Acta Crystallographica Section D
**Biological Crystallography**
ISSN 1399-0047

# Structure of human Bloom's syndrome helicase in complex with ADP and duplex DNA

**Michael K. Swan, Valerie Legris, Adam Tanner, Philip M. Reaper, Sarah Vial, Rebecca Bordas, John R. Pollard, Peter A. Charlton, Julian M. C. Golec and Jay A. Bertrand***

Vertex Pharmaceuticals (Europe), Abingdon, Oxfordshire, England

Correspondence e-mail: jay_bertrand@vrtx.com

Received 19 December 2013
Accepted 5 March 2014

**PDB reference:** BLM, complex with ADP and duplex DNA, 4o3m

Bloom's syndrome is an autosomal recessive genome-instability disorder associated with a predisposition to cancer, premature aging and developmental abnormalities. It is caused by mutations that inactivate the DNA helicase activity of the BLM protein or nullify protein expression. The BLM helicase has been implicated in the alternative lengthening of telomeres (ALT) pathway, which is essential for the limitless replication of some cancer cells. This pathway is used by 10–15% of cancers, where inhibitors of BLM are expected to facilitate telomere shortening, leading to apoptosis or senescence. Here, the crystal structure of the human BLM helicase in complex with ADP and a 3′-overhang DNA duplex is reported. In addition to the helicase core, the BLM construct used for crystallization (residues 640–1298) includes the RecQ C-terminal (RQC) and the helicase and ribonuclease D C-terminal (HRDC) domains. Analysis of the structure provides detailed information on the interactions of the protein with DNA and helps to explain the mechanism coupling ATP hydrolysis and DNA unwinding. In addition, mapping of the missense mutations onto the structure provides insights into the molecular basis of Bloom's syndrome.

## 1. Introduction

Mutations in the human *BLM* gene give rise to Bloom's syndrome (BS), a rare autosomal recessive disorder characterized by a predisposition to cancer, growth retardation, sunlight sensitivity, immunodeficiency and fertility defects (Ellis *et al.*, 1995; German, 1997). Individuals with BS are highly susceptible to most types of cancer and diagnosis occurs at an average age of 24 years. Cells from BS patients display marked chromosome instability, including chromatid breaks and gaps, gross structural rearrangements and a characteristic surge in the frequency of sister chromatid exchanges (German *et al.*, 1965; Chaganti *et al.*, 1974; Neff *et al.*, 1999). The *BLM* gene product, BLM, is a member of the RecQ family of DNA helicases and catalyzes single-stranded DNA-dependent ATP hydrolysis and ATP-dependent 3′–5′ unwinding of duplex DNA (Ellis *et al.*, 1995; Karow *et al.*, 1997). BLM has been implicated in a number of important DNA-repair processes that occur at DNA double-strand breaks (DSBs) and sites of DNA-replication stress (Chu & Hickson, 2009). A key role for BLM is regulation of the homologous recombination (HR) pathway of DSB repair to prevent the formation of potentially deleterious recombination products. This is in part achieved by driving convergent branch migration and the 'dissolution' of double Holliday junctions, which commonly arise during HR (Karow *et al.*, 2000; Wu & Hickson, 2003). Related roles for BLM in the restart of stalled replication forks and the decatenation of replication intermediates also appear to be

very significant for maintenance of genome integrity (Chan *et al.*, 2007; Davies *et al.*, 2007). In keeping with these cellular functions, the helicase activity of BLM can resolve a variety of specific DNA substrates *in vitro* that mimic structures arising during DSB repair and DNA replication, including 3′-tailed duplexes and four-way Holliday junctions (Sun *et al.*, 1998; Mohaghegh *et al.*, 2001; Bachrati *et al.*, 2006).

Native BLM functions in complex with a number of proteins that are likely to modulate its cellular activity, including topoisomerase III$\alpha$, RMI1 and RMI2 (Chu & Hickson, 2009). However, even for BLM alone the structural picture of its mechanism of action is incomplete. For example, in addition to the DEAH helicase domain, BLM harbours a C-terminal extension containing two poorly understood regions: the RecQ carboxy-terminal (RQC) and the helicase and RNaseD C-terminal (HRDC) domains (Morozov *et al.*, 1997; Bernstein & Keck, 2003; Bennett & Keck, 2004; Guo *et al.*, 2005). Both are believed to contribute to DNA binding, yet these domains are not present in all RecQ family members and their exact roles have not been clarified by the available human BLM crystal structures (Kim *et al.*, 2013; PDB entry 4cdg). Furthermore, although most cases of BS result from nonsense mutations that severely truncate BLM, roughly 15% of cases are caused by missense mutations (German *et al.*, 2007). Finally, despite its association with BS, there has been growing interest in BLM as an anticancer drug target. The roles of BLM in the cellular response to DNA damage suggest that BLM inhibition could provide a novel means to sensitize tumour cells to DNA-damaging chemotherapy (Imamura *et al.*, 2001; Davies *et al.*, 2007). Additional results indicate that BLM may be particularly important for the survival of tumours using the recombination-based alternate lengthening of telomeres (ALT) pathway for telomere maintenance (Durant, 2012; Rezazadeh, 2013). The first small-molecule BLM inhibitors have recently been disclosed (Nguyen *et al.*, 2013; Rosenthal *et al.*, 2013), but there is currently no detailed understanding of their mechanism of action. In addition, any structure-aided optimization of lead compounds is certain to be hampered by the absence of high-resolution crystallographic data for the drug target in complex with DNA.

The growing interest in BLM and its relationship to human disease prompted us to seek structural and mechanistic insight into this helicase. Here, we describe the 2.3 Å resolution structure of the BLM helicase solved in a ternary complex form bound to a 3′-overhang duplex DNA and a molecule of ADP. This structure provides insight into both the positioning of the domains and the mechanism of strand separation employed by the enzyme.

## 2. Materials and methods

### 2.1. Protein and DNA preparation

Three BLM fragments ($BLM_{636–1195}$, $BLM_{636–1298}$ and $BLM_{640–1298}$) were cloned into pET-15b using *Xho*I and *Bam*HI restriction sites. The resulting proteins contained N-terminal 6×His tags and TEV protease cleavage sites. The expression plasmids were transformed into *Escherichia coli* Rosetta 2(DE3) competent cells (Millipore). The expression and purification protocols were identical for the different BLM fragments. 5 ml from a 100 ml overnight culture in Terrific Broth containing 50 µg ml$^{-1}$ kanamycin and 34 µg ml$^{-1}$ chloramphenicol was used to inoculate 100 ml TB/kanamycin/chloramphenicol. After 4 h of growth, 20 ml of this culture was used to inoculate 1 l TB/kanamycin/chloramphenicol. The culture was grown at 37°C to an $OD_{600}$ of 1.8. The temperature was reduced to 18°C and IPTG was added to 0.5 m*M*, followed by overnight incubation at 18°C. Selenomethionine-derivatized protein was produced using the Overnight Express Autoinduction System 2 following the manufacturer's protocol (Millipore, UK). Selenium incorporation was checked using mass spectrophotometry and revealed that 14 out of 18 methionines had incorporated selenium.

Frozen cell pellets were lysed three times using an M-110Y Microfluidiser (Microfluidics) in ten volumes of lysis buffer (20 m*M* Tris pH 7.9, 0.5 *M* NaCl, 10% glycerol, 0.1% Triton X-100, 20 m*M* imidazole) supplemented with cOmplete Protease Inhibitor Cocktail Tablets (Roche). Cell debris was removed by centrifugation (30 min, 24 000 rev min$^{-1}$ at 4°C) and the supernatant (200 ml) was bound to 10 ml Ni Beads (HisPur, Thermo Scientific) for 2 h at 4°C with gentle mixing. The beads were washed twice in lysis buffer and loaded onto a gravity-flow column. The column was washed with 200 ml lysis buffer and eluted at 5 ml min$^{-1}$ with 60 ml elution buffer (lysis buffer containing 250 m*M* imidazole). The eluted fractions were then diluted with an equal volume of 20 m*M* Tris pH 7.9, 5% glycerol, loaded onto a heparin column (Sepharose Fast Flow, GE Healthcare) on an ÄKTApurifier system at 5 ml min$^{-1}$ and eluted with a linear gradient of NaCl (0.25–0.8 *M*). Fractions containing BLM were combined and the affinity tag was cleaved by TEV protease (at a ratio of 1:50) at room temperature for 4 h with gentle mixing. Cleaved protein was then concentrated using Amicon Ultra concentrators (30K MWCO; Millipore) and loaded onto a GFS200 26/60 gel-filtration column equilibrated with gel-filtration buffer (20 m*M* Tris pH 7.9, 0.5 *M* NaCl, 5% glycerol). Fractions were collected at 2 ml min$^{-1}$ using an ÄKTApurifier system. Fractions containing BLM protein were combined, concentrated to 22 mg ml$^{-1}$ using Vivaspin concentrators (30K MWCO; Sartorius), divided into 50 µl aliquots and flash-frozen in liquid nitrogen.

