



Published in final edited form as:

Nat Commun. ; 5: 5556. doi:10.1038/ncomms6556.

Mechanistic Insight into the Interaction of BLM helicase with Intra-strand G-quadruplex Structures

Sujoy Chatterjee^{1,†}, Jennifer Zagelbaum^{1,†}, Pavel Savitsky², Andreas Sturzenegger⁵, Diana Huttner^{3,4}, Pavel Janscak⁵, Ian D Hickson³, Opher Gilaedi², and Eli Rothenberg^{1,*}

¹Department of Biochemistry and Molecular Pharmacology, New York University School of Medicine, NY 10016, NY, USA ²Genome Integrity group, Structural Genomics Consortium, University of Oxford, Oxford, UK ³NovoNordisk Foundation Center for Protein Research, University of Copenhagen, 2200 Copenhagen, Denmark ⁴Department of Cellular and Molecular Medicine, University of Copenhagen, 2200 Copenhagen, Denmark ⁵Institute of Molecular Cancer Research, University of Zurich, CH-8057 Zurich, Switzerland

Abstract

Bloom syndrome is an autosomal recessive disorder caused by mutations in the RecQ family helicase BLM that is associated with growth retardation and predisposition to cancer. BLM helicase has a high specificity for non-canonical G-quadruplex (G4) DNA structures, which are formed by G-rich DNA strands and play an important role in the maintenance of genomic integrity. Here, we used single-molecule FRET to define the mechanism of interaction of BLM helicase with intra-stranded G4 structures. We show that the activity of BLM is substrate dependent, and highly regulated by a short ssDNA segment that separates the G4 motif from dsDNA. We demonstrate cooperativity between the RQC and HRDC domains of BLM during binding and unfolding of the G4 structure, where the RQC domain interaction with G4 is stabilized by HRDC binding to ssDNA. We present a model that proposes a unique role for G4 structures in modulating the activity of DNA processing enzymes.

Introduction

G-quadruplexes (G4) are noncanonical structures formed by Hoogsteen base-pairing in guanine-rich DNA sequences^{1,2}. There are more than 375,000 predicted G4 motifs in the human genome and more than 1,400 G4 motifs have been predicted in *S. cerevisiae*^{3,4}. G4-forming sequences are found in telomeres and regulatory regions of the genome such as gene promoter regions, and can spontaneously form when ssDNA is generated during DNA

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

Corresponding author: Eli Rothenberg, (eli.rothenberg@nyumc.org).

[†]Equal contribution

Author contributions: S.C., E.R. and J.Z. designed the study. P.S., D.H., A.S., P.J., I.D.H and O.G. generated reagents. S.C., J.Z., and E.R. performed experiments. S.C. and J.Z. analyzed the data. E.R. and S.C. wrote the manuscript, which was reviewed and approved by all co-authors.

Competing financial interests: The authors declare no competing financial interests.

metabolic processes^{5,6}. The formation of G4 structures poses challenges to progression through replication, transcription and repair, and various ssDNA binding proteins and helicases are utilized for removal of G4 structures during these processes^{7–13}. Specifically, it was previously shown that WRN, BLM, Pif1 and FANCD1 helicases unwind G4 structures preferentially over duplex DNA *in vitro*^{7–9,12,14–16}. Mutations in these helicases result in genomic instability, suggesting that their role in processing of G4 structures is central for maintaining genome integrity^{17,18}.

BLM and WRN belong to a highly conserved RecQ 3'–5' helicase family implicated in DNA damage repair and the maintenance of genomic integrity¹⁹. A characteristic feature of RecQ helicases is their ability to recognize and unwind a wide variety of DNA structures. The diverse functionalities and often weak unwinding activity of RecQ family member proteins in a standard DNA helicase unwinding assay, together with their intrinsic strand-annealing activity, suggest a complex role for these enzymes that goes beyond a simple dsDNA unwinding function⁹. In particular, BLM and WRN have been suggested to play a role in the metabolism of G4 structures during chromosomal and telomeric replication, as well as in transcription^{20–29}. Previous studies of the interaction of BLM with DNA substrates containing G4 structures have shown that BLM binds G4 structures with high specificity, but at the same time G4 structures can hinder BLM ssDNA translocation activity and decrease its dsDNA unwinding efficiency^{30,31}.

Here we define the interaction of BLM with G4 DNA through the use of single-molecule fluorescence resonance energy transfer (smFRET) analysis^{32,33}. Single-molecule methods are powerful approaches capable of resolving the conformational states and dynamics of DNA and proteins in real-time that are otherwise masked by ensemble averaging. Recent single-molecule studies have provided important insights into the mechanisms of RecQ helicases with their DNA substrates and interaction of ssDNA binding proteins with G4 structures^{34–37}. We used smFRET assays to monitor the conformational states of a human intramolecular G4 structure in the presence of BLM, and to analyze BLM unwinding of dsDNA located downstream of G4. Our data reveal that in substrates where a G4 motif is immediately adjacent to duplex DNA, BLM does not unfold the G4 motif, and is consequently unable to unwind the duplex DNA. On the other hand, introduction of a short ssDNA between the G4 motif and the duplex region restores BLM's activity, enabling ATP-independent unfolding of the G4 structure and permitting helicase-mediated unwinding of the downstream duplex. Analysis of G4 unfolding using BLM truncation fragments suggests a specific binding configuration where unfolding of the G4 motif by the RQC domain of BLM is facilitated by binding of the HRDC domain of BLM to the ssDNA between a G4 and a duplex region. Our experiments also reveal the existence of a different BLM mediated and ATP-dependent G4 unfolding pathway in which unfolding is assisted by external ssDNA that is homologous to the G4 sequence. These findings define a novel mechanism of interaction between BLM and G4 structures whereby the association between BLM and the G4 structure is DNA substrate selective, and results in regulation of BLM unwinding activity via the G4 folding conformation. Based on our data, we propose that intramolecular G4 sequences confer an additional level of regulation through DNA substrate selectivity and modulation of the activity of DNA metabolic enzymes (e.g. helicases) in their vicinity.

These sequences consequently play a pivotal role in key biological process such as DNA replication, transcription and repair.

Results

DNA Substrate Specificity of BLM in of G4 Unfolding

To probe the binding of BLM to G4 structures, we used a partial DNA duplex FRET substrates having a 3'-tail containing G4-forming human telomere or c-kit2 sequences ($G_3(T_2AG_3)_3$ and $(CG_3)_2CGCG(AG_3)_2G$ respectively, Supplementary Table 1, sequences 2 – 4 and 12 – 15). In these FRET substrates (illustrated in Fig. 1a–c and Supplementary Fig. 1a), the donor and acceptor dyes are placed at either end of the G4 forming sequence such that a folded G4 structure results in high FRET, while G4 unfolding would yield a low FRET signal. This arrangement enabled us to directly monitor G4 conformational changes. Previous studies have shown that monovalent cations have a pronounced effect on the stability of G4 structures; among these, K^+ has the highest stabilizing effect³⁸. To establish a baseline for K^+ dependent G4 folding in our DNA substrates, we carried out a titration of our surface-tethered FRET probe with K^+ and derived the population percentage of folded G4 structures from the resulting FRET histograms (Supplementary Fig. 1b). At low K^+ ion concentration, the G4 structure is predominantly unfolded (Supplementary Fig. 1c), whereas at 50 mM K^+ the G4 structure is predominantly in a folded conformation³⁸.

To investigate the interaction of BLM with G4 DNA, surface tethered G4 FRET substrates were incubated with 50 nM BLM in the presence of 50 mM K^+ . We used three different partial duplex DNA substrates, as illustrated in Fig. 1a–c and Supplementary Table 1, with the G4 motif either immediately adjacent to the duplex region, or separated by either 5 nt or 10 nt (substrates termed G4, G4 (5 nt) and G4 (10 nt), respectively). The histograms for these substrates before and after addition of 50 nM BLM are shown in Fig. 1a–c for telomeric G4 (upper and lower panels, respectively) and in Supplementary Fig. 2a–d for c-kit G4. Substrate G4 showed very little change in G4 folding in the presence of 50 nM BLM (Fig. 1a and Supplementary Fig. 2a), while substrate G4 (5 nt) displayed a change in FRET population upon addition of BLM (Fig. 1b). Substrate G4 (10 nt) showed a substantial change in FRET distribution upon addition of BLM, consistent with an efficient and ATP-independent unfolding of the G4 structure by BLM (Fig. 1c and Supplementary Fig. 2b). The unfolding of G4 structures in these substrates was found to be independent of the sequence of the ssDNA spacer used, displaying comparable unfolding for poly-dT and a mixed ssDNA spacer (Supplementary table 1 and Supplementary Fig. 3a–b). To further establish that the observed unfolding of G4 structures by BLM is ATP independent, we tested the G4 unfolding activity of the helicase-dead BLM mutant (K695M)³⁹. Similar to BLM, the helicase-dead BLM mutant showed no unfolding activity for the G4 substrate, but robust unfolding of the G4 (10 nt) substrate (Supplementary Fig. 4a–b). These findings demonstrate a unique and ATP independent BLM - G4 substrate interaction, where BLM can efficiently unfold the G4 structure only when a ssDNA segment is present between the G4 motif and the duplex region.

