J Biomol NMR (2014) 58:141–147
DOI 10.1007/s10858-014-9812-8

NMR STRUCTURE NOTE

Solution structure of the RecQ C-terminal domain of human Bloom syndrome protein

Chin-Ju Park · Junsang Ko · Kyoung-Seok Ryu · Byong-Seok Choi

Received: 8 December 2013 / Accepted: 9 January 2014 / Published online: 17 January 2014

Abstract RecQ C-terminal (RQC) domain is known as the main DNA binding module of RecQ helicases such as Bloom syndrome protein (BLM) and Werner syndrome protein (WRN) that recognizes various DNA structures. Even though BLM is able to resolve various DNA structures similarly to WRN, BLM has different binding preferences for DNA substrates from WRN. In this study, we determined the solution structure of the RQC domain of human BLM. The structure shares the common winged-helix motif with other RQC domains. However, half of the N-terminal has unstructured regions (α1–α2 loop and α3 region), and the aromatic side chain on the top of the β-hairpin, which is important for DNA duplex strand separation in other RQC domains, is substituted with a negatively charged residue (D1165) followed by the polar residue (Q1166). The structurally distinctive features of the RQC domain of human BLM suggest that the DNA binding modes of the BLM RQC domain may be different from those of other RQC domains.

Keywords RecQ helicase · BLM · RecQ C-terminal (RQC) domain · NMR structure

C.-J. Park (✉)
Division of Liberal Arts and Sciences, Department of Chemistry, Gwangju Institute of Science and Technology, Oryong-dong, Buk-gu, Gwangju 500-712, Republic of Korea
e-mail: cjpark@gist.ac.kr

J. Ko · B.-S. Choi
Department of Chemistry, KAIST, 373-1, Guseong-dong, Yuseong-gu, Daejeon 305-701, Republic of Korea

K.-S. Ryu
Division of Magnetic Resonance, Korea Basic Science Institute Ochang Campus, Cheongwon-Gun, Ochang-Eup, Yangcheong-Ri 804-1, Chungcheongbuk-Do 363-883, Republic of Korea

Biological context

Bloom syndrome protein (BLM) is a member of a family of five human RecQ helicases [Werner syndrome protein (WRN), RecQ1, RecQ4, RecQ5 and BLM] that are evolutionarily well conserved. The RecQ helicases play crucial roles in multiple DNA metabolic processes, such as DNA recombination, replication and repair, by resolving various DNA structures with ATP-dependent helicase activities (Bohr 2008). Mutations in BLM gene cause Bloom syndrome, a recessive genetic disorder, which is characterized as the early onset of several types of cancer. Similarly, deficiencies of WRN and RecQ4 induce Werner syndrome and Rothmund–Thomson syndrome, respectively. Both genetic disorders are distinguished by premature aging. All three genetic disorders clearly show that RecQ helicases are important for maintaining genomic stability, and each RecQ helicase has distinctive roles in DNA metabolism (Monnat 2010).

All of the RecQ helicases contain multiple domains that provide specificity and functionality. Besides the common helicase domains, BLM and WRN contain a RecQ carboxy-terminal (RQC) domain, a helicase and RNase D C-terminal (HRDC) domain. Several studies have shown that the HRDC domain regulates helicase activity by interacting with other proteins as well as DNA substrates (Kitano et al. 2007, Kim and Choi 2010). The RQC domain is known as the main DNA binding module of RecQ helicases that recognizes various DNA structures (Huber et al. 2006). For example, WRN has been shown to bind and resolve the bubble, forked, splayed arm, and G-quadruplex. Even though BLM is able to resolve various DNA structures similarly to WRN, it has been shown that BLM has different binding preferences for DNA substrates from WRN. BLM has a specific binding preference for the G-quadruplex structure (Kamath-Loeb et al. 2012). It has