The $BLM^{636–1298}$ Asn1164Ala mutant was expressed and purified using the same protocols as for the other constructs except that the protein failed to bind to the heparin column.

### 2.2. Enzyme assays and binding data

BLM helicase activity was measured using a forked duplex DNA substrate labelled with a tetramethylrhodamine fluorophore (TAMRA) at the 3′ end of one strand and a non-fluorescent Black Hole Quencher 2 (BHQ2) at the 5′ end of the complementary strand (Nguyen *et al.*, 2013). ATP-dependent unwinding of the forked duplex DNA by BLM

**Table 1**
Data-collection, phasing and refinement statistics.

Values in parentheses are for the highest resolution shell.

| | | SeMet | | |
|---|---|---|---|---|
| | Native | Peak | Inflection | Remote |
| Data collection | | | | |
| Space group | *C*2 | *C*2 | *C*2 | *C*2 |
| Unit-cell parameters | | | | |
| *a* (Å) | 100.93 | 99.51 | 99.71 | 99.79 |
| *b* (Å) | 164.68 | 166.13 | 166.56 | 166.79 |
| *c* (Å) | 150.98 | 51.00 | 51.08 | 51.11 |
| $\alpha = \gamma$ (°) | 90.00 | 90.00 | 90.00 | 90.00 |
| $\beta$ (°) | 90.14 | 90.85 | 90.92 | 91.00 |
| Wavelength (Å) | 0.9200 | 0.9793 | 0.9795 | 0.9678 |
| Resolution (Å) | 2.30 | 3.1 | 3.3 | 3.5 |
| $R_{sym}$ or $R_{merge}$ | 0.030 (0.695) | 0.050 (0.633) | 0.047 (0.650) | 0.050 (0.640) |
| $\langle I/\sigma(I)\rangle$ | 18.4 (2.1) | 14.2 (2.3) | 14.9 (2.2) | 13.9 (2.1) |
| Completeness (%) | 98.8 (99.0) | 99.1 (98.7) | 99.4 (99.8) | 99.3 (98.6) |
| Multiplicity | 3.4 (3.4) | 3.8 (3.9) | 3.8 (3.7) | 3.8 (3.8) |
| Refinement | | | | |
| Resolution (Å) | 2.30 | | | |
| No. of reflections | 36392 | | | |
| $R_{work}/R_{free}$ | 0.181/0.215 | | | |
| No. of atoms | | | | |
| Protein | 4764 | | | |
| Nucleic acid | 570 | | | |
| Ligand/ion | 29 | | | |
| Water | 60 | | | |
| *B* factors (Å$^2$) | | | | |
| Protein | 80.2 | | | |
| Nucleic acid | 132.4 | | | |
| Ligand/ion | 63.5 | | | |
| Water | 69.5 | | | |
| R.m.s. deviations | | | | |
| Bond lengths (Å) | 0.010 | | | |
| Bond angles (°) | 1.07 | | | |

results in an increase in fluorescence intensity owing to the loss of quenching. Helicase assays were carried out in a Corning 3676 Low Volume Black 384-Well plate containing 50 m*M* Tris pH 7.5, 5 m*M* $MgCl_2$, 50 m*M* NaCl, 0.01% Tween 20, 1 m*M* DTT, 2.5 µg ml$^{-1}$ poly(dI–dC), 3 m*M* phosphoenolpyruvate, 70 µg ml$^{-1}$ pyruvate kinase and either 20 n*M* $BLM_{636–1298}$ or 100 n*M* $BLM_{636–1195}$. Forked duplex DNA (5′-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCGTACCCGATGTGTTCGTTC-3′-TAMRA, BHQ2-5′-GAACGAACACATCGGGTACGTTTTTTTTTTTTTTTTTTTTTTTTTTTTT-3′) and ATP were added at various concentrations for $K_m$ determinations in a final volume of 20 µl. Reactions were initiated with ATP and immediately transferred to a BMG PHERAstar plate reader to measure the increase in fluorescence for 30 min in kinetic mode using 540 nm excitation and 580 nm emission filters. $K_m$ values were calculated from linear initial rate data by nonlinear regression using the *GraphPad Prism* software package (v.4.0a; GraphPad Software, San Diego, California, USA).

The ATPase activity of BLM was measured in an enzyme-coupled spectrophotometric assay. Assays were carried out in Corning 3706 White 384-Well plates containing 50 m*M* Tris pH 7.5, 5 m*M* $MgCl_2$, 50 m*M* NaCl, 0.01% Tween 20, 1 m*M* DTT, 3 m*M* phosphoenolpyruvate, 70 µg ml$^{-1}$ pyruvate kinase, 360 µ*M* NADH, 24 µg ml$^{-1}$ lactate dehydrogenase, 200 µ*M* ATP and either 40 n*M* $BLM_{636–1298}$ or 10 n*M* $BLM_{636–1195}$. DNA oligomers were added at various concentrations for $K_m$ determination in a final volume of 32 µl. Reactions were initiated with ATP and were immediately transferred to a SpectraMax plate reader (Molecular Devices, Sunnyvale, California, USA) to measure the change of absorbance at 340 n*M* (corresponding to stoichiometric consumption of NADH) over 15 min at 25°C. $K_m$ values were calculated from linear initial rate data by nonlinear regression using the *GraphPad Prism* software package (v.4.0a). The DNA oligomers used in the ATPase assay include the forked duplex described above, a single-stranded DNA, a 12/18-nucleotide 3′-overhang duplex and a 16/24-nucleotide 3′-overhang duplex. The 16/24-nucleotide 3′-overhang duplex contains a 16-nucleotide oligonucleotide primer (3′-GAACCTAGAGCTGCGA-5′) annealed to a 24-nucleotide template (5′-CTTGGATCTCGACGCTCTCCCTTA-3′). Oligonucleotides were obtained from Eurofins MWG Operon in desalted form.

For label-free MST experiments, the concentrations of the BLM constructs were kept constant (500 n*M* for $BLM_{636–1195}$ and 200 n*M* for $BLM_{636–1298}$) and the concentration of the 16/24-nucleotide 3′-overhang duplex DNA was varied between 1.25 µ*M* and 1.2 n*M*. Assays were performed in 20 m*M* Tris pH 7.2, 0.25 *M* NaCl, 5% glycerol. After a short incubation and centrifugation, the samples were loaded into MST NT.LabelFree standard glass capillaries and the MST analysis was performed using a Monolith NT.LabelFree (NanoTemper Technologies GmbH).

### 2.3. Co-crystallization

The native ternary complex was prepared by incubating $BLM_{640–1298}$ with the 16/24-nucleotide 3′-overhang duplex suspended in 10 m*M* Tris–HCl pH 8.0 with 2 m*M* ADP at a molar ratio of 1:1.2 and was then crystallized from solutions containing either 10% 2-propanol or 2–6%(*w*/*v*) PEG 400, 50 m*M* sodium cacodylate pH 5.5–6.0 and 5–40 m*M* magnesium chloride. Selenomethionine-incorporated ternary-complex crystals were prepared using $BLM_{636–1298}$ as per the native complex and grew in equivalent conditions. The co-crystals grew as single crystals in the form of rhombohedral rods to a length of ∼100 µm over a period of 4–7 d. They belonged to space group *C*2, with unit-cell parameters *a* = 100.1, *b* = 164.8,

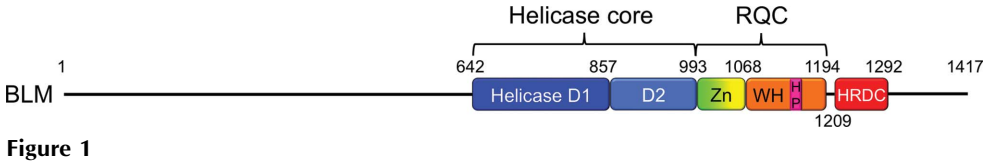


**Figure 1**
Schematic representation of the BLM structure illustrating the helicase core, RQC and HRDC domains. Both the helicase core and the RQC domains can be further divided into two separate subdomains. Domains are coloured to conform to the subsequent structural figures. Abbreviations are as follows: Zn, zinc-binding subdomain; WH, winged helix; HP, strand-separating hairpin.

