8742–8753 *Nucleic Acids Research, 2016, Vol. 44, No. 18* *Published online 24 June 2016*
doi: 10.1093/nar/gkw574

G-quadruplex recognition and remodeling by the FANCJ helicase

Colin G. Wu and Maria Spies*

Department of Biochemistry, University of Iowa Carver College of Medicine, 51 Newton Rd., 4-532 BSB, Iowa City, IA 52242, USA

Received May 06, 2016; Revised June 14, 2016; Accepted June 15, 2016

ABSTRACT

Guanine rich nucleic acid sequences can form G-quadruplex (G4) structures that interfere with DNA replication, repair and RNA transcription. The human FANCJ helicase contributes to maintaining genomic integrity by promoting DNA replication through G4-forming DNA regions. Here, we combined single-molecule and ensemble biochemical analysis to show that FANCJ possesses a G4-specific recognition site. Through this interaction, FANCJ targets G4-containing DNA where its helicase and G4-binding activities enable repeated rounds of stepwise G4-unfolding and refolding. In contrast to other G4-remodeling enzymes, FANCJ partially stabilizes the G-quadruplex. This would preserve the substrate for the REV1 translesion DNA synthesis polymerase to incorporate cytosine across from a replication-stalling G-quadruplex. The residues responsible for G-quadruplex recognition also participate in interaction with MLH1 mismatch-repair protein, suggesting that the FANCJ activity supporting replication and its participation in DNA interstrand crosslink repair and/or heteroduplex rejection are mutually exclusive. Our findings not only describe the mechanism by which FANCJ recognizes G-quadruplexes and mediates their stepwise unfolding, but also explain how FANCJ chooses between supporting DNA repair versus promoting DNA replication through G-rich sequences.

INTRODUCTION

G-quadruplexes (G4s) are four-stranded structures formed by guanine-rich nucleic acids. Any single-stranded (ss) DNA sequence containing four stretches of three or more consecutive guanines can fold into a G-quadruplex through Hoogsteen hydrogen bonding between guanines from each run; these interactions are additionally stabilized by monovalent cations such as sodium and potassium (1). G4-forming sequences are highly abundant and fold into stable structures in human cells, with as many as 716 310 unique G-quadruplexes identified within the human genome (2). This staggering number of structures is not formed simultaneously but any that persist can interfere with DNA metabolism. Small molecules that enhance G-quadruplex stability can disrupt DNA replication and RNA transcription by stalling the respective polymerases (3–5). For these essential cellular processes to continue unperturbed, the help of specialized proteins, many of which are helicases, is needed to unfold G-quadruplexes (6,7).

Helicases are motor proteins that use adenosine triphosphate (ATP) to fuel two important biochemical activities: (i) duplex unwinding, where double-stranded (ds) nucleic acids are separated into the intermediates of DNA replication, recombination and repair, RNA transcription and splicing; (ii) translocation or directional movement along nucleic acids, which can be coupled to the removal of nucleoprotein complexes and the remodeling of unconventional DNA structures (8,9). Many human helicases, including FANCJ (Fanconi Anemia Complementation group J), coordinate these fundamental activities to support multiple genome maintenance pathways (1,7,8,10). FANCJ (or BACH1) is a Superfamily-2 (SF2) helicase that not only facilitates DNA replication through G4-forming sequences, but also participates in homologous recombination (HR) and interstrand DNA crosslink (ICL) repair (4,11–14). Defects in FANCJ can lead to Fanconi anemia, a chromosome instability disorder and to increased susceptibility to various cancers (15,16).

FANCJ has 5′→3′ directionality and belongs to a group of human XPD-like DNA helicases, which include XPD, RTEL1 and CHLR1 (17). Crystal structures of archaeal XPD revealed that in addition to the canonical SF2 helicase domains (HD1 and HD2) that form the motor core, an iron-sulfur (FeS) cluster-containing domain and an ARCH domain are inserted into HD1 (Figure 1A) (18–20). This architecture is shared by all XPD-like helicases. FANCJ, however, has two additional elements that are absent from XPD. The first is the C-terminal domain that, upon phosphorylation on serine 990, binds to the BRCT domain of the BRCA1 tumor suppressor (21,22). The second region is

*To whom correspondence should be addressed. Tel: +1 319 335 3221; Fax: +1 319 335 9570; Email: maria-spies@uiowa.edu

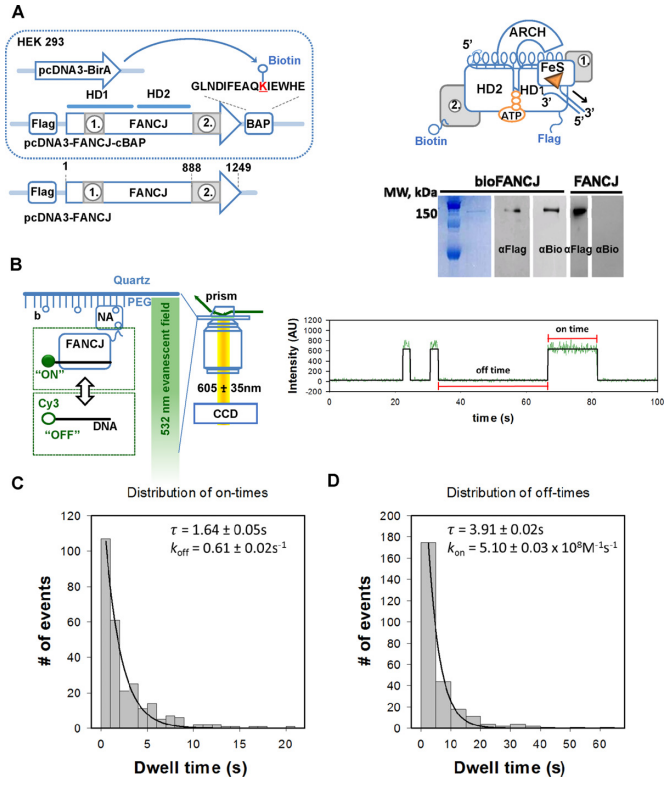

Figure 1. FANCJ for TIRFM experiments. (**A**) N-terminally FLAG-tagged FANCJ helicase was produced in HEK293T cells following transient transfection. The bioFANCJ expression vector also contained a biotin acceptor peptide (GLNDIFEAQKIEWHE) at the C-terminus while the biotin-free FANCJ construct lacked this modification. To produce bioFANCJ, HEK293 cells were co-transfected with pcDNA–BirA. BirA ligase selectively biotinylates the lysine residue within the biotin acceptor peptide. In the cartoon representation of FANCJ, HD1 and HD2 comprise the FANCJ motor core that binds ssDNA and ATP. The FeS containing domain and the ARCH domain are inserted into HD1. The two FANCJ-specific features, a C-terminal domain and a modular insertion in HD1, are shown in gray. Coomasie staining and Western blot analysis of purified bioFANCJ and FANCJ (∼150 kDa) is shown. A full summary including the $FANCJ^{HD}$ and $bioFANCJ^{K141/K142A}$ proteins is provided in Supplementary Figure S1A. (**B**) TIRFM experiment following the binding of freely diffusing Cy3-labled dT_{42} to surface-tethered bioFANCJ. A custom TIRF system was used to generate an evanescent wave for Cy3 illumination. Cy3 emission was collected into an EMCCD camera. A representative dT_{42} binding trajectory (green) from a single bioFANCJ is overlaid with an idealized two-state model (black). (**C**) Dwell-time histogram of the on-times for bioFANCJ- dT_{42} binding was fit to a single exponential decay to determine the time constant, τ_{ON}. The dissociation rate constant for ssDNA binding, k_{off}, was calculated as the inverse of the time constant and was independent of ssDNA concentration (Supplementary Figure S1B). (**D**) Dwell-time distribution histogram of the off-times for bioFANCJ-ssDNA binding was fit to a single exponential decay to determine the time constant, τ_{OFF}. The association rate, V_{on}, was calculated as the inverse of the time constant. The association rate constant, k_{on}, was determined by monitoring the dependence of V_{on} on ssDNA concentration (Supplementary Figure S1B).

a modular insertion in HD1 that contains a nuclear localization sequence (NLS), as well as a binding site for the MLH1 DNA mismatch repair protein (23,24). These unique features may contribute to different biochemical properties ofy FANCJ and XPD.

FANCJ unwinds both intramolecular and intermolecular G-quadruplexes *in vitro* whereas *Thermoplasma acidophilum* (*Ta*) XPD shows no G-quadruplex melting activity; instead, *Ta*XPD prefers to unwind DNA forks and recognizes damaged DNA (5,25–27). Notably, another archaeal XPD helicase, *Sulfolobus acidocaldarius* (Sa) XPD was shown to have a G4 unwinding activity (6). Several other human helicases are also known to unwind G-quadruplexes: another FeS helicase RTEL1; the RecQ-family helicases BLM, WRN and RECQ4; and PIF1 (7,16,28). FANCJ activity is tightly coupled to G-quadruplex maintenance. G4-containing DNA can be efficiently unwound and replicated in the presence of FANCJ, while its depletion sensitizes human cells to G4-stabilizing compounds and causes persistent DNA replication stalling at G-quadruplexes (4). It is possible that one of its specific structural features allows FANCJ to identify and to process G-quadruplex structure. To understand how FANCJ supports DNA replication through G4-containing sequences, we have employed single-molecule Total Internal Reflection Fluorescence Microscopy (TIRFM) and ensemble G4-unfolding assays to investigate the mechanism by which FANCJ mediates G4-remodeling.

MATERIALS AND METHODS

Buffers and reagents

All solutions were prepared using reagent-grade chemicals and double distilled water that was further purified with a Barnstead GenPure system (Thermo Scientific, Waltham, MA, USA). Buffers and reagents were filtered through a 0.2 μm filter after preparation, and all experiments were performed in Buffer H at 25°C (25 mM HEPES (pH 7.0), 100 mM KCl, 10 mM $MgCl_2$, 5% (v/v) glycerol, 5 mM TCEP).