been speculated that the distinctive clinical and biochemical properties of the RecQ helicases stem from the RQC domain (Harami et al. 2013). The solution structure of the RQC domain of human WRN, and the crystal structures of the RQC domain of human WRN with double stranded DNA duplex, the human RecQ1 helicases and the *Escherichia coli* RecQ helicase showed that they are a winged-helix motif, which belongs to the helix-turn-helix superfamily (Bernstein et al. 2003; Hu et al. 2005; Pike et al. 2009; Kitano et al. 2010). In this study, we present the solution structure of the RQC domain of human BLM. The structure shares the common winged-helix motif with other RQC domains. However, half of the N-terminal domain has unstructured regions (α1–α2 loop and α3 region), and the aromatic side chain on the top of the β-hairpin, which is important for DNA duplex strand separation in other RQC domains, is substituted with a negatively charged residue (D1165) followed by a polar residue (Q1166). The structurally distinctive features of the RQC domain of human BLM suggest that the DNA binding modes of the BLM RQC domain may be different from those of other RQC domains.

Methods and results

Protein expression and purification

The gene encoding the RQC domain of BLM protein (residues 1,067–1,210) was subcloned into the pET15b vector as an N-terminal histidine-tagged fusion and the construct was transformed into *E. coli* BL21(DE3) cells (Novagen). To obtain uniformly ^{13}C, ^{15}N-labeled proteins, transformed *E. coli* cells were cultivated in minimal M9 medium containing $^{15}NH_4Cl$ and [$^{13}C_6$]-D-glucose as the sole nitrogen and carbon sources, respectively, at 37 °C. The cells were induced by addition of 1 mM isopropylthio-β-D-galactoside (IPTG) when the OD_{600} of the culture reached 0.5. After an additional 5 h of incubation at 37 °C, cells were harvested by centrifugation at 5,000 rpm and 4 °C for 20 min. The cell pellets were resuspended in 50 mM sodium phosphate and 300 mM NaCl, pH 8.0, and sonicated. The labeled proteins were initially purified with a Ni–NTA affinity column. After a thrombin digestion reaction, samples were loaded onto an SP Sepharose FF column (GE healthcare) and purified with a

