing in crystals of space group $P222_1$. The left side of panel f orients the channels of all tetramers toward the viewer. **(g)** Packing in crystals of space group $P4_322$. The channel formed by a tetramer is highlighted by a black rectangle in the right side of the panel.

weight of ~ 47.2 kDa, most consistent with a non-compact GapR tetramer. Analytical size exclusion chromatogram (SEC) of GapR in the presence and absence of the 19 bp DNA used for crystallization demonstrated that DNA binding does not alter the oligomeric state of GapR (Fig. 3b). Thus the analytical SEC and SEC-MALS show that GapR can exist as a tetramer in solution, even in the absence of DNA.

We next tested if the conserved lysines (Fig. 2) did contribute to DNA binding. We mutated all six of these lysines to aspartic acids, and purified the resulting protein (GapR_{6KtoE}). GapR_{6KtoE} showed low solubility, which

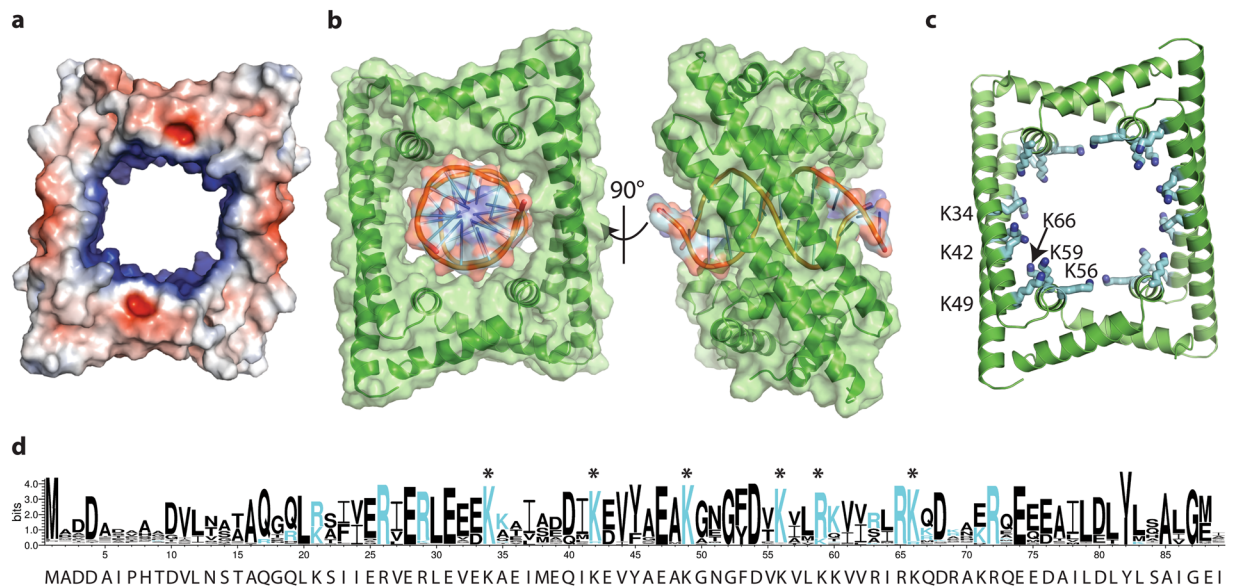


Figure 2. Identification and analysis of a DNA binding channel. (a) Electrostatic surface representation of GapR with electropositive and electronegative regions colored blue and red respectively. (b) The GapR tetramer (green cartoon with transparent surface representation overlaid) channel of the I4,32 crystal form can accommodate B-DNA (shown as sticks overlaid with transparent surface representation). B-DNA (PDBID: 1BNA)¹² was manually placed into the GapR channel using PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC) and then subject to a round of energy-minimization in Phenix²⁴. (c) Six highly conserved channel-lining lysine residues were identified and targeted for mutation. (d) Weblogo³²⁶ showing sequence conservation in the 250 proteins in the NCBI protein database which share highest percentage identity with GapR from *C. crescentus* (shown beneath). The lysine residues drawn as sticks in panel C are marked with an asterisk.

has also been reported for wild type GapR in the absence of DNA⁴, and it could not be concentrated above 0.4 mg/ml. To ensure the mutations did not disrupt tetramerization, we compared the oligomeric state of the mutant and wild type proteins by migration in size exclusion chromatography. GapR_{6KtoE} migrated slightly slower than wild type GapR (Fig. 3c), but comparison with proteins of known molecular weights shows this small difference is not indicative of a change in oligomeric state. To assess DNA binding of wild type and mutant proteins, we performed an electrophoretic mobility shift assay (EMSA) (Fig. 3d). DNA binding was observed for wild type GapR, but not for GapR_{6KtoE}. Like wild type protein, GapR_{6KtoE} migrated through the native gel as a single band (Fig. 3d) but slightly faster than wild type GapR. The small differences in migration is likely caused by the difference in charge introduced by the six lysine to aspartic acid mutations per monomer (24 per tetramer). The inability of GapR_{6KtoE} to bind DNA supports the conclusion that GapR binds DNA through its central pore⁵.