**Table 2**
Kinetic and binding parameters of $BLM_{636–1195}$ compared with those of $BLM_{636–1298}$.

$V^*_{max}$ values have been normalized to 10 n*M* enzyme concentration.

| | Assay | DNA substrate | $BLM_{636–1195}$ | $BLM_{636–1298}$ |
|---|---|---|---|---|
| $V^*_{max}$ | DNA unwinding | Forked duplex | 3.8 ± 0.1 fluorescence units $s^{-1}$ | 34 ± 1 fluorescence units $s^{-1}$ |
| $K_m$, ATP | DNA unwinding | Forked duplex | 21 ± 3 μ*M* | 81 ± 5 μ*M* |
| $K_m$, DNA | DNA unwinding | Forked duplex | 96 ± 20 n*M* | 9 ± 2 n*M* |
| $V^*_{max}$ | ATPase | Forked duplex | −12.8 ± 0.4 mAbs $min^{-1}$ | −3.4 ± 0.1 mAbs $min^{-1}$ |
| $K_m$, DNA | ATPase | Forked duplex | 2.3 ± 0.4 n*M* | 2.9 ± 0.4 n*M* |
| $K_m$, DNA | ATPase | Single strand | 1.5 ± 0.1 n*M* | 3.7 ± 0.3 n*M* |
| $K_m$, DNA | ATPase | 12/18-Nucleotide 3′-overhang | 25 ± 4 n*M* | 31 ± 3 n*M* |
| $K_m$, DNA | ATPase | 16/24-Nucleotide 3′-overhang | 3.2 ± 0.3 n*M* | 6.6 ± 0.6 n*M* |
| $K_d$, DNA | MST | 16/24-Nucleotide 3′-overhang | 16.2 ± 1.9 n*M* | 12.8 ± 2.2 n*M* |

$c$ = 50.7 Å, $\beta$ = 90.2°. For data collection, the co-crystals were cryoprotected by rapid soaking in mother-liquor solutions plus 25% ethylene glycol and were then flash-cooled in liquid nitrogen. Native crystals used to generate the final refinement data set were grown from conditions containing 20 m*M* calcium chloride in place of magnesium chloride and were harvested after 3 d.

### 2.4. Data collection and processing

The best native data set used for refinement was measured on Diamond Light Source (DLS) beamline I04-1 at λ = 0.92 Å and extended to 2.3 Å resolution in space group *C*2, with one molecule per asymmetric unit. Data collections from SeMet-derivative complex crystals were performed on DLS beamline I03; the crystals were observed to be significantly more radiation-sensitive than the native crystals. MAD data were measured at three wavelengths: peak, inflection and high-energy remote as determined by fluorescence scans around the Se *K* absorption edge. Data were obtained from crystals that processed in both space groups *C*2 and $P2_1$; however, the *C*2 data were used to determine the initial phases. These data extended to 3.1, 3.3 and 3.5 Å resolution for the peak, inflection and high-remote wavelengths, respectively.

### 2.5. Structure determination

The program *autoSHARP* (Vonrhein *et al.*, 2007) was used to locate and refine the selenium sites and to calculate phases. The resulting electron-density map calculated at 3.1 Å resolution after solvent flattening with *SOLOMON* (Abrahams & Leslie, 1996) showed clear secondary-structural features and facilitated subsequent positioning of the initial model by molecular replacement. This was carried out using a model prepared by *CHAINSAW* from the RecQ1 structure (PDB entry 2v1x; Pike *et al.*, 2009) to leave only residues corresponding to the BLM sequence that were divided into three domains: 639–856, 857–1068 and 1069–1193. The domains were sequentially positioned into the experimental electron-density map using *MOLREP*; the DNA model derived from the RecQ1 complex structure (PDB entry 2wwy; Pike *et al.*, 2009) was then manually positioned using *Coot*. This preliminary model was used as a search model for input into *MOLREP* using the 2.3 Å resolution native data, enabling positioning of the protein and bound DNA duplex. Subsequent iterative rounds of refinement with *BUSTER* and model building with *Coot* (Emsley *et al.*, 2010) enabled building of the HRDC domain, bound ADP molecule and metal ions; the final addition of water molecules then resulted in an $R_{cryst}$ of 18.1% and an $R_{free}$ of 21.5%. The final model includes residues 640–798, 809–950, 954–1010, 1014–1092, 1107–1194 and 1209–1292 of the protein and residues 2–15 and 4–18 of the DNA duplex. A bound molecule of ADP is clearly visible within the active site coordinated by a $Ca^{2+}$ ion together with 60 water molecules. There is also a $Zn^{2+}$ ion in the zinc-binding motif. Validation was carried out with *BUSTER-REPORT* and *MolProbity* (Chen *et al.*, 2010). Data-collection, phasing and refinement statistics are given in Table 1.

## 3. Results and discussion

### 3.1. Protein production

Three truncated constructs of human BLM were produced for enzymatic and structural studies that include the helicase and RQC domains either with or without the HRDC domain (Fig. 1). The shorter construct is based on the structural homology within the RecQ family ($BLM_{636–1195}$) for the helicase and RQC domains (Bernstein *et al.*, 2003; Pike *et al.*, 2009; Kitano *et al.*, 2010). The longer constructs that include the HRDC domain ($BLM_{636–1298}$ and $BLM_{640–1298}$) are based on the protein used by Nguyen and coworkers in their HTS assay (Nguyen *et al.*, 2013). The recombinant proteins were purified to homogeneity using nickel-affinity, heparin and gel-filtration chromatography and, based on retention times in gel filtration, are both predicted to be monomers.

### 3.2. BLM helicase and ATPase assays

Full-length human BLM is a DNA-stimulated ATPase and an ATP-dependent helicase that unwinds DNA in the 3′–5′ direction (Karow *et al.*, 1997). Although enzymatic characterizations have been performed with truncated constructs similar to $BLM_{636–1298}$ (Janscak *et al.*, 2003), comparative information for a pure homogenous protein without the HRDC domain was lacking. Therefore, we decided to perform a parallel enzymatic characterization of the BLM constructs with and without the HRDC using both a helicase assay and an ATPase assay (Table 2, Fig. 2 and Supplementary Fig. S1[1]). The helicase assay follows the unwinding of a forked duplex DNA using a fluorescent quenching system. As expected, the

[1] Supporting information has been deposited in the IUCr electronic archive (Reference: CB5052).