To determine the effect of BLM binding in these substrates we examined the temporal behavior of individual trajectories for each DNA substrate. Fig. 1d–e displays representative

smFRET trajectories from the G4 and G4 (10 nt) substrates, obtained in the presence of 50 nM BLM, where the top panels show the donor (green) and acceptor (red) intensities and the bottom panel the resulting FRET efficiencies. Trajectories from the G4 substrate (Fig. 1d) showed a persistent signal indicative of a stable G4 folded conformation, whereas trajectories from both the G4 (5 nt) and G4 (10 nt) substrates (Fig. 1e and Supplementary Fig. 5) showed rapid fluctuations, indicative of G4 transitioning between folded and unfolded conformations. These repetitive unfolding-refolding transitions suggest that G4 unfolding by BLM is meta-stable such that G4 is capable of returning to its folded conformation. To determine the effect of BLM on the different substrates, we quantified the population percentage of trajectories showing fluctuations, evaluating a minimum of 300 molecules used for each substrate (Fig. 1f). This quantification demonstrated a clear trend in which the percentage of trajectories exhibiting fluctuations increased with increasing distance of the G4 motif from the duplex region.

A more detailed analysis of the fluctuating trajectories revealed that G4 unfolding and refolding does not switch directly from the folded to unfolded conformations (high-FRET to low-FRET), but occurs via intermediate states in a step-wise manner. To map the specific BLM-G4 unfolding-refolding pathways, we conducted Hidden Markov Model (HMM) analysis on the smFRET trajectories we obtained. HMM analysis is a powerful tool for analysis of single-event trajectories, and provides important characteristics of the specific states and corresponding transition probabilities⁴⁰. The resulting HMM fit to the G4 (10 nt) trajectory was superimposed on the smFRET trajectory, and showed both a good fit and well defined states (Fig 1e-bottom, Cyan trace). This analysis enabled us to extract the frequency and transition rates for each FRET state and to construct a two-dimensional transition density plot (TDP). Fig. 1g-i shows the TDP plots for the G4, G4 (5 nt) and G4 (10 nt) substrates, respectively, where the y-axis represents the initial FRET value before the transition and the x-axis is the final FRET value after the transition. As expected, the TDP for the G4 substrate (Fig. 1g) shows only two states, both at high FRET values, corresponding to the anti-parallel and parallel conformations of the G4 structures, which are the two main folded conformations of G4. The TDP for the G4 (5 nt) and G4 (10 nt) substrates shows a number of transitions starting from high FRET values down to intermediate and low FRET states, which correspond to transitions between G4 folded and unfolded states and vice-versa. Here the transitions residing below the diagonal of the TDP represent G4 refolding transitions (from lower FRET to higher FRET), whereas the transitions above the diagonal represent unfolding of the G4 structure (from higher FRET to lower FRET). The observed transition and spacing between the FRET states indicates that the G4 unfolding and refolding pathway occurs via small sequential sub-steps. This is consistent with a model in which unfolding and refolding of the four strands that participate in the G4 structure occurs one strand at a time, as illustrated in the cartoon in Fig. 1k. We note that a similar observation of sequential stepwise unfolding of G4 has recently been reported for POT1-TPP1⁴¹.

A comparison of the mean unfolding and refolding rates for the G4 (5 nt) and G4 (10 nt) substrates revealed that refolding rates for both substrates are higher than their unfolding rates (see Supplementary Methods for rate calculation). However, the refolding rates for the G4 (5 nt) substrate were higher than the G4 (10 nt) substrate ($2.59 \pm 0.57 \text{ sec}^{-1}$ and 1.96 ± 0.3

sec^{-1} , respectively). We interpret this result to mean that BLM's interaction and unfolding of the G4 motif in the G4 (10 nt) substrate is more stable than in the case of the G4 (5 nt) motif. This conclusion is in agreement with the overall increase in G4 unfolding efficiency in the case of the G4 (10 nt) substrate, as seen in the FRET histograms (Fig. 1b and c) and in the corresponding population percentage of trajectories exhibiting fluctuations (Fig. 1f). We ascribe the improved stability of BLM G4 binding and unfolding efficiency in the G4 (10 nt) substrate to the placement of a longer ssDNA segment between the G4 motif and dsDNA.

G4 unfolding is Mediated by BLM's RQC and HRDC domains

We sought to define the specific domains of BLM that participate in G4 unfolding. The full length BLM is a 1417 aa protein, whose functional core (core BLM) is located between amino acids (aa) 640 and 1298, and consists of three distinct domains: the helicase core, the RecQ C-terminal (RQC) domain, and the helicase and RNaseD C-terminal (HRDC) domain⁴². While the helicase domain contains the catalytic ATP and DNA binding activities, the RQC domain and HRDC domains have been identified as being important for binding non-canonical DNA substrates including G4^{39,42-44}. We used truncated protein fragments containing the RQC and HRDC domains (RQC-HRDC: 858–1298aa) or HRDC domain only (HRDC: 1069–1298aa) and compared their effect on the conformations of the G4 structure in the G4 (10 nt) substrate (Fig. 2a). Fig. 2b shows the resulting smFRET histograms for the G4 (10 nt) substrate in the presence of 50 nM HRDC or 50 nM RQC-HRDC (mid and bottom panels, respectively). Both fragments resulted in altered smFRET histograms, as compared to the G4 (10 nt) DNA only histogram (top panel). However, whereas the HRDC fragment resulted in a change in FRET populations within the G4 folded conformation, the RQC-HRDC fragment resulted in a broader distribution corresponding to unfolded conformations of G4. Analysis of individual smFRET trajectories for each of the fragments revealed that the RQC-HRDC fragment resulted in trajectories undergoing rapid fluctuations (Fig. 2c) similar to the dynamics observed for BLM, indicative of unfolding and refolding dynamics, whereas the HRDC fragment displayed a uniform and stable signal corresponding to a folded G4 conformation. Fig. 2d shows a quantification of the population percentage of fluctuating trajectories for each fragment, where both BLM and RQC-HRDC showed a considerable number of fluctuating trajectories, while HRDC did not (Supplementary Fig. 6a–b). The characteristic behavior observed for trajectories with the RQC-HRDC fragment revealed transitions between discrete steps, similar to those seen with BLM. We applied HMM analysis to obtain the states and corresponding rates for the observed transitions, and determined how these compared with the metrics obtained for BLM. The resulting TDP for the G4 (10 nt) substrate in the presence of 50 nM RQC-HRDC shows discrete transitions corresponding to sequential sub-steps of unfolding and refolding of G4 (Fig. 2e). We note that these transitions are comparable to the transitions observed for the same substrate in the presence of BLM. We next quantified the G4 unfolding and refolding rates for RQC-HRDC, and compared them with the rates obtained for BLM (Fig. 2f). While the G4 refolding rates for both BLM and RQC-HRDC were found to be similar ($1.96 \pm 0.3 \text{ sec}^{-1}$ and $2.08 \pm 0.1 \text{ sec}^{-1}$, respectively), the unfolding rate of RQC-HRDC was higher than the unfolding rate of BLM ($2.12 \pm 0.4 \text{ sec}^{-1}$ and $1.5 \pm 0.01 \text{ sec}^{-1}$, respectively), possibly due to an enhanced overall stability of BLM binding to DNA, as compared to RQC-HRDC⁴². We infer from the observation that both HRDC and RQC-HRDC bind the

G4 substrate, while only RQC-HRDC can efficiently unfold it, that the HRDC domain assists in stabilizing the RQC domain interaction with G4, possibly by binding to the ssDNA segment downstream of G4.

To further test the hypothesis that the observed ATP-independent unfolding of G4 is unique and mediated by RQC-HRDC domains, we tested whether G4 disruption in these substrates can be carried out by other DNA helicases. We used two RecQ family helicases, WRN and RecQL5: WRN has RQC and HRDC domains similar to BLM, while RecQL5 has an RQC domain but no HRDC domain²⁷. As with BLM, WRN was able to unfold the G4 structure in the G4 (10 nt) substrate, but not in the G4 substrate alone (Supplementary Fig. 7a–b); on the other hand, RecQL5 did not display any G4 unfolding activity with either of these substrates (Supplementary Fig. 7c). We also tested the Pif1 5′-3′ helicase, previously shown to have robust G4 unwinding activity and to be implicated in the metabolism of G4 structures^{10,17,45,46}. In a recent smFRET study, Pif1 was shown to bind at the 3′ tailed DNA junction, reeling in the 3′ ssDNA tail to unfold intramolecular G4 structures in an ATP-dependent manner⁴⁶. In agreement with that study, we found that binding of Pif1 did not result in unfolding of the G4 structure in either substrate (Supplementary Fig. 7d). These results indicate that ATP-independent G4 unfolding is carried by HRDC-RQC domains and that this is a function that is unique to RecQ helicases.

G4 Structure Regulates Activity of BLM at Proximal dsDNA

We sought to determine whether BLM's substrate selectivity in unfolding the G4 motif also affects its unwinding activity of a duplex located downstream of the G4 motif. Recent studies have shown that G4 inhibits the ssDNA translocation activity of BLM and reduces the overall efficiency of BLM-mediated duplex unwinding³¹. We first tested the regular G4 substrate (i.e. having no ssDNA between the G4 and duplex region); this showed no G4 unfolding upon addition of BLM. Full unwinding of the duplex part in these FRET substrates will result in a complete loss of the donor (Cy3) labeled DNA strand, and consequently loss of the fluorescence signal³². For our unwinding assay, BLM and ATP were added to the perfusion chamber containing surface tethered DNA FRET substrates and incubated for two minutes, followed by a wash with high salt (500 mM NaCl) buffer to terminate the reaction. To quantify the unwinding yield, surface tethered FRET molecules were counted before and after the unwinding reaction. Fig. 3a–b shows images of individual FRET molecules obtained before and after the unwinding reaction with G4 DNA substrate (Fig. 3a) and regular partial duplex DNA substrate (Fig. 3b), where both substrates contain a duplex region of the same length and sequence (see Supplementary Table 1). These images clearly show that BLM is unable to unwind the G4 substrates, which persist after the unwinding reaction, whereas the majority of partial duplex DNA molecules were unwound and removed from the surface even in the presence of 50 nM BLM and 100 μM ATP.