FANCJ expression and purification

BioFANCJ. Full-length human FANCJ helicase was produced and biotinylated in HEK293T cells (25,29,30). Cells were cultured in presence of 5% CO_2 at 37°C in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% fetal bovine serum (FBS), 1 mM sodium pyruvate, 1% penicillin and 1% streptomycin. Upon 90% confluence, cells were transferred to reduced serum media containing 10 μM biotin and were co-transfected with pcDNA3-FANCJ-cBAP and pcDNA3-BirA plasmids using branched polyethylenimine (PEI) (MW: 25 000; Sigma Aldrich, St. Louis, MO, USA). A 1:3 ratio of total plasmid DNA to PEI was used for transient transfection (30 μg pcDNA3-FANCJ-cBAP, 30 μg pcDNA3-BirA, 180 μg PEI per 175 cm^2 growth area). After 24 h, transfected cells were harvested by trypsinization (0.25% Trypsin-EDTA, Sigma Aldrich), washed with ice-cold DPBS (Thermo Scientific) and the cell pellet was stored at −80°C until lysis by freeze-thaw. The cell paste was resuspended in lysis buffer (50 mM HEPES (pH 7.5), 250 mM NaCl, 1 mM EDTA, 0.1% Tween20 and 10% (v/v) glycerol) containing 1 mM PMSF and protease inhibitor cocktail (Roche, San Francisco, CA, USA). The lysate was clarified by centrifugation (20 000 *xg*, 4°C) and anti-FLAG magnetic beads (Sigma Aldrich) were added to the supernatant and incubated with rotation at 4°C for 2 h. The beads were subsequently washed with lysis buffer and bioFANCJ was eluted with Buffer H containing

150 μg/ml 3xFLAG peptide (Sigma Aldrich). Purified bioFANCJ was passed through a Bio-Gel P6 column (Bio-Rad, Hercules, CA, USA) and concentrated with an Amicon Ultra 0.5 ml centrifugal filter (EMD Millipore, Billerica, MA, USA). Protein aliquots were snap frozen in liquid nitrogen and stored at −80°C until use. BioFANCJ was visualized by coomassie staining and western blot analysis after SDS-PAGE. Biotin incorporation was detected with anti-biotin-alkaline phosphatase antibody (Sigma Aldrich) monitoring reaction with ECF substrate (Thermo Scientific) on a ChemiDoc MP imaging system (Bio-Rad). The FLAG epitope was visualized by incubating bioFANCJ with mouse anti-FLAG primary antibody (Sigma Aldrich), followed by anti-mouse-alkaline phosphatase secondary antibody (Sigma Aldrich) and reaction with ECF.

FANCJ, FANCJ^{HD} and bioFANCJ$^{K141/K142A}$. Non-biotinylated full-length human FANCJ helicase and the FANCJ helicase domain were produced in a similar way by transfecting HEK293T cells with either pcDNA3-FANCJ or pcDNA3-FANCJ^{HD} plasmids, both of which lack the biotin acceptor peptide sequence. The pcDNA3-FANCJ^{HD} vector was generated by mutating amino acids 889, 890 and 891 simultaneously to stop codons using pcDNA3-FANCJ as a template, thereby terminating FANCJ synthesis after residue 888. To produce bioFANCJ$^{K141/K142A}$, codons encoding lysine residues 141 and 142 were both mutated to alanine using site directed mutagenesis with pcDNA3-FANCJ-cBAP as a template, and HEK293T cells were co-transfected with pcDNA3-FANCJ$^{K141/K142A}$-cBAP and pcDNA3-BirA plasmids. Purification of FANCJ, FANCJ^{HD} and bioFANCJ$^{K141/K142A}$ was carried out as described above for bioFANCJ.

DNA substrates

DNA oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA, USA). The sequences of the individual oligos and additional modifications are listed in Supplementary Table S1. DNA duplexes were formed by mixing the complementary strands at equimolar ratios in Buffer H, and the resulting mixture was first heated to 95°C for 10 min and then slowly cooled to 25°C over 3 h to allow annealing. Fully annealed DNA substrates were gel purified and stored at 4°C until use.

Ensemble measurements

G-quadruplex unfolding. FANCJ at the indicated concentrations was incubated with 25 nM of a Cy3 labeled G-quadruplex in Buffer H for 5 min at 25°C to allow the protein–DNA complex to form. G4 unfolding was initiated by adding ATP to 10 mM. Simultaneously with ATP, a ssDNA trap strand was introduced in excess (200 nM) over the DNA substrate to hybridize with the unwound quadruplex thereby preventing it from reforming. The unfolding reaction was quenched after 5 min at 25°C with 0.4 M EDTA and 10% SDS, and the products were separated by 15% PAGE. The individual bands were visualized on a ChemiDoc MP imager (Bio-Rad) and the band intensities were determined using Image Lab software (Bio-Rad).

DNA unwinding. FANCJ helicase, FANCJ^{HD} or bioFANCJ$^{K141/K142A}$ was pre-mixed with 25 nM of a Cy3 labeled DNA fork at the indicated concentrations for 5 min at 25°C in Buffer H to allow DNA binding. DNA unwinding was initiated by adding ATP to a final concentration of 10 mM and the reaction was quenched after 5 min at 25°C with 0.4 M EDTA and 10% SDS. The duplex DNA and unwound ssDNA were separated on a 10% polyacrylamide gel, and the individual bands were visualized on a ChemiDoc MP imaging system and the band intensities were quantified using Image Lab software.

DNA binding. FANCJ-DNA binding was monitored by electrophoretic mobility shift assay (EMSA) and by FeS mediated fluorescence quenching. For EMSA experiments, FANCJ, FANCJ^{HD} or bioFANCJ$^{K141/K142A}$ was mixed with 25 nM of a Cy3 labeled G4 at the indicated concentrations for five minutes at 25°C in Buffer H. Protein–DNA complexes were separated from free DNA on a 4% polyacrylamide gel and the individual bands were visualized on a ChemiDoc MP imaging system and analyzed with Image Lab software. Fluorescence DNA binding experiments were carried out on a Cary Eclipse Fluorescence Spectrophotometer (Agilent Technologies, Santa Clara, CA, USA) at 25°C. A total of 40 nM of Cy3 labeled DNA (fork substrate or G-quadruplex) was placed in the cuvette and FANCJ, FANCJ^{HD} or bioFANCJ$^{K141/K142A}$ was titrated into the sample. Cy3 was excited at 520 nm with 10 nm slit widths while Cy3 fluorescence emission was collected at 570 nm with 10 nm slit widths. The relative Cy3 signal change and the extent of fluorescence quenching by the FANCJ FeS cluster was calculated from the following equation, where F_0 is the initial signal from the Cy3 labeled DNA substrate alone and F_i is the fluorescence signal upon the ith addition of protein after correcting for dilution.

$$\Delta F_{obs} = \frac{F_0 - F_i}{F_0}$$

$$\%\ Quenching = (1 - \Delta F_{obs}) \times 100$$

Single-molecule TIRF microscopy

Single-molecule imaging was performed on a custom built prism-type TIRFM system as described previously (26,29–31). Microscope slides and cover slips were prepared as described to sparsely coat the surfaces with biotinylated PEG and neutravidin (26,29–31). A 45 mW 532 nm diode-pumped solid state (DPSS) laser (Coherent) was guided through a Pellin–Broca prism to generate an evanescent field for illuminating Cy3. For single-molecule ssDNA binding experiments, 50 pM bioFANCJ was immobilized on the slide surface and dT_{42}-3′Cy3 was added to the imaging chambers at 100, 200 and 500 pM. For single-molecule DNA unwinding and G-quadruplex unfolding experiments, either 50 pM bioFANCJ or biotinylated DNA substrate was placed on the slide surface. The FANCJ–DNA complex was formed with 100 pM of freely diffusing DNA or FANCJ, and the reaction was initiated by adding 1 mM ATP. All TIRFM experiments were performed in Buffer H containing an oxygen scavenging system (1 mg/ml glucose oxidase, 0.4% (w/v) D-glucose, 0.04

mg/ml catalase, 1% (v/v) 2-mercaptoethanol) in addition to 12 mM Trolox (6-hydroxy-2,5,7,8-tetramethylchromane-2-carboxylic acid). Cy3 and Cy5 (where appropriate) fluorescence emission was collected through a water immersion objective and passed through a Cy3/Cy5 dual bandpass filter into an EMCCD camera. Images were recorded at 10 frames per second and amplification gain of 250 without binning. Single-molecule fluorescence trajectories were extracted from the recorded videos as previously described (26,30). Data collected from two independent experiments were pooled together for analysis.

Analysis of single-molecule data

The individual ssDNA and G-quadruplex binding trajectories were fit to a two-state binding model using QuB software (University of Buffalo). The resulting dwell times for the 'on-states' and 'off-states' were binned and plotted as histograms (30). The resulting dwell time distribution histograms were fit to a single-exponential function using Sigmaplot (Systat Software Inc, San Jose, CA, USA). For single-molecule FRET (smFRET) G-quadruplex unfolding measurements, FRET was calculated from the corrected Cy3 donor and Cy5 acceptor fluorescence intensities using the equation below, where I_A^* and I_D^* are the background corrected acceptor and donor intensities, and γ is the ratio between the change in acceptor fluorescence intensity to the change in donor fluorescence intensity upon photobleaching $\gamma = \Delta I_A^* / \Delta I_D^*$

$$FRET = \frac{\left(I_A^* - \beta\right)}{\left(I_A^* - \beta\right) + \gamma\left(I_D^* - \beta\right)}$$

FRET trajectories collected from a particular FANCJ construct were globally analyzed with hidden Markov modeling using an empirical Bayesian approach (ebFRET Matlab Suite) to detect subpopulation transitions (32). The ebFRET Matlab library was used to statistically determine the number of FRET states observed during the course of G4 melting, and the mean lower bound was used to evaluate the number of states that best described each time series.

