

Structural insights into the functions of the FANCM-FAAP24 complex in DNA repair

Hui Yang^{1,2}, Tianlong Zhang¹, Ye Tao^{1,2}, Fang Wang¹, Liang Tong³ and Jianping Ding^{1,*}

¹State Key Laboratory of Molecular Biology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 320 Yue-Yang Road, Shanghai 200031, China, ²Graduate School of Chinese Academy of Sciences, 320 Yue-Yang Road, Shanghai 200031, China and ³Department of Biological Sciences, Columbia University, New York, NY 10027, USA

Received June 26, 2013; Revised and Accepted August 12, 2013

ABSTRACT

Fanconi anemia (FA) is a genetically heterogeneous disorder associated with deficiencies in the FA complementation group network. FA complementation group M (FANCM) and FA-associated protein 24 kDa (FAAP24) form a stable complex to anchor the FA core complex to chromatin in repairing DNA interstrand crosslinks. Here, we report the first crystal structure of the C-terminal segment of FANCM in complex with FAAP24. The C-terminal segment of FANCM and FAAP24 both consist of a nuclease domain at the N-terminus and a tandem helix-hairpin-helix (HhH)₂ domain at the C-terminus. The FANCM-FAAP24 complex exhibits a similar architecture as that of ApXPF. However, the variations of several key residues and the electrostatic property at the active-site region render a catalytically inactive nuclease domain of FANCM, accounting for the lack of nuclease activity. We also show that the first HhH motif of FAAP24 is a potential binding site for DNA, which plays a critical role in targeting FANCM-FAAP24 to chromatin. These results reveal the mechanistic insights into the functions of FANCM-FAAP24 in DNA repair.

INTRODUCTION

Fanconi anemia (FA) is a rare genetic disease characterized by bone marrow failure, developmental defects and increased incidence of cancers (1,2). FA is genetically heterogeneous and can be caused by defects in any of 15 FA genes. The FA proteins function in resolving DNA interstrand crosslinks (ICLs) and stalled replication forks during replication via the FA pathway (1,2). As the central event of the FA pathway, the eight upstream FA proteins (FANCA, FANCB, FANCC,

FANCE, FANCF, FANCG, FANCL and FANCM) cooperate with five FA-associated proteins (FAAP20, FAAP24, FAAP100, MHF1 and MHF2) to form the FA core complex, which acts as an E3 ubiquitin ligase to monoubiquitinate the FANCI-FANCD2 complex and then activates the downstream DNA repair pathways (3–14).

FANCM is a key component of the FA core complex, which contains a DEAH helicase domain at the N-terminus and a putative nuclease domain and a tandem helix-hairpin-helix (HhH)₂ motif domain at the C-terminus. The DEAH helicase domain harbors an ATP-dependent DNA-remodeling translocase activity, which is important for the FA core complex to monoubiquitinate FANCI-FANCD2 (15–19). FANCM can interact with the FANCM-associated histone-fold proteins 1 and 2 (MHF1-MHF2) complex via the region following the helicase domain (residues 661–800) and with FAAP24 via the C-terminal domains to form two stable complexes to bind different DNA substrates, which are crucial for FANCM to facilitate repair of ICLs (3–5). FANCM can also interact with the RMI1-RMI2 complex via its MM2 motif (residues 1218–1251) to recruit the BLM-RMI1-Topo III α dissolvasome to ICL-stalled replication forks (20–22).

The FANCM-FAAP24 complex binds preferentially to single-stranded DNA (ssDNA). It plays a key role in the recruitment of the FA core complex to damaged DNA and the monoubiquitylation of FANCD2 (5,18). It is also essential for ataxia-telangiectasia and Rad-3-related (ATR)-mediated S phase checkpoint signaling (23–25). Moreover, in addition to their coordinated functions, the two proteins have non-overlapping and distinct functions: FAAP24 promotes ATR-mediated checkpoint activation particular in response to DNA ICL agents, whereas FANCM participates in recombination-independent ICL repair by facilitating recruitment of lesion incision activities (26). The FANCM-MHF complex binds preferentially to branched DNA and double-stranded DNA (dsDNA) *in vitro* and stimulates the DNA branch

*To whom correspondence should be addressed. Tel: +86 21 5492 1619; Fax: +86 21 5492 1116; Email: jpding@sibcb.ac.cn

migration and replication fork reversal activities of FANCM (3,4). The MHF1-MHF2 complex assumes a heterotetrameric architecture similar to that of the histone (H3-H4)₂ heterotetramer and interacts with FANCM to form a binding site for DNA (27,28).

FANCM and FAAP24 are suggested to belong to the XPF family (5,16). All members of this family contain an excision repair cross-complementation group 4 (ERCC4) nuclease domain and a tandem helix-hairpin-helix (HhH)₂ motif domain and exist as heterodimers in eukaryotes and homodimers in archaea. These proteins can be divided into two groups: catalytic members and non-catalytic members. The catalytic members such as XPF and Mus81 contain a conserved GDX_nERKX₃D motif at the active site of the nuclease domain and exhibit nuclease activity, and the non-catalytic members such as Emel, Eme2 and ERCC1 lack the catalytic motif (29,30). Previous data showed that FANCM has a divergent sequence CDX_nERRX₃E at the equivalent region of the catalytic motif and FAAP24 lacks the catalytic motif, and the FANCM-FAAP24 complex has no detectable nuclease activity *in vitro* (5,16). The (HhH)₂ domain is composed of two tandem HhH motifs, which can mediate the binding of different DNA structures and dimerization of the two proteins (31–34).

We report here the crystal structure of human FANCM-FAAP24 complex, which together with the biological data provide the molecular insights into its functions in DNA repair. The FANCM-FAAP24 heterodimer structurally resembles the ApXPF homodimer. Although the putative active site of FANCM has a similar structure as that of XPF, differences of several key residues and the electrostatic surface of the surrounding region render an inactive nuclease domain, accounting for the lack of nuclease activity. The first HhH motif of FAAP24 is a potential DNA binding site, which is essential for proper localization of FANCM-FAAP24 to chromatin in DNA repair.

MATERIALS AND METHODS

Cloning, expression and purification

The gene fragments encoding the C-terminal regions of human FANCM (residues 1727–2048, 1799–2048 and 1813–2031) and human FAAP24 (full-length and residues 17–215) were inserted into the pET-Duet plasmid (Novagen) and the pET-28a plasmid (Novagen), respectively; the latter attaches a His₆ tag at the C-terminal of FAAP24. The two plasmids were co-transformed into *Escherichia coli* BL21 (DE3) Codon-Plus strain (Novagen). The transformed cells were grown at 37°C in lysogeny broth (LB) medium until OD₆₀₀ reached 0.8 and then induced with 0.2 mM isopropyl β-D-1-thiogalactopyranoside. The cells were collected by centrifugation, suspended in buffer A [20 mM Tris-HCl (pH 8.0), 500 mM NaCl, 10% glycerol and 1 mM phenylmethylsulfonyl fluoride] and lysed by sonication. The FANCM-FAAP24 complex was purified by affinity chromatography using a Ni-NTA column (Qiagen) with buffer A supplemented with 20 mM imidazole and

200 mM imidazole serving as washing buffer and elution buffer, respectively. The elution sample was further purified by gel filtration using a Superdex 200 16/60 column (GE Healthcare) pre-equilibrated with buffer B [10 mM HEPES (pH 7.5), 100 mM NaCl and 2 mM MgCl₂]. For the Se-Met derivative proteins, the cells were grown in M9 medium supplemented with amino acids Lys, Thr, Phe, Leu, Ile, Val, Se-Met and 1% lactose. The FANCM and FAAP24 mutants containing different point mutations were generated using the QuikChange[®] Site-Directed Mutagenesis Kit (Stratagene). The resultant protein samples were of >95% purity as analyzed by SDS-PAGE.