To establish that the observed low yield of unwinding is indeed due to formation of G4 structures, we carried out two additional unwinding experiments with the G4 substrate, but under buffer conditions where a G4 structure cannot readily form. Instead of 50 mM KCl, we used either 2 mM KCl, where the G4 structure is predominantly in its unfolded state, or 50 mM LiCl, as Li⁺ does not stabilize G4 structures⁴⁷. Under these conditions, BLM

showed a comparable unwinding efficiency to that observed on the regular partial duplex DNA substrate (Fig. 3c), confirming that G4 structure formation indeed inhibits BLM-mediated unwinding of dsDNA when it is positioned downstream. To confirm that inhibition of unwinding is specific for the G4 containing substrate, rather than universal for substrates containing secondary structures, we used a hairpin substrate having a high GC base content ($T_m \sim 50^\circ\text{C}$) with approximately the same length of G4 region (Supplementary Table 1 and Supplementary Fig. 1a). We found that unwinding of the hairpin substrate by BLM under the same buffer conditions used for the G4 substrate was highly efficient (Fig. 3c), confirming that the G4 structure specifically inhibits BLM unwinding. To further establish the effect of G4 folding on BLM unwinding activity in the G4 substrate, we quantified BLM's unwinding yield as a function of KCl concentration (Fig. 3d). This experiment showed that the unwinding yield steeply decreased with increasing KCl concentration and is fully inhibited at ~ 10 mM KCl and higher. The KCl-dependent reduction in unwinding yield correlates well with the increase in G4 folding stability at those concentrations (Supplementary Fig. 1b).

Next, we characterized BLM's unwinding activity for the G4 (5 nt) and G4 (10 nt) substrates, which showed elevated unfolding of G4 structures by BLM as compared to regular G4 substrate containing no ssDNA spacer (Fig. 1). Quantification of the BLM unwinding yield for these substrates and a regular G4 substrate is plotted in Fig. 4a and Supplementary Fig. 8a–b; the data show that BLM can efficiently unwind these substrates compared to the G4 substrate. Substrate G4 (10 nt) had a higher unwinding efficiency than G4 (5 nt), suggesting that initial unfolding of the G4 structure determines BLM's ability to unwind the downstream duplex. To rule out the possibility that BLM initiates unwinding of the downstream duplex directly from the 10 nt ssDNA, we used a gapped tailed substrate (Supplementary Fig. 9 and Supplementary Table 1) which showed low unwinding efficiency compared to the G4 (10nt) substrate. This observation confirms that unwinding is not initiated from the 10 nt ssDNA segment, but occurs after BLM unfolds and passes the G4 motif. We note that the reduced unwinding efficiency we observe for the gapped substrate is consistent with previous reports, and stem from BLM's substrate affinity rather than its binding footprint⁴⁸. To further establish the correlation between substrate dependent G4 unfolding and the unwinding of downstream duplex, we tested the unwinding activity of WRN and RECQL5 in these substrates. Consistent with the ATP-independent G4 unfolding observed for these helicases (Supplementary Fig. 7a–b), WRN was able to unwind the G4 (10) but not the G4 substrate, whereas RECQL5 was unable to unwind either substrate (Supplementary Fig. 10). We conclude that the RQC-HRDC mediated unfolding of the G4 structure in these substrates governs the ability of the helicase to unwind downstream dsDNA.

To determine how ATP affects BLM's G4 unfolding activity in these substrates, we added BLM together with ATP, and then monitored G4 FRET efficiency as a function of ATP concentration. We used ATP concentrations of 1, 10 and 100 μM because at these concentrations unwinding of the downstream duplex does not dominate the observed dynamics; thus the G4 refolding–unfolding transition can be extracted and analyzed (Fig. 4b–c). While both substrates displayed a change in the FRET distribution upon addition of ATP, the G4 (10 nt) substrate showed the most substantial redistribution of the FRET

populations with ATP. Analysis of individual trajectories for the two substrates showed rapid unfolding–refolding transitions for the G4 (5 nt) and G4 (10 nt) substrates, respectively (Fig. 4d–e). These fluctuations are more dynamic than the fluctuations observed for BLM unfolding–refolding transitions in the absence of ATP; this may be due to BLM’s ATP-dependent translocation dynamics away from the G4 region. To obtain the specific transition states and the refolding and unfolding rates as a function of ATP, we carried out HMM analysis on individual trajectories from each substrate and generated TDP images. Fig. 4f–g shows the TDP for G4 (5 nt) and G4 (10 nt) substrates in the presence of 50 nM BLM and 10 μ M ATP. The TDP for the G4 (10 nt) substrate showed broadening of the transition peaks, whereas transitions for the G4 (5 nt) substrate formed defined peaks (see also Supplementary Fig. 11a–d). To determine whether the TDP reflects a specific ATP-induced G4 folding response, we plotted the unfolding and refolding rates for each substrate as a function of [ATP] for both G4 (5 nt) and G4 (10 nt) substrates (Fig. 4h–i, respectively). While the unfolding and refolding rates for the G4 (5 nt) substrate were largely unchanged as a function of [ATP], the rates for the G4 (10 nt) substrates followed a clear trend of increase with increasing ATP concentration. The increase in both unfolding and refolding rates with ATP is indicative of BLM translocating through the G4 region following G4 unfolding, which would result in an increase in both unfolding (into the G4 structure) and refolding (away from G4 toward the duplex region). The fact that this dependence is not observed for the G4 (5 nt) substrate are consistent with the dependence of this behavior on the availability of the longer ssDNA segment between G4 and the duplex. Specifically, the length of the ssDNA spacer will govern both G4 unfolding efficiency and consequently initiation of helicase activity, so that unwinding is observed on G4 substrates where the G4 is efficiently unfolded by BLM and enough ssDNA is produced. In this event, the observed inhibition of unwinding can be ascribed to steric occlusion induced by G4 folding.

G4 Complementary Strand Assists Downstream Duplex Unwinding

We next sought to determine whether BLM’s unwinding activity in the G4 substrate could be restored via external disruption of G4 folding. To do this, we used either replication protein A (RPA) or a short ssDNA complementary to the G4 sequence; these were added at various concentrations along with BLM (50 nM) and ATP (2 mM). Contrary to our expectations, RPA (1 or 10 nM) did not assist core BLM unwinding of the G4 substrate either when added together with BLM and ATP, or when pre-incubated with the DNA substrate prior to addition of BLM and ATP. Since core BLM lacks the RPA interaction domain⁴⁹, we also carried out these experiments with full-length BLM, which, in a similar manner to core BLM, did not show enhanced unwinding in the presence of RPA (Supplementary Fig. 12a). In contrast, addition of the complementary strand (CS) along with BLM and ATP resulted in efficient unwinding of the downstream duplex. Fig. 5A shows the unwinding yield of the G4 substrate by 50 nM BLM and 2 mM ATP either added alone or in the presence of 1 nM or 10 nM CS, where the addition of CS enabled BLM to unwind the downstream duplex, and the unwinding efficiency increased with increasing CS concentration. To determine whether the external ssDNA assisted G4 disruption and unwinding is homology dependent, we measured the unwinding of G4 substrate in the presence of ssDNA of similar length to CS, but with a sequence that was not complementary (NC) to G4. This resulted in a considerably lower unwinding efficiency (Fig. 5a),

demonstrating that external ssDNA assisted unwinding of the G4 substrate is homology dependent. Fig. 5b shows the resulting smFRET histograms for each concentration of CS compared to no CS, where a substantial shift and redistribution of FRET values is seen for both 1nM and 10 nM CS. Fig. 5c shows a representative smFRET trajectory of the G4 substrate in the presence of 10 nM CS, 50 nM BLM and 2 mM ATP: here the data reveal a display of dynamic fluctuations, representing rapid unfolding and refolding transitions. These dynamics are in contrast to the stable FRET trajectory obtained for the G4 substrate in the presence of 50 nM BLM and 2 mM ATP, when no CS was added (Supplementary Fig. 12b). To determine the role of ATP in G4 unfolding, we quantified the number of trajectories showing fluctuations in the presence and absence of ATP (Fig. 5d). In contrast to the ATP-independent unfolding in the G4 (5 nt) and G4 (10 nt) substrates, unfolding-refolding transitions were not observed in the absence of ATP, confirming that CS assisted G4 unfolding is ATP dependent. To characterize the observed transitions, we carried out HMM analysis of trajectories at two concentrations of CS, and generated their TDP (Fig. 5e–f). The resulting TDP for both CS concentrations showed similar transition probabilities. This verifies that the occurrence of G4 unfolding, but not the unfolding pathway itself, depends upon CS concentration. Moreover, this finding suggests that CS assisted unwinding is mediated by a single CS per G4 substrate. This conclusion is further supported by the unfolding-refolding rates (Fig. 5g), which remained unchanged with increased concentration of CS.

We note that when CS was added to the G4 substrate alone, no change was observed in G4 conformation, confirming that CS interaction with G4 is not random and is facilitated by BLM (Supplementary Fig. 12c). While BLM has strand-annealing activity, this interaction is not carried out via its annealing activity, as the core BLM lacks the domain for this activity⁵⁰. This was also verified by testing CS assisted unwinding of the G4 substrate using full length BLM, which did not result in an increase in unwinding yield as compared to core BLM (Supplementary Fig. 12d). We conclude that the HRDC domain may play a role here as well, in which the helicase core and the RQC domain are in contact with the G4 substrate while the HRDC domain captures the CS and facilitates the destabilization of G4 and unwinding of the downstream duplex.