The GapR channel can accommodate B-form DNA. Modelling of DNA into our GapR structures revealed that our original 19-bp oligonucleotide was not ideal for structural studies, as GapR would be able to bind at various locations along the 19-mer. To determine the optimal length of DNA required for GapR binding we performed EMSA experiments. We reasoned that when GapR is in excess over DNA, multiple copies of the GapR tetramer should bind to DNA if its length is sufficient. By using increasing lengths of oligonucleotides in EMSA experiments it should be possible to measure the transition that accommodates extra GapR binding as the threshold length is crossed. Accordingly, we performed EMSA with excess GapR and 14, 16 and 18-bp DNA oligonucleotides. GapR bound all DNA oligomers, but a second shift was observed only with the 18-bp DNA (Fig. 4a). Thus, a segment of 18 base pairs of DNA is just sufficient to bind two GapR tetramers, suggesting that a single GapR binds to ~9 base pairs of DNA.

We next repeated our crystallization protocols with GapR in the presence of DNA oligos of 8, 9 and 10 base pairs. Crystals with identical morphology to those grown with 19-bp DNA were obtained for the 8-bp DNA, but despite the inclusion of excess DNA oligonucleotide during cryoprotection, no density for DNA could be seen in the electron density maps of the resulting structure. A native gel of washed, dissolved crystals demonstrated that DNA was in fact present (Fig. 4b). Thus, the absence of DNA in the crystal structures suggests that the DNA binds in a continuum of positions within the channel. Indeed, the lysine residues are fairly evenly spread out in the channel, and a mechanism to keep the DNA at any one particular position within the channel is not evident.

Discussion

During the course of this work, a structure of GapR in complex with DNA was published⁵. GapR forms essentially the same tetramer seen in our structures, with DNA bound in the central channel, as we anticipated. The six lysines that we identified, plus an additional arginine, all make contacts with the DNA backbone. In marked contrast to our study, which used DNA with an AT content ranging from 40 to 53%, the DNA used by Guo *et al.*