Crystallization, data collection and structure determination

Crystallization was performed using the hanging drop vapor diffusion method at 20°C. Crystals of both native and Se-Met FANCM-FAAP24 were grown from drops consisting of 1 μl of protein solution (~6 mg/ml) and 1 μl of reservoir solution containing 1.0 M (NH₄)₂SO₄ and 0.1 M HEPES (pH 7.2). For diffraction data collection, the crystals were cryoprotected using the reservoir solution supplemented with 30% glycol and then flash-cooled into liquid nitrogen. Three Se-derivative datasets of 3.4 Å resolution and a native data set of 2.9 Å resolution were collected at 100 K at beamline 17U of Shanghai Synchrotron Radiation Facility, China. The diffraction data were processed with HKL2000 (35). The statistics of the diffraction data are summarized in Table 1.

The structure of FANCM-FAAP24 was solved using the multi-wavelength anomalous dispersion method using Phenix (36), which identified 7 Se atoms and yielded a figure of merit of 0.44. The structure model was refined against the native data set using Phenix (36) and Refmac5 (37). Model building was performed using Coot (38). The final structure model contains residues 1815–2030 of FANCM and residues 18–213 of FAAP24, except that three surface exposed loops (residues 1903–1915 and 1965–1969 of FANCM, and residues 147–152 of FAAP24) could not be modeled owing to poor electron density. Structure analysis was carried out using programs in CCP4 (39) and the PISA server (40). All the structure figures were prepared using Pymol (<http://www.pymol.org>). The statistics of the structure refinement and the quality of the final structure model are also summarized in Table 1.

Electrophoretic mobility shift assay

Electrophoretic mobility shift assays (EMSA) were performed as described previously (5). The ssDNA, dsDNA and splayed-arm DNA substrates were also prepared as described previously (5,41). For comparison, all DNAs were labeled with 5'-biotin in oligo 1 and oligo 3. Reaction mixture (20 μl) contained 10 nM 5'-biotin-labeled DNA in a buffer containing 25 mM Tris-HCl (pH 7.5), 5 mM EDTA and 6% glycerol. Reactions were initiated by adding different amounts of FANCM-FAAP24 and then incubated for 30 min at 4°C.

Table 1. Summary of diffraction data and structure refinement statistics

	Peak	Infection	Remote	Native
Diffraction data				
Wavelength (Å)	0.9795	0.9792	0.9568	0.9796
Space group	<i>I</i> 222	<i>I</i> 222	<i>I</i> 222	<i>I</i> 222
Cell parameters				
<i>a</i> (Å)	80.1	80.3	80.0	79.2
<i>b</i> (Å)	98.2	98.2	98.1	98.1
<i>c</i> (Å)	139.1	139.1	139.0	140.2
Resolution (Å)	50.0–3.40 (3.52–3.40) ^a	50.0–3.40 (3.52–3.40)	50.0–3.40 (3.52–3.40)	50.0–2.90 (3.00–2.90)
Observed reflections	38 469	38 197	38 209	40 533
Unique reflections ($I/\sigma(I) > 0$)	7996	7965	7942	12 155
Average redundancy	4.8 (4.9)	4.8 (4.8)	4.8 (4.9)	3.3 (3.4)
Average $I/\sigma(I)$	14.4 (3.4)	13.8 (2.7)	14.4 (4.3)	20.5 (2.8)
Completeness (%)	99.8 (100.0)	99.8 (100.0)	99.7 (99.9)	97.6 (97.5)
R_{merge} (%) ^b	15.1 (67.3)	15.6 (85.6)	13.8 (53.4)	6.8 (50.2)
Refinement and structure model				
Reflections ($F_o \geq 0\sigma(F_o)$)				
Working set				11 554
Test set				598
R factor/Free R factor (%) ^c				25.4/29.9
No. of Protein atoms				3,058
No. of Water atoms				8
Average B factor (Å ²)				
All atoms				69.7
FANCM main chain/side chain				62.2/64.6
FAAP24 main chain/side chain				75.7/77.2
Water				42.6
RMS deviations				
Bond lengths (Å)				0.006
Bond angles (°)				1.1
Ramachandran plot (%)				
Most favoured				90.5
Allowed				8.4
Generously allowed				1.1

^aNumbers in parentheses represent the highest resolution shell.

^b $R_{\text{merge}} = \frac{\sum_{\text{hkl}} \sum_i |I_i(\text{hkl}) - \langle I(\text{hkl}) \rangle|}{\sum_{\text{hkl}} \sum_i I_i(\text{hkl})}$.

^c $R = \frac{\sum_{\text{hkl}} ||F_o| - |F_c||}{\sum_{\text{hkl}} |F_o|}$.

Reaction mixture was loaded on a 6% neutral polyacrylamide gel prepared in 0.5x TBE buffer (45mM Tris–borate (pH 8.0) and 1mM EDTA). Gels were run in 0.5x TBE buffer at 100 V for 100–120 min at 4°C and then transferred to a positively charged nylon membrane (GE Healthcare) in 0.5x TEB buffer at 100 V for 60 min at 4°C. The nylon membrane was cross-linked at 120 mJ/cm² using UVC 5000 Crosslinker (Hofer). The biotin-labeled DNA was detected by LightShift Chemiluminescent EMSA Kit (Thermo Scientific).

Cell culture, transfection and immunofluorescence

The genes of the full-length FANCM and the C-terminal segment of FANCM (residues 1799–2048, FANCM_{L2}) in wild-type, truncated and mutant forms were subcloned into the pEGFP-C3 vector (Clontech), which attaches an enhanced green fluorescent protein (EGFP) tag at the N-terminus, and the genes of the full-length FAAP24 in wild-type, truncated and mutant forms into the pcDNA3 vector (Invitrogen), which attaches a Flag tag at the N-terminus. HEK 293 T cells were cultured on coverslips in Dulbecco's modified Eagle's medium (Hyclone) supplemented with 10% fetal bovine serum (Biochrom AG)

and transfected with 2 μg of plasmids using Lipofectamine 2000 (Invitrogen). Twenty-four hours after transfection, the cells were fixed with 4% paraformaldehyde (Sigma), washed three times with phosphate buffered saline (PBS) and permeabilized using PBS supplemented with 0.1% Triton X-100 and then blocked using 2% bovine serum albumin in PBS supplemented with 0.1% Triton X-100. Flag-FAAP24 was stained with mouse anti-Flag M2 mAb (1:1000, Sigma) followed by Alexa Fluor 647 goat anti-mouse mAb (1:1000, Invitrogen). DNA was stained with 4',6-diamidino-2-phenylindole (DAPI, 1:2000, Sigma). Images were acquired utilizing a 63X oil immersion lens on a Leica TCS SP5 II confocal microscope (Leica).

RESULTS AND DISCUSSION

Structure of the FANCM-FAAP24 complex

It was reported previously that two forms of the C-terminal fragment of human FANCM (residues 1727–2048, FANCM_{L1}; residues 1799–2048, FANCM_{L2}) can form stable complexes with full-length FAAP24 (residues 1–215, FAAP24_{FL}), which exhibit binding