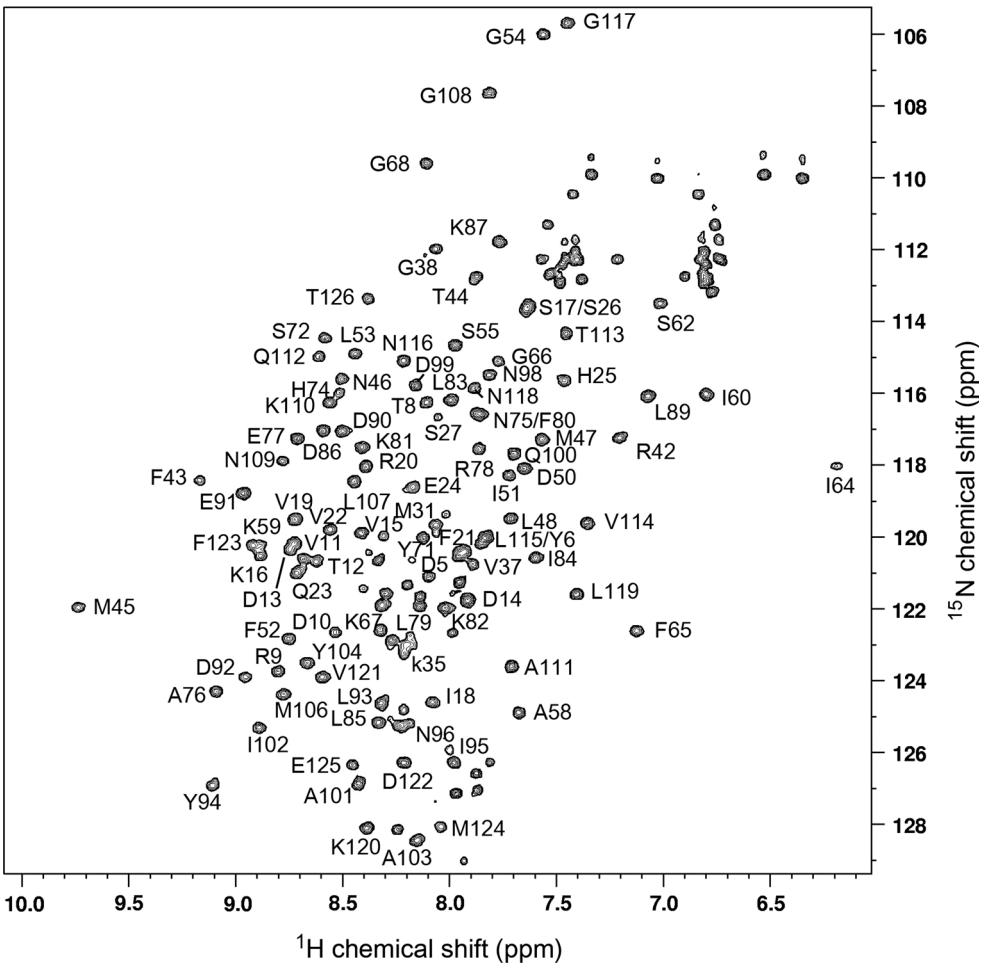

Fig. 1 The ^{1}H–^{15}N HSQC spectra of the RQC domain of human BLM with the backbone amide resonance assignments

linear gradient of 0–1 M NaCl in 20 mM Tris, 100 mM NaCl and 1 mM DTT, pH 7.0. Further purification was performed with a HiLoad 16/60 Superdex 75 gel-filtration column (GE healthcare) with 20 mM Tris, 100 mM NaCl and 1 mM DTT, pH 7.0. The purity and homogeneity of all samples were confirmed by SDS polyacrylamide gel electrophoresis.

NMR spectroscopy

The NMR sample contained 0.8 mM protein in 20 mM Tris, 100 mM NaCl and 1 mM DTT, pH 7.0. NMR experiments were recorded on Bruker Avance 900 spectrometer equipped with a cryoprobe. All NMR experiments were carried out at 300 K. The backbone resonances were assigned using a combination of standard triple resonance experiments including 3D HNCO, HN(CA)CO, CBCA (CO)NH and HNCACB experiments. HBHA(CO)NH with HCCH-TOCSY experiments were used to unambiguously assign the side chain spin system. $^{1}H–^{15}N$ NOESY-HSQC, $^{1}H–^{13}C$ NOESY-HSQC and $^{1}H–^{13}C$ NOESY-HSQC for aromatic side chains were also analyzed. All 3D NOE experiments were performed with a mixing time of 150 ms. The NMR data were processed using Topspin and analyzed using Sparky (T.D. Goddard and D.G. Kneller, Sparky 3.114, University of California, San Francisco).

Structure calculation

The $^{1}H–^{15}N$ HSQC spectra of the RQC domain of human BLM at pH 7.0 show a relatively well dispersed single set

