Ring Structure of the Escherichia coli DNA-binding Protein RdgC Associated with Recombination and Replication Fork Repair*

Received for publication, February 6, 2007, and in revised form, February 16, 2007
Published, JBC Papers in Press, February 16, 2007
DOI 10.1074/jbc.C700023200
Geoffrey S. Briggs 1, Paul A. McEwan 1, Jing Yu 1, Timothy Moore 1, Jonas Emsley 2, and Robert G. Lloyd 1

From the 1Institute of Genetics, University of Nottingham, Queen’s Medical Centre, Nottingham NG7 2UH and the 2Centre for Biomolecular Sciences, School of Pharmacy, University of Nottingham, University Park, Nottingham NG7 2RD, United Kingdom

The DNA-binding protein, RdgC, is associated with recombination and replication fork repair in Escherichia coli and with the virulence-associated, pilin antigenic variation mediated by RecA and other recombination proteins in Neisseria species. We solved the structure of the E. coli protein and refined it to 2.4 Å. RdgC crystallizes as a dimer with a head-to-head, tail-to-tail organization forming a ring with a 30 Å diameter hole at the center. The protein fold is unique and reminiscent of a horse-shoe with twin gates closing the open end. The central hole is lined with positively charged residues and provides a highly plausible DNA binding channel consistent with the nonspecific mode of binding detected in vitro and with the ability of RdgC to modulate RecA function in vivo.

Proteins associated with homologous recombination play important roles in maintaining the structural and functional integrity of the genome. They facilitate DNA repair and aid the rescue damaged replication forks. A key feature of recombination is the process of homologous DNA strand exchange. This reaction is catalyzed in bacteria by RecA protein and in eukaryotes and archaea by the RecA homologues Rad51 and RadA, respectively. These proteins assemble on ssDNA to form helical nucleoprotein filaments that initiate the search for homology (1, 2). However, their activity has to be controlled to avoid unnecessary exchanges, which can be highly detrimental (3). RecA activity in E. coli is limited by the need for specialized loading factors (RecBCD and RecFOR) to overcome the propensity for any ssDNA to be rapidly sequestered by SSB protein and is tempered by other proteins (DinI and RecX) that affect the stability of the assembled filament (2). A number of recent studies have indicated that RdgC may be another factor that acts as a negative regulator of RecA function.

The RdgC protein is restricted to the β- and γ-subsections of the Proteobacteria (4) that includes E. coli and Neisseria spp. In the latter, RdgC promotes the recombination-mediated, pilin antigenic variation associated with virulence (5). However, it was first characterized as a factor essential for growth of recombination-deficient mutants of E. coli (6) and later as a DNA-binding protein needed to maintain viability in strains lacking PriA (4), a primosome assembly factor that loads the DnaB replicative helicase at stalled replication forks. In particular, the presence of RdgC appears necessary to counter or survive RecFOR-mediated loading of RecA (4). Biochemical studies with RdgC from Escherichia coli and Neisseria meningitidis showed it to be a dimer in solution and to bind nonspecifically to linear and circular DNA, although it displays a higher affinity for double-stranded over single stranded substrates (4, 7, 8). Recently, it has been shown to inhibit RecA-mediated DNA strand exchange reactions (9). Taken together with the estimate of ~1000 copies of the dimer in an E. coli cell during exponential phase growth (4), these observations are consistent with a role as a negative regulator of RecA. The crystal structure described here provides a number of clues as to how RdgC might bind DNA to achieve this effect.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification—Native RdgC was expressed from E. coli BL21(DE3) carrying pLysS and pGS853, as described (4). A selenomethionyl (Se-Met) derivative was expressed from the same strain using feedback inhibition of methionine biosynthesis (10) by growth in fully supplemented 56/2 minimal salts medium with Se-Met added to 50 mg/ml. After expression, resuspended cell pellets were sonicated and clarified. RdgC was purified by ammonium sulfate precipitation, heparin, and Q ion exchange chromatography followed by gel filtration into 20 mM Tris:HCl, pH 7.5, 100 mM NaCl. Finally, the purified proteins were concentrated to 25–30 mg/ml. Incorporation of Se-Met was confirmed by mass spectrometry.