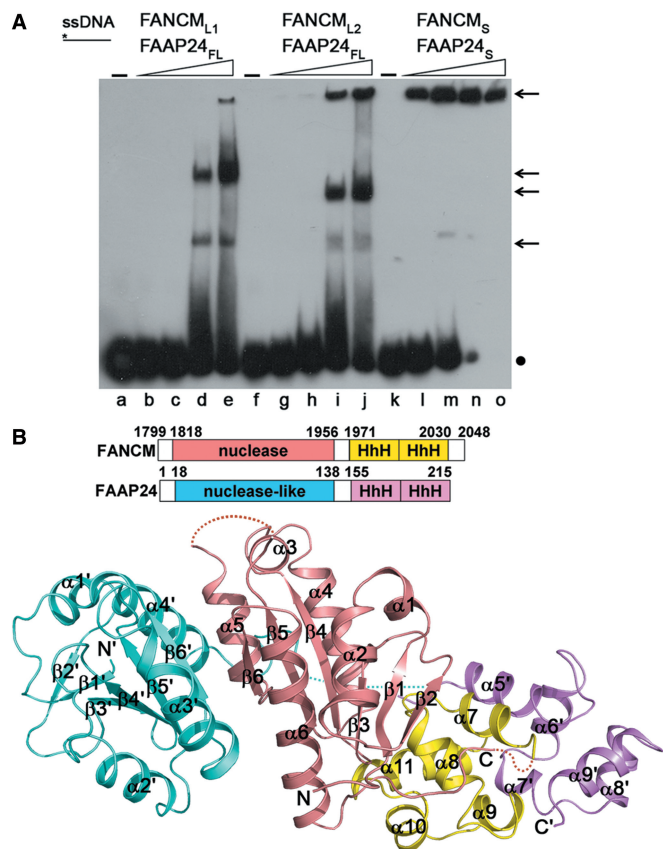


Figure 1. Structure of the FANCM-FAAP24 complex. (A) EMSA of the DNA-binding ability of FANCM-FAAP24 on ssDNA. The reaction contained the biotin-labeled ssDNA (10 nM) without or with increasing amounts of FANCM-FAAP24 (0.0 μ M, lanes a, f and k; 0.05 μ M, lanes b, g and l; 0.1 μ M, lanes c, h and m; 0.4 μ M, lanes d, i and n; and 0.8 μ M, lanes e, j and o). The asterisk represents the biotin label at the DNA 5' end. The arrows indicate the shifted bands of the protein-DNA complexes, and the dot represents the free DNA probe. (B) Ribbon representation of the overall structure of FANCM-FAAP24. The nuclease and the (HhH)₂ domains of FANCM are colored in salmon and yellow, and those of FAAP24 in cyan and violet, respectively. The secondary structure elements are labeled with those of FAAP24 designated with apostrophes. The disordered regions are indicated with dotted lines.

activities to different types of DNA structures with preferences for ssDNA, splayed-arm DNA and a 3'-flap DNA 5–10-fold better than for dsDNA (5). We obtained these two forms of the FANCM-FAAP24 complex with high purity, stability and homogeneity and confirmed their binding to different types of DNA structures (Figure 1A and Supplementary Figure S1), but failed to obtain any crystals. We then constructed a slightly shorter form of FANCM (residues 1813–2031, FANCM_S) and an N-terminal truncated form of FAAP24 (residues 17–215, FAAP24_S), and the resultant FANCM_S-FAAP24_S complex led to successful crystallization and structure determination of the complex (Figure 1B). We also confirm that the FANCM_S-FAAP24_S complex maintains the capability to bind different types of DNA substrates (Figure 1B and Supplementary Figure S1). It is noteworthy that there are several shifted bands in the EMSA results. As the FANCM-FAAP24 complex can bind

different types of DNA structures in a non-sequence-specific way, and the DNA substrates are relatively long, it is possible that a single DNA substrate might bind one or more protein molecules under the assay condition, leading to multiple shifted bands. This speculation is supported in part by the observation that the bands with larger molecular masses appear to be more intense at higher protein concentration. The FANCM_S-FAAP24_S complex appears to bind the DNA substrates to form large protein-DNA complexes with higher molecular masses, which barely move in the gels. For simplicity, we will refer to this complex as the FANCM-FAAP24 complex hereafter unless otherwise specified.

The crystal structure of the FANCM-FAAP24 complex was solved using the multi-wavelength anomalous dispersion method and refined to 2.9 Å resolution (Figure 1B and Table 1). There is one FANCM-FAAP24 complex in the asymmetric unit, which contains residues 1815–2030 of FANCM and residues 18–213 of FAAP24 except that three solvent-exposed loops (residues 1903–1915 and 1965–1969 of FANCM and residues 147–152 of FAAP24) are disordered. The N-terminal ends of both FANCM and FAAP24 are not located in the protein-protein interface (Figure 1B), consistent with our biochemical data showing that the N-terminal truncations of these proteins do not affect the stability of the complex and the binding to DNA (Figure 1A and Supplementary Figure S1).

Both FANCM and FAAP24 are composed of a nuclease domain at the N-terminus and an (HhH)₂ domain at the C-terminus, although they share only 20% sequence identity. The nuclease domains of both FANCM and FAAP24 contain a central six-stranded β -sheet flanked by helices on both sides (Figure 1B). The (HhH)₂ domains of both FANCM and FAAP24 are composed of five helices, namely, α 7– α 11 and α 5'– α 9', respectively (components of FAAP24 are designated by apostrophes) (Figure 1B). The overall structures of the equivalent domains in the two proteins are similar with root-mean-square deviation of 1.9 Å and 1.6 Å, respectively (Figure 1B and Supplementary Figure S2A and B). However, the relative position and orientation of the two domains are significantly different: the nuclease domain of FANCM is sandwiched between and interacts with the (HhH)₂ domain of FANCM and the nuclease domain of FAAP24, whereas the nuclease domain of FAAP24 is far apart from and has no interaction with the (HhH)₂ domain of FAAP24 (Figure 1B and Supplementary Figure S2C). The two nuclease domains and the two (HhH)₂ domains form two compact heterodimers, respectively.

Interactions between FANCM and FAAP24

The interactions between the nuclease domain and the (HhH)₂ domain of FANCM are mediated mainly by residues from α 4 and the β 2– β 3, β 3– β 4, α 4– α 5 and α 6– α 7 loops of the nuclease domain, and α 8 and the α 7– α 8 and α 10– α 11 loops of the (HhH)₂ domain, which involve numerous hydrophobic contacts and two hydrogen bonds and bury a total of 655 Å² solvent-accessible

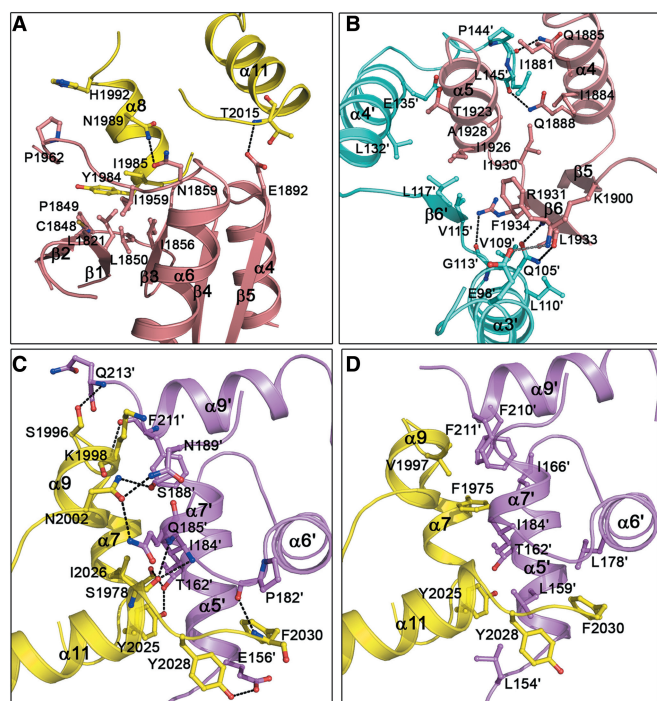


Figure 2. Interactions between FANCM and FAAP24. (A) Interactions between the nuclease and the (HhH)₂ domains of FANCM. The interface is mainly mediated by residues from $\alpha 4$ and the $\beta 2$ - $\beta 3$, $\beta 3$ - $\beta 4$, $\alpha 4$ - $\beta 5$ and $\alpha 6$ - $\alpha 7$ loops of the nuclease domain, and $\alpha 8$ and the $\alpha 7$ - $\alpha 8$ and $\alpha 10$ - $\alpha 11$ loops of the (HhH)₂ domain. (B) Interactions between the nuclease domains of FANCM and FAAP24. The interface is mainly mediated by residues from $\alpha 4$, $\alpha 5$ and $\beta 6$ of FANCM and $\alpha 3'$, $\alpha 4'$, $\beta 6'$ and the $\alpha 4'$ - $\alpha 5'$ loop of FAAP24. (C) Hydrophilic interactions and (D) hydrophobic interactions between the (HhH)₂ domains of FANCM and FAAP24. The interface is mainly mediated by residues from $\alpha 7$, $\alpha 9$ and $\alpha 11$ of FANCM and $\alpha 5'$, $\alpha 6'$, $\alpha 7'$ and $\alpha 9'$ of FAAP24. The nuclease and the (HhH)₂ domains of FANCM are colored in salmon and yellow, and those of FAAP24 in cyan and violet, respectively. The interacting residues are shown with ball-and-stick models. The hydrogen bonds are indicated with black dashed lines and the salt bridge with gray dashed line.

surface area (Figures 1B and 2A and Supplementary Table S1). The N-terminal of $\alpha 8$ of the (HhH)₂ domain is embedded in a hydrophobic surface patch formed by the $\beta 2$ - $\beta 3$, $\beta 3$ - $\beta 4$, $\alpha 6$ - $\alpha 7$ loops of the nuclease domain, which contribute more than half of the hydrophobic contacts and one hydrogen bond at this interface (Supplementary Table S1). Helix $\alpha 4$ and the following $\alpha 4$ - $\beta 5$ loop of the nuclease domain mainly interact with the $\alpha 7$ - $\alpha 8$ and $\alpha 10$ - $\alpha 11$ loops of the (HhH)₂ domain, which also contribute nearly half of the hydrophobic contacts and one hydrogen bond at this interface (Supplementary Table S1).