RESULTS

FANCJ recognizes ssDNA and G-quadruplexes

Full-length human FANCJ helicase, with or without a biotin label as depicted in Figure 1A (and Supplementary Figure S1A), was produced in HEK293T cells and purified as described in Materials and Methods. To monitor ssDNA binding to individual biotinylated FANCJ molecules (bioFANCJ), equilibrium TIRFM experiments were carried out with bioFANCJ immobilized on a neutravidin-coated slide and Cy3-labeled dT_{42} in the solution. Because FANCJ is monomeric at low concentration (33), the observed activities of the surface-tethered bioFANCJ should reflect those of FANCJ monomers. The association of the Cy3-dT_{42} to the immobilized bioFANCJ and its subsequent dissociation from the helicase resulted in the appearance and disappearance of Cy3 fluorescence intensity. The corresponding trajectories (time-based changes in Cy3 signal at a particular bioFANCJ locus on the slide) collected from individual bioFANCJ molecules contained multiple ssDNA binding and dissociation events (Figure 1B). Each trajectory was fit to a two-state model (black line) that described bioFANCJ as either free (background Cy3 fluorescence) or bound to the ssDNA (high Cy3 fluorescence). The idealized data yielded the on-times and off-times for every binding event. Consistent with this simple model, the dwell-time distribution histograms of the on-times and off-times were both well-described by a single exponential decay function. The time constant, τ_{ON}, obtained from analysis of the on-time distribution histogram (Figure 1C), reflected the average time the ssDNA oligo remained associated with bioFANCJ and was inversely related to the dissociation rate constant, k_{off}. As expected for a unimolecular reaction, k_{off} was independent of Cy3-dT_{42} concentration (Supplementary Figure S1B). In contrast, the time constant determined from analysis of the off-time distribution, τ_{OFF}, measured the average time between successive ssDNA binding events and was related to the on rate, V_{on} (Figure 1D). Binding experiments were repeated at multiple Cy3-dT_{42} concentrations to obtain the association rate constant, k_{on} (Supplementary Figure S1B). The equilibrium dissociation constant, K_D (k_{off}/k_{on}), for the bioFANCJ-dT_{42} interaction was estimated to be 1.2 ± 0.3 nM, which is in excellent agreement with previous ensemble measurements (33).

When the same TIRFM experiments were performed in the presence of 1 mM ATPγS, a slowly hydrolyzed ATP analog, the bioFANCJ-dT_{42} complexes were short lived (Figure 2A), suggesting that ATPγS promoted rapid dissociation of the bioFANCJ-dT_{42} complex. In the presence of ATP, however, bioFANCJ moved along Cy3-dT_{42} with expected 5′→3′ directionality. As bioFANCJ approached the Cy3-labeled 3′ terminus, we observed a gradual decrease in Cy3 signal as a result of FeS-mediated quenching of Cy3 fluorescence as a function of distance (Supplementary Figure S2A) (31,34). The translocation trajectories were sensitive to ATP concentration (Supplementary Figure S2B) and non-processive movements producing partial quenching of Cy3 signal were also detected. Together, these TIRFM data indicate that surface-immobilized bioFANCJ were active in both ssDNA binding and translocation.

Next, we used TIRFM to study the binding of a Cy3-labeled G-quadruplex to bioFANCJ. We first chose the human telomeric repeat sequence (5′-$(TTAGGG)_4$-Cy3-3′) for the following reasons. This sequence had been extensively characterized biochemically and structurally; furthermore, this sequence formed a stable hybrid antiparallel G-quadruplex under our experimental conditions containing 100 mM KCl (35–39). Our G4-binding experiments produced two surprising observations. First, in the absence of a flanking ssDNA overhang, bioFANCJ bound to the G-quadruplex directly (Figure 2B). Second, while ATPγS destabilized the bioFANCJ-dT_{42} complex, the bioFANCJ-G4 interaction was not influenced by ATPγS (Figure 2B and C). We attributed these unexpected results to a potential secondary binding site on FANCJ that can recognize a G-quadruplex structure, distinct from the primary ssDNA binding pocket within its canonical SF2 helicase core. Kinetic analysis of the bioFANCJ-G4 binding trajectories (Figure 2D and E) showed that the affinities of bioFANCJ

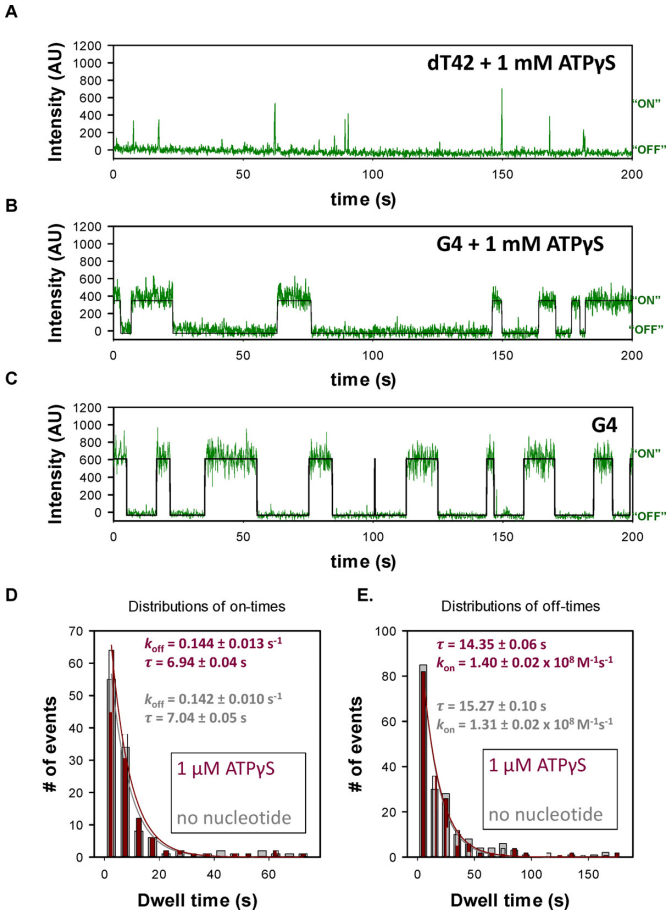

Figure 2. Evidence for a G4-specific binding site in FANCJ. (**A**) TIRFM bioFANCJ-dT_{42} binding trajectory in presence of 1 mM ATPγS shows short lived FANCJ-ssDNA complexes. (**B**) A representative TIRFM bioFANCJ-G4 binding trajectory in the presence of 1 mM ATPγS. All trajectories were fit to a two state model to extract the on-times and off-times associated with each G-quadruplex binding event (black lines). (**C**) A TIRFM bioFANCJ-G4 binding trajectory in the absence of ATPγS. (**D**) Dwell-time distribution histogram of the on-times for G-quadruplex binding either in the presence (red) or absence (gray) of nucleotide co-factor. The distribution was fit to a single exponential decay function to determine the time constants from which the kinetic parameters were calculated. (**E**) Dwell-time distribution histogram of the off-times for G-quadruplex binding either in the presence (red) or absence (gray) of nucleotide co-factor.

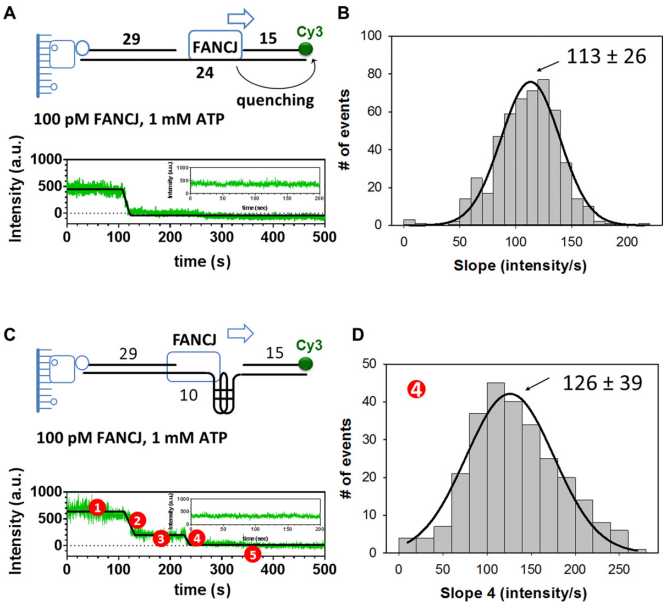

Figure 3. FANCJ unwinds DNA downstream of the G-quadruplex. (**A**) Representative fluorescence trajectory showing FANCJ-catalyzed unwinding of a surface-tethered Cy3-labeled reference dsDNA in the presence of 1 mM ATP. The unwinding trajectory was fit to a three segmented line (black trace) which determined the slope of the transition as well as the intercepts of the plateau regions. The inset shows a representative trajectory in the absence of ATP. (**B**) The distribution of the slopes collected from 505 unwinding trajectories were fit to a Gaussian distribution with a mean of 113 ± 26 intensity/s. (**C**) Representative trajectory demonstrating FANCJ-catalyzed unwinding of a surface-tethered Cy3-labled gaped DNA containing an intervening G-quadruplex in the presence of 1 mM ATP. A trajectory collected without ATP is shown in the inset. The unwinding time-course was fit to a five segmented line (black trace) which provided the values of the plateau regions 1, 3 and 5, as well as the slopes of the two observed transitions 2 and 4. (**D**) The distribution of slopes corresponding to the second transition (slope 4) was analyzed by a Gaussian distribution. A total of 266 traces were examined and provided a mean of 126 ± 26 intensity/s.

for G4 and dT_{42} were similar ($K_D \approx 1$ nM); however, the individual rate constants determined from the two sets of experiments differed considerably (Supplementary Figure S1B versus S2C and S2D). This was consistent with different recognition mechanisms for the two substrates, possibly originating from separate DNA binding sites. Alternatively, ssDNA and G4-binding could occur at the same pocket; however, this would require bioFANCJ to unfold the G-quadruplex upon binding and load the newly released DNA onto the conserved ssDNA interaction motifs.