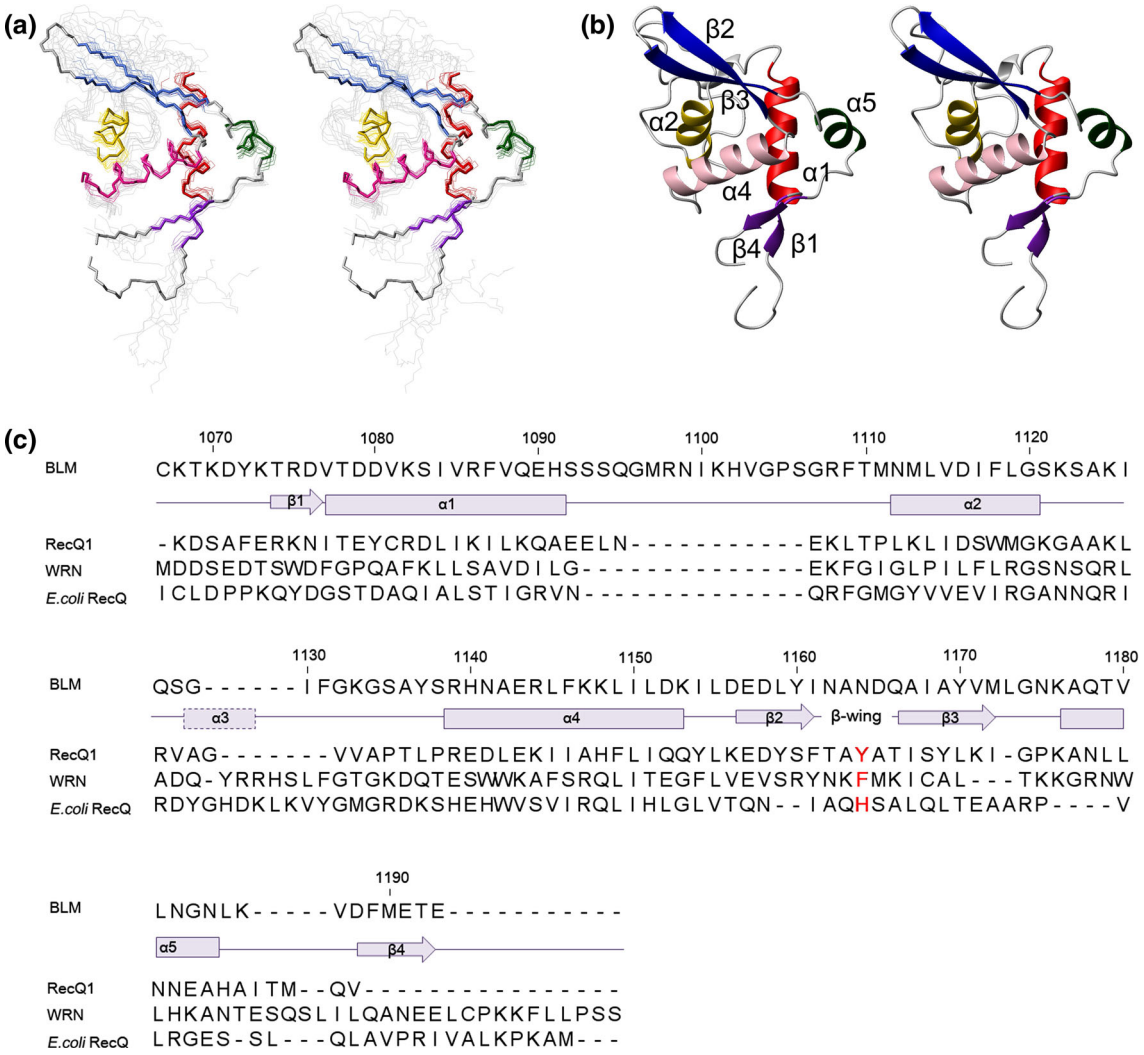

Fig. 2 **a** Stereo views of the overlay of the ensemble of the ten lowest energy structures of the RQC domain of human BLM. The secondary structured regions are shown with *different colors*. **b** *Ribbon* representation of the lowest energy structure of the RQC domain of human BLM. **c** Structure-based sequence alignments of the RQC domain of human BLM, human RecQ1, human WRN and *E. coli* RecQ. The common secondary structures are shown. α3, which is shown with a *dashed line*, was not observed in the solution structure of the RQC domain of human BLM

of peaks (Fig. 1). Based on the monomeric size from the gel filtration chromatography data, the RQC domain of human BLM exists as a monomer in these experimental conditions. The ^{1}H, ^{15}N and ^{13}C resonances of the RQC domain (1,067–1,193, 127 residues) of human BLM have been assigned and deposited in the Biological Magnetic Resonance Data Bank under accession number 19530.

The NOE distance restraints applied for structure determination were extracted from the ^{1}H–^{15}N NOESY-HSQC, ^{1}H–^{13}C NOESY-HSQC and ^{1}H–^{13}C NOESY-HSQC spectra for aromatic side chains. NOEs were assigned using the automated NOE assignment procedure of CYANA 2.1 (Guntert 2004). The phi(φ) and psi(ψ) torsion angle restraints for protein backbone were predicted using the TALOS+ software program (Shen et al. 2009). Only phi(φ) and psi(ψ) angle restraints, which were classified as "good" prediction by the TALOS prediction, were used in the structure calculation. Hydrogen bonds were introduced as a pair of distance restraints based on NOE analysis in combination with the prediction of protein secondary structural elements using the CSI software (Wishart and Sykes 1994).