The two nuclease domains form a heterodimer with pseudo 2-fold symmetry such that the C-terminal edges of the two β -sheets ($\beta 6$) are in contact with each other (Figures 1B and 2B). The interaction interface involves mainly residues from $\alpha 4$, $\alpha 5$ and $\beta 6$ of FANCM and $\alpha 3'$, $\alpha 4'$, $\beta 6'$ and the $\alpha 4'$ - $\alpha 5'$ loop of FAAP24 and buries a total of 1021 Å² solvent-accessible surface area (Figures 1B and 2B and Supplementary Table S2). The interactions are largely hydrophobic but also comprise 5 hydrogen

bonds and 1 salt bridge. Specifically, the interactions between $\beta 6$ and $\alpha 5$ of FANCM and $\beta 6'$, $\alpha 3'$ and $\alpha 4'$ of FAAP24 contribute more than two-thirds of the hydrophobic contacts and four hydrophilic interactions at this interface. The interactions between $\alpha 4$ of FANCM and the $\alpha 4'$ - $\alpha 5'$ loop of FAAP24 contribute the rest of the hydrophobic contacts and two hydrogen-bonding interactions.

The two (HhH)₂ domains also form a heterodimer with pseudo 2-fold symmetry, and the interaction interface involves mainly residues from $\alpha 7$, $\alpha 9$ and $\alpha 11$, and the connecting loops of FANCM and $\alpha 5'$, $\alpha 6'$, $\alpha 7'$ and $\alpha 9'$, and the connecting loops of FAAP24, and buries a total of 950 Å² solvent-accessible surface area (Figures 1B and 2C and D and Supplementary Table S2). Similarly, the interactions are largely hydrophobic but also comprise 11 hydrogen bonds. Helix $\alpha 9$ and the preceding loop of FANCM mainly interact with $\alpha 7'$, $\alpha 9'$, the $\alpha 7'$ - $\alpha 8'$ loop, and the C-terminal of FAAP24, which contribute one fourth of the hydrophobic contacts and five hydrogen bonds at this interface. Helix $\alpha 11$ and the C-terminal of FANCM mainly interact with $\alpha 5'$, $\alpha 7'$, and the $\alpha 4'$ - $\alpha 5'$ and $\alpha 6'$ - $\alpha 7'$ loops of FAAP24, which contribute more than half of the hydrophobic contacts and five hydrogen bonds. The extensive hydrophilic and hydrophobic interactions between the two (HhH)₂ domains are in agreement with the observation that the (HhH)₂ domains mediate the dimerization in other XPF family members (31,32,34).

Structural comparisons with other XPF family members

Previously, the structures of ApXPF (*Aeropyrum pernix*, Ap) and DrMus81-HsEme1 (*Danio rerio*, Dr; *Homo sapiens*, Hs) containing both the nuclease and the (HhH)₂ domains, the dimeric (HhH)₂ domains of human XPF and human XPF-ERCC1, and the central domain of human ERCC1 have been determined (31–34,42,43). Structural comparisons show that the nuclease and the (HhH)₂ domains of both FANCM and FAAP24 resemble those of ApXPF, DrMus81 and HsEme1, with root-mean-square deviations of 1.5–2.5 Å (Supplementary Figure S3A and B). The structure of the FANCM-FAAP24 heterodimer has the highest similarity to that of the ApXPF homodimer (Supplementary Figure S3C and D), even though they share ~25% sequence identity (Supplementary Figure S4A). It was suggested previously that FANCM and XPF may have a common ancestor with a helicase domain at the N-terminal and a nuclease domain at the C-terminal (16). Although the structure of human XPF-ERCC1 containing both the nuclease and the (HhH)₂ domains is unavailable yet, the structural similarities between FANCM-FAAP24 and ApXPF and between the dimeric (HhH)₂ domains of FANCM-FAAP24 and human XPF-ERCC1 support this notion and further suggest that the overall structure of XPF-ERCC1 should be similar to that of FANCM-FAAP24.

Structural comparisons also show that the position of the dimeric (HhH)₂ domains relative to the dimeric nuclease domains in FANCM-FAAP24 is different from those in ApXPF and DrMus81-HsEme1. When these complexes are superimposed based on the dimeric nuclease domains, the position of the (HhH)₂ domains

in FANCM-FAAP24 differs by a rotation of 60° from to that in the dsDNA-bound ApXPF, a rotation of 125° from to that in the apo ApXPF, and a rotation of 125° from to that in DrMus81-HsEme1 (Supplementary Figure S3E–G). The previous structural data have shown that when dsDNA binds to ApXPF, the (HhH)₂ domains are rotated toward the nuclease domains, and the dsDNA is bound in the surface groove formed between the nuclease and the (HhH)₂ domains (31). The variations in the relative arrangements of the nuclease and the (HhH)₂ domains in these complexes might reflect their differed binding abilities and/or specificities for different DNA structures.

Molecular basis for the lack of nuclease activity

The catalytic members of the XPF family contain a conserved GDX_nERKX₃D motif in the active site of the nuclease domain, which is responsible for the nuclease activity. The previous mutagenesis studies of HsXPF showed that the first two acidic residues of the catalytic motif participate in metal binding; however, the functional roles of the two basic residues Arg and Lys and the last acidic residue Asp in catalysis have not been fully understood yet (29). Consistently, the structure of ApXPF-dsDNA showed that the side chains of the first two acidic residues (Asp52 and Glu62) and the main chain of the first basic residue (Arg63) are involved in the binding of Mg²⁺, but the side chains of the two basic residues (Arg63 and Lys64) and the last Asp (Asp68) are not (31). Sequence comparison of FAAP24 from different vertebrates shows that the catalytic motif is completely absent in FAAP24 (Supplementary Figure S4B). Although the nuclease domain of FAAP24 has a similar structure as that of ApXPF and DrMus81, the corresponding region of the catalytic motif becomes PDX_nLYVX₃D in FAAP24 (Figure 3A and Supplementary Figure S4A), suggesting the lack of capability to bind a metal ion that is required for the nuclease activity.

On the other hand, sequence comparison of FANCM from different vertebrates shows that the catalytic motif has evolved to CDX_nERRX₃E in FANCM; in particular, the conserved Lys of the motif is substituted with Arg (Arg1866) and the last Asp is substituted with Glu (Glu1870) in human FANCM, both of which are shown to be involved in catalysis (Supplementary Figure S4A and C). Structural comparison shows that FANCM also has a similar active-site structure as that of ApXPF and DrMus81, and particularly all the key residues of the catalytic motif including the altered two residues assume similar side-chain conformations (Figure 3A). Nevertheless, no metal ion is found at the active site of FANCM, despite the fact that there is 2 mM MgCl₂ in the protein buffer solution. We speculate that the altered residues in the catalytic motif of FANCM might change the coordination environment of the metal ion and/or the precise geometry of the active site, which could contribute in part to the lack of the nuclease activity of FANCM.