DNA Substrates—The oligonucleotides YJ7 (5’-TGTAATCGTCTATGACGT-3’) and YJ8 (5’-AACGTCATAGAC-GATTACA-3’) were annealed to make the 19-mer duplex, YJ7/8. The duplex was purified by gel electrophoresis and eluted into 10 mM Tris:HCl, pH 8.5, before concentration to ~300 mM.

Cryocryoprotectant—Cryoprotection was by the dropwise addition of 100% 2,3-butandiol to the crystallization
drop until the volume was 130% of the original. Crystals were harvested and plunged into liquid nitrogen. Data for the Se-Met derivative were collected on ID14-4 at the European Synchrotron Radiation Facility (ESRF) with a wavelength of 0.98 Å to allow collection of anomalous data. Data were indexed and reduced with MOSFLM (11) and SCALA (12) before using SOLVE (13) for initial phase determination. A mean figure of merit of 0.64 with data to 2.5 Å was obtained and produced maps of interpretable quality. An initial model comprising 30% of the structure was built using ARP/wARP. Coot (14) was used for model building with Se-Met residues modeled appropriately, and refinement was carried out to 2.4 Å. Refinement used CNS (15) and REFMAC (16), maintaining a consistent test set throughout. Additional manipulations used CCP4 (17). The final model is complete for both chains in the dimer, apart from residues at the extreme C terminus.

RESULTS AND DISCUSSION

Crystal Structure—We crystallized native RdgC and the Se-Met derivative with and without DNA. The structure presented is for the Se-Met derivative co-crystallized with a 19-bp duplex, YJ7/8, which was solved to 2.4 Å (Table 1). Data were also collected for native RdgC, but the resulting model was less complete. Crystallization was enhanced in the presence of either ssDNA or dsDNA, with crystals reaching 400 × 200 × 200 μm. Crystals formed in the absence of DNA were smaller, non-isomorphous, and diffracted to only 2.8 Å.

In agreement with biochemical data (4), RdgC crystallizes as a dimer with a head-to-head, tail-to-tail organization (Fig. 1, A and B). The dimer forms a ring with a 30 Å diameter hole (Cα to Cα) at the center. Interestingly, the packing of RdgC molecules in the crystal lattice causes these holes to form long, continuous pores through the crystal (Fig. 1C). Analysis of a washed and dissolved crystal by gel electrophoresis confirmed the presence of DNA. Unfortunately, no additional density was observed in the $F_o - F_c$ difference map that could be used to locate the DNA. We assume that the DNA is disordered, which may be related to the mode of binding.

Protein Fold—A search of the structure data base with the Dali server (18) revealed no matches to RdgC. The overall pro-
tein fold is unique and reminiscent of a horseshoe in shape, with twin gates closing the open end (Fig. 1A). The inner face consists of eight anti-parallel β-strands with the dimer interface falling between the fourth and fifth strands (Fig. 2B). This interface is completed by two anti-parallel helices on the outer surface. It is held together by hydrogen bonds between the β-strands and also between a triad of highly conserved residues at the ends of the helices; Gln-212 and Glu-218 from one monomer and Lys-227 from the other (Fig. 2, A and B). The conserved histidine, His-222, completes the dimer interface. The second dimer interface is created where the two regions of extended β-strand forming the gates overlap. Unlike the horsehoe dimer interface, the interface closing the gate appears to be mediated primarily by a double hydrophobic interaction. A conserved phenylalanine, Phe-120, on one chain sits in a pocket formed by highly conserved residues on the other chain: Ile-74, Leu-75, Pro-76, Val-79, Leu-115, Leu-116, Arg-118, Ala-119, and Phe-120, and vice versa (Fig. 2, C and D). Hydrogen bonding between the guanidinium and carbonyl groups of the Arg-118 residues stabilize the dimer interaction (Fig. 2D).