FANCJ bypasses G-quadruplexes and unwinds downstream DNA

To test whether FANCJ can bypass a G-quadruplex, we first used TIRFM to monitor the unwinding of a surface-tethered reference duplex, and then examined if FANCJ can continue beyond an intervening G-quadruplex and unwind a downstream target. FANCJ-catalyzed unwinding of the reference duplex is shown in the representative trajectory in Figure 3A. In the absence of ATP, all trajectories originating from the immobilized Cy3-labeled substrate showed a stable fluorescence signal (Figure 3A inset). Upon binding to the DNA substrate and in the presence of ATP, FANCJ initiated DNA unwinding by moving with 5′→3′ directionality toward the labeled 3′ terminus. As FANCJ progressed through the duplex, its FeS cluster gradually quenched Cy3 fluorescence. The resulting unwinding trajectories showed a linear decrease in Cy3 signal as FANCJ unwound this reference substrate followed by a plateau at the near background Cy3 level and then a complete loss of Cy3 fluorescence indicative of the dissociation of the unwound oligo or Cy3 photobleaching. Note that none of the trajectories where the quenching was observed showed the FANCJ dissociation prior to the complete duplex unwinding or prior to the Cy3 photobleaching; i.e. the fluorescence recovery has not been observed. Each time course was fit to a three segmented line, from which the slope of the transition as well as the total duration for each unwinding event were

determined, binned and plotted. The slopes of these transitions, which were related to the unwinding rate, were best described by a Gaussian distribution with a mean of 113 ± 26 intensity units per second (Figure 3B). Based on the total starting Cy3 signal and the duplex length, FANCJ unwinding occurred with an upper bound of ∼3–4 bp/s. Notably, DNA unwinding activity was also achieved when the experiment was carried out with surface-tethered bioFANCJ and biotin-free DNA (Supplementary Figure S3A).

When a G-quadruplex is introduced between the FANCJ loading site and the duplex, FANCJ must gain entrance to the dsDNA before it can unwind the reporter. As seen from an unwinding trajectory (Figure 3C), the FeS-mediated quenching of Cy3 signal occurred in two discrete phases instead of a single monotonic decrease. The individual trajectories were fit to a five segmented line consisting of three plateau regions 1, 3 and 5, and slopes 2 and 4, which described the rates of the two characteristic transitions. While slope 2 had a broad and complex distribution (Supplementary Figure S3B), slope 4 was described by a simple Gaussian with a mean of 126 ± 39 intensity units per second (Figure 3D). Interestingly, this value was the same within error as that of the reference duplex. Experiments performed with a G-quadruplex directly adjacent to the dsDNA, without an ssDNA overhang, detected no unwinding activity (Supplementary Figure S3C). This was not due to a lack of DNA binding since surface-immobilized bioFANCJ bound to this substrate (Supplementary Figure S3D). Hence, direct interaction with the G-quadruplex alone was not sufficient for G4-bypass nor DNA unwinding. We conclude from this result that the second observed transition, as indicated by the distribution of slope 4, reflected the unwinding of the downstream duplex while the initial phase (slope 2) contributed to G4-bypass.

FANCJ unfolds G-quadruplex DNA

Next, we monitored the unfolding of a G-quadruplex in an ensemble experiment (5). FANCJ was pre-incubated with a Cy3-labeled quadruplex and G4-unfolding was initiated by adding ATP. An unlabeled 12 nucleotide ssDNA trap (5′-CCCTAACCCTAA-3′) complementary to the quadruplex sequence was introduced along with ATP to prevent the unfolded DNA substrate from refolding. Since the unfolded G-quadruplex hybridized with the ssDNA trap, the resulting partial duplex showed reduced mobility on a polyacrylamide gel. Neither ATP nor ATPγS enabled FANCJ-mediated unfolding of a G-quadruplex in the absence of an ssDNA overhang ((TTAGGG)$_4$-Cy3-3′) (Figure 4A). When FANCJ was given the same DNA substrate with a 5′-dT$_{10}$ overhang, the G-quadruplex was unfolded in an ATP-dependent reaction (Figure 4B, Supplementary Figure S4A), suggesting that motor engagement on the ssDNA as well as motor activity were essential for G4-unfolding.

To test whether the FANCJ C-terminal domain was involved in G-quadruplex recognition, we purified a truncated protein consisting of the helicase domain only (residues 1–888) and performed ensemble G4-unfolding experiments with FANCJ^{HD}. FANCJ^{HD} unfolded G-quadruplexes in the presence of ATP only when the DNA substrate contained a 5′-dT$_{10}$ tail (Figure 4A and B)

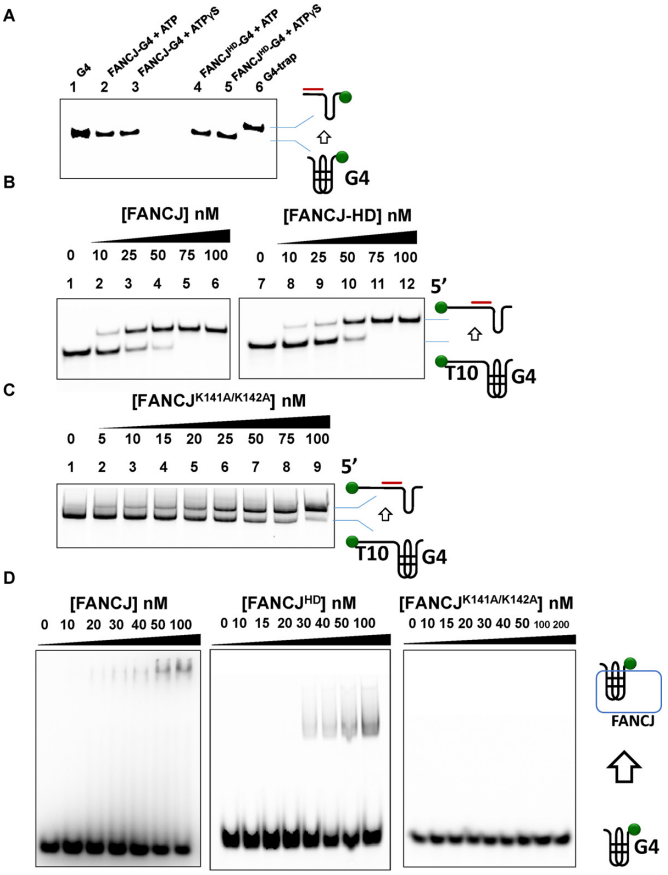

Figure 4. K141/K142A mutations in FANCJ abolished G-quadruplex binding but not unfolding. (**A**) Ensemble G-quadruplex unfolding experiments with FANCJ and FANCJ^{HD}. Proteins were incubated with a Cy3-labeled G-quadruplex DNA and unfolding was initiated by adding nucleotide and ssDNA trap. The reaction was quenched after 5 min and the products are separated by PAGE. The mobility of the quadruplex alone is shown in lane 1. FANCJ was included with ATP (lane 2) or ATPγS (lane 3) and FANCJ^{HD} was included with ATP (lane 4) or ATPγS (lane 5). A heat denatured G-quadruplex that is annealed to the ssDNA trap is shown in lane 6 to illustrate the expected mobility of the unfolded G-quadruplex. (**B**) Quadruplex unfolding experiments were repeated with FANCJ and FANCJ^{HD} using a G-quadruplex possessing a 5′-dT$_{10}$ ssDNA overhang. Lanes 1 and 7 show the folded G4 DNA only while lanes 6 and 12 indicate the mobility of the heat denatured G-quadruplex hybridized with the ssDNA trap. A total of 1 mM ATP was added along with FANCJ or FANCJ^{HD} at the concentrations indicated in all other lanes. (**C**) Quadruplex unfolding experiments using bioFANCJ$^{K141/K142A}$ and 5′T10-G4. The DNA alone is shown in lane 1, while ATP (1 mM) and bioFANCJ$^{K141/K142A}$ were added at the indicated concentrations at all other lanes. (**D**) Formation of the FANCJ-G4 complex was monitored by EMSA. FANCJ, FANCJ^{HD} and bioFANCJ$^{K141/K142A}$ at the indicated concentrations were mixed with 25 nM Cy3-labeled G-quadruplex without a ssDNA overhang.

but formed a complex with a substrate lacking the overhang (Figure 4D). Hence, a G-quadruplex recognition site resided somewhere within FANCJ^{HD} and not the C-terminal domain. Another distinguishing feature of FANCJ is a modular insertion in HD1 that contains a NLS as well as an interaction site for MLH1 (23,24). We purified FANCJ$^{K141/K142A}$, which was deficient in MLH1 binding but not in DNA unwinding (24). As expected, our purified FANCJ, FANCJ^{HD} and FANCJ$^{K141/K142A}$ retained

DNA binding and unwinding activities (Supplementary Figure S4B and S4C). Remarkably, FANCJ$^{K141/K142A}$ did not bind to a G-quadruplex without a 5′-dT$_{10}$ overhang while FANCJ and FANCJ^{HD} formed a stable complex with this substrate (Figure 4D). Therefore, lysine residues K141 and/or K142 formed direct contacts with the G-quadruplex. This observation was unexpected as it suggests that the G-quadruplex and the MLH1 binding are mutually exclusive. FANCJ$^{K141/K142A}$, however, maintained G4-unfolding activity when the DNA substrate possessed an ssDNA overhang for FANCJ loading (Figure 4C). This showed that although we uncovered a FANCJ-G4 recognition site, this interaction may be used only to target FANCJ to G4-containing DNA regions. G-quadruplex unfolding was enabled purely by the motor activity of the helicase.