Structure calculations were initially performed using CYANA 2.1, which combines automated assignment of NOE cross-peaks and structure calculations. On the basis of distance restraints derived from direct CYANA output, structure calculations were also carried out using CNS 1.3 in explicit solvent using the RECOORD protocol (Brunger et al. 1998; Nederveen et al. 2005; Brunger 2007). The 10 lowest energy structures were validated by the PROCHECK-NMR (Laskowski et al. 1996). Structures were visualized using the MOLMOL program (Koradi et al. 1996).

Structure determination

The solution structure of the RQC domain of human BLM was calculated with 1,641 NOE distance restraints, 178 dihedral angle restraints, and 36 × 2 hydrogen bond distance restraints. The superimposition of backbone traces for the ensemble of the ten lowest energy conformers selected from 100 calculated structures is shown in Fig. 2a, revealing a good agreement among applied NMR restraints. Table 1 shows a summary of structure statistics for the RQC domain of human BLM. The root-mean-square deviation value for the backbone atoms in the structured regions was calculated to be 0.77 ± 0.18 Å. The long and flexible loop between α1 and α2 (α1–α2 loop, from H1091 to T1100), which has several missed N–H peaks in the ^{1}H–^{15}N HSQC (R1098, N1099, S1106 and G1107), is not well converged. The Ramachandran plot analysis showed that 80.9 % of the residues are in the most favored region, 16.5 % in the additionally allowed region, 1.7 % in the generously allowed region and 0.9 % in the disallowed regions. The residue located in the disallowed regions of the Ramachandran plot is Q1095, which lacks dihedral angle restraints and is in the α1–α2 loop. The NMR restraints and coordinates of the 10 lowest energy structures of the RQC domain have been deposited in the Protein Data Bank (PDB) with ID 2MH9.

Overview of solution structure

Figure 2b shows that the RQC domain of human BLM is a winged-helix motif, which is a common structural motif found in RQC domains of other RecQ helicases such as human RecQ1, WRN and *E. coli* RecQ helicase (Bernstein et al. 2003, Hu et al. 2005, Pike et al. 2009, Kitano et al. 2010). It is composed of four α-helices and four β-strands. While the other RQC domains have an additional short α-helix after the α2, BLM does not have the equivalent α-helix in the solution structure (Figs. 2c, 3a). Several missed N–H peaks and significantly decreased backbone ^{1}H–^{15}N heteroNOE values show that the α1–α2 loop is flexible (Data is not shown). With the insertion of the long, flexible and disordered α1–α2 loop, the RQC domain of BLM has a relatively high portion of unstructured regions compared with the other RQC domains. α2 and α4 form a layer, and α1, α5 and the residues between α2 and α4 (putative α3) form another layer. The antiparallel β-sheet formed with β2 and β3 is located between α4 and α5, and it tilts toward the α2/α4 layer.

Table 1 Structural statistics for the RQC domain of the human BLM protein

NOE upper distance limits	1,641
Intra-residual	456
Short-range ($\|i - j\| = 1$)	461
Medium-range ($1 < \|i - j\| < 5$)	320
Long-range ($\|i - j\| \geq 5$)	404
Dihedral angle constraints	178
Hydrogen bonds	36 × 2
Violations	
Distances >0.5 Å	none
Dihedral angles >5°	none
Ramachandran analysis (%)	
Most favored regions	80.9
Additionally allowed regions	16.5
Generously allowed regions	1.7
Disallowed regions	0.9
R.M.S.D. from mean structure (*structured region*)[a]	
Backbone (Å)	0.77 ± 0.18
Heavy atom (Å)	1.34 ± 0.13