Two anti-parallel α-helices project perpendicular to the ring on either side of the dimer interface at the gate. These “finger” domains run from Ser-77 to Leu-115. They contain a high proportion of conserved lysines that cause the upper surface to be highly electropositive (Fig. 2, A and E). Of particular note is the turn between the two helices, consisting of an arginine (Arg-97) and three lysines (Lys-98, Lys-100, Lys-101), which form a basic tip (Fig. 2E). Although the exact positioning of the fingertip residues is not conserved across all species, the propensity for basic residues in this region is retained (Fig. 2A). The two finger domains do not have identical conformations in the crystal structure; this is probably due to their relative positions in the
crystal lattice, which allows them to make different crystallographic contacts. It may also be indicative of their high B factors, and it is possible that some flexibility is necessary for protein function.

Although RdgC has a novel fold, an electrostatic surface representation (Fig. 3) invites comparisons with the sliding clamp proteins proliferating cell nuclear antigen (19) and β-clamp (20), and with the SARS-CoV nsp7-nsp8 hexadecamer (21). Like RdgC, these have a ring structure with a hole lined with positive potential and an outer negative surface. Sliding clamps are processivity factors that encircle DNA at replication forks. The SARS-CoV protein is thought to be involved in replication and encircles dsRNA. Thus RdgC could bind DNA by encircling it with the positively charged channel interacting with the DNA phosphate backbone (Fig. 3, B–D). The dimensions of the hole (30 Å) are similar to that in SARS-CoV and would suggest that the positively charged side chains can interact directly with the DNA (25 Å diameter). This is in contrast to the sliding clamp proteins, which have a 35 Å hole and are thought to maintain a topological interaction with DNA, whereas not directly interacting with it (22).

The ring structure, with a positively charged hole, would explain the limited sequence and structural specificity seen in biochemical studies. The proposed DNA binding channel is 40 Å from the front to the back of the ring and 55 Å from fingertip to fingertip (Figs. 1B and 3C). This area could interact with ~18 bp of DNA, which agrees well with the 15–20 bp required to make a stable complex in vitro (4, 7). If the protein is primarily interacting with the negatively charged backbone, then as long as the DNA molecule has a region of the correct dimensions (a duplex of at least 20 bp), the binding of different structures will be indistinguishable. Backbone contacts would explain the slight preference for dsDNA. With ssDNA, the backbone phosphates would be less likely to be positioned for optimum binding. The failure to observe electron density for any of the DNA substrates tested (single-stranded, blunt duplexes, 5’ or 3’ overhangs, hairpins, mismatches, insertion-induced bends), despite evidence for its presence in the crystals, is likely to reflect the nonspecific nature of DNA binding.

Another similarity with the sliding clamp family of proteins can be seen in the structure at the horsehoe dimer interface. Although the overall topologies are different, the dimer interfaces are analogous. Both are held together by a continuous run of β-sheet across the interface with a pair of anti-parallel α-helices providing additional interdomain bonds. RdgC has the helices on the outside of the protein, whereas β-clamp has them on the inner face of the ring. In the case of β-clamp, the stability of the dimer interfaces necessitates active loading and unloading on DNA via the clamp loader (23). RdgC is able to bind to circular DNA molecules without a requirement for additional factors (7, 8), suggesting that it can load by itself, possibly by opening of the second dimer interface in the gate region. The highly electropositive finger domain (Figs. 2E and 3, C and D) is an obvious site for the initial recognition of, and interaction with, the DNA phosphate backbone. Once bound, a change in conformation of the two helices forming the finger could disrupt the hydrophobic pocket at the base, which forms the dimer interface, causing the gate to open. The DNA, held in position by the finger, could then enter the opened gate and become encircled by the RdgC protein.

The narrowness of the central hole would suggest that RdgC cannot slide easily along a length of DNA and so is unlikely to be a processivity factor. The tight binding of small oligonucleotides in biochemical assays supports this view. If RdgC could readily slide along DNA, then we would expect these complexes to be inherently unstable. β-Clamp, for example, will not form stable complexes with linear DNA in vitro (24). The lack of specificity with regard to DNA binding substrates would limit its role as a recruitment protein because, although it could clearly form protein-protein interactions, there would be no specific feature of the DNA that could be targeted. Of course, the lack of DNA specificity displayed by RdgC in vitro does not preclude the possibility that RdgC itself is recruited to specific regions by interactions with proteins that do demonstrate DNA structure or sequence specificity.
The function that has been suggested for RdgC is as a modulator of RecA activity. The structure determined suggests how RdgC might achieve this effect. If RdgC binds to DNA in the manner predicted here, the main consequence will be the stabilization of the DNA duplex to unwinding as the DNA is completely encircled. Such a complex could curb the activity of RecA either by forming an obstacle to filament assembly or by preventing subsequent strand exchange or both. We think of RdgC not as a complete block to RecA but rather as a temporary halt that prevents inappropriate or unnecessary activity.