G-quadruplex unfolding occurs stepwise in repeating cycles

We performed TIRFM experiments with the DNA substrate shown in Figure 5A to examine conformational transitions within the G-quadruplex during FANCJ-catalyzed G4-unfolding. This DNA substrate was derived from Figure 3B except a Cy3/Cy5 dye pair, which can undergo Förster resonance energy transfer (FRET), was incorporated across the G4-containing gap. When the G-quadruplex was folded, the fluorophores were in close proximity and had an average FRET value of ∼0.65. Stable signals from Cy3 and Cy5 were detected with the surface-tethered DNA alone and when FANCJ was added to this substrate (Supplementary Figure S5A). When FANCJ and ATP were both included, G4-unfolding was seen as anti-correlated changes in Cy3 and Cy5 fluorescence intensities, where an increase in Cy3 signal was accompanied by a concomitant decrease in Cy5 signal (Figure 5A). Interestingly, the G-quadruplex in each DNA observed was repeatedly unfolded and refolded. We were surprised by this result since FANCJ was able to unwind a duplex downstream of the G-quadruplex (Figure 3B). However, that experiment was only sensitive to the complete unwinding of the duplex with the Cy3 positioned at the terminus. Indeed when single-molecule unwinding experiments were carried out with a DNA substrate possessing the Cy3 at the FANCJ entrance site, multiple unwinding initiation attempts were observed from the same FANCJ molecule (Supplementary Figure S5B).

An analogous G4-unfolding experiment was performed with bioFANCJ tethered to the slide and freely diffusing DNA (Figure 5B). Without ATP, the association of the DNA substrate to bioFANCJ and its dissociation from the helicase resulted in the simultaneous appearances and disappearances of Cy3/Cy5 signals and G4-unfolding was not detected (Supplementary Figure S5C). When ATP was included, anti-correlated changes in Cy3 and Cy5 fluorescence intensities were observed. This showed that an individual FANCJ helicase molecule was sufficient to unfold and refold a G-quadruplex. Consistent with the ensemble results, FANCJ^{HD} and bioFANCJ$^{K141/K142A}$ also facilitated G4-unfolding and refolding (Figure 5C and D). The smFRET trajectories were analyzed with hidden Markov modeling using empirical Bayesian methods (32). This approach provided a statistically rigorous determination of the number of discrete FRET states observed over the course of FANCJ-catalyzed G4-unfolding, along with their relative occupancies and associated transitions. These results are summarized in the transition density plots (TDPs) in Figure 6A. The G4-unfolding experiments performed with freely diffusing FANCJ as well as surface-tethered bioFANCJ yielded similar TDPs with four characteristic FRET states with average values of 0.13 (1), 0.32 (2), 0.49 (3) and, 0.64 (4). Therefore, surface immobilization did not adversely influence the G4-unfolding activity of FANCJ. Formation of a parallel G-quadruplex structure was also observed albeit infrequently as the high FRET peak at 0.73 (40,41). The four major FRET states indicated that FANCJ-mediated

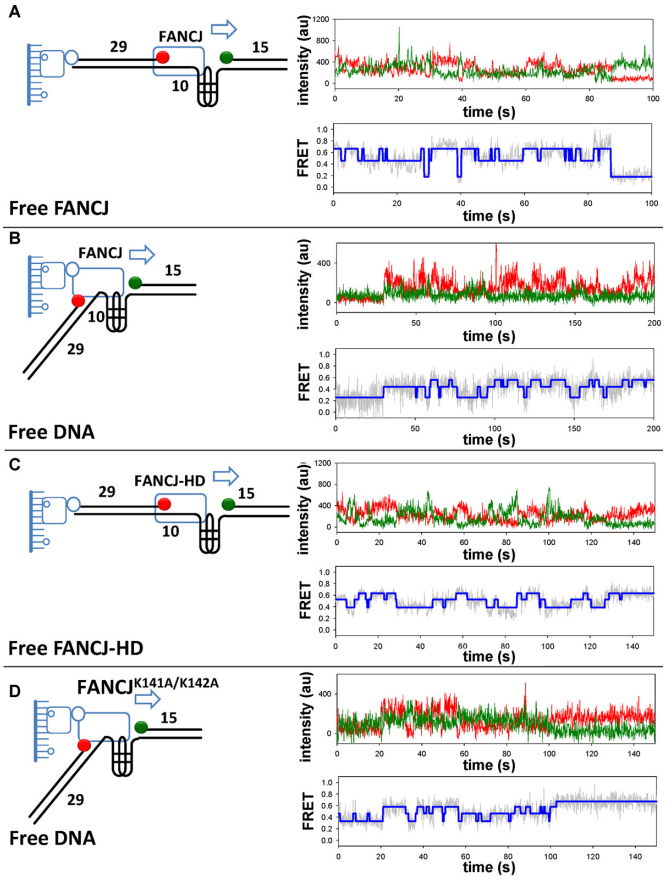

Figure 5. FANCJ-mediated G4-unfolding monitored by smFRET. (**A**) Single-molecule G4 unfolding was examined using surface-immobilized DNA substrate and freely diffusing FANCJ (100 pM) and ATP (1 mM). A representative Cy3 (green) and Cy5 (red) fluorescence trajectory along with the corresponding calculated FRET trajectory (grey) is shown. A total of 72 trajectories were analyzed with hidden Markov modeling (blue idealized trajectory). (**B**) Representative fluorescence (green and red) and FRET (grey) trajectories from the experiment where the G4 unfolding was examined with surface-immobilized bioFANCJ and freely diffusing DNA substrate (100 pM) and ATP (1 mM). A total of 66 traces were analyzed with hidden Markov modeling (blue idealized trajectory). (**C**) FANCJ^{HD}-mediated G4 unfolding was examined with surface immobilized DNA substrate and freely diffusing FANCJ^{HD} (100 pM) and ATP (1 mM). A total of 41 trajectories were analyzed with hidden Markov modeling (blue trace). (**D**) bioFANCJ$^{K141/K142A}$-mediated G4 unfolding was examined with surface-immobilized helicase and freely diffusing DNA substrate (100 pM) and ATP (1 mM). A total of 58 traces were analyzed with hidden Markov modeling (blue trace).

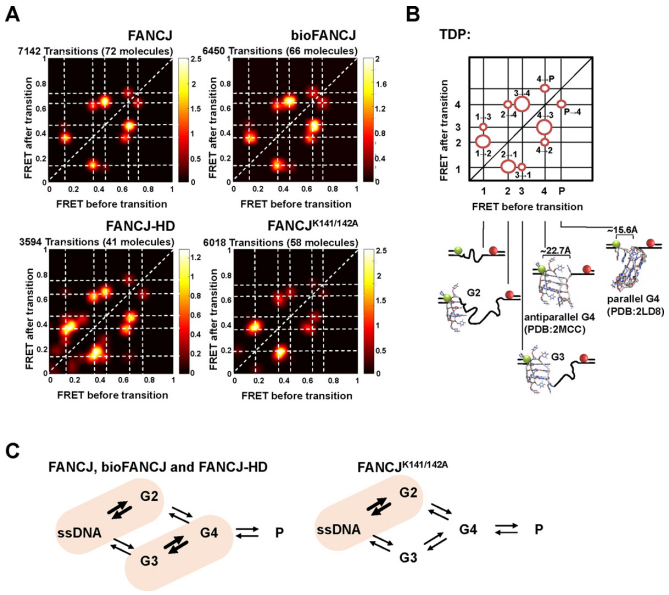

Figure 6. Mechanism of FANCJ-mediated G4-unfolding and refolding. (**A**) Regions of the single-molecule G4-unfolding FRET time-courses for each FANCJ construct containing FRET transitions were globally analyzed by the ebFRET Matlab suite. The resulting transition density plots (TDPs) for each FANCJ construct show four major FRET states with mean FRET values of 0.13 (ssDNA; 1), 0.32 (G2), 0.49 (G3) and 0.64 (G4). Additional, albeit infrequent state (FRET ≈ 0.73) represents a parallel G-quadruplex (P). The scale bar for transition frequency is shown on the right of each respective TDP where the brighter peaks correspond to more frequent transitions. (**B**) Schematic representation of the TDP, observed FRET states and their corresponding DNA conformations, and transitions. (**C**) FANCJ-, bioFANCJ- and FANCJ^{HD}-mediated G4-unfolding and refolding proceeded through the same intermediates and mechanism based on the TDPs. The TDP for the bioFANCJ$^{K141/K142A}$-mediated process shows a reduction in the G2⇔G4 and G3⇔G4 interconversions.

G4-unfolding proceeded through two intermediates (Figure 6B). Each of the transitions likely signifies the removal or rebinding of one (transitions between adjacent states) or two (3⇔1 and 4⇔2 transitions) of the G4 strands (Figure 6C). The TDP of FANCJ^{HD} contained all characteristic states and transitions of the full-length protein, confirming that the C-terminal domain of FANCJ was not involved in processing G-quadruplexes. FANCJ$^{K141/K142A}$-mediated G4-unfolding also showed the same characteristic FRET states; however, the interconversions between the two higher states were severely diminished (G2⇔G4 and G3⇔G4 interconversions). In other words, the fully folded G-quadruplex was not as readily reformed with FANCJ$^{K141/K142A}$. This was consistent with residues K141 and K142 participating in G4-binding since disrupting these interactions reduced the stability of the quadruplex and hindered its refolding. As expected, TIRFM experiments with FANCJ$^{K141/K142A}$ did not detect binding to a G-quadruplex without an ssDNA overhang (Supplementary Figure S6A). FANCJ$^{K141/K142A}$ retained ssDNA binding activity and the FANCJ$^{K141/K142A}$–DNA complex showed the same sensitivity to ATPγS as the wild-type FANCJ (Supplementary Figure S6B).