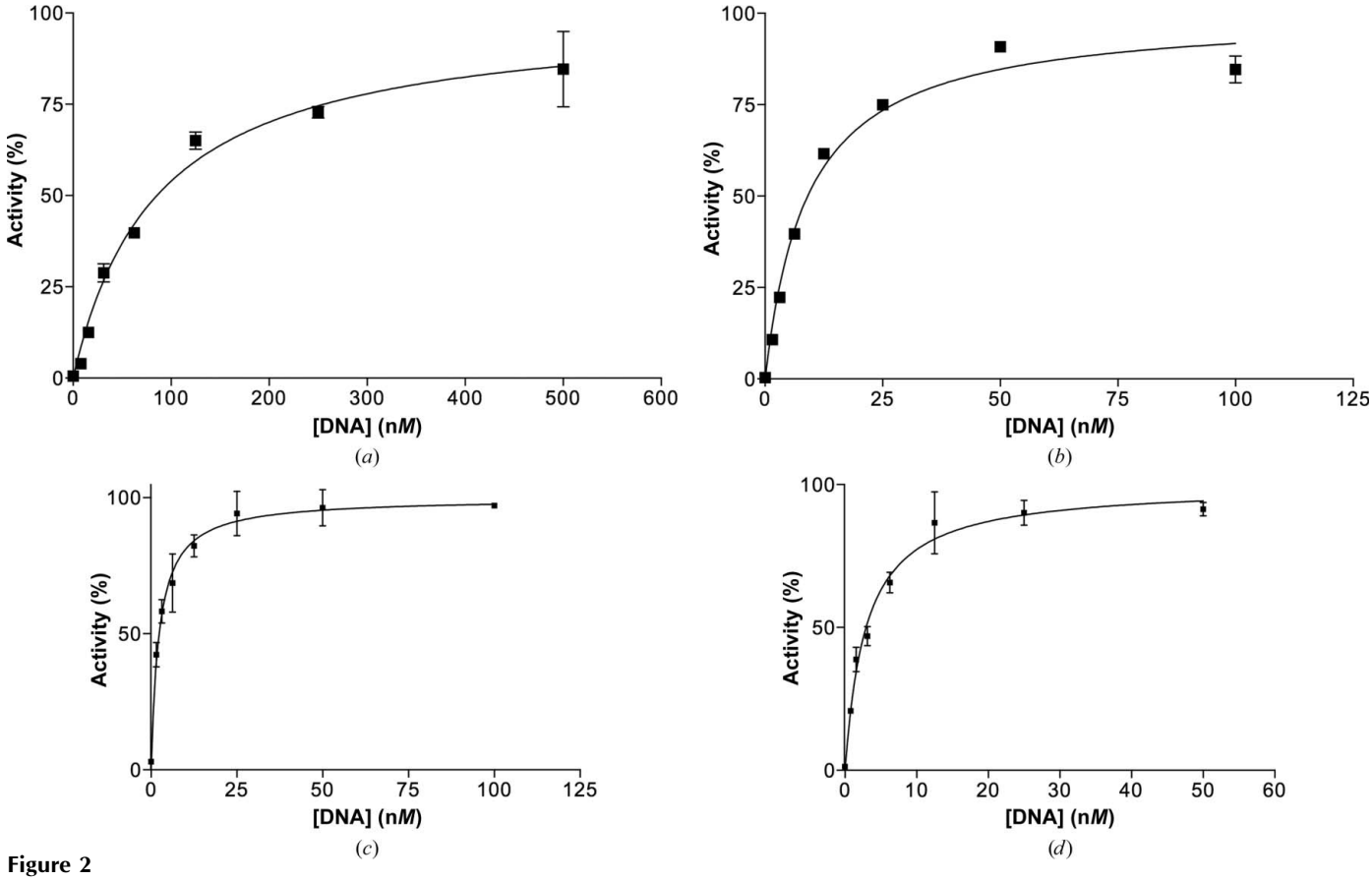


**Figure 2**
Kinetic data for the BLM constructs with and without the HRDC using a helicase and an ATPase assay. $K_m$ plots for DNA in the helicase assay with (*a*) $BLM_{636–1195}$ or (*b*) $BLM_{636–1298}$ and in the ATPase assay with (*c*) $BLM_{636–1195}$ or (*d*) $BLM_{636–1298}$.

activity in the helicase assay is higher for the construct with the HRDC domain, with a $V_{max}$ value for $BLM_{636–1298}$ that is approximately ninefold greater than that of $BLM_{636–1195}$. Also, the average $K_m$ value for DNA is significantly lower when the HRDC domain is present, with values of 96 n*M* for $BLM_{636–1195}$ and 9 n*M* for $BLM_{636–1298}$. In contrast, the average $K_m$ value for ATP is slightly higher for the construct with the HRDC, with values of 21 µ*M* for $BLM_{636–1195}$ and 81 µ*M* for $BLM_{636–1298}$. Given that the ATP concentration in cells is >1 m*M*, ATP would be saturating in both cases, so the differences in the ATP $K_m$ values would be unlikely to have any biological relevance. However, in the ATPase assay with the same forked duplex DNA the activity is lower for the construct with the HRDC domain, with a $V_{max}$ value for $BLM_{636–1298}$ that is approximately fourfold less than that for $BLM_{636–1195}$. Interestingly, the $K_m$ values for the forked duplex DNA in the ATPase assay are similar for $BLM_{636–1195}$ and $BLM_{636–1298}$. As a consequence of this unexpected result, we varied the DNA substrates in the ATPase assay to determine whether the effect was specific to the forked duplex. ATPase assays with three additional DNA substrates (single strand, 12/18-nucleotide 3′-overhang duplex and 16/24-nucleotide 3′overhang duplex) gave similar $K_m$ values for DNA with the BLM constructs with and without the HRDC. Furthermore, three of the four DNA substrates gave $K_m$ values for DNA in a similar range between 1.5 and 6.6 n*M*, with the exception being the shorter 12/18-nucleotide 3′-overhang duplex, with values between 25 and 31 n*M*. Therefore, although the presence of the HRDC domain improves the efficiency in the helicase assay, it also decreases the efficiency in the ATPase assay. To better understand the discrepancies in the $K_m$ values for DNA in the helicase and ATPase assays, the $K_d$ values for the two BLM constructs binding the 16/24-nucleotide 3′-overhang duplex DNA were measured using microscale thermophoresis (MST). This gave $K_d$ values of 16.2 n*M* for $BLM_{636–1195}$ and 12.8 n*M* for $BLM_{636–1298}$ (Supplementary Fig. S2), values which correlate well with the $K_m$ values from the ATPase assay. Together, these results suggest that the HRDC plays a key role in the mechanism coupling the ATP and helicase activities for human BLM.

### 3.3. Crystallization of the $BLM_{636–1298}$ ternary complex with AMP-PNP and a DNA duplex

While structures were available for select members of the RecQ family as well as for the isolated BLM HRDC domain (Bernstein *et al.*, 2003; Pike *et al.*, 2009; Kim & Choi, 2010; Kitano *et al.*, 2010; Sato *et al.*, 2010), information on larger fragments of human BLM that include the helicase, RQC and HRDC domains was lacking. To address this and to provide information on BLM in complex with DNA to help our understanding of the unwinding mechanism, we initiated

efforts to obtain crystals of human BLM. Crystallization screening focused on BLM constructs ($BLM_{636–1195}$, $BLM_{636–1298}$ or $BLM_{640–1298}$) in complex with ATP analogues (ADP or AMP-PNP) either with or without a DNA duplex. Interestingly, the longer constructs gave crystals both with and without a DNA duplex, while only micro-crystals were obtained with the shorter construct in the ternary-complex form. Optimization of the crystals of either $BLM_{636–1298}$ or $BLM_{640–1298}$ with ADP and a DNA duplex gave diffraction-quality crystals, while only low-resolution diffraction was obtained with the crystal grown in the absence of DNA. The structure of the BLM ternary complex with ADP and a DNA duplex was solved to 3.1 Å resolution by SeMet MAD, and molecular replacement was subsequently used to solve the 2.3 Å resolution structure.

### 3.4. Overall structure of the $BLM_{640–1298}$ ternary complex

The BLM ternary-complex structure with both a DNA duplex and bound nucleotide reveals a V-shaped molecule with the DNA near the point (Fig. 3 and Supplementary Fig. S3). The protein is roughly 47 × 68 × 79 Å in size and can be divided into three conserved elements that are characteristic of members of the RecQ DNA helicase family. These include the helicase core, the RQC domain [which includes $Zn^{2+}$-binding and winged-helix (WH) subdomains] and the HRDC domain. The helicase core domain can be divided into two subdomains. The first subdomain consists of residues 642–857 and is made up of a seven-stranded parallel $\beta$-sheet sandwiched between five $\alpha$-helices on the solvent-exposed side and four $\alpha$-helices on the ATP-binding face. The second subdomain includes residues 858–993 and is composed of a six-stranded parallel $\beta$-sheet sandwiched between two $\alpha$-helices on either side. This core helicase/ATPase domain is highly conserved within the RecQ helicase family and also in the RNA helicase superfamily. Downstream of the core helicase domain is the RQC domain (residues 994–1194) that consists of two conserved subdomains. The first subdomain contains two antiparallel $\alpha$-helices (994–1032) followed by a $Zn^{2+}$-binding motif (1033–1068) that coordinates the $Zn^{2+}$ ion *via* four cysteine residues that reside in short helical segments. The second RQC subdomain (residues 1069–1194) adopts a WH fold that includes the DNA strand-separating hairpin (1154–1173). The subdomain begins with a short antiparallel double-stranded $\beta$-sheet followed by a four-helix bundle with the central hairpin interacting with the fork of the 3′-overhang DNA. Residues 1195–1208, which are disordered in the structure, make an extended linker connecting the WH subdomain and the HRDC domain. Finally, the HRDC domain, consisting of residues 1209–1292, is composed of five $\alpha$-helices and a $3_{10}$-helical loop. This domain sits above the ATP-binding site in the cleft between the helicase core subdomains, where it makes extensive contacts with both.