Discussion

Using smFRET we show that BLM requires a ssDNA segment between the G4 motif and duplex region to efficiently unfold intramolecular G4 DNA structures. The unfolding of G4 is ATP-independent, is mediated by the RQC and HRDC domains, and occurs in a sequential manner, where G4 does not transition directly to the unfolded conformation, but proceeds via sub-steps. Moreover, BLM unfolding of G4 is metastable, which allows for repetitive refolding and unfolding transitions to occur. The length of ssDNA segment determines the magnitude of G4 destabilization and consequently unwinding of a downstream duplex. Taken together, our observations demonstrate that the BLM-G4 interaction constitutes a unique substrate-dependent activity.

Based on our findings, we propose an integrated model for the cooperative binding and interaction of BLM with an intramolecular G4 structure (Fig. 6). Unfolding of G4 by BLM

is mediated via cooperative binding of the RQC and HRDC domains, where the RQC interaction with G4 is stabilized by HRDC binding to the ssDNA segment between G4 and the duplex (Fig. 6a). Additionally, the ssDNA binding activity of the HRDC domain may also play a role in capturing the CS to promote BLM's CS-assisted unwinding of the regular G4 substrate, as illustrated in Fig. 6b. Previous studies have identified the RQC domain as having a high affinity for G4 structures, whereas the HRDC domain was shown to play a regulatory role in HJ dissolution and binding to ssDNA^{39,43,44,51}. Recent structure-based studies have suggested a model for BLM HJ interaction in which HRDC binds ssDNA at the junctions, and operates at a distance from the RQC and helicase core^{52,53}. This distant interaction is afforded by a long flexible linker (13 a.a. ~ 10 nm) between the RQC and HRDC, which is sufficient for binding ssDNA over the folded G4 structure (~ 5 nm)⁵². Cooperative binding at a distance may explain helicase selectivity in unfolding and processing specific G4 substrates.

There is a growing body of evidence that G4 structures and G4 associated proteins play a central role in chromosome metabolism. Several hundred thousand DNA sequences with the potential to form intra-stranded G4 structures are present in the human genome^{5,6,54,55}. These sequences will spontaneously form intra-stranded G4 structures when ssDNA is generated during DNA replication, transcription and repair, and if not removed will obstruct these processes^{18,56}. Several human helicases have been implicated in the removal of G4 structures, including FANCI, RTEL1 and Pif1 (having a 5'-3' directionality), and two RecQ family helicases, BLM and WRN (each with a 3'-5' directionality)⁵⁷. Importantly, amongst the five human RecQ family helicases, only BLM and WRN have both RQC and HRDC domains²⁷. With respect to their function in the removal of G4 structures, BLM and WRN have been found to associate with the replication machinery in both chromosomal and telomere replication^{20,22-26,35,58}. During replication, intra-molecular G4 structures can form after dsDNA unwinding by the replicative helicase and impede leading or lagging strand synthesis, resulting in an accumulation of ssDNA and the collapse of the replication fork. Based on our findings and proposed model, BLM and WRN may assist in the removal of characteristic replication fork G4 structures, as previously suggested¹⁹. Additionally, the substrate selectivity and removal of G4 structures by BLM and WRN may play an important role during transcription. Recent studies have identified a novel correlation between the presence of G4 forming sequences in the regulatory regions of genes and the regulation of gene expression by BLM and WRN helicases^{28,59}. The regulation of gene expression by BLM and WRN may thus be mediated by removal of G4 structures formed on ssDNA within the transcription R-loop, which would otherwise impede gene expression.

In a recent smFRET study of Pif1 helicase it was shown that removal of G4 structures is ATP-dependent and mediated by reeling of a ssDNA tail⁴⁶. Importantly, this study found that the binding of Pif1 and reeling of ssDNA is only permitted in substrates having a ssDNA spacer between the dsDNA and G4. While the underlining mechanisms for G4 removal by Pif1 and BLM are distinct, the striking similarity in the requirement for a ssDNA spacer suggests a unique DNA substrate selectivity amongst G4 specific helicases. Such selectivity may be jointly utilized in G4 metabolism. For example, BLM has been shown to interact with FANCI, a helicase like BLM implicated in DNA repair but with opposite directionality⁶⁰. FANCI has been shown to unwind G4 structures in a unique

manner compared to representative Fe-S cluster helicases⁶¹, and has been implicated with BLM in a common pathway that maintains epigenetic stability at G4 DNA⁶². We predict that G4 structural selectivity and cooperation among other DNA processing enzymes provides an additional level of functional regulation that plays a critical role in genomic metabolism.

Methods

DNA Preparations

All oligonucleotides (Supplementary Table 1) were purchased from Integrated DNA Technology (IDT). For annealing, the appropriate oligonucleotides were mixed and heated for 3 minutes at 94°C followed by slow cooling.

Protein Purification

To investigate the interaction of BLM with G4 DNA, we used the core BLM helicase, a truncated BLM variant (BLM^{636–1298}) containing RQC and HRDC domains that retain the unwinding and translocation activity of the wild type BLM^{30,36,42}. In addition we used core-WRN (418–1232) containing the RQC and HRDC domains, human Pif1 helicase domain (206–641) shown to binding G4 structures and unwind inter-stranded G4, helicase-dead core-BLM (K695M) and RecQL5. Full length BLM was used to verify our observations relating to core-BLM.

Core BLM (a.a. 636–1298) was expressed in *E. coli* system and purified using tag affinity capture⁶³. After cleavage with TEV protease, the protein retains the vector-derived sequences SM and ENLYFQ at the N- and C-termini, respectively. Truncated fragments (RQC-HRDC: aa 858 – 1298, and HRDC: aa 1069 – 1298) were cloned into vector pNIC28-Bsa4 and purified on HisTrap FF and Superdex 75 columns⁶⁴; the proteins were used without removal of the N-terminal His₆ tags. All purified proteins maintained their native state as determined by their gel filtration absorption profile, which migrates as a single symmetrical peak (Supplementary Fig. 13 for RQC-HRDC fragment). Core-WRN, Pif1 HD were expressed and purified from a Baculovirus expression vector system^{65,66}. For expression and purification RecQL5 and helicase dead core BLM we used *E. coli* expression system and purified using an affinity tag^{42,67}. Expression and purification of the recombinant full-length BLM protein was performed using a three-step protocol⁶⁸: Nickel affinity, Heparin affinity and gel filtration chromatography. In brief, yeast cell pellets expressing human BLM (JEL1+pJK1_BLM) were re-suspended in an equal volume of lysis buffer, consisting of 50 mM Hepes-KOH pH 7.4, 1M KCl, 1 mM DTT, 10% glycerol, 1 mM PMSF and supplemented with Complete EDTA-free protease inhibitor (Roche). Benzonase (E1014-25kU, Sigma) was added at 25 U/ml. An equal volume of glass beads (425–600 μm, Sigma) was added to the suspension and the cells were lysed by vigorous shaking for 10 × 1 min with incubations of 30 s on ice after each burst. The lysate was clarified at 350,000Xg for 1 hour at 4°C. The supernatant was adjusted to contain 15 mM imidazole, filtered through 45 μm filter and loaded onto HisTrap column (GE Healthcare). The column was washed with lysis buffer containing 15 mM imidazole. Subsequently, the washing buffer was exchanged to the same buffer with reduced salt concentration (500 mM KCl instead of

1M). Several step gradients of this buffer, containing different concentrations up to 80 mM imidazole, were used to further wash the column. The column was eluted using a 10-column volume gradient between 80 mM and 500 mM imidazole. Eluted fractions containing BLM were pooled, diluted to adjust the final pH and KCl concentration to 7.5 and 125 mM, respectively, and were loaded onto a Heparin column (GE Healthcare). The column was washed with a buffer containing 50 mM Hepes-KOH pH 6.8, 125 mM KCl, 1 mM DTT, 10% glycerol and supplemented with Complete EDTA-free protease inhibitor (Roche). The BLM protein was eluted using a 20-column volume gradient between 150 mM and 1 M KCl. Eluted fractions containing the BLM protein were pooled and concentrated up to 10 fold using an Amicon Ultra-4 centrifugal filter device (50K cutoff, Millipore). The pooled eluate was loaded onto a gel filtration column (Superose 6 10/300 GL; GE Healthcare) using a buffer containing 50 mM Hepes-KOH pH 7.4, 10% glycerol, 500 mM KCl and 1 mM DTT. Fractions containing recombinant BLM, as determined by Coomassie stained SDS-PAGE, were pooled and the protein was dialyzed against a buffer containing 50 mM Hepes-KOH pH 7.2, 250 mM KCl, 10% glycerol and 1 mM DTT. Finally, the protein was stored at -80°C in small aliquots.

Single-molecule Fluorescence Resonance Energy Transfer (smFRET)

Reactions were carried out at room temperature in a standard buffer composed of 50 mM Tris-HCl (pH 8.0), 2 mM MgCl_2 , and an oxygen scavenging system (1 mg/ml glucose oxidase, 0.4% (w/v) D-glucose, 0.02 mg/ml catalase, and 2 mM Trolox)³². The concentration of K^+ in the buffer was varied and is indicated accordingly in the text or in the figure legends. All experiments had 100 pM DNA immobilized on a PEG-coated quartz surface with biotin-neutravidin linkage.