Furthermore, we found that the electrostatic surface surrounding the putative active site in FANCM-FAAP24 is substantially different from that in ApXPF

and DrMus81-HsEme1 (Figure 3B and Supplementary Figure S5). The electrostatic surfaces surrounding the active site in ApXPF and DrMus81-HsEme1 are largely positively charged. In ApXPF-dsDNA, the dsDNA is bound in the surface groove between the nuclease and the (HhH)₂ domains and interacts mainly with several basic residues and the G-I-G hairpin from the (HhH)₂ domain of monomer A, but it is positioned distantly from the active site of the nuclease domain (Figure 3B and Supplementary Figure S3E). Structural analysis of ApXPF-dsDNA showed that there is a pseudo 2-fold symmetry between the two (HhH)₂ domains, implying that the (HhH)₂ domain of monomer B could also bind DNA via the equivalent region. The positively charged surface region formed by the equivalent basic residues of monomer B is located in the vicinity of the active site, and a modeled dsDNA could bind to this region and is positioned closely to the active site (Figure 3B) (31). In addition, structural analysis of the HsXPF (HhH)₂-ssDNA complex showed that the basic residues in the equivalent regions of both (HhH)₂ domains participate in the binding of ssDNA (Supplementary Figure S5) (31,33). Moreover, mutagenesis studies of DrMus81-HsEme1 suggested that several basic residues surrounding the active site are involved in DNA binding (Supplementary Figure S5) (32). These data together indicate that the positively charged surface surrounding the active site is involved in the binding of DNA, which is also essential for the nuclease activity.

In sharp contrast, structural analysis of FANCM-FAAP24 shows that the equivalent region surrounding the active site is relatively hydrophobic (Figure 3B). In addition, although the nuclease and the (HhH)₂ domains form a surface groove on the top side and a surface cleft on the bottom side, both of them are also composed of largely acidic and hydrophobic residues. These results suggest that the active-site region in FANCM-FAAP24 is unable to bind a DNA substrate, which is in agreement with the previous biochemical data showing that the C-terminal segment of FANCM (FANCM_{L1}) exhibits little DNA-binding ability to ssDNA, dsDNA or splayed-arm DNA (5). Furthermore, as all members of the XPF family in eukaryotes exist as heterodimers, we also examined the DNA-binding ability of the dimeric nuclease domains of FANCM-FAAP24. Our EMSA results show that the dimeric nuclease domains of FANCM-FAAP24 have no detectable DNA-binding ability to different types of DNA structures (Supplementary Figure S6), which is also consistent with the structural data. Taken together, we suggest that the variation of two key residues at the active site and the lack of DNA binding ability of the surrounding region render a catalytically inactive nuclease domain of FANCM and account for the lack of nuclease activity of FANCM-FAAP24.

As reported previously, FANCM-FAAP24 targets the FA core complex to stalled replication forks during DNA replication at S phase (5,18). For repairing stalled replication forks, Mus81-Eme1 is required for the first incision to form a double-strand break (DSB) (44), and XPF-ERCC1 is required for the second incision (45–47). Thus, it seems that the nuclease activity of FANCM-FAAP24 is not

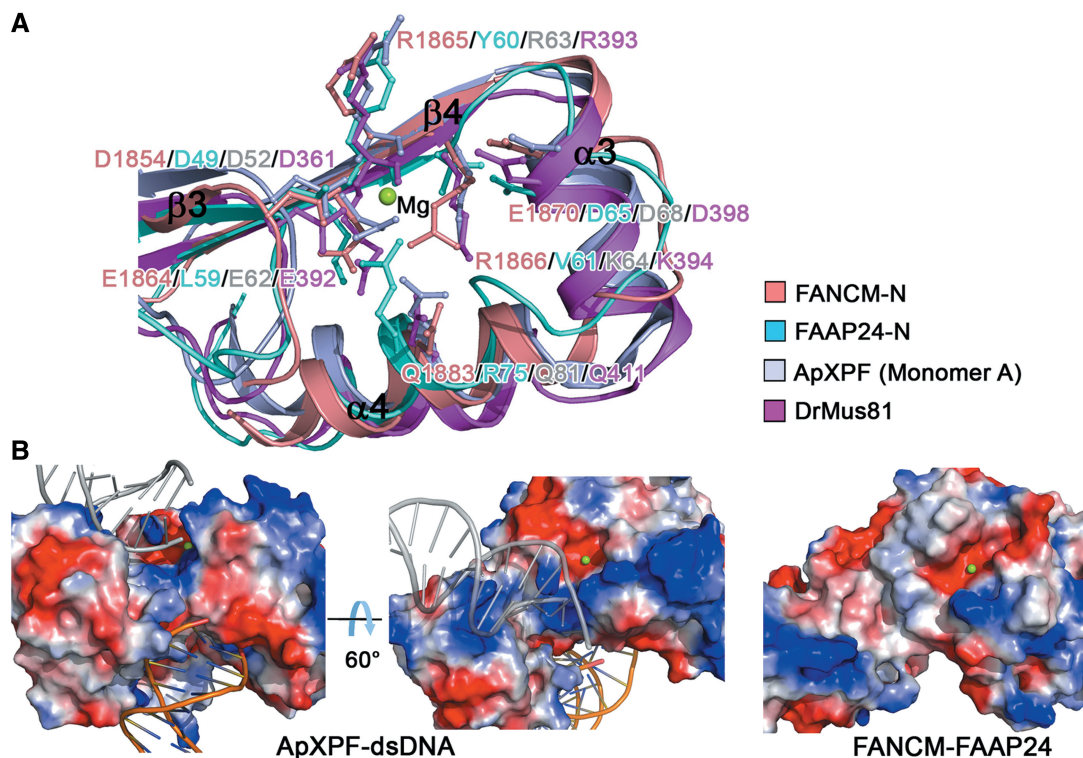


Figure 3. Structural comparison of the FANCM-FAAP24 complex with other XPF family members. (A) Comparison of the active sites of FANCM (salmon), FAAP24 (cyan), ApXPF (PDF code 2BGW, light blue) and DrMus81 (PDB code 2ZIU, magenta). Residues corresponding to the catalytic GDX_nERKX_3D motif in FANCM, FAAP24, ApXPF and DrMus81 are shown with ball-and-stick models. Only the secondary structure elements of ApXPF are labeled. In the color coding scheme, the nuclease domains are abbreviated as N. Mg^{2+} in ApXPF-dsDNA is shown with a green sphere. (B) Electrostatic surface representations of the active sites in ApXPF-dsDNA and FANCM-FAAP24. The bound dsDNA in ApXPF-dsDNA is shown with orange ribbon and the modeled dsDNA with gray ribbon. ApXPF-dsDNA in the right panel and FANCM-FAAP24 are shown in the same orientation as these in Figure 3A. Mg^{2+} in FANCM-FAAP24 was modeled based on ApXPF-dsDNA and is shown with a green sphere.

required in the fork cleavage reaction. It is possible that the nuclease domain of FANCM was degenerated and became catalytically inactive during evolution to avoid redundant function, and FANCM may function mainly via its translocase activity in DNA repair. Indeed, it was recently shown that in addition to its coordinated function with FAAP24 in the activation of the FA pathway, FANCM can function alone in recombination-independent ICL repair by facilitating recruitment of lesion incision activities via its translocase activity (26).

Potential DNA binding site

The previous and our own biochemical data have shown that FANCM-FAAP24 could bind different types of DNA structures (5) (Figure 1A and Supplementary Figure S1). The previous biochemical data have shown that the (HhH)₂ domain of FAAP24 is essential for targeting FANCM-FAAP24 to ICL DNA (24). Thus, we first investigated whether the (HhH)₂ domains of FANCM and FAAP24 are essential for their localization to chromatin (Figure 4A). Our results show that the wild-type FANCM and FAAP24 co-localize to chromatin; however, deletion of the (HhH)₂ domain of either FANCM (FANCM₁₋₁₉₇₀, FANCM Δ H) or FAAP24 (FAAP24₁₋₁₅₃, FAAP24 Δ H) significantly reduces the localization to chromatin. In contrast, deletion of the

nuclease domain of either FANCM (FANCM $\Delta_{1813-1970}$, FANCM Δ N) or FAAP24 (FAAP24₁₅₃₋₂₁₅, FAAP24 Δ N) shows invisible or slight effect on the localization. Deletion of both nuclease and (HhH)₂ domains of FANCM (FANCM₁₋₁₈₁₃, FANCM Δ N Δ H) disrupts the interactions with FAAP24 (Supplementary Figure S7B) and shows similar effect on chromatin localization as that of FANCM Δ H. These results indicate that the (HhH)₂ domains of FANCM and FAAP24 are critical for their binding to DNA and localization to chromatin.