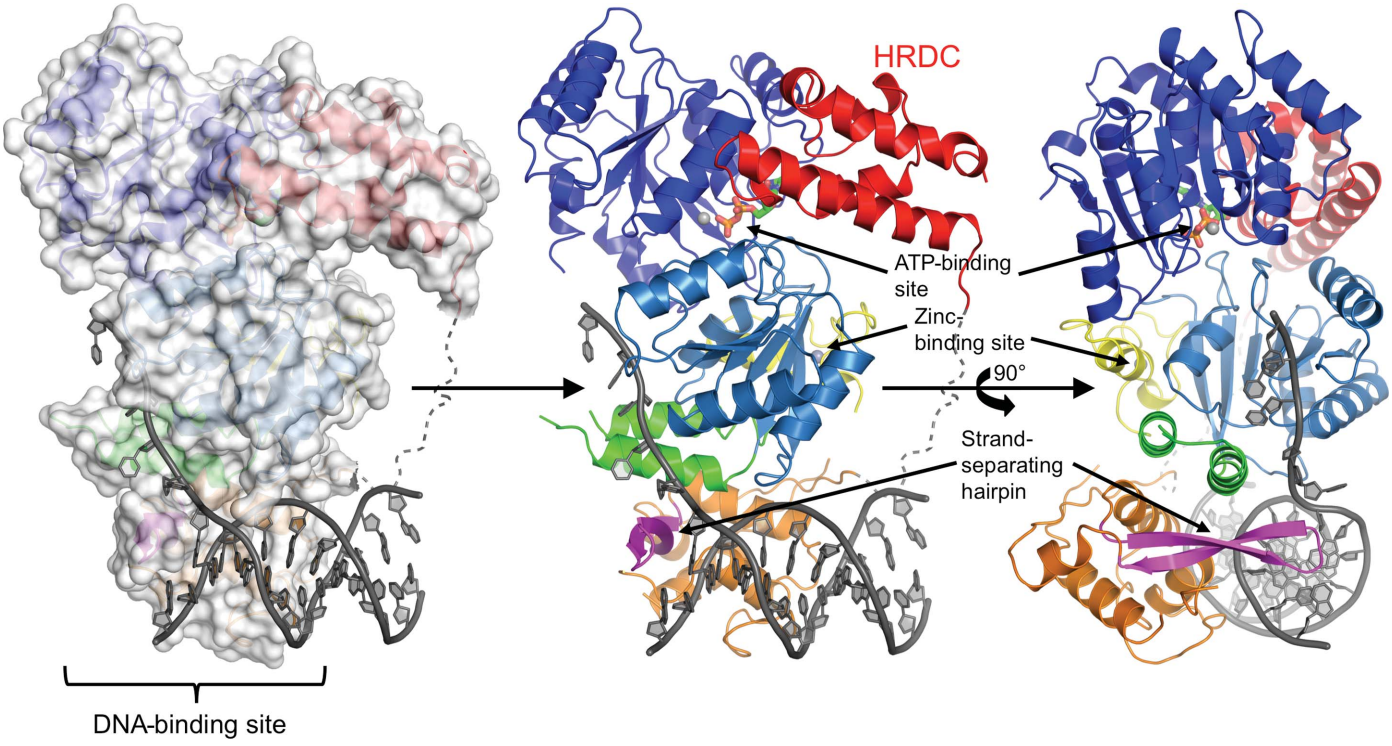


**Figure 3**
Representation of the ternary-complex structure of BLM helicase bound to a 3′-overhang DNA duplex and ADP. The protein is shown in cartoon form and is coloured blue and slate blue for the D1 and D2 subdomains of the helicase core; yellow and green for the zinc-binding motif and the antiparallel $\alpha$-helices region, respectively, which are both part of the $Zn^{2+}$-binding subdomain; orange for the WH subdomain, of which the strand-separating hairpin (shown in magenta) is a part; and red for the HRDC domain. The ADP is shown in stick form with green C atoms and the DNA in cartoon form with the phosphate backbone coloured dark grey. The coordinated calcium and zinc ions are shown as light grey spheres.

### 3.5. ADP is bound in the BLM active site

As shown in Fig. 4, the bound molecule of ADP is most closely associated with the first subdomain of the BLM helicase core. The active site contains the highly conserved glutamine residue (Gln672) that makes two hydrogen bonds to the adenine moiety of the nucleotide and is stabilized by a hydrogen bond to the Asn667 carbonyl O atom and stacking interactions with Leu665 and Arg669. In addition, the guanidine group of Arg669 hydrogen bonds to the ribose sugar. The P-loop ($G_{692}GGKS_{696}$) coordinates the phosphates of the nucleotide with a mixture of backbone and side-chain interactions. The catalytic residues from the DEVH motif, Asp795 and Glu796, then make a water-mediated interaction with the calcium ion in the magnesium-binding site and a pair of side-chain hydrogen bonds with the N atom of Ala831 and the $O^{\gamma}$ atom of Thr830, respectively. Interestingly, the side chain of Arg982 sits slightly above ADP in a position that could allow it to interact with the missing $\gamma$-phosphate. Thus, the structure supports the hypothesis presented by Ren and coworkers that Arg982 functions as an arginine finger to detect the presence of $\gamma$-phosphate and communicate between the two domains (Ren *et al.*, 2007).

### 3.6. Interactions between BLM and DNA duplex

The majority of the interactions of BLM with the 16/24-nucleotide 3′-overhang DNA are with the single-strand region of the template (Supplementary Fig. S4). However, the $\beta$-hairpin of the WH subdomain also makes a number of key interactions with the duplex region and presumably causes the separation of the first two bases. The conformation of the 3′-overhang DNA is predominantly in a standard B-form duplex except for the single-stranded regions, of which only four bases are visible from the template (15–18) and one base from the primer (2). Bases from both ends of the primer (1 and 16) and template (1–3 and 19–24) are disordered in the electron density. Interactions between the protein and the bound DNA are focused on the single-stranded template end, with no contacts observed past base 7 of the primer. The single-strand portion of the template sits in a cleft formed by the second helicase subdomain, the antiparallel $\alpha$-helices at the N-terminal end of the $Zn^{2+}$-binding subdomain and the hairpin of the WH subdomain. As expected with a nonsequence-specific helicase, the majority of the contacts are with the phosphate backbone and are either direct or water-mediated. The side chains of three residues, Thr946, Arg898 and Arg1000, make key interactions with the template phosphate backbone, while only the Thr1110 side chain interacts with the primer phosphate backbone. As mentioned above, the $\beta$-hairpin of the WH subdomain appears to play an important role in DNA strand separation based on its position in the structure (Fig. 5). Interestingly, BLM positions Asn1164 of the $\beta$-hairpin within hydrogen-bonding distance of the terminal base of the primer (G2) in a position that mimics the template base (C15) that has been flipped out of the helix, where it is exposed to solvent. It is worth noting that the loop of the $\beta$-hairpin only contains two residues, Ala1163 and Asn1164, and the loop geometry is stabilized by hydrogen bonds between the Asn1162 side chain and the backbone N atoms of Asn1164 and Gln1166. This suggests that the sequence of the entire loop is important since the adjacent residues help to stabilize the position of Asn1164. To better assess the importance of this residue, the $BLM_{636–1298}$ Asn1164Ala mutant was produced and evaluated in the helicase assay. As expected, the activity in the helicase assay is lower for the Asn1164Ala mutant, with a $V_{max}$ value that is approximately tenfold lower than that of $BLM_{636–1298}$ (Supplementary Fig. S5). Together,

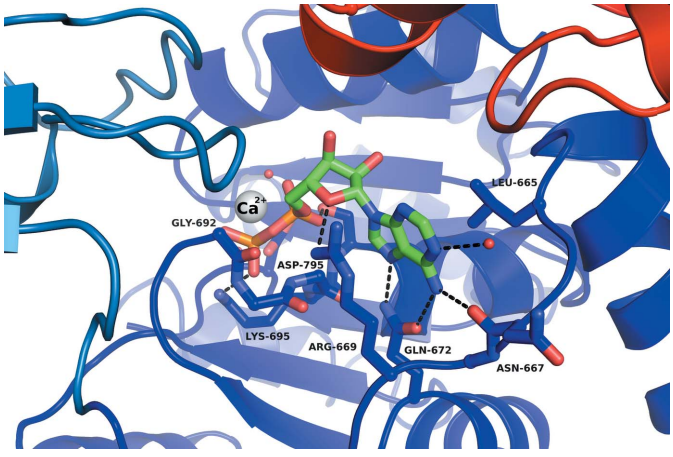


**Figure 4**
The active site of the BLM core helicase domain involved in ATP hydrolysis is shown with a bound molecule of ADP in stick form. Hydrogen bonds are displayed as dashed lines.

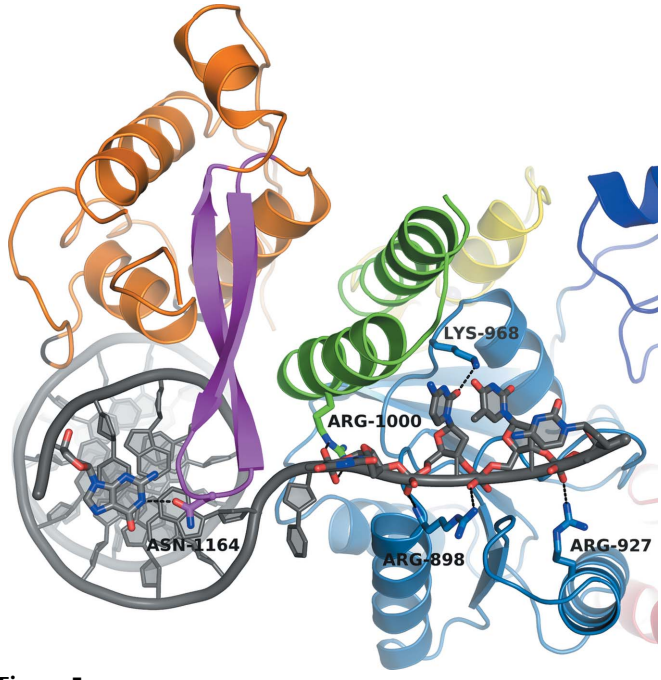


**Figure 5**
The interaction of the strand-separating hairpin of BLM with the terminus of the 3′-overhang DNA is shown. Asn1164 at the apex of the $\beta$-hairpin makes a hydrogen bond to a guanine base of the primer strand, pushing the complementary cytosine base out of the Watson–Crick base pair.

the structural and mutational data confirm that Asn1164 is a critical residue for the DNA strand separation by BLM.