Microscopy setup and single-molecule analysis

We used a custom-built microscopy setup based on a Leica DMI3000 microscope equipped with an HCX PL APO 63x NA=1.47 OIL CORR TIRF objective followed by achromatic 2x tube lens magnification. The microscope was coupled to 532nm and 640nm solid-state lasers to excite the sample at Total-Internal-Reflection (TIRF) illumination mode for improved signal to noise ratio and reject out of plane fluorescence. Sample emission was collected and split into two channels through the use of proper dichroic and emission narrow-band bandpass filters (filter for green channel 580/60, filter for red channel 680/40, Semrock) in conjunction with the use of a Dual View (DV2-Photometrics) to image two colors simultaneously, side-by-side, onto a single EM-CCD camera (Andor iXon+ 897) acquiring at 33 Hz. For accurate alignment and mapping of the two color channels, we first imaged diffraction-limited fluorescent beads that have wide emission spectra spanning both channels (Invitrogen). The location of the beads was matched for both channels, and a mapping matrix was generated using an IDL (Exelis Visual Information Solutions) custom mapping routine. Briefly, this routine is based on the use of a polynomial morph-type mapping function, whereby mapping coefficients are generated by Gaussian and centroid fits to the sub-diffraction limit point-spread functions of the fluorescence beads. Another IDL code was used, along with the mapping matrix, to extract corresponding single-molecule donor and acceptor spots into single molecule trajectories. Programs written in MATLAB were used to view and analyze single-molecule fluorescence time trajectories.

FRET efficiency, E_{FRET} , was approximated as the ratio between the acceptor intensity and the sum of acceptor and donor intensities. Each smFRET histogram was generated by a minimum of a hundred (100) trajectories, and the contribution of each trajectory was normalized in the histogram so that the weight of the FRET states from each molecule in the histogram will be identical. For unwinding yield quantification dual images were analyzed using a threshold and cluster counting routine, to obtain the number of molecules observed in the field of view. Hidden Markov Modeling (HMM) analysis, transition density plot and derived statistics were done using freely available HMM and TDP software (Ha lab)⁴⁰. The resulting TDP plots showed a number of symmetrical peaks for folding and refolding transitions, along with a low intensity singular peak for refolding into higher FRET state. We interpret this peak as possible ssDNA looping mediated by the HRDC domain occurring after unfolding of the G4 structure. To generate the mean rates for folding (unfolding) we averaged the folding (unfolding) rates obtained from each TDP peak/transition.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Nick Cowan and Jim Borowiec for commenting on the manuscript. Research in E.R. laboratory is supported by the NIH (GM057691/Delmar and GM110385/Rothenberg) the Arnold and Mabel Beckman Foundation and the Shifrin-Meyer Breast Cancer Discovery award. D.H. and I.D.H. were supported by The Nordea Foundation and the Danish Medical Research Council (FSS). P.J. and A.S. were supported by the Swiss National Science Foundation (31003A_146206). O.G. and P.S. were supported by the SGC, a registered charity (number 1097737) that receives funds from AbbVie, Bayer Pharma AG, Boehringer Ingelheim, the Canada Foundation for Innovation, Genome Canada, GlaxoSmithKline, Janssen, Lilly Canada, the Novartis Research Foundation, the Ontario Ministry of Economic Development and Innovation, Pfizer, Takeda, and the Wellcome Trust [092809/Z/10/Z]

References

1. Gellert M, Lipsett MN, Davies DR. Helix formation by guanylic acid. *Proceedings of the National Academy of Sciences of the United States of America*. 1962; 48:2013–2018. [PubMed: 13947099]
2. Sen D, Gilbert W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature*. 1988; 334:364–366.10.1038/334364a0 [PubMed: 3393228]
3. Hershman SG, et al. Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in *Saccharomyces cerevisiae*. *Nucleic acids research*. 2008; 36:144–156.10.1093/nar/gkm986 [PubMed: 17999996]
4. Todd AK, Johnston M, Neidle S. Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic acids research*. 2005; 33:2901–2907.10.1093/nar/gki553 [PubMed: 15914666]
5. Huppert JL. Structure, location and interactions of G-quadruplexes. *The FEBS journal*. 2010; 277:3452–3458.10.1111/j.1742-4658.2010.07758.x [PubMed: 20670279]
6. Qin Y, Hurley LH. Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. *Biochimie*. 2008; 90:1149–1171.10.1016/j.biochi.2008.02.020 [PubMed: 18355457]
7. Huber MD, Lee DC, Maizels N. G4 DNA unwinding by BLM and Sgs1p: substrate specificity and substrate-specific inhibition. *Nucleic acids research*. 2002; 30:3954–3961. [PubMed: 12235379]
8. London TB, et al. FANCI is a structure-specific DNA helicase associated with the maintenance of genomic G/C tracts. *The Journal of biological chemistry*. 2008; 283:36132–36139.10.1074/jbc.M808152200 [PubMed: 18978354]

9. Mohaghegh P, Karow JK, Brosh RM Jr, Bohr VA, Hickson ID. The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases. *Nucleic acids research*. 2001; 29:2843–2849. [PubMed: 11433031]
10. Paeschke K, Capra JA, Zakian VA. DNA replication through G-quadruplex motifs is promoted by the *Saccharomyces cerevisiae* Pif1 DNA helicase. *Cell*. 2011; 145:678–691.10.1016/j.cell.2011.04.015 [PubMed: 21620135]
11. Salas TR, et al. Human replication protein A unfolds telomeric G-quadruplexes. *Nucleic acids research*. 2006; 34:4857–4865.10.1093/nar/gkl564 [PubMed: 16973897]
12. Sun H, Karow JK, Hickson ID, Maizels N. The Bloom's syndrome helicase unwinds G4 DNA. *The Journal of biological chemistry*. 1998; 273:27587–27592. [PubMed: 9765292]
13. Zaug AJ, Podell ER, Cech TR. Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro. *Proceedings of the National Academy of Sciences of the United States of America*. 2005; 102:10864–10869.10.1073/pnas.0504744102 [PubMed: 16043710]
14. Fry M, Loeb LA. Human werner syndrome DNA helicase unwinds tetrahelical structures of the fragile X syndrome repeat sequence d(CGG)n. *The Journal of biological chemistry*. 1999; 274:12797–12802. [PubMed: 10212265]
15. Sun H, Bennett RJ, Maizels N. The *Saccharomyces cerevisiae* Sgs1 helicase efficiently unwinds G–G paired DNAs. *Nucleic acids research*. 1999; 27:1978–1984. [PubMed: 10198430]
16. Ribeyre C, et al. The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences in vivo. *PLoS genetics*. 2009; 5:e1000475. [PubMed: 19424434]
17. Paeschke K, et al. Pif1 family helicases suppress genome instability at G-quadruplex motifs. *Nature*. 2013; 497:458–462.10.1038/nature12149 [PubMed: 23657261]
18. Tarsounas M, Tijsterman M. Genomes and G-quadruplexes: for better or for worse. *Journal of molecular biology*. 2013; 425:4782–4789.10.1016/j.jmb.2013.09.026 [PubMed: 24076189]
19. Croteau DL, Popuri V, Opresko PL, Bohr VA. Human RecQ helicases in DNA repair, recombination, and replication. *Annual review of biochemistry*. 2014; 83:519–552.10.1146/annurev-biochem-060713-035428
20. Edwards DN, Orren DK, Machwe A. Strand exchange of telomeric DNA catalyzed by the Werner syndrome protein (WRN) is specifically stimulated by TRF2. *Nucleic acids research*. 2014; 42:7748–7761.10.1093/nar/gku454 [PubMed: 24880691]
21. Li K, et al. Regulation of WRN protein cellular localization and enzymatic activities by SIRT1-mediated deacetylation. *The Journal of biological chemistry*. 2008; 283:7590–7598.10.1074/jbc.M709707200 [PubMed: 18203716]
22. Lillard-Wetherell K, et al. Association and regulation of the BLM helicase by the telomere proteins TRF1 and TRF2. *Human molecular genetics*. 2004; 13:1919–1932.10.1093/hmg/ddh193 [PubMed: 15229185]
23. Machwe A, Karale R, Xu X, Liu Y, Orren DK. The Werner and Bloom syndrome proteins help resolve replication blockage by converting (regressed) holliday junctions to functional replication forks. *Biochemistry*. 2011; 50:6774–6788.10.1021/bi2001054 [PubMed: 21736299]
24. Machwe A, Xiao L, Orren DK. TRF2 recruits the Werner syndrome (WRN) exonuclease for processing of telomeric DNA. *Oncogene*. 2004; 23:149–156.10.1038/sj.onc.1206906 [PubMed: 14712220]
25. Chan KL, North PS, Hickson ID. BLM is required for faithful chromosome segregation and its localization defines a class of ultrafine anaphase bridges. *The EMBO journal*. 2007; 26:3397–3409.10.1038/sj.emboj.7601777 [PubMed: 17599064]
26. Davies SL, North PS, Hickson ID. Role for BLM in replication-fork restart and suppression of origin firing after replicative stress. *Nature structural & molecular biology*. 2007; 14:677–679.10.1038/nsmb1267
27. Larsen NB, Hickson ID. RecQ Helicases: Conserved Guardians of Genomic Integrity. *Advances in experimental medicine and biology*. 2013; 767:161–184.10.1007/978-1-4614-5037-5_8 [PubMed: 23161011]
28. Nguyen GH, et al. Regulation of gene expression by the BLM helicase correlates with the presence of G-quadruplex DNA motifs. *Proceedings of the National Academy of Sciences of the United States of America*. 2014; 111:9905–9910.10.1073/pnas.1404807111 [PubMed: 24958861]