We next tried to identify the potential DNA binding site(s) in the (HhH)₂ domains of FANCM-FAAP24. Analysis of the electrostatic surface of the (HhH)₂ domains reveals two positively charged surface patches (Figure 4B), and residues composing these patches are conserved in different vertebrate species (Supplementary Figure S4B and C). Region I is mainly composed of residues from helix α_9 of FANCM and helix α_9' of FAAP24, both of which belong to the second HhH motif. Region II is mainly composed of residues from helix α_5' and the α_5' - α_6' loop of the first HhH motif of FAAP24. Particularly, two basic residues (Lys171 and Lys173) in region II are highly conserved, and the strictly conserved Gly168-Val169-Gly170 residues form a GhG hairpin (where h indicates a hydrophobic residue) with their main-chain amino groups exposed on the

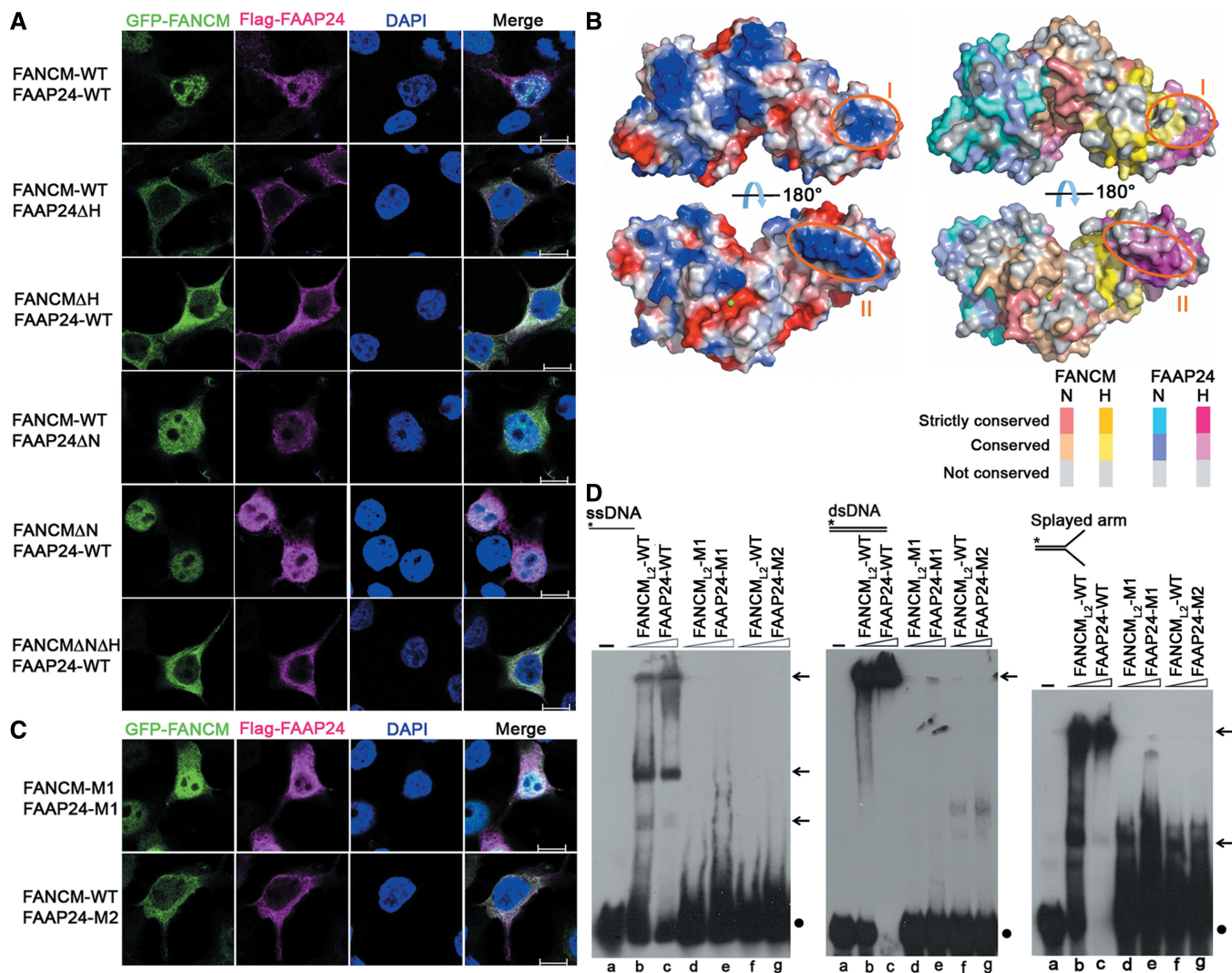


Figure 4. A potential DNA-binding site in FAAP24. (A) Localization of the wild-type and truncated FANCM-FAAP24. Scale bars: 10 μ m. (B) Electrostatic surface (left panels) and distribution of the conserved residues on the surface (right panels) of FANCM-FAAP24. The two positively charged surface patches in the (HhH)₂ domains are indicated with orange circles. The active site of FANCM is indicated by a modeled Mg²⁺ shown with a green sphere. The color coding scheme of the residue conservation is indicated. In the color coding scheme, the nuclease and the (HhH)₂ domains are abbreviated as N and H, respectively. (C) Localization of the mutant FANCM-FAAP24. Scale bars: 10 μ m. (D) EMSA of the wild-type and mutant FANCM-FAAP24. The reaction contained the biotin-labeled ssDNA, dsDNA and splayed-arm DNA (10 nM) without or with FANCM-FAAP24 (0.0 μ M, lane a; 0.4 μ M, lanes b, d and f; 0.8 μ M, lanes c, e and g). The asterisks represent the biotin labels at the DNA 5' end. The arrows indicate the shifted bands of the protein-DNA complexes, and the dots represent the free DNA probes.

surface (Supplementary Figure S7A), which may be equivalent to the GhG hairpin in ERCC1 and XPF that is involved in DNA binding (31,34,43).

To assess the functional importance of these two regions, we examined whether mutations in the two regions affect the localization of the FANCM-FAAP24 complex to chromatin and the binding ability to DNA (Figure 4C and D). The mutant complex containing mutations in region I (K1998A/R1999A of FANCM, FANCM-M1 and H208A of FAAP24, FAAP24-M1) shows slightly reduced localization to chromatin and an impaired DNA-binding activity, whereas the mutant complex containing mutations in region II (R161A/Q164A/K171A of FAAP24, FAAP24-M2) fails to localize to chromatin and abolishes the DNA-binding

activity. The mutant complex containing the quadruple mutations R161A/Q164A/K171A/K173A of FAAP24 in region II exhibits similar effects as that containing the triple mutations (data not shown). We further confirmed that the mutations we introduced do not affect the proper folding of the proteins and the formation of the FANCM-FAAP24 complex both *in vivo* and *in vitro*. Our co-immunoprecipitation results show that the wild-type and mutant GFP-FANCM and Flag-FAAP24 form stable complexes (Supplementary Figure S7B), and our size-exclusion chromatography results show that the wild-type and mutant FANCM-FAAP24 complexes used in EMSA exist as heterodimers in solution (Supplementary Figure S7C).