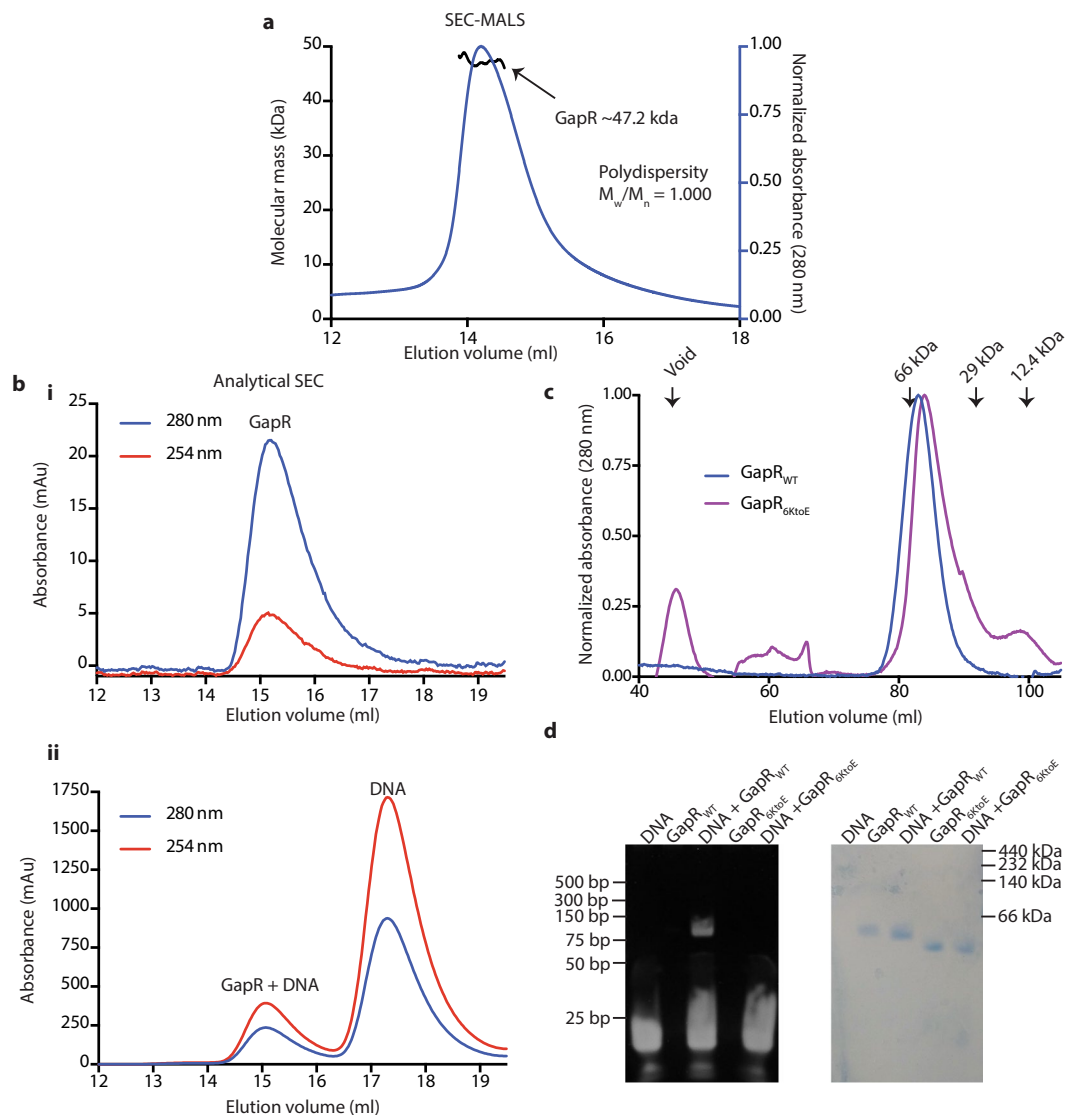


Figure 3. Mutation of channel lysines abolishes DNA binding but does not disrupt tetramer formation. **(a)** SEC-MALS of apo GapR. Absorbance at 280 nm (blue) and the calculated scattering mass across the peak (black) are shown. The expected mass of tetrameric GapR is 40.9 kDa. **(b)** Analytical size exclusion chromatography of GapR (6.25 μ M tetramer, $\epsilon_{280} = 2980 \text{ M}^{-1} \text{ cm}^{-1}$) in the (i) absence and (ii) presence of 19 bp DNA (20 μ M, $\epsilon_{260} = 309,776 \text{ M}^{-1} \text{ cm}^{-1}$) using a Superdex 200 Increase 10 300 GL column. **(c)** Overlay of size exclusion chromatograms for GapR (blue) and GapR_{6KtoE} (purple) GapR. Samples of 6.25 μ M (if calculated using the tetrameric molecular weights) protein was applied to a HiLoad 16/600 Superdex 200 pg column. Arrows indicate the void volume of the column and elution positions of proteins standards of known molecular weight (66 kDa – bovine serum albumin, 29 kDa – carbonic anhydrase, 12.4 kDa – cytochrome *c*). **(d)** EMSA experiment demonstrating that DNA binding by GapR is mediated by conserved lysine residues lining the tetrameric channel. Wild type GapR or GapR_{6KtoE} (0.425 μ M tetramer) was incubated in the presence or absence of 19 bp oligonucleotide (1.7 μ M) and used for non-denaturing gel electrophoresis. The gel was stained with SYBR Gold to detect DNA (left side of panel), then with InstantBlue to detect protein (right side of panel). Note that GapR has an overall negative charge. Uncropped gels are presented in Supplementary Figure 1.

was 100% AT. This DNA adopted an over-twisted conformation, with a wider major groove and narrower minor groove than seen in B-DNA. The length of this DNA oligomer is 11 base pairs and it forms head-to-tail contacts with the adjacent symmetry-related DNA molecule to form extended noncovalent DNA polymers. The length and AT content of the DNA were likely both important for crystallization and the ordering that allowed the DNA to be visualized.

Over-twisting of DNA slightly narrows its diameter relative to B-DNA and the DNA-bound GapR tetramer⁵ is constricted relative to our structure, with a narrower DNA channel (Fig. 4c). Guo *et al.* report being unable to fit B-DNA into their GapR tetramer without clashes⁵, although these appear to be minor clashes with mobile side chains. In contrast, our analysis of our structure indicates the tetramer we observe could comfortably accommodate B-DNA (Fig. 2b). In addition to crystallizing tetrameric GapR bound to DNA, Guo *et al.* were also able to