We next tested if FANCJ and FANCJ$^{K141/K142A}$ can bind and unfold a G-quadruplex formed by the DNA sequence $(GGGT)_4$. This G4 structure is very stable, as it has a melting temperature of over 85°C, compared to $(TTAGGG)_4$ that has a T_m of 60°C (42,43). Single-molecule FRET experiments indicate that both FANCJ (Supplementary Figure S7A) and FANCJ$^{K141/K142A}$ (Supplementary Figure S7B) can repeatedly unfold and refold this high T_m substrate, but through different mechanisms when compared to the human telomeric G-quadruplex. Hidden Markov modeling of the FRET transitions for this G-quadruplex shows five major FRET states: 0.13 (1), 0.21 (2), 0.33 (3), 0.43 (4) and 0.61 (5). Although an initial NMR study suggested that $(GGGT)_4$ forms an antiparallel G-quadruplex, there is increasing recent evidence indicating that this sequence adopts a parallel conformation (44–46). We assume that states 1 and 5 correspond to the fully unfolded and folded G-quadruplex, respectively, but the presence of an additional intermediate state could indicate the folding and unfolding of a mixture of antiparallel and parallel structures. The TDP obtained from wild-type FANCJ shows interconversions from state 4 to state 1; hence, multiple strands of the G-quadruplex can be unfolded and refolded simultaneously (Supplementary Figure S7A). FANCJ$^{K141/K142A}$, on the other hand, unfolds this high T_m G-quadruplex in a step-wise manner (Supplementary Figure S8A). However, there is an accumulation of states 3 and 4, while states 1, 2 and 5 are rarely observed. This suggests that the complete unfolding and complete refolding of this substrate are both impaired when the G-quadruplex interaction site is mutated in FANCJ. Gel mobility shift experiments using $(GGGT)_4$ showed that FANCJ retained G-quadruplex binding activity while FANCJ$^{K141/K142A}$ did not bind to this G-quadruplex (Supplementary Figure S8A and S8B). Taken together, the FANCJ G-quadruplex recognition site is dispensable for G-quadruplex unfolding activity unless the helicase encounters a very stable G4 structure, but is necessary to selectively target different G-quadruplexes and to promote their refolding.

DISCUSSION

The vast majority of helicases examined are able to bind and/or unwind G-quadruplex structures *in vitro* (1). While these observations suggest that G4-unfolding may be a feature shared among helicases, additional information is needed to determine which of these enzymes could have a preference for G4 substrates and participate in G4-remodeling *in vivo*. There is strong experimental evidence that FANCJ directly participates in G4-processing. FANCJ deficiency increases the number of G-quadruplexes formed in human cells; cell lines from Fanconi anemia patients carrying *fancj* mutations have deletions overlapping G4-forming regions (11). Likewise, disruption of the *fancj* homolog *dog-1* in *C.elegans*, also results in genetic deletions upstream of guanine-rich DNA (47,48). Consistent with these results, FANCJ activity is essential in cellular extracts for efficient DNA replication through G-quadruplexes (4). In FANCJ depleted cells, small molecules that stabilize G-quadruplexes restrict replication fork movements, uncouple leading and lagging strand synthesis, and generate small ssDNA gaps (4,10,49). These studies represent the best-documented links between G-quadruplex unfolding in the human genome and the onset of a disease that is connected

to helicase malfunctions. Several other human helicases, including WRN, BLM, RECQ4, RTEL1 and PIF1, have been implicated in the unwinding of G-quadruplexes as well, and mutations in these helicases are similarly associated with genome instability and genetic disorders (28,50–53). Without a detailed understanding of their G4-targeting and unfolding mechanisms, it is impossible to distinguish whether these helicases are functionally redundant, specific to distinctive types of G4 structures or used in separate cellular pathways.

A molecular hot pocket for G4 recognition

Since helicases are molecular motors, directional movement along the nucleic acid lattice alone (i.e. ssDNA translocase activity) can result in the removal of nucleoprotein complexes and even in the partial unwinding of dsDNA (54). This potentially explains why many helicases exhibit G4-unfolding activity *in vitro*, especially when the enzymes are included in vast excess over the DNA substrate. Here, we have shown that in addition to the helicase core that binds and translocates along ssDNA with 5′→3′ directionality, FANCJ also possesses a G4-recognition site. Notably, the two lysine residues (K141/K142) that are involved in G-quadruplex recognition also interact with the MHL1 DNA repair protein. Having the same structural motif involved in both, the nucleic acid and protein recognition is unprecedented. The positively charged lysine residues in the recognition site can form essential electrostatic interactions with the negatively charged DNA backbone. But why does this binding pocket have a preference for the G-quadruplex? Structural, biochemical and single-molecule studies of the closely related XPD helicase have provided valuable insight to how ssDNA and forked DNA are loaded onto FANCJ as well as how its subdomains are spatially arranged (18–20,55,56). However, the FANCJ subdomain containing this G-quadruplex binding site is not present in XPD. An extensive search through available structures and sequences of G4-binding proteins revealed that a peptide derived from the RHAU (or DHX36) protein in complex with a G-quadruplex has several similarities to our findings with FANCJ (57). The G4-binding peptide, Rhau18 (SMHPGHLKGREIGMWYAKKQ), adopts an L-shaped structure, in which a proline residue initiates the formation of an α-helix that spans between G5 and A17. The positively charged K8, R10 and K18/19 are observed to sit this helix against the G-quadruplex (57). The terminating amino acids, including the two consecutive lysines, form a loop that anchors into the G-quadruplex and locks the α-helix in place, while the N-terminal residues are seen pointing away from the DNA, thereby forming an 'L' shape. Lysines 18/19 in Rhau18 are needed for G4-binding, which is analogous to our data with FANCJ$^{K141/K142A}$. The FANCJ amino acid sequence around K141 and K142 (129-PEKTTLAAKLSAKKQ-143) resembles the Rhau18 peptide and even possesses the same 'AKKQ' anchoring motif. We expect proline 129 in FANCJ also initiates the formation of an α-helix that clips onto G4 DNA with a similar mechanism via K131, K137 and K141/142. This is supported by secondary structure prediction analysis of this FANCJ peptide sequence. Notably, this motif is absent in other G4-unwinding DNA helicases, such as Pif1, RTEL1 and the RecQ-family helicases. The discovery of a G4-specific interaction site explains why replication through G4-forming regions relies so strongly on FANCJ, which uses G4-binding along with its motor activity to precisely target the removal of G4 structures.

Mechanism of G-quadruplex unfolding

Having a G4-specific DNA binding pocket implies that FANCJ can also stabilize a folded G-quadruplex. Our smFRET results show that FANCJ can undergo repetitive cycles of G4-unfolding and refolding, during which four major FRET states are observed. These states likely represent a fully folded antiparallel G-quadruplex (FRET ≈ 0.64), an unfolded ssDNA (FRET ≈ 0.13) and two intermediates: three-strand (G3; FRET ≈ 0.49) and two-strand (G2; FRET ≈ 0.32) partially folded structures. The strands are peeled from the G-quadruplex one or two at a time in reversible steps as G4⇔G3, G4⇔G2, G3⇔ssDNA and G2⇔ssDNA (Figure 6). The refolding reaction is presumably aided by direct interactions between FANCJ and the G-quadruplex. This is consistent with our FANCJ$^{K141/142A}$ results, which in contrast with the wild-type protein, showed impaired transitions between the higher FRET states (diminished G2⇔G4 and G3⇔G4 interconversions) as a result of losing its G4-interaction. While stabilizing a G4 seemingly contradicts with its G4-unfolding function, the binding free energy obtained from this interaction provides a potential mechanism for FANCJ to identify the appropriate DNA substrate. Considering a G4-dsDNA junction would be produced during DNA replication as a result of the replicative polymerase stalling at the G4, the interactions between FANCJ and this junction may have important implications to how FANCJ supports DNA replication through G4-forming sequences. While the *S. cerevisiae* Pif1 helicase also unfolds G-quadruplexes in repeating cycles (53), unlike FANCJ, it has specificity for a ds/ssDNA junction and patrols the nearby ssDNA to remove obstacles such as dsDNA, nucleoprotein complexes and G-quadruplexes (53,58). FANCJ, on the other hand, targets G4-containing ssDNA gaps and stabilizes the folded G-quadruplex during its unfolding-refolding cycles. This key distinction and its mechanistic implications are discussed in the model below.

Model for DNA synthesis through G-quadruplexes

Our finding that K141/K142 are involved in FANCJ-G4 binding is unexpected since these lysine residues also participate in interaction with the MLH1 DNA mismatch repair protein. Based on the Rhau18-G4 structure, FANCJ cannot use its AKKQ motif to bind a G-quadruplex and to form protein–protein interactions with MLH1 simultaneously (Figure 7A). Consequently, FANCJ function at a G-quadruplex will be incompatible with its activity that involves MLH1, such as facilitating ICL repair and ensuring the fidelity of homologous recombination via heteroduplex rejection (23,59). This exclusion model could explain how FANCJ partitions between its repair activities and in assisting DNA replication through G4-forming sequences.

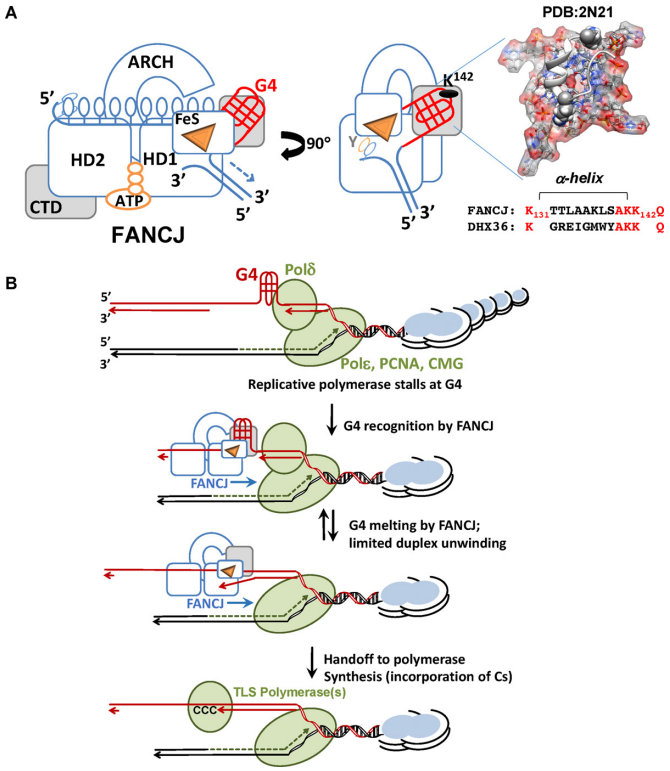

Figure 7. Model for DNA synthesis through G4-forming regions. (**A**) FANCJ possesses a G-quadruplex recognition site within a modular insertion in its HD1. The amino acid sequence at this region is similar to the G-quadruplex recognition peptide seen in the NMR structure of the Rhau18 G4-recognizing peptide (PDB:2N21). The Rhau18 peptide adopts an α-helix that sits on top of the G-quadruplex and is followed by a G4-anchoring loop formed by an 'AKKQ' motif. This motif is also present in FANCJ at K141/K142. (**B**) The AKKQ motif in FANCJ can bind to either MLH1 or to a G-quadruplex. Interaction with MLH1 would allow FANCJ to participate in ICL repair as well as heteroduplex rejection to ensure the fidelity of HR, while G4-binding directs FANCJ to support DNA replication through G4-forming sequences. Lagging strand synthesis is stalled when Polδ encounters a G-quadruplex. FANCJ is targeted to the resulting G4-ssDNA gap and undergoes G4-unfolding and refolding. This process is repeated until FANCJ recruits a translesion DNA synthesis polymerase such as REV1 either through direct interactions via their respective PIP-RIR sequences or through mediated interactions with PCNA.

