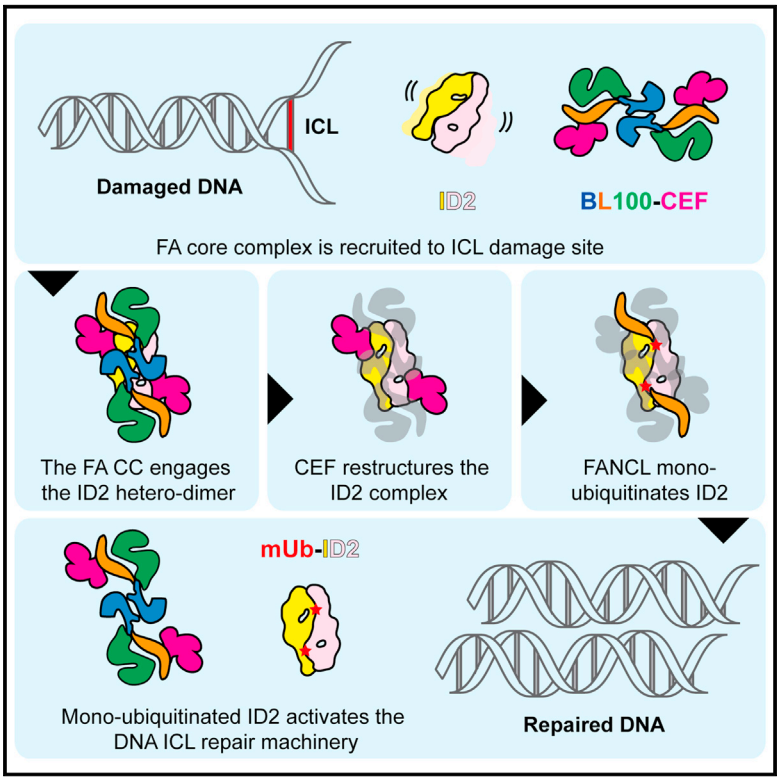
Article

Cell Reports

The FA Core Complex Contains a Homo-dimeric Catalytic Module for the Symmetric Mono-ubiquitination of FANCI-FANCD2

Graphical Abstract



Highlights

- FANCB, FANCL, and FAAP100 form a symmetric dimer of trimers
- FANCL is ideally poised for the symmetric mono-ubiquitination of FANCI-FANCD2
- Two separate FANCC-FANCE-FANCF complexes bind to the opposing poles of FANCB-FANCL-FAAP100
- FANCC-FANCE-FANCF stabilizes FANCI-FANCD2 for efficient mono-ubiquitination

Authors

Paolo Swuec, Ludovic Renault, Aaron Borg, ..., Ambrosius P. Snijders, Andrew J. Deans, Alessandro Costa

Correspondence

alessandro.costa@crick.ac.uk

In Brief

Mono-ubiquitination of FANCI-FANCD2 by the Fanconi anemia core complex activates a major DNA interstrand-crosslink repair pathway important for genome stability maintenance. Here, Swuec et al. reveal the structural basis of this reaction by showing that the core complex exists as a dimeric catalytic module for the symmetric mono-ubiquitination of FANCI-FANCD2.

Swuec et al., 2017, Cell Reports *18*, 611–623
January 17, 2017
http://dx.doi.org/10.1016/j.celrep.2016.11.013

Cell Reports
Article

OPEN ACCESS
CellPress

The FA Core Complex Contains a Homo-dimeric Catalytic Module for the Symmetric Mono-ubiquitination of FANCI-FANCD2

Paolo Swuec,[1] Ludovic Renault,[1,4] Aaron Borg,[2] Fenil Shah,[3,5] Vincent J. Murphy,[3] Sylvie van Twest,[3] Ambrosius P. Snijders,[2] Andrew J. Deans,[3] and Alessandro Costa[1,6,*]

[1]Macromolecular Machines Laboratory, Clare Hall Laboratory, The Francis Crick Institute, Blanche Lane, South Mimms, EN6 3LD, UK
[2]Mass Spectrometry Proteomics and Metabolomics, Clare Hall Laboratory, The Francis Crick Institute, Blanche Lane, South Mimms, EN6 3LD, UK
[3]Genome Stability Unit, St. Vincent's Institute of Medical Research, 9 Princes St Fitzroy, Victoria, VIC 3065, Australia
[4]Present address: NeCEN, Gorlaeus Laboratory, Einsteinweg 55, Leiden 2333, the Netherlands
[5]Present address: Herman B Wells Center for Pediatric Research, Indiana University School of Medicine, 1044 West Walnut, Indianapolis, IN 46202, USA
[6]Lead Contact
*Correspondence: alessandro.costa@crick.ac.uk
http://dx.doi.org/10.1016/j.celrep.2016.11.013