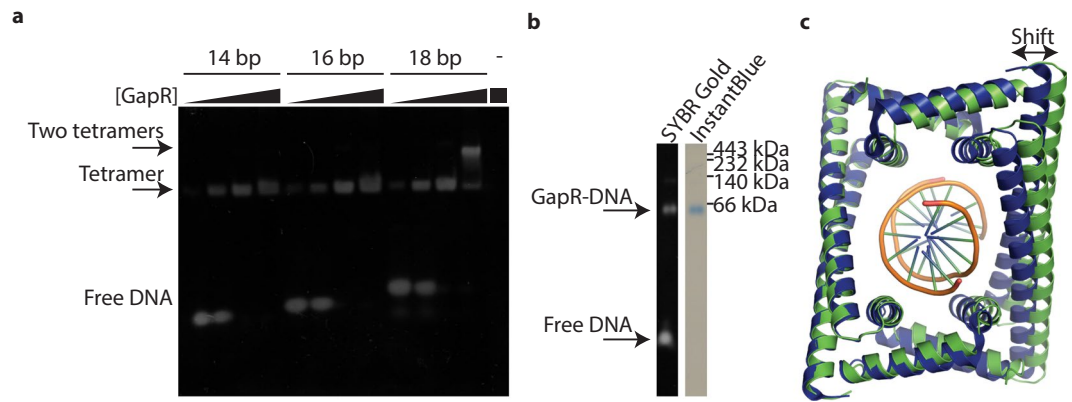


Figure 4. GapR DNA-binding analysis. (a) EMSA analysis of GapR binding to DNA oligonucleotides of different lengths. $0.4 \mu\text{M}$ DNA oligonucleotide (Table 1) was incubated with GapR (of concentrations 0.4, 1.2, 3.6 and $10.8 \mu\text{M}$, if calculated using monomeric molecular weight or 0.1, 0.3, 0.9 and $2.7 \mu\text{M}$ if tetramer) prior to non-denaturing gel electrophoresis. GapR ($2.7 \mu\text{M}$ tetramer) without oligonucleotide was analysed in the rightmost lane. The uncropped gel is presented in Supplementary Figure 1. (b) Crystals of GapR contain DNA. Crystals of space group $I4_132$ crystals, grown from a sample of GapR with 8 bp oligonucleotide were subject to gel electrophoresis and stained for the presence of DNA (left side of panel) and protein (right side of panel). It is likely the free DNA observed here results from disassociation of the oligo after harvesting & washing the crystal and when loading onto the gel, when it is no longer in excess. The uncropped gel is presented in Supplementary Figure 1. (c) Overlay of the crystal structure determined from crystals of space group $I4_132$ (green) with the GapR-DNA structure of Guo *et al.*⁵ (blue with DNA shown in orange, PDB ID: 6CG8). The Guo *et al.* structure features a marginally narrower channel.

crystallize GapR from the closely related *Bosea* sp. Root381, in the presence and absence of DNA. In the presence of DNA, which was visible in electron density maps but could not be reliably modelled, GapR was tetrameric. In the absence of DNA, GapR crystallized as a dimer, with α helices 2 and 3 rearranging to form a single extended helix⁵. On the basis of these structures, the authors proposed a model for DNA binding whereby dimeric GapR uses its two extended helices to track along DNA, monitoring the size of the major and minor grooves. Upon encountering over-twisted DNA, the extended α -helix re-organizes into two shorter helices and forms the stable tetramer with a co-translocating dimer. In our studies with GapR from *C. crescentus*, we saw no evidence of the dimer form in the presence or absence of DNA (Fig. 3). However, we did note some subtle difference in position of helix three in each of our crystal forms (Fig. 1d), suggesting that plasticity of this region could possibly be important for formation and size of the central channel.

That the current structures show GapR tetramers can accommodate B-DNA in the channel, and that we observe tetrameric GapR in absence of DNA, has important implications for the model of DNA binding (Fig. 5). GapR could first associate with B-DNA as a dimer, with the region of helices 2 and 3 extended into a single long helix, like in the Guo model, or as a tetramer, which would open to allow DNA binding by splaying of two interacting copies of helices 2 and 3. In both proposed pathways, the tetramer would close around DNA by rearrangement into the α helices 2 and 3 into the conformation observed in our structures, with the wider GapR tetramer bound to B-DNA. GapR with this wider channel would scan along B-DNA until it encounters over-twisted, AT-rich DNA. The over-twisted DNA would induce the transition to the slightly constricted tetramer, allowing it to bind the over-twisted, AT-rich DNA with higher affinity⁵, thus localizing GapR to this site.

A key point of the updated model is the scanning of DNA by the GapR tetramer. GapR encircling and scanning B-DNA could be more processive and efficient than scanning as a dimer, as it could proceed by one-dimensional diffusion rather than repeated association and disassociation. The model also allows GapR tetramers to form on B-form DNA in the bacterial genome, and not just at over-twisted DNA, which obviates the less probable event of two GapR dimers being coincidentally attracted to the same over-twisted DNA. It is also not clear how a GapR dimer would be preferentially attracted to over-twisted DNA. Scanning of B-DNA by tetrameric GapR remains to be directly shown, but passive movement along double stranded DNA has ample precedence in proteins which encircle DNA such as sliding clamps, Muts-ATP, Ku70/80 and proliferating cell nuclear antigen^{13–18}. It is also possible that interactions with other proteins in the cell could affect the size of the channel and the sliding rate along the DNA.

The GapR binding and scanning model also help explain previous observations that seemed significant for biological function yet lacked a mechanistic basis. Having GapR tetramers sliding across a predominantly CG-rich genome and occasionally encountering AT-rich patches where presumably they pause with a tighter conformation and a narrower channel could help reconcile the differential reporting of a strong² or weak preference for AT-rich DNA *in vivo*^{3–5}. It is possible that the FLAG-tagged GapR used by Ricci *et al.* somehow favored the tighter conformation, accentuating the preference for AT-rich DNA.