Efficient replication of G-quadruplexes was shown to require FANCJ working in concert with either the REV1 translesion DNA synthesis (TLS) polymerase, or the WRN and BLM DNA helicases although it is not known why there are two FANCJ-dependent pathways (60). Since mutations in *C. elegans* DOG-1, a FANCJ-like helicase, led to defects in lagging strand synthesis (47), we envision that Polδ is stalled when it encounters a G-quadruplex in the lagging strand (Figure 7B). Note, that while the cartoon representing our model is drawn with the lagging strand G4, a similar scenario can be envisioned for the leading strand. FANCJ is then targeted to this DNA substrate via direct interactions with the G-quadruplex and the resulting ssDNA gap at the 5′-side of the block. G4-unfolding then proceeds in the 5′→3′ direction since it is driven by FANCJ helicase/translocase activity. The FeS domain is thought to facilitate strand separation during DNA unwinding and is positioned in front of the two motor domains (55) while the G-quadruplex is held by the AKKQ motif. Once FANCJ unfolds the last remaining strand of the G-quadruplex, a DNA polymerase can use the newly released ssDNA as a template for synthesis. The dependence on REV1 and FANCJ to maintain epigenetic stability near G-quadruplexes suggests that REV1 may be the DNA polymerase for this reaction (60). Moreover, since REV1 has a specificity for incorporating cytosine, it would efficiently replicate DNA across from the unfolded G-quadruplex (60,61). The partial stabilization of the G-quadruplex as a result of FANCJ binding may be needed to keep FANCJ nearby until REV1 is recruited to the DNA, which may prevent the refolding of G4 DNA structures (62). FANCJ likely recruits REV1 either directly or through interactions with PCNA, since residues 1001–1017 of FANCJ (SWSSFNSLGQYFTGKIP) form a putative PCNA-interacting (PIP) motif, which can also serve as a REV1-interacting (RIR) regions, as shown for other PIP sequences (63). It is likely that after the DNA synthesis across the first few Gs, REV1 hands the substrate over to other TLS or replicative polymerase (62). Our model for G-quadruplex recognition and remodeling by FANCJ is based on the structure formed by the human telomeric repeat sequence and by a more stable $(GGGT)_4$ G4. The unfolding of other G-quadruplex arrangements formed by different DNA sequences may involve more complex mechanisms and necessitate a division of labor among the G4-unfolding helicases, such as the simultaneous activities of the 5′→3′ FANCJ helicase and the 3′→5′ WRN or BLM helicases (16,60).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Dr Marc S. Wold and Dr Todd M. Washington, and members of the Spies lab for valuable scientific discussions and comments on the manuscript.

FUNDING

National Institutes of Health [NIH R01 GM101167 to M.S]; American Cancer Society [PF-11-243-01-DMC to C.G.W]. Funding for open access charge: National Institutes of Health [NIH R01 GM108617].
Conflict of interest statement. None declared.

REFERENCES

1. Bochman,M.L., Paeschke,K. and Zakian,V.A. (2012) DNA secondary structures: stability and function of G-quadruplex structures. *Nat. Rev. Genet.*, **13**, 770–780.
2. Chambers,V.S., Marsico,G., Boutell,J.M., Di Antonio,M., Smith,G.P. and Balasubramanian,S. (2015) High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.*, **33**, 877–881.
3. Piazza,A., Boule,J.B., Lopes,J., Mingo,K., Largy,E., Teulade-Fichou,M.P. and Nicolas,A. (2010) Genetic instability triggered by G-quadruplex interacting Phen-DC compounds in Saccharomyces cerevisiae. *Nucleic Acids Res.*, **38**, 4337–4348.