Structures of two homologous RecQ family members have been solved in complex with DNA: human RecQ1 (PDB entry 2wwy; Pike *et al.*, 2009) and the DNA-binding domain of human WRN (PDB entry 3aaf; Kitano *et al.*, 2010). Interestingly, both use the $\beta$-hairpin to directly catalyze duplex strand separation. However, rather than using the Asn to mimic the templating base as in BLM, both RECQ1 and WRN wedge an aromatic residue from the $\beta$-hairpin between the last paired and first unpaired bases to disrupt base stacking (Pike *et al.*, 2009; Kitano *et al.*, 2010). In addition, mutational studies on the specific residues in both RECQ1 (Tyr564) and WRN (Phe1037) confirm the importance of these residues to the helicase activity.

### 3.7. Interaction between the HRDC and helicase domains

Unexpectedly, given the high level of affinity of the isolated RecQ HRDC domain for different DNA substrates (Huber *et al.*, 2006), the HRDC domain of BLM sits away from the DNA-binding region of the complex (Fig. 6). Based on the significant differences in $K_m$ values for $BLM_{636–1298}$ and $BLM_{636–1195}$ in the helicase assay with the forked duplex DNA, it seemed possible that the HRDC domain would make a direct interaction with the DNA. However, the structure reveals that the HRDC domain and DNA are ~28 Å apart and are situated on opposite sides of the D2 subdomain. Therefore, even when the disordered single-strand template bases are taken into account, it appears unfeasible that the HRDC domain and DNA interact directly. This result explains why Sato and coworkers failed to detect interactions between the BLM HRDC domain and DNA substrates using fluorescence polarization assays (Sato *et al.*, 2010). Although the HRDC domain does not interact with DNA, it makes numerous contacts with the D1 and D2 subdomains of the helicase core. There are ten hydrogen bonds and two salt bridges with the D1 subdomain and two hydrogen bonds and one salt bridge with the D2 subdomain. Surprisingly, there are no direct polar interactions between the D1 and D2 subdomains, so the bound nucleotide and the HRDC domain may play a role in coupling the subdomains during the enzymatic cycle. A bridging water molecule links the side chain of Asp1264 to N1 of adenine.

Wu and coworkers showed that truncation of the BLM HRDC domain inhibits the dissolution of double Holliday junctions (Wu *et al.*, 2005). The BLM constructs compared were $BLM_{213–1417}$ and $BLM_{213–1267}$ and the truncation removes the helical segments $\alpha$25 and $\eta$4 immediately after Asp1264. In our structure, this segment contains three polar interactions with the helicase subdomain D1: two salt bridges (Asp1269–Lys662 and Lys1273–Asp731) and one hydrogen bond (Tyr1274–Ser729). Wu and coworkers concluded that the inability of $BLM_{213–1267}$ to catalyze the dissolution of double Holliday junctions was likely to be owing to the absence of a functional HRDC domain. Our structure adds credence to this idea since it shows that truncation would remove critical interactions linking the D1 subdomain and the HRDC domain. Furthermore, Wu and coworkers produced and characterized the $BLM_{213–1417}$ mutation Lys1270Val based on a homology-modelling prediction that this residue would be on the protein surface, where it might make interactions with DNA (Wu *et al.*, 2005). Their results showed that the Lys1270Val mutation significantly reduces the ability of BLM to support double Holliday junction dissolution. In our structure Lys1270 is buried at the interface between the HRDC domain and the D1 helicase subdomain, where the hydrophobic portion of the side chain interacts with Leu730,

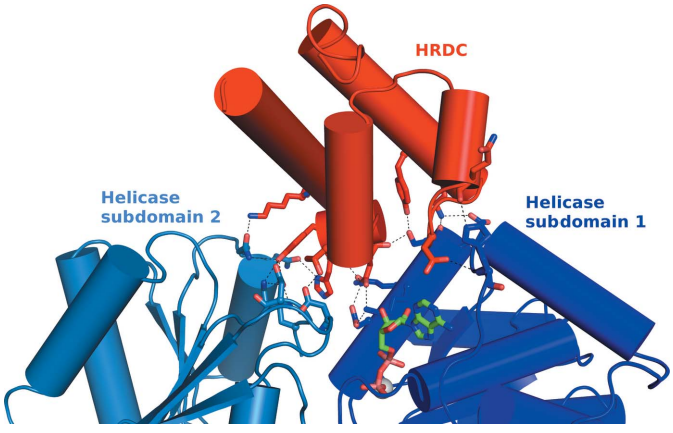


**Figure 6**
The interface between the HRDC domain and the D1 and D2 subdomains of the helicase core is shown with the key residues involved in hydrogen bonding or hydrophobic interactions shown in stick form.

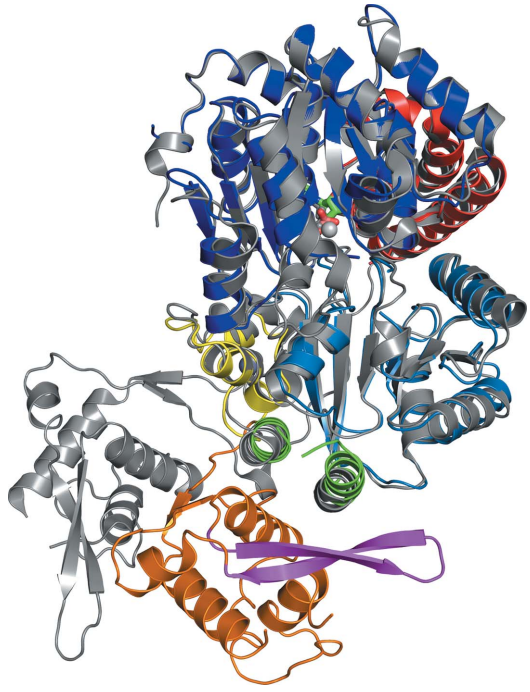

**Figure 7**
The superposition of the two BLM structures, the ternary and nanobody complexes, highlights the shift in position of the WH subdomain. BLM in the ternary complex is shown in the colour scheme described in Fig. 3 and BLM from the nanobody complex structure is shown in grey. DNA and the nanobody have been omitted for clarity.

**Table 3**
Bloom's syndrome mutations within the current structural construct.

| Mutation | Phenotype | Domain location | Structural effect | Reference |
|---|---|---|---|---|
| Gln672Arg | Bloom's syndrome | Helicase subdomain 1 | ATP-binding disruption | Ellis *et al.* (1995) |
| Ile841Thr | | | Helicase domain hydrophobic core destabilization | |
| Cys878Arg | | Helicase subdomain 2 | | |
| Gly891Glu | | | | German *et al.* (2007) |
| Cys901Tyr | | | | |
| Gly952Val | | | Possible disruption of ssDNA interaction | |
| His963Tyr | | | Helicase core destabilization | |
| Cys1036Phe | | RQC $Zn^{2+}$-binding motif | $Zn^{2+}$-binding disruption/domain destabilization | Foucault *et al.* (1997) |
| Cys1055Ser/Gly/Arg | | | | Ellis *et al.* (1995), German *et al.* (2007) |
| Asp1064Val | | | | German *et al.* (2007) |
| Cys1066Tyr | | | | |
| Pro690Leu | Non-Bloom associated total loss of function | Helicase subdomain 1 | P-loop disruption | Mirzaei & Schmidt (2012) |
| Arg717Thr | | | Possible disruption of ssDNA interaction | |
| Trp803Arg | | | | |
| Tyr811Cys | | | Helicase domain hydrophobic core destabilization | |
| Phe857Leu | | | Helicase subdomain hinge destabilization | |
| Gly972Val | | Helicase subdomain 2 | Disruption of downstream nucleotide interactions | |
| Arg791Cys | Non-Bloom associated partial loss of function | Helicase subdomain 1 | Helicase core destabilization | |
| Pro868Leu | | Helicase subdomain 2 | Disruption of downstream dsDNA interaction | |
| Gly1120Arg | | RQC winged-helix subdomain | Possible disruption of downstream dsDNA interaction | |

Ile1240, Phe1241, Thr1267 and Tyr1274 (Supplementary Fig. S6). $N^{\zeta}$ of Lys1270 then interacts with the carbonyl O atom of Ile1240 and two bridging water molecules. Surprisingly, $N^{\zeta}$ of Lys1270 is located ∼8 Å from the ribose O2′ and helps to make up the surface of the nucleotide cavity. Presumably, a valine substitution of Lys1270 would be unable to mimic the specific interactions and the result would be a destabilization of the interface between the HRDC domain and the D1 subdomain.