29. Selak N, et al. The Bloom's syndrome helicase (BLM) interacts physically and functionally with p12, the smallest subunit of human DNA polymerase delta. *Nucleic acids research*. 2008; 36:5166–5179.10.1093/nar/gkn498 [PubMed: 18682526]
30. Liu JQ, Chen CY, Xue Y, Hao YH, Tan Z. G-quadruplex hinders translocation of BLM helicase on DNA: a real-time fluorescence spectroscopic unwinding study and comparison with duplex substrates. *Journal of the American Chemical Society*. 2010; 132:10521–10527.10.1021/ja1038165 [PubMed: 20614884]
31. Wang Q, et al. G-quadruplex formation at the 3' end of telomere DNA inhibits its extension by telomerase, polymerase and unwinding by helicase. *Nucleic acids research*. 2011; 39:6229–6237.10.1093/nar/gkr164 [PubMed: 21441540]
32. Rothenberg E, Ha T. Single-molecule FRET analysis of helicase functions. *Methods in molecular biology*. 2010; 587:29–43.10.1007/978-1-60327-355-8_3 [PubMed: 20225140]
33. Yodh JG, Schlierf M, Ha T. Insight into helicase mechanism and function revealed through single-molecule approaches. *Quarterly Reviews of Biophysics*. 2010; 43:185–217.10.1017/s0033583510000107 [PubMed: 20682090]
34. Ray S, Bandaria JN, Qureshi MH, Yildiz A, Balci H. G-quadruplex formation in telomeres enhances POT1/TPP1 protection against RPA binding. *Proceedings of the National Academy of Sciences of the United States of America*. 2014; 111:2990–2995.10.1073/pnas.1321436111 [PubMed: 24516170]
35. Klaue D, et al. Fork sensing and strand switching control antagonistic activities of RecQ helicases. *Nature communications*. 2013; 4:2024.10.1038/ncomms3024
36. Yodh JG, Stevens BC, Kanagaraj R, Janscak P, Ha T. BLM helicase measures DNA unwound before switching strands and hRPA promotes unwinding reinitiation. *The EMBO journal*. 2009; 28:405–416.10.1038/emboj.2008.298 [PubMed: 19165145]
37. Budhathoki JB, et al. RecQ-core of BLM unfolds telomeric G-quadruplex in the absence of ATP. *Nucleic acids research*. 2014.10.1093/nar/gku856
38. Lee JY, Okumus B, Kim DS, Ha T. Extreme conformational diversity in human telomeric DNA. *Proceedings of the National Academy of Sciences of the United States of America*. 2005; 102:18938–18943.10.1073/pnas.0506144102 [PubMed: 16365301]
39. Wu L, et al. The HRDC domain of BLM is required for the dissolution of double Holliday junctions. *The EMBO journal*. 2005; 24:2679–2687.10.1038/sj.emboj.7600740 [PubMed: 15990871]
40. McKinney SA, Joo C, Ha T. Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophysical journal*. 2006; 91:1941–1951.10.1529/biophysj.106.082487 [PubMed: 16766620]
41. Hwang H, Buncher N, Opresko PL, Myong S. POT1-TPP1 regulates telomeric overhang structural dynamics. *Structure*. 2012; 20:1872–1880.10.1016/j.str.2012.08.018 [PubMed: 22981946]
42. Janscak P, et al. Characterization and mutational analysis of the RecQ core of the bloom syndrome protein. *Journal of molecular biology*. 2003; 330:29–42. [PubMed: 12818200]
43. Huber MD, Duquette ML, Shiels JC, Maizels N. A conserved G4 DNA binding domain in RecQ family helicases. *Journal of molecular biology*. 2006; 358:1071–1080.10.1016/j.jmb.2006.01.077 [PubMed: 16530788]
44. Cahoon LA, Manthei KA, Rotman E, Keck JL, Seifert HS. *Neisseria gonorrhoeae* RecQ helicase HRDC domains are essential for efficient binding and unwinding of the pilE guanine quartet structure required for pilin antigenic variation. *Journal of bacteriology*. 2013; 195:2255–2261.10.1128/JB.02217-12 [PubMed: 23475972]
45. Bochman ML, Sabouri N, Zakian VA. Unwinding the functions of the Pif1 family helicases. *DNA repair*. 2010; 9:237–249.10.1016/j.dnarep.2010.01.008 [PubMed: 20097624]
46. Zhou R, Zhang J, Bochman ML, Zakian VA, Ha T. Periodic DNA patrolling underlies diverse functions of Pif1 on R-loops and G-rich DNA. *eLife*. 2014; 3:e02190.10.7554/eLife.02190 [PubMed: 24843019]
47. Williamson JR, Raghuraman MK, Cech TR. Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell*. 1989; 59:871–880. [PubMed: 2590943]

48. Bachrati CZ, Borts RH, Hickson ID. Mobile D-loops are a preferred substrate for the Bloom's syndrome helicase. *Nucleic acids research*. 2006; 34:2269–2279.10.1093/nar/gkl258 [PubMed: 16670433]
49. Doherty KM, et al. RECQ1 helicase interacts with human mismatch repair factors that regulate genetic recombination. *The Journal of biological chemistry*. 2005; 280:28085–28094.10.1074/jbc.M500265200 [PubMed: 15886194]
50. Cheok CF, Wu L, Garcia PL, Janscak P, Hickson ID. The Bloom's syndrome helicase promotes the annealing of complementary single-stranded DNA. *Nucleic acids research*. 2005; 33:3932–3941.10.1093/nar/gki712 [PubMed: 16024743]
51. Wu X, Maizels N. Substrate-specific inhibition of RecQ helicase. *Nucleic acids research*. 2001; 29:1765–1771. [PubMed: 11292849]
52. Kim SY, Hakoshima T, Kitano K. Structure of the RecQ C-terminal Domain of Human Bloom Syndrome Protein. *Scientific reports*. 2013; 3:3294.10.1038/srep03294 [PubMed: 24257077]
53. Kim YM, Choi BS. Structure and function of the regulatory HRDC domain from human Bloom syndrome protein. *Nucleic acids research*. 2010; 38:7764–7777.10.1093/nar/gkq586 [PubMed: 20639533]
54. Huppert JL, Balasubramanian S. Prevalence of quadruplexes in the human genome. *Nucleic acids research*. 2005; 33:2908–2916.10.1093/nar/gki609 [PubMed: 15914667]
55. Rodriguez R, et al. Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. *Nature chemical biology*. 2012; 8:301–310.10.1038/nchembio.780 [PubMed: 22306580]
56. Bochman ML, Paeschke K, Zakian VA. DNA secondary structures: stability and function of G-quadruplex structures. *Nature reviews. Genetics*. 2012; 13:770–780.10.1038/nrg3296
57. Wu Y, Brosh RM Jr. G-quadruplex nucleic acids and human disease. *The FEBS journal*. 2010; 277:3470–3488.10.1111/j.1742-4658.2010.07760.x [PubMed: 20670277]
58. Machwe A, Xiao L, Lloyd RG, Bolt E, Orren DK. Replication fork regression in vitro by the Werner syndrome protein (WRN): holliday junction formation, the effect of leading arm structure and a potential role for WRN exonuclease activity. *Nucleic acids research*. 2007; 35:5729–5747.10.1093/nar/gkm561 [PubMed: 17717003]
59. Johnson JE, Cao K, Ryvkin P, Wang LS, Johnson FB. Altered gene expression in the Werner and Bloom syndromes is associated with sequences having G-quadruplex forming potential. *Nucleic acids research*. 2010; 38:1114–1122.10.1093/nar/gkp1103 [PubMed: 19966276]
60. Suhasini AN, et al. Interaction between the helicases genetically linked to Fanconi anemia group J and Bloom's syndrome. *The EMBO journal*. 2011; 30:692–705.10.1038/emboj.2010.362 [PubMed: 21240188]
61. Bharti SK, et al. Specialization among iron-sulfur cluster helicases to resolve G-quadruplex DNA structures that threaten genomic stability. *The Journal of biological chemistry*. 2013; 288:28217–28229.10.1074/jbc.M113.496463 [PubMed: 23935105]
62. Sarkies P, et al. FANCF coordinates two pathways that maintain epigenetic stability at G-quadruplex DNA. *Nucleic acids research*. 2012; 40:1485–1498.10.1093/nar/gkr868 [PubMed: 22021381]
63. Nguyen GH, et al. A small molecule inhibitor of the BLM helicase modulates chromosome stability in human cells. *Chemistry & biology*. 2013; 20:55–62.10.1016/j.chembiol.2012.10.016 [PubMed: 23352139]
64. Savitsky P, et al. High-throughput production of human proteins for crystallization: the SGC experience. *Journal of structural biology*. 2010; 172:3–13.10.1016/j.jsb.2010.06.008 [PubMed: 20541610]
65. Gileadi O, et al. High throughput production of recombinant human proteins for crystallography. *Methods in molecular biology*. 2008; 426:221–246.10.1007/978-1-60327-058-8_14 [PubMed: 18542867]
66. Shrestha B, Smee C, Gileadi O. Baculovirus expression vector system: an emerging host for high-throughput eukaryotic protein expression. *Methods in molecular biology*. 2008; 439:269–289.10.1007/978-1-59745-188-8_19 [PubMed: 18370110]

67. Garcia PL, Liu Y, Jiricny J, West SC, Janscak P. Human RECQ5beta, a protein with DNA helicase and strand-annealing activities in a single polypeptide. *The EMBO journal*. 2004; 23:2882–2891.10.1038/sj.emboj.7600301 [PubMed: 15241474]
68. Karow JK, Chakraverty RK, Hickson ID. The Bloom's syndrome gene product is a 3'-5' DNA helicase. *The Journal of biological chemistry*. 1997; 272:30611–30614. [PubMed: 9388193]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