More importantly, the model has implications before and during chromosome replication. The GapR gradient in pre-replication *C. crescentus* “swarmer” cells is easier to explain if GapR tetramers can bind relaxed DNA. Also, *C. crescentus* GapR is required at the start of chromosome replication and during the first stage of

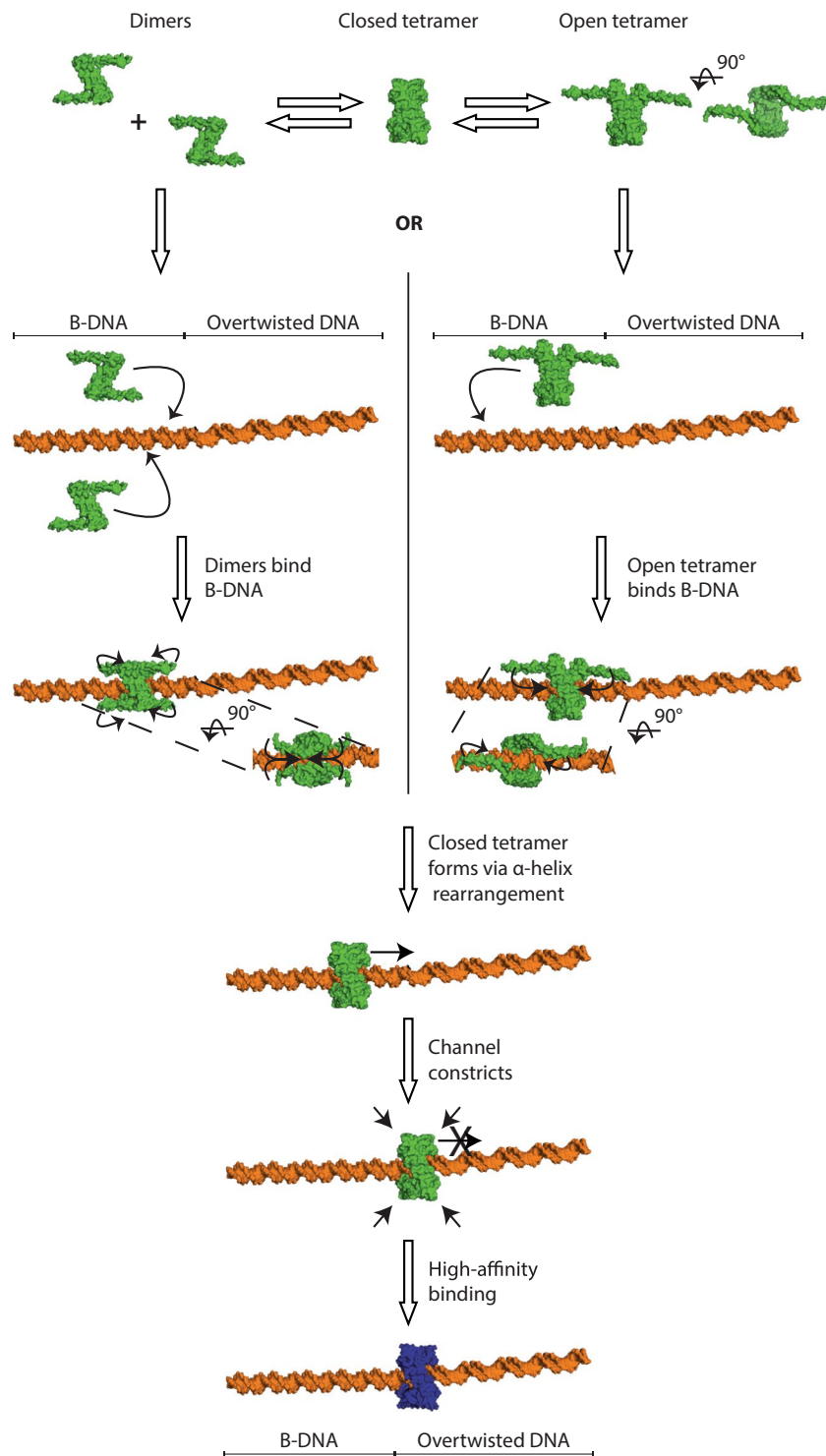


Figure 5. Proposed models for GapR binding to DNA. GapR first associates with DNA via one of two possible pathways. In the pathway shown to the left, GapR dimers⁵ (PDB ID: 6CFY) associate with DNA weakly until encountering a partner. The two dimers then assemble into a GapR tetramer by rearrangements in the C-terminal α helices and encirclement of the DNA. In the pathway shown on the right, a pre-formed GapR tetramer encircles DNA through rearrangement of two C termini on one side of the tetramer. If the tetramer has formed around B-form DNA, it adopts a more open conformation and can freely diffuse along the DNA. Upon encountering over-twisted DNA, GapR constricts slightly (as in PDB ID: 6CG8⁵), narrowing the channel around the DNA and leading to a higher-affinity complex.

chromosome separation that coincides with early DNA duplication³. ChIP experiments in synchronized cells showed distinct high and low patterns of GapR binding at each of the “left”, “middle” and “right” positions within a short (~700 bp) span of the *C. crescentus* chromosome origin of replication. Such localized and DNA-specific binding patterns probably result from GapR encounters with DNA sequence-specific replication proteins as well as RNA and DNA polymerases. GapR binding and sliding would be important for reaching targets inside the origin of replication and similarly important to minimize interference with essential replication proteins on a crowded DNA platform¹⁹.