The previous biological data have shown that FANCM can interact with the MHF1-MHF2 complex via the region following the helicase domain (residues 661–800) and with FAAP24 via the C-terminal region (residues 1799–2048) (3–5). To exclude the possible effect of the intrinsically expressed MHF1-MHF2 on the localization of FANCM to chromatin, we constructed a series of GFP-FANCM_{L2} in wild-type, truncated and mutant forms similar to those of full-length FANCM and examined the localization of FANCM-FAAP24 to chromatin. The immunofluorescence assay results show that the wild-type, truncated and mutant FANCM_{L2} exhibit similar chromatin localization patterns as the corresponding full-length FANCM, indicating that the interaction between the over-expressed GFP-FANCM and the intrinsically expressed MHF1-MHF2 has no notable effect on the localization of FANCM to chromatin (Supplementary Figure S7D). As both GFP-FANCM and Flag-FAAP24 were over-expressed in our assays and detected with the tag-specific antibodies, it is understandable that the signals for GFP-FANCM interacting with other intrinsically expressed protein partner(s) would be comparably very low and buried in the background. Our structural and functional data together indicate that region II (helix $\alpha 5'$ and the $\alpha 5'$ - $\alpha 6'$ loop) of FAAP24 plays a critical role in DNA binding and is essential for the localization of FANCM-FAAP24 to chromatin. Most recently, Wienk *et al.* (48) reported the solution structure of the (HhH)₂ domain of FAAP24 and showed that the first HhH motif is involved in the binding of ssDNA, which is consistent with our results. As region II of FAAP24 is independent of FANCM, we speculate that the DNA-binding ability of this region is not only essential for targeting FANCM-FAAP24 to chromatin but also for the independent function of FAAP24 in promoting ATR-mediated checkpoint activation (26).

The previous biochemical data demonstrated that FANCM-FAAP24 binds preferentially to ssDNA, splayed-arm DNA and a 3'-flap DNA substrate 5–10-fold better than to dsDNA (5), whereas FANCM-MHF binds preferentially to branched DNA and dsDNA (3,4). Both FAAP24 and MHF help to target FANCM to DNA. It is possible that FANCM-MHF-FAAP24 may work together as a molecular sensor, which scans along chromatin to detect stalled replication fork caused by ICLs (4). The two regions of FANCM interacting with MHF and FAAP24 may be located in proximity in the 3D structure such that MHF and FAAP24 could bind to dsDNA and ssDNA regions of one stalled replication fork, respectively, with high affinity and then cooperate to anchor FANCM to the branch point of the replication fork.

In summary, our structural and biological data demonstrate that like other XPF family members, both FANCM and FAAP24 consist of a nuclease domain and an (HhH)₂ domain. The overall structure of FANCM-FAAP24 exhibits a similar architecture as that of ApXPF. Although the active site of FANCM has a similar structure as that of ApXPF and DrMus81, the variation of two key residues at the active site of FANCM and the inability of the surrounding region to

bind DNA render a catalytically inactive nuclease domain. Moreover, our data indicate that helix $\alpha 5'$ and the $\alpha 5'$ - $\alpha 6'$ loop of FAAP24 is a potential binding site for ssDNA and is critical for the localization of FANCM-FAAP24 to chromatin in DNA repair.

ACCESSION NUMBERS

The coordinates and structure factors of the FANCM-FAAP24 complex have been deposited in the RCSB Protein Data Bank with the accession code 4M6W.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank the staff members at Shanghai Synchrotron Radiation Facility of China for technical supports in diffraction data collection, and other members of their group for helpful discussion. H.Y. carried out the structural and functional studies and drafted the manuscript. T.Z. participated in the structural studies. Y.T. and F.W. participated in the functional studies. L.T. participated in the data analysis and discussion. J.D. conceived of the study, participated in the experimental design, data analysis and discussion and wrote the manuscript.

FUNDING

The Ministry of Science and Technology of China [2013CB910404 and 2011CB966301] and the National Natural Science Foundation of China [31230017 and 31221001]. Funding for open access charge: the Ministry of Science and Technology of China [2013CB910404].

Conflict of interest statement. None declared.

REFERENCES

- Kennedy,R.D. and D'Andrea,A.D. (2005) The Fanconi Anemia/BRCA pathway: new faces in the crowd. *Genes Dev.*, **19**, 2925–2940.
- Moldovan,G.L. and D'Andrea,A.D. (2009) How the fanconi anemia pathway guards the genome. *Annu. Rev. Genet.*, **43**, 223–249.
- Singh,T.R., Saro,D., Ali,A.M., Zheng,X.F., Du,C.H., Killen,M.W., Sachpatzidis,A., Wahengbam,K., Pierce,A.J., Xiong,Y. *et al.* (2010) MHF1-MHF2, a histone-fold-containing protein complex, participates in the Fanconi anemia pathway via FANCM. *Mol. Cell*, **37**, 879–886.
- Yan,Z., Delannoy,M., Ling,C., Dae,D., Osman,F., Muniandy,P.A., Shen,X., Oostra,A.B., Du,H., Steltenpool,J. *et al.* (2010) A histone-fold complex and FANCM form a conserved DNA-remodeling complex to maintain genome stability. *Mol. Cell*, **37**, 865–878.
- Ciccio,A., Ling,C., Coulthard,R., Yan,Z., Xue,Y., Meetei,A.R., Laghmani el,H., Joenje,H., McDonald,N., de Winter,J.P. *et al.* (2007) Identification of FAAP24, a Fanconi anemia core complex protein that interacts with FANCM. *Mol. Cell*, **25**, 331–343.
- Ling,C., Ishiai,M., Ali,A.M., Medhurst,A.L., Neveling,K., Kalb,R., Yan,Z., Xue,Y., Oostra,A.B., Auerbach,A.D. *et al.*