Removal of the HRDC domain hampers the unwinding of the forked duplex by reducing the catalytic turnover and increasing the $K_m$ for DNA. However, it also enhances the ATPase activity by increasing the catalytic rate without significantly modifying the $K_m$ for DNA. Effectively, this means that the absence of the HRDC domain leads to an increase in futile ATP hydrolysis. A possible explanation is that the HRDC domain couples the ATPase and helicase activities through its interactions with the helicase subdomains. We hypothesize that the steric presence of the HRDC domain and the associated interactions ought to limit the motions of the D1 and D2 subdomains to those that are required to strongly couple ATP hydrolysis with DNA unwinding. In contrast, when the HRDC domain is absent ATP hydrolysis occurs more easily owing to the increased flexibility between the subdomains. However, the coupling with DNA unwinding would be less efficient and would thus explain the lower enzyme activity in the helicase assay.

### 3.8. Comparison of BLM structures with known homologues

A comparison of our ternary BLM structure and the BLM structure without DNA (PDB entry 4cdg) recently deposited by Gileadi and coworkers (J. A. Newman, P. Savitsky, C. K. Allerston, A. C. W. Pike, E. Pardon, J. Steyaert, C. H. Arrowsmith, F. Von Delft, C. Bountra, A. Edwards & O. Gileadi, unpublished work) highlights the mobility of the WH subdomain in the absence of DNA. The Gileadi group used a nanobody to help form crystallization contacts and, one supposes, to overcome difficulties relating to the flexibility of BLM protein. Nanobodies provide good crystallization 'tools' owing to their ability to bind antigens with one single variable domain (Muyldermans, 2001). A comparison of the two BLM structures reveals significant differences in the position of the WH subdomain (Fig. 7). In our structure, the WH subdomain packs against the first two helices of the $Zn^{2+}$ subdomains ($\alpha$14 and $\alpha$15), making multiple interactions, while in the 4cdg structure it extends roughly 28 Å away from the rest of the protein and interacts exclusively with the nanobody. It therefore appears that the nanobody traps BLM in an open conformation that exists in solution prior to DNA binding. BLM is not the first RecQ family member to be crystallized in an open conformation: the *E. coli* RecQ structure (PDB entry 1oyy; Bernstein *et al.*, 2003) also places the WH subdomain in a position that is dramatically different from that of BLM in the ternary complex. Superimposition of the *E. coli* RecQ structure onto BLM in the ternary complex shows a large shift in the position of the WH subdomain that further illustrates the flexibility between the domains when not bound to DNA (Supplementary Fig. S7). Interestingly, the two available RecQ1 structures, both with and without DNA, position the WH subdomain in a similar manner to that of the ternary BLM structure. This suggests that multiple open conformations can exist within the RecQ family but there is only one closed conformation that is compatible with DNA binding.

Comparisons of the structures of the BLM HRDC domain and other homologous domains from *E. coli* RecQ, *Deinococcus radiodurans* RecQ and human WRN have been described elsewhere (Kim & Choi, 2010). They illustrate the high structural homology between the domains of different species or protein family members in spite of their sequence diversity. It was postulated that these differences determined the DNA substrate specificity of the different helicases and

showed a wide range of ssDNA-binding affinities, with the BLM HRDC domain exhibiting either very low or no binding. However, the structure solved here suggests that the HRDC domain plays a key role in the mechanism coupling the ATP and helicase activities rather than being directly involved in DNA binding. It is clear that several residues initially proposed as involved in ssDNA binding (Kim & Choi, 2010) are in fact positioned at the interface between the HRDC and helicase domains.

### 3.9. Mapping of Bloom's syndrome mutations – implications for function/interaction

BS is caused by various disease-linked mutations in the BLM gene and the most frequent mutations generate frameshifts or are point mutations that severely truncate the protein. However, a small subset of missense mutations have been identified from BS patients that include seven in the helicase core and six in the $Zn^{2+}$-binding subdomain (Table 3). From this subset, five have been characterized *in vitro* (Gln672Arg, Ile841Thr, Cys878Arg, Gly891Glu and Cys901Tyr) and were found to impair the ability of BLM to hydrolyze ATP or bind DNA (Guo *et al.*, 2007). Over the years, efforts have been made to explain the mutations using homology models. Now that the BLM ternary structure is available, it is worthwhile revisiting these mutations and contemplating what effects they might have on the protein stability and enzyme activity. Fig. 8 indicates where these residues are located in the BLM structure. There are two mutations in the first helicase subdomain (D1), the conserved glutamine residue (Gln672) that positions the adenine of the nucleotide and Ile841 that together with other hydrophobic residues helps to form the core of D1. Five of the mutated residues lie within the second subdomain of the helicase (D2): these are Cys878, Gly891, Cys901, Gly952 and His963. Structural considerations would lead us to expect that the mutations of Cys878, Gly891 and Cys901 to amino acids with larger side chains would cause steric clashes with the surrounding residues. The impact of mutating Gly952 is more difficult to assess owing to the lack of electron density for the segment containing this residue in our structure. However, the close proximity of this segment to the single-stranded DNA suggests that mutation may disrupt its ability to interact with DNA. His963 is located at the C-terminal end of $\beta$12, where it makes direct and water-mediated polar interactions with surrounding residues. Mutation of this residue could therefore lead to a steric clash and destroy the hydrogen-bonding network, leading to destabilization of the D2 subdomain. The six missense mutations of the $Zn^{2+}$-binding subdomain involve four residues (Cys1036, Cys1055, Asp1064 and Cys1066) that presumably play a role in stabilizing the subdomain. This includes the three cysteines that are direct ligands of the zinc ion and the aspartate that hydrogen bonds to the backbone N atom of Val866 and makes a salt bridge with Arg1037.

There are also several mutations that are not associated with BS but result in either a total or a partial loss of function (Mirzaei & Schmidt, 2012). Those that completely ablate the function of BLM are found exclusively in the core helicase domain, where the majority of them are found at the interface between the two subdomains (Table 3 and Fig. 7). In particular, Arg717 and Trp803 are located in close proximity to the visible end of the single-stranded DNA, where they may interact with the unwound DNA as it passes over the enzyme. Interestingly, two of the mutated residues (Pro690 and Phe857) make a hydrophobic stacking interaction that may be related to the enzymatic mechanism. This is because Pro690 is one of the loop residues associated with the ATP phosphate groups and Phe857 is positioned within the hinge region between the two subdomains. The mutations that only result in a partial loss of function are found in both helicase subdomains and the WH subdomain of the RQC domain. Interestingly, two of these mutations, of Pro868 and Gly1120, are likely to result in the destabilization of dsDNA interaction, which may be a good indication of the less critical nature of the dsDNA interactions when compared with those involved in contacts with the ssDNA.

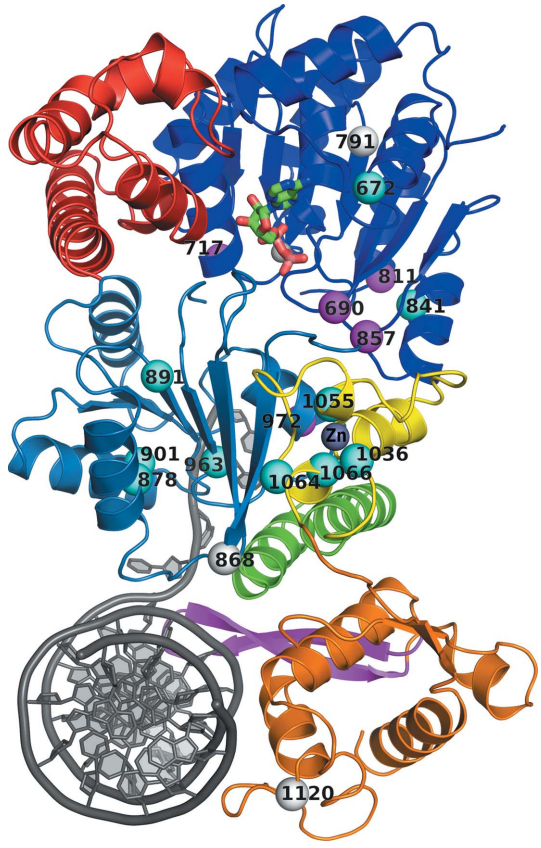


**Figure 8**
The positions of Bloom's syndrome-associated mutations are shown mapped onto the ternary-complex structure as cyan spheres and include Gln672Arg, Ile841Thr, Cys878Arg, Gly891Glu, Cys901Tyr, Gly952Val, His963Tyr, Cys1036Phe, Cys1055Ser, Cys1055Gly, Cys1055Arg, Asp1064Val and Cys1066Tyr. Mutations not associated with Bloom's syndrome but that confer a total loss of function on the protein are shown as magenta spheres and include Pro690Leu, Arg717Thr, Trp803Arg, Tyr811Cys, Phe857Leu and Gly972Val. Those that result in a partial loss of function are illustrated with white spheres and include Arg791Cys, Pro868Leu and Gly1120Arg (Mirzaei & Schmidt, 2012).