SUMMARY

Activation of the main DNA interstrand crosslink repair pathway in higher eukaryotes requires mono-ubiquitination of FANCI and FANCD2 by FANCL, the E3 ligase subunit of the Fanconi anemia core complex. FANCI and FANCD2 form a stable complex; however, the molecular basis of their ubiquitination is ill defined. FANCD2 mono-ubiquitination by FANCL is stimulated by the presence of the FANCB and FAAP100 core complex components, through an unknown mechanism. How FANCI mono-ubiquitination is achieved remains unclear. Here, we use structural electron microscopy, combined with crosslink-coupled mass spectrometry, to find that FANCB, FANCL, and FAAP100 form a dimer of trimers, containing two FANCL molecules that are ideally poised to target both FANCI and FANCD2 for mono-ubiquitination. The FANCC-FANCE-FANCF subunits bridge between FANCB-FANCL-FAAP100 and the FANCI-FANCD2 substrate. A transient interaction with FANCC-FANCE-FANCF alters the FANCI-FANCD2 configuration, stabilizing the dimerization interface. Our data provide a model to explain how equivalent mono-ubiquitination of FANCI and FANCD2 occurs.

INTRODUCTION

Interstrand crosslinks (ICLs) are toxic DNA lesions that prevent the separation of the two filaments in the DNA double helix, impairing DNA replication and transcription. The Fanconi anemia (FA)/BRCA pathway is the major interstrand crosslink repair process in higher eukaryotes (Knipscheer et al., 2009; Kottemann and Smogorzewska, 2013). Activation of this pathway depends on the specific mono-ubiquitination of the FANCI and FANCD2 proteins by the FA core complex (Ceccaldi et al., 2016; Garcia-Higuera et al., 2001; Sims et al., 2007; Smogorzewska et al., 2007; Timmers et al., 2001). Mutations in the FA genes cause a rare genetic disorder characterized by hypersensitivity to crosslinking agents, chromosomal instability, and cancer predisposition (Deans and West, 2011).

Despite the enormous effort of several research groups, a mechanistic understanding of the activation of the FA ICL repair pathway is only starting to emerge (Walden and Deans, 2014). A key player in the recognition of ICLs is the FA anchor complex (consisting of the FANCM DNA translocase, FAAP24, and two histone-like factors named MHF1-2), which, in turn, recruits the FA core complex to the DNA damage site (Ciccia et al., 2007; Coulthard et al., 2013; Kim et al., 2008; Yan et al., 2010). The core complex is composed of three distinct protein assemblies with non-redundant functions: FANCB-FANCL-FAAP100 (hereinafter called BL100), FANCC-FANCE-FANCF (CEF), and FANCA-FANCG-FAAP20 (AG20) (Huang et al., 2014; Rajendra et al., 2014). While the AG20 and CEF play loosely defined ancillary roles (Wang and Smogorzewska, 2015), the BL100 complex is a three-member ubiquitin ligase, with FANCL as the minimal unit required for ubiquitination (Huang et al., 2014; Meetei et al., 2003; Rajendra et al., 2014). FANCL is an E3 ligase containing a RING domain that is essential for the interaction with the E2 enzyme, UBE2T/FANCT, and for the mono-ubiquitination of the FANCI and FANCD2 molecular targets (Cole et al., 2010; Hodson et al., 2011, 2014; Machida et al., 2006). Although FANCL is active in isolation, its catalytic function is greatly stimulated by the interaction with FANCB and FAAP100 (Huang et al., 2014; Rajendra et al., 2014). Whether binding of these two factors causes an activating conformational switch in the FANCL catalytic center, or whether FANCB and FAAP100 play an architectural role important for mono-ubiquitination, is unknown.

Multiple other architectural aspects of the mono-ubiquitination of FANCI and FANCD2 likewise remain poorly understood. First, crystallographic studies on mouse proteins indicate that the

FANCI-FANCD2 (ID2) hetero-dimer forms a pseudo-2-fold symmetric arrangement. However, the mono-ubiquitination sites map at the dimerization interface within the ID2 complex (Joo et al., 2011), suggesting that a reconfiguration of this interface is needed at some point during the mono-ubiquitination reaction to allow for access of the E3 ligase assembly. Second, whether a stable ID2 complex is recruited to the DNA damage site prior to mono-ubiquitination is unclear, although recent cryoelectron microscopy (cryo-EM) studies on the unmodified human proteins support this notion (Liang et al., 2016). Remarkably the symmetric hetero-dimer captured by crystallography cannot be recognized in the electron microscopy (EM) structure, suggesting that the ID2 complex might visit different conformational states throughout the ubiquitination reaction. Finally, although both members of the ID2 dimer need to be mono-ubiquitinated for activation of the ICL repair pathway (Sims et al., 2007; Smogorzewska et al., 2007; Timmers et al., 2001), so far, biochemical reconstitution has mainly focused on the minimal ability of the BL100 E3 ligase to mono-ubiquitinate FANCD2, with little to no FANCI mono-ubiquitination observed (Rajendra et al.,2014). How equivalent mono-ubiquitination of both FANCI and FANCD2 is achieved and whether any of the ancillary proteins in the FA core complex are involved in this process remain unresolved questions.