- (2007) FAAP100 is essential for activation of the Fanconi anemia-associated DNA damage response pathway. *EMBO J.*, **26**, 2104–2114.
7. Ali, A.M., Pradhan, A., Singh, T.R., Du, C., Li, J., Wahengbam, K., Grassman, E., Auerbach, A.D., Pang, Q. and Meetei, A.R. (2012) FAAP20: a novel ubiquitin-binding FA nuclear core-complex protein required for functional integrity of the FA-BRCA DNA repair pathway. *Blood*, **119**, 3285–3294.
 8. Leung, J.W., Wang, Y., Fong, K.W., Huen, M.S., Li, L. and Chen, J. (2012) Fanconi anemia (FA) binding protein FAAP20 stabilizes FA complementation group A (FANCA) and participates in interstrand cross-link repair. *Proc. Natl Acad. Sci. USA*, **109**, 4491–4496.
 9. Yan, Z., Guo, R., Paramasivam, M., Shen, W., Ling, C., Fox, D. III, Wang, Y., Oostra, A.B., Kuehl, J., Lee, D.Y. *et al.* (2012) A ubiquitin-binding protein, FAAP20, links RNF8-mediated ubiquitination to the Fanconi anemia DNA repair network. *Mol. Cell*, **47**, 61–75.
 10. Kim, H., Yang, K., Dejsuphong, D. and D'Andrea, A.D. (2012) Regulation of Rev1 by the Fanconi anemia core complex. *Nat. Struct. Mol. Biol.*, **19**, 164–170.
 11. Meetei, A.R., de Winter, J.P., Medhurst, A.L., Wallisch, M., Waisfisz, Q., van de Vrugt, H.J., Oostra, A.B., Yan, Z., Ling, C., Bishop, C.E. *et al.* (2003) A novel ubiquitin ligase is deficient in Fanconi anemia. *Nat. Genet.*, **35**, 165–170.
 12. Vandenberg, C.J., Gergely, F., Ong, C.Y., Pace, P., Mallery, D.L., Hiom, K. and Patel, K.J. (2003) BRCA1-independent ubiquitination of FANCD2. *Mol. Cell*, **12**, 247–254.
 13. Wang, X., Andreassen, P.R. and D'Andrea, A.D. (2004) Functional interaction of monoubiquitinated FANCD2 and BRCA2/FANCD1 in chromatin. *Mol. Cell Biol.*, **24**, 5850–5862.
 14. Smogorzewska, A., Matsuoka, S., Vinciguerra, P., McDonald, E.R. III, Hurov, K.E., Luo, J., Ballif, B.A., Gygi, S.P., Hofmann, K., D'Andrea, A.D. *et al.* (2007) Identification of the FANCI protein, a monoubiquitinated FANCD2 paralog required for DNA repair. *Cell*, **129**, 289–301.
 15. Gari, K., Decaillet, C., Stasiak, A.Z., Stasiak, A. and Constantinou, A. (2008) The Fanconi anemia protein FANCM can promote branch migration of Holliday junctions and replication forks. *Mol. Cell*, **29**, 141–148.
 16. Meetei, A.R., Medhurst, A.L., Ling, C., Xue, Y., Singh, T.R., Bier, P., Steltenpool, J., Stone, S., Dokal, I., Mathew, C.G. *et al.* (2005) A human ortholog of archaeal DNA repair protein Hef is defective in Fanconi anemia complementation group M. *Nat. Genet.*, **37**, 958–963.
 17. Xue, Y., Li, Y., Guo, R., Ling, C. and Wang, W. (2008) FANCM of the Fanconi anemia core complex is required for both monoubiquitination and DNA repair. *Hum. Mol. Genet.*, **17**, 1641–1652.
 18. Kim, J.M., Kee, Y., Gurtan, A. and D'Andrea, A.D. (2008) Cell cycle-dependent chromatin loading of the Fanconi anemia core complex by FANCM/FAAP24. *Blood*, **111**, 5215–5222.
 19. Mosedale, G., Niedzwiedz, W., Alpi, A., Perrina, F., Pereira-Leal, J.B., Johnson, M., Langevin, F., Pace, P. and Patel, K.J. (2005) The vertebrate Hef ortholog is a component of the Fanconi anemia tumor-suppressor pathway. *Nat. Struct. Mol. Biol.*, **12**, 763–771.
 20. Deans, A.J. and West, S.C. (2009) FANCM connects the genome instability disorders Bloom's Syndrome and Fanconi Anemia. *Mol. Cell*, **36**, 943–953.
 21. Hoadley, K.A., Xue, Y., Ling, C., Takata, M., Wang, W. and Keck, J.L. (2012) Defining the molecular interface that connects the Fanconi anemia protein FANCM to the Bloom syndrome dissolvosome. *Proc. Natl Acad. Sci. USA*, **109**, 4437–4442.
 22. Meetei, A.R., Sechi, S., Wallisch, M., Yang, D., Young, M.K., Joenje, H., Hoatlin, M.E. and Wang, W. (2003) A multiprotein nuclear complex connects Fanconi anemia and Bloom syndrome. *Mol. Cell Biol.*, **23**, 3417–3426.
 23. Collis, S.J., Ciccio, A., Deans, A.J., Horejsi, Z., Martin, J.S., Maslen, S.L., Skehel, J.M., Elledge, S.J., West, S.C. and Boulton, S.J. (2008) FANCM and FAAP24 function in ATR-mediated checkpoint signaling independently of the Fanconi anemia core complex. *Mol. Cell*, **32**, 313–324.
 24. Huang, M., Kim, J.M., Shiotani, B., Yang, K., Zou, L. and D'Andrea, A.D. (2010) The FANCM/FAAP24 complex is required for the DNA interstrand crosslink-induced checkpoint response. *Mol. Cell*, **39**, 259–268.
 25. Schwab, R.A., Blackford, A.N. and Niedzwiedz, W. (2010) ATR activation and replication fork restart are defective in FANCM-deficient cells. *EMBO J.*, **29**, 806–818.
 26. Wang, Y., Leung, J.W., Jiang, Y., Lowery, M.G., Do, H., Vasquez, K.M., Chen, J., Wang, W. and Li, L. (2013) FANCM and FAAP24 maintain genome stability via cooperative as well as unique functions. *Mol. Cell*, **49**, 997–1009.
 27. Tao, Y., Jin, C., Li, X., Qi, S., Chu, L., Niu, L., Yao, X. and Teng, M. (2012) The structure of the FANCM-MHF complex reveals physical features for functional assembly. *Nat. Commun.*, **3**, 782.
 28. Yang, H., Zhang, T., Tao, Y., Wu, L., Li, H.T., Zhou, J.Q., Zhong, C. and Ding, J. (2012) Saccharomyces cerevisiae MHF complex structurally resembles the histones (H3-H4)₂ heterotetramer and functions as a heterotetramer. *Structure*, **20**, 364–370.
 29. Enzlin, J.H. and Scharer, O.D. (2002) The active site of the DNA repair endonuclease XPF-ERCC1 forms a highly conserved nuclease motif. *EMBO J.*, **21**, 2045–2053.
 30. Nishino, T., Komori, K., Ishino, Y. and Morikawa, K. (2003) X-ray and biochemical anatomy of an archaeal XPF/Rad1/Mus81 family nuclease: similarity between its endonuclease domain and restriction enzymes. *Structure*, **11**, 445–457.
 31. Newman, M., Murray-Rust, J., Lally, J., Rudolf, J., Fadden, A., Knowles, P.P., White, M.F. and McDonald, N.Q. (2005) Structure of an XPF endonuclease with and without DNA suggests a model for substrate recognition. *EMBO J.*, **24**, 895–905.
 32. Chang, J.H., Kim, J.J., Choi, J.M., Lee, J.H. and Cho, Y. (2008) Crystal structure of the Mus81-Emel complex. *Genes Dev.*, **22**, 1093–1106.
 33. Das, D., Folkers, G.E., van Dijk, M., Jaspers, N.G., Hoeijmakers, J.H., Kaptein, R. and Boelens, R. (2012) The structure of the XPF-ssDNA complex underscores the distinct roles of the XPF and ERCC1 helix-hairpin-helix domains in ss/ds DNA recognition. *Structure*, **20**, 667–675.
 34. Tripsianes, K., Folkers, G., Ab, E., Das, D., Odijk, H., Jaspers, N.G., Hoeijmakers, J.H., Kaptein, R. and Boelens, R. (2005) The structure of the human ERCC1/XPF interaction domains reveals a complementary role for the two proteins in nucleotide excision repair. *Structure*, **13**, 1849–1858.
 35. Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
 36. Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr.*, **D66**, 213–221.
 37. Murshudov, G.N., Skubak, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F. and Vagin, A.A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr.*, **D67**, 355–367.
 38. Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr.*, **D60**, 2126–2132.
 39. Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A. *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr.*, **D67**, 235–242.
 40. Krissinel, E. and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.*, **372**, 774–797.
 41. Ciccio, A., Constantinou, A. and West, S.C. (2003) Identification and characterization of the human mus81-eme1 endonuclease. *J. Biol. Chem.*, **278**, 25172–25178.
 42. Das, D., Tripsianes, K., Jaspers, N.G., Hoeijmakers, J.H., Kaptein, R., Boelens, R. and Folkers, G.E. (2008) The HhH domain of the human DNA repair protein XPF forms stable homodimers. *Proteins*, **70**, 1551–1563.
 43. Tsodikov, O.V., Enzlin, J.H., Scharer, O.D. and Ellenberger, T. (2005) Crystal structure and DNA binding functions of ERCC1, a subunit of the DNA structure-specific endonuclease XPF-ERCC1. *Proc. Natl Acad. Sci. USA*, **102**, 11236–11241.

44. Hanada, K., Budzowska, M., Modesti, M., Maas, A., Wyman, C., Essers, J. and Kanaar, R. (2006) The structure-specific endonuclease Mus81-Eme1 promotes conversion of interstrand DNA crosslinks into double-strand breaks. *EMBO J.*, **25**, 4921–4932.
45. Bessho, T., Mu, D. and Sancar, A. (1997) Initiation of DNA interstrand cross-link repair in humans: the nucleotide excision repair system makes dual incisions 5' to the cross-linked base and removes a 22- to 28-nucleotide-long damage-free strand. *Mol. Cell Biol.*, **17**, 6822–6830.
46. Kuraoka, I., Kobertz, W.R., Ariza, R.R., Biggerstaff, M., Essigmann, J.M. and Wood, R.D. (2000) Repair of an interstrand DNA cross-link initiated by ERCC1-XPF repair/recombination nuclease. *J. Biol. Chem.*, **275**, 26632–26636.
47. Niedernhofer, L.J., Lalai, A.S. and Hoeijmakers, J.H. (2005) Fanconi anemia (cross)linked to DNA repair. *Cell*, **123**, 1191–1198.
48. Wienk, H., Slootweg, J.C., Speerstra, S., Kaptein, R., Boelens, R. and Folkers, G.E. (2013) The Fanconi anemia associated protein FAAP24 uses two substrate specific binding surfaces for DNA recognition. *Nucleic Acids Res.*, **41**, 6739–6749.