[a] Residues in the structured regions: 9–24, 45–53, 73–86, 90–106, 111–116, 121–123

The DALI search showed that the RQC domain of human BLM is structurally homologous to the RQC domain of human RecQ1 (PDB ID: 1OYY and 1OYW), WRN (PDB ID: 3AAF and 2AXL) and *E. coli* RecQ helicase (PDB ID: 2V1X and 2WWY) (Bernstein et al. 2003; Hu et al. 2005; Pike et al. 2009; Kitano et al. 2010). Figure 2c shows sequence alignments based on the structures, and Fig. 3a shows the structures of the RQC domains of human BLM, WRN and RecQ1. As we already pointed out, only the RQC domain of the human BLM protein has a long loop insertion between α1 and α2. Additionally, the residues of the β-hairpin of the RQC domain of human BLM are distinctive from the other RQC domains. Figure 3a shows that D1165, which has a negatively charged side chain, is located at the equivalent position to crucial aromatic residues (Y564 in RecQ1 and F1037 in WRN). It has been revealed that the stacking interactions between the aromatic side chain and the base at the terminal of the DNA duplex are important for strand separation in the double-stranded DNA duplex (Pike et al. 2009; Kitano et al. 2010).

During the preparation of the manuscript, crystal structures of the RQC domain of human BLM proteins with phosphate and arsenate ions were published (PDB ID: 3WE2 and 3WE3, respectively) (Kim et al. 2013). Furthermore, the crystal structure of BLM helicases containing ATPase domain, the RQC domain with the deletion of the α1–α2 loop, and the HRDC domain were published in the PDB (PDB ID: 4CDG). Except for the orientation of the α1–α2 loop, which is not well converged in our solution structures, the overall structures agree with both crystal structures. Interestingly, the short α3 appeared in one of the crystal structures, but was not observed in the solution structure (Kim et al. 2013). In the ^{1}H–^{15}N NOESY-HSQC spectra, no characteristic NOE peaks indicating α-helix were observed. Moreover, the backbone angle restraints could not be applied to the most residues between α2 and α4 in the structure calculation. In the crystal structure, the phosphate or arsenate ions, which were used for crystallization, have hydrogen bond networks with S1121, K1122 and R1139. We speculate that the interactions could