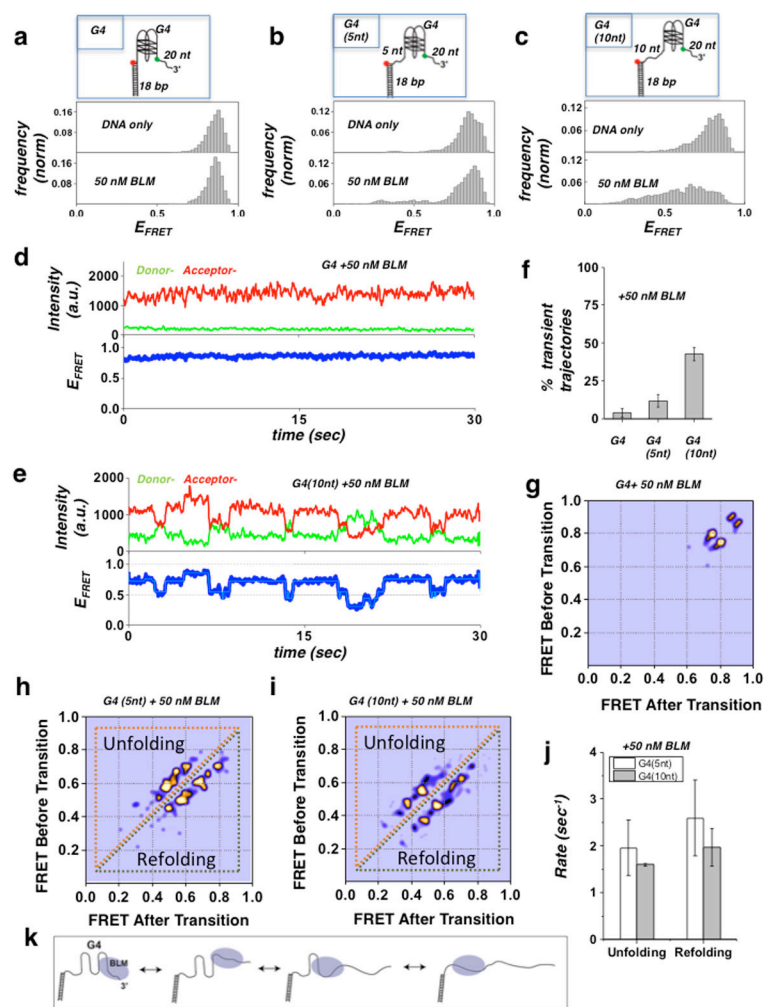


Figure 1. Substrate specific interaction of BLM with G4-forming human telomeric sequence (a–c) BLM mediated unfolding of G4 substrates. The illustration of each G4 DNA substrate is shown above their histograms: (a) G4 substrate, with G4 motif immediately adjacent to duplex region. No change in FRET distribution was observed following addition of 50 nM of BLM. (b) G4 (5nt) substrate with a 5 nt ssDNA between the G4 motif and duplex region. Some change is evident upon addition of 50 nM BLM. (c) G4 (10nt) substrate with a 10 nt ssDNA between the G4 motif and duplex region. A substantial change in the FRET distribution was detected upon addition of 50 nM BLM. Histograms were generated after subtracting the zero FRET values and truncating the photo bleached part from FRET trajectory. A minimum of 100 smFRET trajectories was used to generate the histograms. The concentration of K^+ was kept at 50 mM (d) A representative single molecule trajectory of the G4 substrate in the presence of 50 nM BLM showing persistent high FRET. Top panel: donor (green) - acceptor (red) intensities, bottom panel: corresponding FRET trajectory. (e) Representative smFRET trajectory of the G4 (10nt) substrate in the presence of 50 nM BLM showing dynamic fluctuations in FRET signal (blue). HMM fit is in Cyan.

- (f)** Quantification of observed transient trajectories in population percentage for each G4 substrate in the presence of 50 nM BLM. (Error bars = S.E.M. n=5).
- (g-i)** Generated TDP matrix for each G4 substrate in the presence of 50 nM BLM, where the color intensity corresponds to transition probability. The y-axis is the initial FRET prior to transition, and the x-axis is the final FRET after transition. **(g)** TDP for the G4 substrate with 50 nM BLM showing only the parallel and antiparallel folded conformations of G4. **(h/i)** TDP for the G4 (5 nt) and G4 (10 nt) substrates (respectively), showing peaks corresponding to unfolding (above diagonal) and refolding (below diagonal) transitions of the G4 motif.
- (j)** The calculated mean unfolding and refolding rates for the G4 (5 nt) and G4 (10 nt) substrates. (Error bars = S.E.M. n > 20 for all measurements).
- (k)** Illustration of the BLM-mediated sequential unfolding-refolding transitions of the G4 substrate.

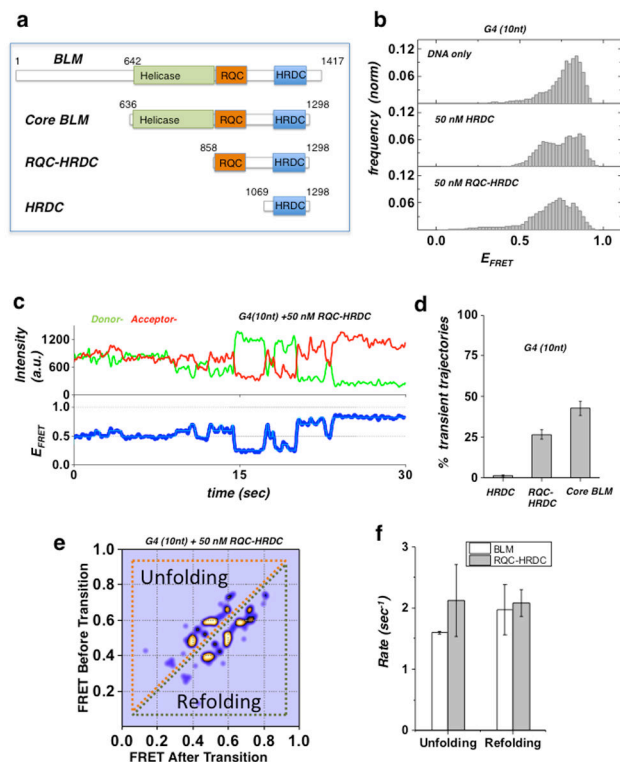


Figure 2. Domain analysis of BLM G4 unfolding activity

- (a) BLM protein functional domains and truncation fragments used in this study: full-length BLM, core BLM, RQC-HRDC fragment and HRDC.
- (b) FRET histogram of the G4 (10nt) substrate showing DNA only substrate (top panel), and in the presence of 50 nM HRDC fragment (mid panel) and RQC-HRDC fragment (bottom panel).
- (c) A representative smFRET trajectory of the G4 (10nt) substrate in the presence of 50 nM RQC-HRDC fragment, showing rapid FRET fluctuations (blue curve). HMM fit is in cyan.
- (d) Quantification of population percentage of transient trajectories for the G4 (10 nt) substrate in the presence of 50 nM of HRDC and RQC-HRDC fragments and core BLM. (Error bar = S.E.M. n=4)
- (e) TDP for the G4 (10 nt) substrate in the presence of 50 nM RQC-HRDC fragment.
- (f) Quantification of the mean unfolding and refolding rates for the G4 motif of the G4 (10 nt) substrate, in the presence of 50 nM RQC-HRDC fragment or core BLM. (Error bar = S.E.M. n > 24 for all measurements)

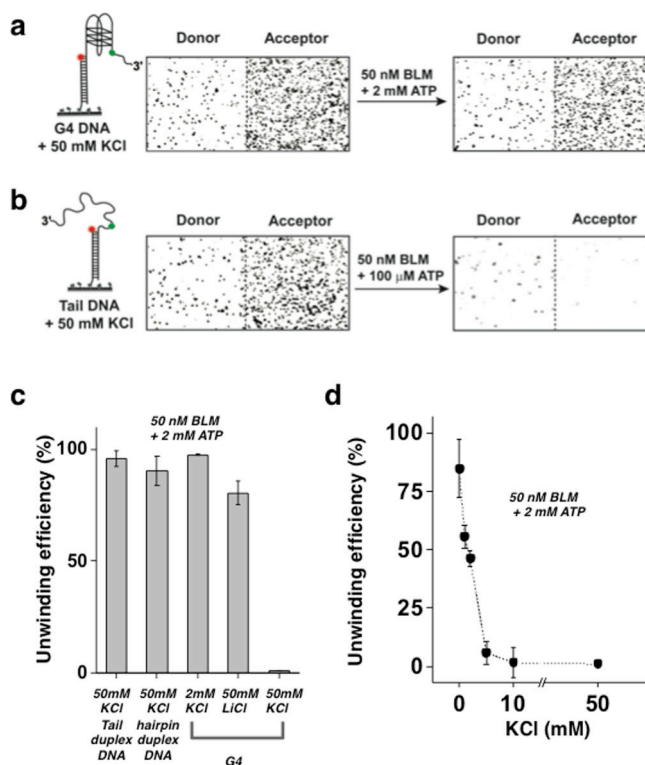


Figure 3. Regulation of BLM unwinding activity by G4 folding

(a/b) Left – Cartoon of the substrates used. Right – Images of individual DNA substrate molecules before and after the unwinding reaction, where donor molecules are shown at the left and corresponding acceptor molecules are at the right. Unwinding was not seen for the G4 substrate (a), whereas complete unwinding was seen for the regular tailed duplex substrate (b).