Materials and Methods

Oligonucleotides. Single stranded DNA oligonucleotides used in this study are listed in Table 1. To anneal them into double stranded DNA, oligonucleotides were resuspended in annealing buffer (10 mM Tris pH 7.5, 50 mM NaCl, 1 mM EDTA) at 100 μ M. Complementary strands were mixed at a 1:1 ratio, incubated at 95 °C for five minutes and then slow-cooled to room temperature.

Cloning, expression and purification of GapR constructs. Full length GapR was amplified from plasmid pJT160³ using primers GapR_Fwd and GapR_Rev (Table 1). The PCR product was digested with *Nco*I and *Eco*RI (New England Biosciences) and ligated into a similarly digested pJ411-derived vector containing an N-terminal TEV cleavable octa-histidine tag. GapR $_{\Delta 1-11}$ was generated by site-directed mutagenesis by deleting the first eleven codons of *gapR* with primers GapRNtDelF and GapRNtDelR. GapR $_{6KtoE}$ was synthesised by ATUM (Newark, California) and cloned into pUC57 between the *Nco*I and *Not*I restriction sites. This region was then subcloned into the *Nco*I/*Not*I digested pJ411-derived vector described above.

Expression of GapR, GapR $_{\Delta 1-11}$ and GapR $_{6KtoE}$ was induced in *Escherichia coli* BL21(DE3) cells grown at 30 °C in LB media to an OD₆₀₀ of ~0.6, before inducing with 1 mM isopropyl β -D-1-thiogalactopyranoside. Cell cultures were grown for five more hours at 30° before harvesting. Cell pellets were stored at –80 °C until required.

Pellets were resuspended in IMAC binding buffer (20 mM HEPES pH 7.4, 200 mM NaCl, 40 mM imidazole, 10% glycerol, 2 mM β -mercaptoethanol (β ME)) supplemented with 1 mM PMSF and several crystals of DNase I. Cells were lysed by sonication (5 minutes total pulse at 50% amplitude, 10 s on, 20 s off) and the lysate cleared by centrifugation at 18 000 g. Cleared lysates were applied to a 5 ml HisTrap IMAC FF column (GE Healthcare) and washed with heparin elution buffer (20 mM HEPES pH 7.4, 2 M NaCl, 10% glycerol, 2 mM β ME). After re-equilibration with IMAC binding buffer, bound protein was eluted with IMAC elution buffer (as binding buffer but 800 mM imidazole). Fractions containing GapR were pooled and applied to a 5 ml HiTrap Heparin HP column (GE Healthcare) and bound protein eluted on a 100 ml gradient to 60% heparin elution buffer. Protein-containing fractions (GapR $_{6KtoE}$ failed to bind the column and protein was recovered from the flow through) were pooled and the His tag cleaved during overnight dialysis against 1 L of IMAC binding buffer in the presence of TEV protease²⁰ at 4 °C. Samples were reapplied to the HisTrap column and the flow-through collected. For formation of GapR-DNA and GapR $_{\Delta 1-11}$ -DNA complexes for crystallization, excess annealed oligonucleotides (Table 1) were added and the resultant complex purified on a HiLoad 16/600 Superdex 200 pg column (GE Healthcare) in SEC buffer (20 mM HEPES pH 7.4, 200 mM NaCl, 0.2 mM TCEP) to remove unbound DNA.

Crystallography. GapR-DNA and GapR $_{\Delta 1-11}$ -DNA complexes were concentrated to 3 mg/ml, as determined by the Bio-Rad Protein Assay, (Bio-Rad Laboratories) using 3 kDa MWCO Amicon Ultra centrifugation devices (EMD Millipore) and subjected to sparse-matrix crystallization against commercially available screens (Qiagen) using sitting drop vapor diffusion at room temperature. Crystals were obtained in multiple conditions, with final conditions optimized in 24-well sitting drop plates with 500 μ l reservoir and 2 μ l protein sample plus 2 μ l reservoir solution in the drop. The final crystallization conditions for the I4₁32 and P222₁ crystal forms of GapR-DNA were 0.16 M ammonium sulfate, 12% PEG 3350 with 1% 1,2-butanediol or 10 mM cadmium chloride respectively. Crystals with morphologies indicative of both crystal forms were seen in both conditions. GapR $_{\Delta 1-11}$ -DNA crystallized in space group P4₃22 in 4% PEG 3350, 0.16 M ammonium sulfate, 0.6% 1,2-butanediol. Crystals were transferred to the reservoir solution used for crystallization that additionally contained all the components of SEC buffer and 20–30% glycerol and then cryo cooled in liquid nitrogen. Datasets were collected on beamline 08ID-1 of the CMCF at the Canadian Light Source, using light of 0.979 Å wavelength in Saskatoon, SK, Canada.