## 4. Concluding remarks

We have successfully solved the ternary structure of human BLM with ADP and a 3′-overhang DNA duplex. The structure reveals how BLM binds the single-strand portion of the DNA and uses a $\beta$-hairpin from the WH domain to separate the first bases of the duplex region. The structural results correlate with a mechanism whereby ATP hydrolysis drives a motor that pulls on the single-stranded DNA and causes the $\beta$-hairpin to wedge apart the DNA duplex. The structure provides the first experimental information on how the HRDC domain interacts with the D1 and D2 subdomains of the helicase core. In addition, the enzymatic characterization of BLM constructs with and without the HRDC domain shows that in the helicase assay the presence of the HRDC domain promotes enzyme turnover and provides an increase in the affinity for DNA. However, in the ATPase assay the presence of the HRDC domain lowers the enzyme turnover while having little or no effect on the affinity for DNA. Together, the enzymatic and structural results suggest that the HRDC domain couples the two activities together in an efficient way so that each ATP hydrolysis cycle also unwinds the DNA. Presumably this is performed by tethering the movements of the D1 and D2 helicase subdomains so that the enzyme is only capable of performing the movements that transfer a maximum of energy for DNA unwinding.

The authors would like to thank Clemens Vonrhein and Gerard Bricogne of Global Phasing for help with the initial phasing of the MAD data and Emilia Danilowicz-Luebert of NanoTemper Technologies for help with the MST data.

## References


Abrahams, J. P. & Leslie, A. G. W. (1996). *Acta Cryst.* D**52**, 30–42.

Bachrati, C. Z., Borts, R. H. & Hickson, I. D. (2006). *Nucleic Acids Res.* **34**, 2269–2279.

Bennett, R. J. & Keck, J. L. (2004). *Crit. Rev. Biochem. Mol. Biol.* **39**, 79–97.

Bernstein, D. A. & Keck, J. L. (2003). *Nucleic Acids Res.* **31**, 2778–2785.

Bernstein, D. A., Zittel, M. C. & Keck, J. L. (2003). *EMBO J.* **22**, 4910–4921.

Chaganti, R. S., Schonberg, S. & German, J. (1974). *Proc. Natl Acad. Sci. USA*, **71**, 4508–4512.

Chan, K. L., North, P. S. & Hickson, I. D. (2007). *EMBO J.* **26**, 3397–3409.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.

Chu, W. K. & Hickson, I. D. (2009). *Nature Rev. Cancer*, **9**, 644–654.

Davies, S. L., North, P. S. & Hickson, I. D. (2007). *Nature Struct. Mol. Biol.* **14**, 677–679.

Durant, S. T. (2012). *J. Cancer*, **3**, 67–82.

Ellis, N. A., Groden, J., Ye, T.-Z., Straughen, J., Lennon, D. J., Ciocci, S., Proytcheva, M. & German, J. (1995). *Cell*, **83**, 655–666.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.

Foucault, F., Vaury, C., Barakat, A., Thibout, D., Planchon, P., Jaulin, C., Praz, F. & Amor-Guéret, M. (1997). *Hum. Mol. Genet.* **6**, 1427–1434.

German, J. (1997). *Cancer Genet. Cytogenet.* **93**, 100–106.

German, J., Archibald, R. & Bloom, D. (1965). *Science*, **148**, 506–507.

German, J., Sanz, M. M., Ciocci, S., Ye, T.-Z. & Ellis, N. A. (2007). *Hum. Mutat.* **28**, 743–753.

Guo, R.-B., Rigolet, P., Ren, H., Zhang, B., Zhang, X. D., Dou, S.-X., Wang, P.-Y., Amor-Gueret, M. & Xi, X.-G. (2007). *Nucleic Acids Res.* **35**, 6297–6310.

Guo, R.-B., Rigolet, P., Zargarian, L., Fermandjian, S. & Xi, X.-G. (2005). *Nucleic Acids Res.* **33**, 3109–3124.

Huber, M. D., Duquette, M. L., Shiels, J. C. & Maizels, N. (2006). *J. Mol. Biol.* **358**, 1071–1080.

Imamura, O., Fujita, K., Shimamoto, A., Tanabe, H., Takeda, S., Furuichi, Y. & Matsumoto, T. (2001). *Oncogene*, **20**, 1143–1151.

Janscak, P., Garcia, P. L., Hamburger, F., Makuta, Y., Shiraishi, K., Imai, Y., Ikeda, H. & Bickle, T. A. (2003). *J. Mol. Biol.* **330**, 29–42.

Karow, J. K., Chakraverty, R. K. & Hickson, I. D. (1997). *J. Biol. Chem.* **272**, 30611–30614.

Karow, J. K., Constantinou, A., Li, J.-L., West, S. C. & Hickson, I. D. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 6504–6508.

Kim, S.-Y., Hakoshima, T. & Kitano, K. (2013). *Sci. Rep.* **3**, 3294.

Kim, Y. M. & Choi, B.-S. (2010). *Nucleic Acids Res.* **38**, 7764–7777.

Kitano, K., Kim, S.-Y. & Hakoshima, T. (2010). *Structure*, **18**, 177–187.

Mirzaei, H. & Schmidt, K. H. (2012). *Proc. Natl Acad. Sci. USA*, **109**, 19357–19362.

Mohaghegh, P., Karow, J. K. Jr, Brosh, R. M., Bohr, V. A. & Hickson, I. D. (2001). *Nucleic Acids Res.* **29**, 2843–2849.

Morozov, V., Mushegian, A. R., Koonin, E. V. & Bork, P. (1997). *Trends Biochem. Sci.* **22**, 417–418.

Muyldermans, S. (2001). *Rev. Mol. Biotechnol.* **74**, 277–302.

Neff, N. F., Ellis, N. A., Ye, T.-Z., Noonan, J., Huang, K., Sanz, M. & Proytcheva, M. (1999). *Mol. Biol. Cell*, **10**, 665–676.

Nguyen, G. H. *et al.* (2013). *Chem. Biol.* **20**, 55–62.

Pike, A. C. W., Shrestha, B., Popuri, V., Burgess-Brown, N., Muzzolini, L., Costantini, S., Vindigni, A. & Gileadi, O. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 1039–1044.

Ren, H., Dou, S.-X., Rigolet, P., Yang, Y., Wang, P.-Y., Amor-Gueret, M. & Xi, X. G. (2007). *Nucleic Acids Res.* **35**, 6029–6041.

Rezazadeh, S. (2013). *Mol. Biol. Rep.* **40**, 3049–3064.

Rosenthal, A. S., Dexheimer, T. S., Gileadi, O., Nguyen, G. H., Chu, W. K., Hickson, I. D., Jadhav, A., Simeonov, A. & Maloney, D. J. (2013). *Bioorg. Med. Chem. Lett.* **23**, 5660–5666.

Sato, A., Mishima, M., Nagai, A., Kim, S.-Y., Ito, Y., Hakoshima, T., Jee, J.-G. & Kitano, K. (2010). *J. Biochem.* **148**, 517–525.

Sun, H., Karow, J. K., Hickson, I. D. & Maizels, N. (1998). *J. Biol. Chem.* **273**, 27587–27592.

Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. (2007). *Methods Mol. Biol.* **364**, 215–230.

Wu, L., Chan, K. L., Ralf, C., Bernstein, D. A., Garcia, P. L., Bohr, V. A., Vindigni, A., Janscak, P., Keck, J. L. & Hickson, I. D. (2005). *EMBO J.* **24**, 2679–2687.

Wu, L. & Hickson, I. D. (2003). *Nature (London)*, **426**, 870–874.