(c) Quantification of BLM unwinding efficiency at 50 nM BLM and 2 mM ATP for the G4 substrate at various buffer conditions. (i) Regular tailed duplex (Fig. 3b) in the presence of 50 mM KCl. (ii) hairpin substrate (Supplementary Fig. 1A) in the presence of 50 mM KCl. (iii–v) G4 substrate in the presence of: (iii) 2 mM KCl; (iv) 50 mM LiCl; (v) 50 mM KCl. (Error bar = S.E.M. n=4)

(d) BLM unwinding activity of the G4 substrate as a function of KCl concentration (Error bar = S.E.M. n=5)

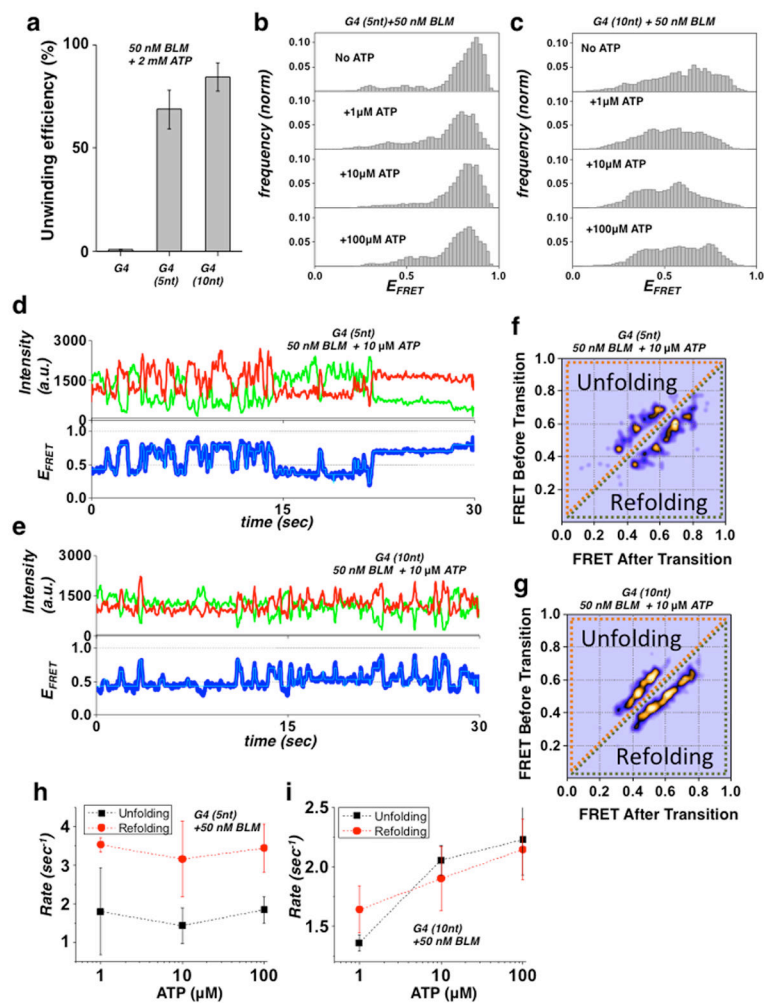


Figure 4. BLM-dependent unwinding activity on a G4 substrate

(a) Quantification of BLM unwinding efficiency at 50 nM BLM and 2 mM ATP for the G4, G4 (5 nt) and G4 (10 nt) substrates (Error bar = S.E.M. n=5)

(b/c) FRET histograms of the G4 (5 nt) and G4 (10 nt) substrates (b and c respectively) in the presence of 50 nM BLM and at different ATP concentrations.

(d/e) Representative single molecule trajectories of the G4 (5 nt) and G4 (10 nt) substrates (d and e respectively) in the presence of 50 nM BLM and 10 μ M ATP. FRET trajectories were fitted with HMM (Cyan).

(f/g) Generated TDP matrix of the G4 (5 nt) and G4 (10 nt) substrates (f and g, respectively) in the presence of 50 nM BLM and 10 μ M ATP.

(h/i) The calculated mean unfolding and refolding rates for the G4 (5 nt) and G4 (10 nt) substrates (h and i respectively) in the presence of 50 nM BLM and at different ATP concentrations. (Error bars = S.E.M. n > 15 for all measurements).

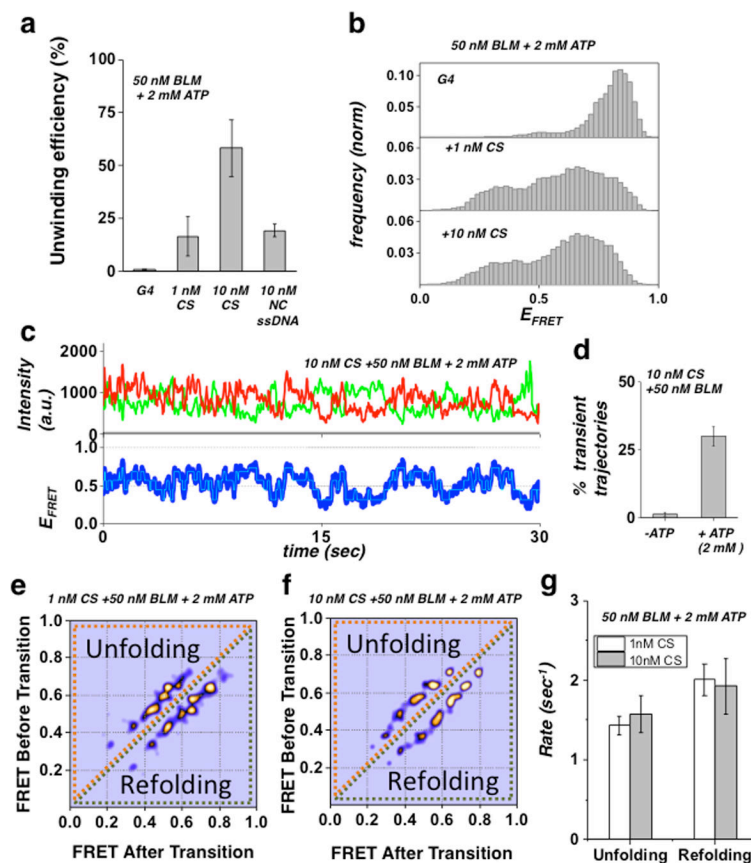


Figure 5. G4 complementary strand-assisted unwinding of the G4 substrate

(a) Quantification of BLM unwinding efficiency at 50 nM BLM and 2 mM ATP for the G4 substrate in the presence of no CS, 1 nM CS or 10 nM CS, or 10 nM NC ssDNA (Error bar = S.E.M. $n=5$)

(b) FRET histograms of the G4 substrate in the presence of 50 nM BLM and 2 mM ATP with either no CS (top panel), 1 nM CS (mid panel) or 10 nM CS (bottom panel).

(c) Representative single molecule trajectories of the G4 substrate in the presence of 50 nM BLM, 2 mM ATP and 10 nM CS. FRET trajectories were fitted with HMM (Cyan).

(d) Quantification of population percentage of transient trajectories for the G4 substrate in the presence of 50 nM of BLM and 10 nM CS and with or without 2 mM ATP. (Error bar = S.E.M. $n=4$)

(e/f) Generated TDP matrix of the G4 substrate in the presence of 50 nM BLM and 2 mM ATP as well as either 1 nM CS or 10 nM CS (panels d and e, respectively).

(g) The calculated mean unfolding and refolding rates for the G4 substrate in the presence of 50 nM BLM, 2 mM ATP and at 1 nM or 10 nM CS concentration. (Error bars = S.E.M. $n > 34$ for all measurements).

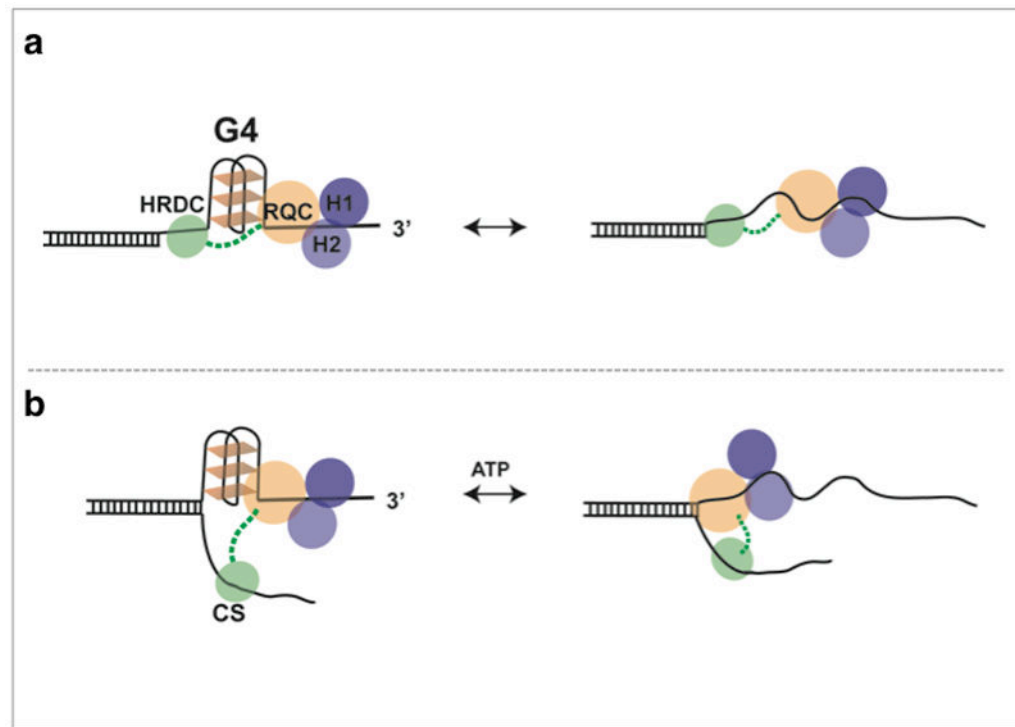


Figure 6. Proposed model for BLM G4 substrate selectivity, unfolding of a G4 motif and regulation of BLM activity via binding cooperativity of RQC-HRDC domains

(a) In binding to substrates containing ssDNA between G4 and duplex DNA, the HRDC domain, via its flexible linker, binds the ssDNA between G4 and the duplex region thus anchoring and stabilizing RQC interaction with G4, facilitating its unfolding. (b) ATP dependent complementary strand assisted unfolding of G4, the HRDC binds CS assisting to destabilize G4 structure.