Datasets were indexed with the program iMosflm²¹ and scaled with the program AIMLES²². The P222₁ and P4₃22 datasets displayed high levels of anisotropy and were re-indexed and re-scaled in XDS²³ prior to submission to the diffraction anisotropy server^{10,11}. Ellipsoidal resolution boundaries of 2.6, 2.0 and 2.3 Å for the P222₁ dataset and 3.8, 3.8 and 3.3 Å for the P4₃22 dataset were applied along the a*, b* and c* axes respectively and anisotropically scaled using the webserver. Data collection statistics for the pre- and post-anisotropically corrected datasets are presented in Table 1.

Structure determination of GapR in the I4₁32 space group was performed by molecular replacement in the Phaser module of PHENIX²⁴ using models generated by the Ro β etta server⁸ from the GapR sequence. This produced a partial solution for residues 35–66. After using these residues alone as a molecular replacement model, clear density for the remaining residues was visible in the resulting maps. GapR was manually built into the maps in the program Coot²⁵ followed by refinement in PHENIX to produce the final model (Table 1). This structure was then used as a search model to determine the structure of GapR in space groups P222₁ and P4₃22 by molecular replacement using Phaser followed by iterative rounds of model building and refinement in Coot and PHENIX. Figures were generated using PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC).

SEC-MALS. A sample of 50 μ l of GapR at 0.5 mM concentration (when calculated using tetrameric molecular weight) was applied to a Superdex-200 Increase 10 300 GL column (GE Healthcare) equilibrated in SEC buffer, attached to in-line miniDAWN TREOS (Wyatt Technologies) and Optilab rEX (Wyatt Technologies) SEC-MALS instruments. Molecular mass from SEC-MALS data was calculated with the ASTRA 5.3.4.20 (Wyatt Technologies) software. Lower concentrations of GapR eluted at the same volume but gave noisier scattering. Bovine serum albumin was used as a standard.

Gel electrophoresis. For the EMSA experiments, GapR was incubated with DNA in annealing buffer and subjected to electrophoresis in an 8%, 0.5 X TBE gel at 100 V for 60–80 minutes in 0.5 X TBE buffer pH 8.3 at 4 °C. DNA was visualized with SYBR Gold (Thermo Fisher Scientific) and imaged with an AlphaDigiDoc gel documentation system (Alpha Innotech). Protein complexes were stained with InstantBlue (Sigma).

For gel electrophoresis analysis of GapR-DNA crystals, 10 crystals were each looped into 1 μ l stabilization buffer (20 mM HEPES pH 7.4, 200 mM NaCl, 0.2 mM TCEP, 4% PEG 3350, 0.16 M ammonium sulfate, 1% 1,2 butanediol), then looped into 20 μ l of annealing buffer prior to electrophoresis as described above for EMSA.

Data availability

GapR structures determined in this study are available from the Protein Data Base (PDB) under the accession codes 6OZX, 6OZY and 6OZZ.

Received: 30 May 2019; Accepted: 25 October 2019;