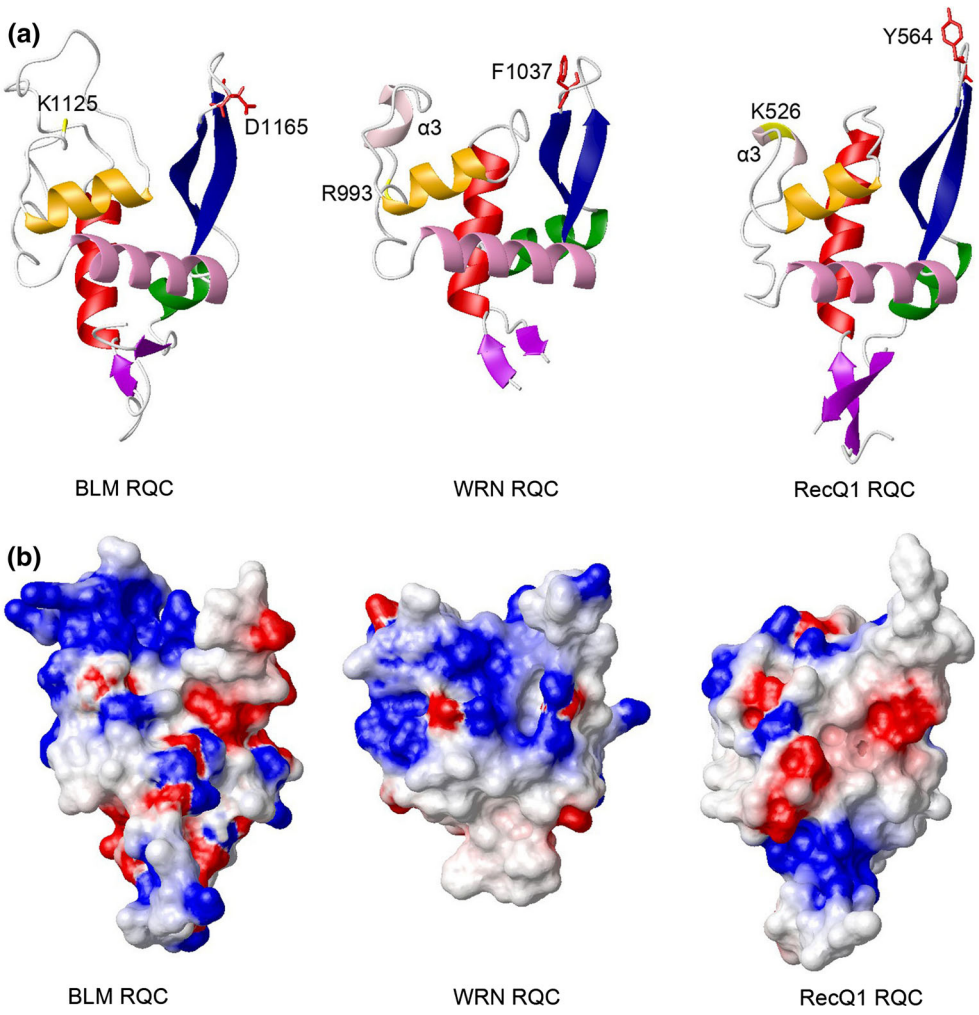

Fig. 3 **a** *Ribbon* representation of the RQC domain of human BLM (*left*), human WRN (*center*) and human RecQ1 (*right*). The residues on the top of β-hairpin are shown in *red*. R993 in WRN RQC and its equivalent residues (K1125 in BLM RQC and K526 in RecQ1 RQC) are shown in *yellow*. **b** Electrostatic potential surface of the RQC domain of human BLM (*left*), human WRN (*center*) and human RecQ1 (*right*). The orientation is the same as in the structures in Fig. 3a. Positive charges are shown in *blue* and negative charges are shown in *red*

stabilize the region between α2 and α4, and induce the formation of α3.

Discussion and conclusions

The structure of the RQC domain of human BLM is a winged-helix motif like the equivalent domains of other RecQ helicases. It is known that the amino acid sequences of the RQC domains are not significantly homologous, even though they show great structural similarity (Harami et al. 2013; Kim et al. 2013).

It is remarkable that the RQC domain of BLM shows several distinctive structural features while maintaining a common fold with other RQC domains. The insertion of a long and flexible α1–α2 loop and an unstructured α3 region, which both contain highly positively charged residues, gives the protein flexibility and potential surfaces for DNA substrate binding through ionic interactions with phosphate backbones (Fig. 3b). The induced α3 in the crystal structure implies that both regions may interact with the DNA phosphate backbones and get stabilized. It has been pointed out that R993 in WRN RQC forms a salt bridge with the phosphate on the 5′ strand of the DNA duplex (Kitano et al. 2010). The equivalent K1125 in BLM RQC may similarly contribute to DNA binding (Kim et al. 2013).

It is reasonable to speculate that the RQC domain of human BLM has different binding modes from the same domain of human WRN. The negatively charged D1165 substitutes the aromatic side chain on the top of the β-hairpin, which has been shown to have crucial roles in DNA duplex strand separation in the crystal structure of the RQC domain of WRN with duplex DNA (Kitano et al. 2010). It has been shown that human BLM and WRN have different binding preferences for various DNA structures, and their RQC domains are responsible for that (Hu et al. 2005, Kamath-Loeb et al. 2012). Interestingly, the RQC domain of human BLM has a high preference for G-quadruplex and low binding affinity for other structures such as partial DNA duplexes and bubbled DNA, which are good binding substrates for the RQC domain of human WRN (Kamath-Loeb et al. 2012). Further structural studies of the RQC domains with DNA substrates are required to elucidate the unique role of the domain in each RecQ helicase.