4. Castillo Bosch,P., Segura-Bayona,S., Koole,W., van Heteren,J.T., Dewar,J.M., Tijsterman,M. and Knipscheer,P. (2014) FANCJ promotes DNA synthesis through G-quadruplex structures. *EMBO J.*, **33**, 2521–2533.
5. Bharti,S.K., Sommers,J.A., George,F., Kuper,J., Hamon,F., Shin-ya,K., Teulade-Fichou,M.P., Kisker,C. and Brosh,R.M. Jr (2013) Specialization among iron-sulfur cluster helicases to resolve G-quadruplex DNA structures that threaten genomic stability. *J. Biol. Chem.*, **288**, 28217–28229.
6. Gray,L.T., Vallur,A.C., Eddy,J. and Maizels,N. (2014) G quadruplexes are genomewide targets of transcriptional helicases XPB and XPD. *Nat. Chem. Biol.*, **10**, 313–318.
7. Mendoza,O., Bourdoncle,A., Boule,J.B., Brosh,R.M. Jr and Mergny,J.L. (2016) G-quadruplexes and helicases. *Nucleic Acids Res.*, **44**, 1989–2006.
8. Lohman,T.M., Tomko,E.J. and Wu,C.G. (2008) Non-hexameric DNA helicases and translocases: mechanisms and regulation. *Nat. Rev. Mol. Cell Biol.*, **9**, 391–401.
9. Wu,C.G. and Spies,M. (2013) Overview: What are helicases? *Adv. Exp. Med. Biol.*, **973**, 1–16.
10. Brosh,R.M. Jr and Cantor,S.B. (2014) Molecular and cellular functions of the FANCJ DNA helicase defective in cancer and in Fanconi anemia. *Front. Genet.*, **5**, 1–14.
11. London,T.B., Barber,L.J., Mosedale,G., Kelly,G.P., Balasubramanian,S., Hickson,I.D., Boulton,S.J. and Hiom,K. (2008) FANCJ is a structure-specific DNA helicase associated with the maintenance of genomic G/C tracts. *J. Biol. Chem.*, **283**, 36132–36139.
12. Wu,Y., Shin-ya,K. and Brosh,R.M. Jr (2008) FANCJ helicase defective in Fanconia anemia and breast cancer unwinds G-quadruplex DNA to defend genomic stability. *Mol. Cell. Biol.*, **28**, 4116–4128.
13. Hiom,K. (2010) FANCJ: solving problems in DNA replication. *DNA Repair (Amst)*, **9**, 250–256.
14. Beyer,D.C., Ghoneim,M.K. and Spies,M. (2013) Structure and Mechanisms of SF2 DNA Helicases. *Adv. Exp. Med. Biol.*, **767**, 47–73.
15. Litman,R., Peng,M., Jin,Z., Zhang,F., Zhang,J., Powell,S., Andreassen,P.R. and Cantor,S.B. (2005) BACH1 is critical for homologous recombination and appears to be the Fanconi anemia gene product FANCJ. *Cancer Cell*, **8**, 255–265.
16. Brosh,R.M. Jr (2013) DNA helicases involved in DNA repair and their roles in cancer. *Nat. Rev. Cancer*, **13**, 542–558.
17. White,M.F. (2009) Structure, function and evolution of the XPD family of iron-sulfur-containing 5′–>3′ DNA helicases. *Biochem. Soc. Trans.*, **37**, 547–551.
18. Wolski,S.C., Kuper,J., Hanzelmann,P., Truglio,J.J., Croteau,D.L., Van Houten,B. and Kisker,C. (2008) Crystal structure of the FeS cluster-containing nucleotide excision repair helicase XPD. *PLoS Biol.*, **6**, e149.
19. Fan,L., Fuss,J.O., Cheng,Q.J., Arvai,A.S., Hammel,M., Roberts,V.A., Cooper,P.K. and Tainer,J.A. (2008) XPD helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations. *Cell*, **133**, 789–800.
20. Liu,H., Rudolf,J., Johnson,K.A., McMahon,S.A., Oke,M., Carter,L., McRobbie,A.M., Brown,S.E., Naismith,J.H. and White,M.F. (2008) Structure of the DNA repair helicase XPD. *Cell*, **133**, 801–812.
21. Cantor,S.B., Bell,D.W., Ganesan,S., Kass,E.M., Drapkin,R., Grossman,S., Wahrer,D.C., Sgroi,D.C., Lane,W.S., Haber,D.A. *et al.* (2001) BACH1, a novel helicase-like protein, interacts directly with BRCA1 and contributes to its DNA repair function. *Cell*, **105**, 149–160.
22. Clapperton,J.A., Manke,I.A., Lowery,D.M., Ho,T., Haire,L.F., Yaffe,M.B. and Smerdon,S.J. (2004) Structure and mechanism of BRCA1 BRCT domain recognition of phosphorylated BACH1 with implications for cancer. *Nat. Struct. Mol. Biol.*, **11**, 512–518.
23. Cantor,S.B. and Xie,J. (2010) Assessing the link between BACH1/FANCJ and MLH1 in DNA crosslink repair. *Environ. Mol. Mutagen.*, **51**, 500–507.
24. Peng,M., Litman,R., Xie,J., Sharma,S., Brosh,R.M. Jr and Cantor,S.B. (2007) The FANCJ/MutLalpha interaction is required for correction of the cross-link response in FA-J cells. *EMBO J.*, **26**, 3238–3249.
25. Pugh,R.A., Lin,Y., Eller,C., Leesley,H., Cann,I.K. and Spies,M. (2008) Ferroplasma acidarmanus RPA2 facilitates efficient unwinding of forked DNA substrates by monomers of FacXPD helicase. *J. Mol. Biol.*, **383**, 982–998.
26. Ghoneim,M. and Spies,M. (2014) Direct correlation of DNA binding and single protein domain motion via dual illumination fluorescence microscopy. *Nano Lett.*, **14**, 5920–5931.
27. Mathieu,N., Kaczmarek,N. and Naegeli,H. (2010) Strand- and site-specific DNA lesion demarcation by the xeroderma pigmentosum group D helicase. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 17545–17550.
28. Vannier,J.B., Pavicic-Kaltenbrunner,V., Petalcorin,M.I., Ding,H. and Boulton,S.J. (2012) RTEL1 dismantles T loops and counteracts telomeric G4-DNA to maintain telomere integrity. *Cell*, **149**, 795–806.
29. Masuda-Ozawa,T., Hoang,T., Seo,Y.S., Chen,L.F. and Spies,M. (2013) Single-molecule sorting reveals how ubiquitylation affects substrate recognition and activities of FBH1 helicase. *Nucleic Acids Res.*, **41**, 3576–3587.
30. Bain,F.E., Wu,C.G. and Spies,M. (2016) Single-molecule sorting of DNA helicases. *Methods*, doi:10.1016/j.ymeth.2016.05.009.
31. Honda,M., Park,J., Pugh,R.A., Ha,T. and Spies,M. (2009) Single-molecule analysis reveals differential effect of ssDNA-binding proteins on DNA translocation by XPD helicase. *Mol. Cell*, **35**, 694–703.
32. van de Meent,J.W., Bronson,J.E., Wiggins,C.H. and Gonzalez,R.L. Jr (2014) Empirical Bayes methods enable advanced population-level analyses of single-molecule FRET experiments. *Biophys. J.*, **106**, 1327–1337.
33. Wu,Y., Sommers,J.A., Loiland,J.A., Kitao,H., Kuper,J., Kisker,C. and Brosh,R.M. Jr (2012) The Q motif of Fanconi anemia group J protein (FANCJ) DNA helicase regulates its dimerization, DNA binding, and DNA repair function. *J. Biol. Chem.*, **287**, 21699–21716.
34. Pugh,R.A., Honda,M. and Spies,M. (2010) Ensemble and single-molecule fluorescence-based assays to monitor DNA binding, translocation, and unwinding by iron-sulfur cluster containing helicases. *Methods*, **51**, 313–321.
35. Long,X. and Stone,M.D. (2013) Kinetic partitioning modulates human telomere DNA G-quadruplex structural polymorphism. *PLoS One*, **8**, e83420.
36. Petraccone,L., Erra,E., Esposito,V., Randazzo,A., Mayol,L., Nasti,L., Barone,G. and Giancola,C. (2004) Stability and structure of telomeric DNA sequences forming quadruplexes containing four G-tetrads with different topological arrangements. *Biochemistry*, **43**, 4877–4884.
37. Tran,P.L., Mergny,J.L. and Alberti,P. (2011) Stability of telomeric G-quadruplexes. *Nucleic Acids Res.*, **39**, 3282–3294.
38. Balagurumoorthy,P. and Brahmachari,S.K. (1994) Structure and stability of human telomeric sequence. *J. Biol. Chem.*, **269**, 21858–21869.
39. Heddi,B. and Phan,A.T. (2011) Structure of human telomeric DNA in crowded solution. *J. Am. Chem. Soc.*, **133**, 9824–9833.
40. Gray,R.D., Trent,J.O. and Chaires,J.B. (2014) Folding and unfolding pathways of the human telomeric G-quadruplex. *J. Mol. Biol.*, **426**, 1629–1650.
41. Lee,J.Y., Okumus,B., Kim,D.S. and Ha,T. (2005) Extreme conformational diversity in human telomeric DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 18938–18943.
42. Ogloblina,A.M., Bannikova,V.A., Khristich,A.N., Oretskaya,T.S., Yakubovskaya,M.G. and Dolinnaya,N.G. (2015) Parallel G-quadruplexes formed by guanine-rich microsatellite repeats inhibit human topoisomerase I. *Biochemistry (Mosc.)*, **80**, 1026–1038.
43. Petraccone,L., Spink,C., Trent,J.O., Garbett,N.C., Mekmaysy,C.S., Giancola,C. and Chaires,J.B. (2011) Structure and stability of higher-order human telomeric quadruplexes. *J. Am. Chem. Soc.*, **133**, 20951–20961.
44. Jing,N., Gao,X., Rando,R.F. and Hogan,M.E. (1997) Potassium-induced loop conformational transition of a potent anti-HIV oligonucleotide. *J. Biomol. Struct. Dyn.*, **15**, 573–585.
45. Johnson,J., Okyere,R., Joseph,A., Musier-Forsyth,K. and Kankia,B. (2013) Quadruplex formation as a molecular switch to turn on intrinsically fluorescent nucleotide analogs. *Nucleic Acids Res.*, **41**, 220–228.
46. Do,N.Q., Lim,K.W., Teo,M.H., Heddi,B. and Phan,A.T. (2011) Stacking of G-quadruplexes: NMR structure of a G-rich

oligonucleotide with potential anti-HIV and anticancer activity. *Nucleic Acids Res.*, **39**, 9448–9457.
47. Cheung,I., Schertzer,M., Rose,A. and Lansdorp,P.M. (2002) Disruption of dog-1 in Caenorhabditis elegans triggers deletions upstream of guanine-rich DNA. *Nat. Genet.*, **31**, 405–409.
48. Kruisselbrink,E., Guryev,V., Brouwer,K., Pontier,D.B., Cuppen,E. and Tijsterman,M. (2008) Mutagenic capacity of endogenous G4 DNA underlies genome instability in FANCJ-defective C. elegans. *Curr. Biol.*, **18**, 900–905.
49. Schwab,R.A., Nieminuszczy,J., Shin-ya,K. and Niedzwiedz,W. (2013) FANCJ couples replication past natural fork barriers with maintenance of chromatin structure. *J. Cell Biol.*, **201**, 33–48.
50. Budhathoki,J.B., Ray,S., Urban,V., Janscak,P., Yodh,J.G. and Balci,H. (2014) RecQ-core of BLM unfolds telomeric G-quadruplex in the absence of ATP. *Nucleic Acids Res.*, **42**, 11528–11545.
51. Sanders,C.M. (2010) Human Pif1 helicase is a G-quadruplex DNA-binding protein with G-quadruplex DNA-unwinding activity. *Biochem. J.*, **430**, 119–128.
52. Paeschke,K., Capra,J.A. and Zakian,V.A. (2011) DNA replication through G-quadruplex motifs is promoted by the Saccharomyces cerevisiae Pif1 DNA helicase. *Cell*, **145**, 678–691.
53. Zhou,R., Zhang,J., Bochman,M.L., Zakian,V.A. and Ha,T. (2014) Periodic DNA patrolling underlies diverse functions of Pif1 on R-loops and G-rich DNA. *Elife*, **3**, e02190.
54. Betterton,M.D. and Julicher,F. (2005) Opening of nucleic-acid double strands by helicases: active versus passive opening. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, **71**, 011904.
55. Kuper,J., Wolski,S.C., Michels,G. and Kisker,C. (2012) Functional and structural studies of the nucleotide excision repair helicase XPD suggest a polarity for DNA translocation. *EMBO J.*, **31**, 494–502.
56. Pugh,R.A., Wu,C.G. and Spies,M. (2012) Regulation of translocation polarity by helicase domain 1 in SF2B helicases. *EMBO J*, **31**, 503–514.
57. Heddi,B., Cheong,V.V., Martadinata,H. and Phan,A.T. (2015) Insights into G-quadruplex specific recognition by the DEAH-box helicase RHAU: Solution structure of a peptide-quadruplex complex. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 9608–9613.
58. Paeschke,K., Bochman,M.L., Garcia,P.D., Cejka,P., Friedman,K.L., Kowalczykowski,S.C. and Zakian,V.A. (2013) Pif1 family helicases suppress genome instability at G-quadruplex motifs. *Nature*, **497**, 458–462.
59. Honda,M., Okuno,Y., Hengel,S.R., Martin-Lopez,J.V., Cook,C.P., Amunugama,R., Soukup,R.J., Subramanyam,S., Fishel,R. and Spies,M. (2014) Mismatch repair protein hMSH2-hMSH6 recognizes mismatches and forms sliding clamps within a D-loop recombination intermediate. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, E316–E325.
60. Sarkies,P., Murat,P., Phillips,L.G., Patel,K.J., Balasubramanian,S. and Sale,J.E. (2012) FANCJ coordinates two pathways that maintain epigenetic stability at G-quadruplex DNA. *Nucleic Acids Res.*, **40**, 1485–1498.
61. Sarkies,P., Reams,C., Simpson,L.J. and Sale,J.E. (2010) Epigenetic instability due to defective replication of structured DNA. *Mol. Cell*, **40**, 703–713.
62. Eddy,S., Ketkar,A., Zafar,M.K., Maddukuri,L., Choi,J.Y. and Eoff,R.L. (2014) Human Rev1 polymerase disrupts G-quadruplex DNA. *Nucleic Acids Res.*, **42**, 3272–3285.
63. Boehm,E.M., Powers,K.T., Kondratick,C.M., Spies,M., Houtman,J.C. and Washington,M.T. (2016) The proliferating cell nuclear antigen (PCNA)-interacting protein (PIP) motif of DNA polymerase eta mediates its interaction with the C-terminal domain of Rev1. *J. Biol. Chem.*, **291**, 8735–8744.