Acknowledgments—We thank Karen Bunting for many helpful comments on the structure and the manuscript. Beam time was provided at the ESRF, Grenoble, France as part of BAG MX-405.

REFERENCES

1. Kowalczykowski, S. C. (2000) Trends Biochem. Sci. 25, 156–165
2. Cox, M. M. (2007) Nat. Rev. Mol. Cell. Biol. 8, 127–138
3. Mahdi, A. A., Buckman, C., Harris, L., and Lloyd, R. G. (2006) Genes Dev. 20, 2135–2147
4. Moore, T., McGlynn, P., Ngo, H. P., Sharples, G. J., and Lloyd, R. G. (2003) EMBO J. 22, 735–745
5. Mehr, I. J., Long, C. D., Serkin, C. D., and Seifert, H. S. (2000) Genetics 154, 523–532
6. Ryder, L., Sharples, G. J., and Lloyd, R. G. (1996) Genetics 143, 1101–1114
7. Moore, T., Sharples, G. J., and Lloyd, R. G. (2004) J. Bacteriol. 186, 870–874
8. Tessmer, I., Moore, T., Lloyd, R. G., Wilson, A., Erie, D. A., Allen, S., and Tendler, S. J. (2005) J. Mol. Biol. 350, 254–262
9. Drees, J. C., Chitteni-Pattu, S., McCaslin, D. R., Inman, R. B., and Cox, M. M. (2006) J. Biol. Chem. 281, 4708–4717
10. Double, S. (1997) Methods Enzymol. 276, 523–530
11. Leslie, A. G. W. (1992) Joint CCP4 + ESRF-EAMCB Newsletter on Protein Crystallography 26
12. Evans, P. R. (1993) in CCP4 Study Weekend: Data Collection and Processing (Sawyer, L., Isaacs, N., and Bailey, S., eds) pp. 114–122, Daresbury Laboratory, Cheshire, UK
13. Terwilliger, T. C., and Berendzen, J. (1999) Acta Crystallogr. Sect. D Biol. Crystallogr. 55, 849–861
14. Emsley, P., and Cowtan, K. (2004) Acta Crystallogr. Sect. D Biol. Crystallogr. 60, 2126–2132
15. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Acta Crystallogr. Sect. D Biol. Crystallogr. 54, 905–921
16. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Acta Crystallogr. Sect. D Biol. Crystallogr. 53, 240–255
17. Collaborative Computational Project, Number 4 (1994) Acta Crystallogr. Sect. D Biol. Crystallogr. 50, 760–763
18. Holm, L., and Sander, C. (1993) J. Mol. Biol. 233, 123–138
19. Krishna, T. S., Kong, X. P., Gary, S., Burgers, P. M., and Kuriyan, J. (1994) Cell 79, 1233–1243
20. Kong, X. P., Onrust, R., O'Donnell, M., and Kuriyan, J. (1992) Cell 69, 425–437
21. Zhai, Y., Sun, F., Li, X., Pang, H., Xu, X., Bartlam, M., and Rao, Z. (2005) Nat. Struct. Mol. Biol. 12, 980–986
22. Kelman, Z., and O'Donnell, M. (1995) Nucleic Acids Res. 23, 3613–3620
23. Yao, N., Turner, J., Kelman, Z., Stukenberg, P. T., Dean, F., Shechter, D., Pan, Z. Q., Hurwitz, J., and O'Donnell, M. (1996) Genes Cells 1, 101–113
24. Stukenberg, P. T., Studwell-Vaughan, P. S., and O'Donnell, M. (1991) J. Biol. Chem. 266, 11328–11334