Acknowledgments This work was supported by the National Research Foundation of Korea grant funded by the Korean Government (2011-0014498 for C.-J.P. and 2009-0092818, 2011-0020322 for B.-S.C.).

Conflict of interest The authors declare that they have no conflict of interest.

References

Bernstein DA, Zittel MC, Keck JL (2003) High-resolution structure of the *E. coli* RecQ helicase catalytic core. EMBO J 22(19):4910–4921. doi:10.1093/emboj/cdg500

Bohr VA (2008) Rising from the RecQ-age: the role of human RecQ helicases in genome maintenance. Trends Biochem Sci 33(12): 609–620. doi:10.1016/j.tibs.2008.09.003

Brunger AT (2007) Version 1.2 of the crystallography and NMR system. Nat Protoc 2(11):2728–2733. doi:10.1038/nprot.2007.406

Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 54(Pt 5):905–921

Guntert P (2004) Automated NMR structure calculation with CYANA. Methods Mol Biol 278:353–378. doi:10.1385/1-59259-809-9:353

Harami GM, Gyimesi M, Kovacs M (2013) From keys to bulldozers: expanding roles for winged helix domains in nucleic-acid-binding proteins. Trends Biochem Sci 38(7):364–371. doi:10.1016/j.tibs.2013.04.006

Hu JS, Feng H, Zeng W, Lin GX, Xi XG (2005) Solution structure of a multifunctional DNA- and protein-binding motif of human Werner syndrome protein. Proc Natl Acad Sci U S A 102(51):18379–18384. doi:10.1073/pnas.0509380102

Huber MD, Duquette ML, Shiels JC, Maizels N (2006) A conserved G4 DNA binding domain in RecQ family helicases. J Mol Biol 358(4):1071–1080. doi:10.1016/j.jmb.2006.01.077

Kamath-Loeb A, Loeb LA, Fry M (2012) The Werner syndrome protein is distinguished from the Bloom syndrome protein by its capacity to tightly bind diverse DNA structures. PLoS One 7(1):e30189. doi:10.1371/journal.pone.0030189

Kim YM, Choi BS (2010) Structure and function of the regulatory HRDC domain from human Bloom syndrome protein. Nucl Acids Res 38(21):7764–7777. doi:10.1093/nar/gkq586

Kim SY, Hakoshima T, Kitano K (2013) Structure of the RecQ C-terminal domain of human bloom syndrome protein. Sci Rep 3:3294. doi:10.1038/srep03294

Kitano K, Yoshihara N, Hakoshima T (2007) Crystal structure of the HRDC domain of human Werner syndrome protein, WRN. J Biol Chem 282(4):2717–2728. doi:10.1074/jbc.M610142200

Kitano K, Kim SY, Hakoshima T (2010) Structural basis for DNA strand separation by the unconventional winged-helix domain of RecQ helicase WRN. Structure 18(2):177–187. doi:10.1016/j.str.2009.12.011

Koradi R, Billeter M, Wuthrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph 14(1):51–55, 29–32

Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 8(4):477–486

Monnat RJ Jr (2010) Human RECQ helicases: roles in DNA metabolism, mutagenesis and cancer biology. Semin Cancer Biol 20(5):329–339. doi:10.1016/j.semcancer.2010.10.002

Nederveen AJ, Doreleijers JF, Vranken W, Miller Z, Spronk CA, Nabuurs SB, Guntert P, Livny M, Markley JL, Nilges M, Ulrich EL, Kaptein R, Bonvin AM (2005) RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. Proteins 59(4):662–672. doi:10.1002/prot.20408

Pike AC, Shrestha B, Popuri V, Burgess-Brown N, Muzzolini L, Costantini S, Vindigni A, Gileadi O (2009) Structure of the

human RECQ1 helicase reveals a putative strand-separation pin. Proc Natl Acad Sci U S A 106(4):1039–1044. doi:10.1073/pnas.0806908106

Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR 44(4):213–223. doi:10.1007/s10858-009-9333-z

Wishart DS, Sykes BD (1994) The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. J Biomol NMR 4(2):171–180