Published online: 13 November 2019

References

- Dillon, S. C. & Dorman, C. J. Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat Rev Microbiol* **8**, 185–195, <https://doi.org/10.1038/nrmicro2261> (2010).
- Ricci, D. P. *et al.* Cell cycle progression in *Caulobacter* requires a nucleoid-associated protein with high AT sequence recognition. *Proc Natl Acad Sci USA* **113**, E5952–E5961, <https://doi.org/10.1073/pnas.1612579113> (2016).
- Taylor, J. A., Panis, G., Viollier, P. H. & Marczyński, G. T. A novel nucleoid-associated protein coordinates chromosome replication and chromosome partition. *Nucleic Acids Res* **45**, 8916–8929, <https://doi.org/10.1093/nar/gkx596> (2017).
- Arias-Cartin, R. *et al.* Replication fork passage drives asymmetric dynamics of a critical nucleoid-associated protein in *Caulobacter*. *EMBO J* **36**, 301–318, <https://doi.org/10.15252/embj.201695513> (2017).
- Guo, M. S., Haakonsen, D. L., Zeng, W., Schumacher, M. A. & Laub, M. T. A Bacterial Chromosome Structuring Protein Binds Overtwisted DNA to Stimulate Type II Topoisomerases and Enable DNA Replication. *Cell* **175**, 583–597 e523, <https://doi.org/10.1016/j.cell.2018.08.029> (2018).
- Liu, L. F. & Wang, J. C. Supercoiling of the DNA template during transcription. *Proc Natl Acad Sci USA* **84**, 7024–7027 (1987).
- Rigden, D. J. Ab initio modeling led annotation suggests nucleic acid binding function for many DUFs. *OMICS* **15**, 431–438, <https://doi.org/10.1089/omi.2010.0122> (2011).
- Kim, D. E., Chivian, D. & Baker, D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res* **32**, W526–531, <https://doi.org/10.1093/nar/gkh468> (2004).
- McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658–674, <https://doi.org/10.1107/S0021889807021206> (2007).
- Strong, M. *et al.* Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *Proc Natl Acad Sci USA* **103**, 8060–8065, <https://doi.org/10.1073/pnas.0602606103> (2006).
- Sawaya, M. R. Methods to refine macromolecular structures in cases of severe diffraction anisotropy. *Methods Mol Biol* **1091**, 205–214, https://doi.org/10.1007/978-1-62703-691-7_15 (2014).
- Drew, H. R. *et al.* Structure of a B-DNA dodecamer: conformation and dynamics. *Proc Natl Acad Sci USA* **78**, 2179–2183 (1981).
- Gradia, S. *et al.* hMSH2–hMSH6 forms a hydrolysis-independent sliding clamp on mismatched DNA. *Mol Cell* **3**, 255–261 (1999).
- Jeong, C. *et al.* MutS switches between two fundamentally distinct clamps during mismatch repair. *Nat Struct Mol Biol* **18**, 379–U174, <https://doi.org/10.1038/nsmb.2009> (2011).
- Paillard, S. & Strauss, F. Analysis of the Mechanism of Interaction of Simian Ku Protein with DNA. *Nucleic Acids Research* **19**, 5619–5624, <https://doi.org/10.1093/nar/19.20.5619> (1991).
- Devries, E., Vandriel, W., Bergsma, W. G., Arnberg, A. C. & Vandervliet, P. C. Helix Nuclear-Protein Recognizing DNA Termini and Translocating on DNA Forming a Regular DNA Multimeric Protein Complex. *J Mol Biol* **208**, 65–78, [https://doi.org/10.1016/0022-2836\(89\)90088-0](https://doi.org/10.1016/0022-2836(89)90088-0) (1989).
- Tinker, R. L., Kavassavetis, G. A. & Geiduschek, E. P. Detecting the Ability of Viral, Bacterial and Eukaryotic Replication Proteins to Track Along DNA. *Embo Journal* **13**, 5330–5337, <https://doi.org/10.1002/j.1460-2075.1994.tb06867.x> (1994).
- Burgers, P. M. J. & Yoder, B. L. Atp-Independent Loading of the Proliferating Cell Nuclear Antigen Requires DNA Ends. *J Biol Chem* **268**, 19923–19926 (1993).
- Marczyński, G. T., Rolain, T. & Taylor, J. A. Redefining bacterial origins of replication as centralized information processors. *Front Microbiol* **6**, 610, <https://doi.org/10.3389/fmicb.2015.00610> (2015).
- van den Berg, S., Lofdahl, P. A., Hard, T. & Berglund, H. Improved solubility of TEV protease by directed evolution. *J Biotechnol* **121**, 291–298, <https://doi.org/10.1016/j.jbiotec.2005.08.006> (2006).
- Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D* **67**, 271–281, <https://doi.org/10.1107/S0907444910048675> (2011).
- Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **69**, 1204–1214, <https://doi.org/10.1107/S0907444913000061> (2013).
- Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125–132, <https://doi.org/10.1107/S0907444909047337> (2010).
- Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D* **66**, 213–221, <https://doi.org/10.1107/S0907444909052925> (2010).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D* **66**, 486–501, <https://doi.org/10.1107/S0907444910007493> (2010).
- Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: A sequence logo generator. *Genome Res* **14**, 1188–1190, <https://doi.org/10.1101/gr.849004> (2004).

Acknowledgements

The authors thanks Alba Guarne and Bhushan Nagar for helpful discussions, Svenja de Buhr, Sarah Roffeis, Kim Tran and Jorge Luis Flores for help with preliminary experiments, Diego Alonzo for help with data collection, Maximilian Eivaskhani for assistance with data processing and Yu Seby Chen for help with SEC-MALS. Research described in this paper was performed using beamline 08ID-1 at the Canadian Light Source (CLS), which is supported by the Canada Foundation for Innovation, Natural Sciences and Engineering Research Council of Canada, the University of Saskatchewan, the Government of Saskatchewan, Western Economic Diversification Canada, the National Research Council Canada, and the Canadian Institutes of Health Research. We are grateful to Shaun Labuik and Pawel Grochulski of the CLS. This work was funded from Canadian Institutes of Health Research (FDN-148472) (TMS) and Canada Research Chairs (TMS). JAT and GTM were supported by the Canadian Institutes of Health Research (CIHR, MOP-125998).

Author contributions

G.T.M. and T.M.S. conceived and designed the project. C.H. and J.A.T. performed initial cloning, purification and crystallization studies. M.J.T. purified, crystallized and determined the structure of GapR and performed DNA binding, SEC, and SEC-MALS experiments. M.J.T., T.M.S. and G.T.M. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-019-52964-2>.

Correspondence and requests for materials should be addressed to G.T.M. or T.M.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019