To address these issues, we have used EM, combined with crosslinking and mass spectrometry (XL/MS), to analyze the architecture of the BL100 complex and its substrate ID2, either in isolation or with other components of the FA core complex. We found that BL100 forms a symmetric dimer of trimers, inviting a molecular model for the equivalent ubiquitination of the dimeric ID2 assembly. The CEF component of the FA core complex binds to the two extremities in the BL100 dimer, probably acting as a molecular matchmaker that mediates the BL100 and ID2 interaction. We also uncovered an unexpected structural function for CEF, which stabilizes the dimerization interface of the ID2 complex. We postulate that this function might have a direct role in the efficient mono-ubiquitination of the ID2 hetero-dimeric substrate. Our findings provide key insights into the structural basis for the mono-ubiquitination of the ID2 complex, the defining step in the activation of the FA/BRCA ICL repair pathway.

RESULTS

BL100 Assembles in a 2-Fold Symmetric Dimer

Although the FANCL subunit of the core complex is an active E3 ligase in isolation (Alpi et al., 2008), its mono-ubiquitination function is stimulated by 6-fold in the presence of FANCB and FAAP100 (Rajendra et al., 2014). To understand whether the B100 association provides a structural framework for efficient mono-ubiquitination, or whether it causes a conformational change in FANCL that stimulates its activity, we have expressed and purified human BL100 from insect cells. Several steps of purification yielded a pure trimeric complex that eluted with a sharp peak under gel filtration and with an apparent molecular weight of 440 kDa (Figures 1A and S1). EM imaging confirmed that our preparation is highly homogeneous and monodisperse (Figures 1B and 1C). Two-dimensional classification of negative stain data revealed an elongated bow-tie structure containing apparent 2-fold symmetry (Figure 1D). The particle consists of a compact center and two globular modules, mapping at the two extremities of the assembly. Cryo-EM analysis yielded classes that were highly reminiscent of those obtained after negative stain imaging; however, no significant improvement in resolution could be detected, hinting at an inherent flexible nature in the BL100 complex (Figure 1E). To test our hypothesis that BL100 is mobile and can adopt different structural states, we have performed focused classification on the central core of the structure, followed by classification on the whole particle without further alignment. This approach revealed that the globular domains at the two poles in the structure move as rigid bodies, either tilting or stretching, relative to the central module (Figure S1; Movie S1). Despite this flexibility, the apparent 2-fold symmetry is maintained in the central core and at the two poles in the structure. From the data shown so far, it remains unclear whether BL100 is, indeed, a 2-fold symmetric assembly or, rather, contains a pseudo-2-fold symmetry axis, similar to ID2 (Figure 1F). To address this issue, we have repeated the two-dimensional analysis, using a FANCL derivative that contains maltose-binding protein (MBP) fused to the N terminus (Figures 2A and S2). MBP tags provide an electron-dense feature that can be averaged in negative stain EM, and they serve as pointers to localize individual proteins in a macromolecular assembly (Chen et al., 2008). We reasoned that a bona fide 2-fold symmetric complex should show two (FANCL-fused) MBP densities that are dyad related. Indeed, as evident from two-dimensional class averages, B-(MBP)L100 is decorated with two symmetry-related MBP densities, establishing that the three factors assemble into a homo-dimer (B:L:100 = 2:2:2 stoichiometry, with an expected molecular weight of 468 kDa, coherent with our analytical gel filtration profile) (Figures 2B and 2C). A dimeric nature of the FA core complex was unexpected and invites a mechanistic model for the ID2 mono-ubiquitination reaction that is further detailed in the Discussion section.

FANCB Is the Sole Candidate for the Homo-dimerization Interface in the FA Core Complex

Previous reports indicate that RING domains can dimerize (Brzovic et al., 2001; Liew et al., 2010; Lorick et al., 1999; Pruneda et al., 2011), and our MBP analysis shows that the RING-containing FANCL subunit maps in proximity of the dimerization interface of BL100 (Figures 2B and 2C). To test whether the FA core complex dimerization requires FANCL, we have expressed and purified a subunit dropout assembly that only contains FANCB and FAAP100 for single-particle analysis (Figures 2D and S3). In the absence of FANCL, the dimerization interface remains intact, while the globular poles in the assembly become markedly flexible (Figures 2E and S3; Movie S2), revealing a structural role for FANCL in stabilizing the BL100 assembly (Figure 2F), in line with previous reports (Alpi et al., 2008). To identify the dimerization component in the BL100 complex, we have attempted to purify the isolated FANCB subunit, but we found that this factor is insoluble when expressed in the absence of FAAP100 (data not shown). Conversely, the FAAP100 subunit is soluble and monodisperse in the absence of FANCB, according to analytical gel filtration and single-particle EM (Figures 2G

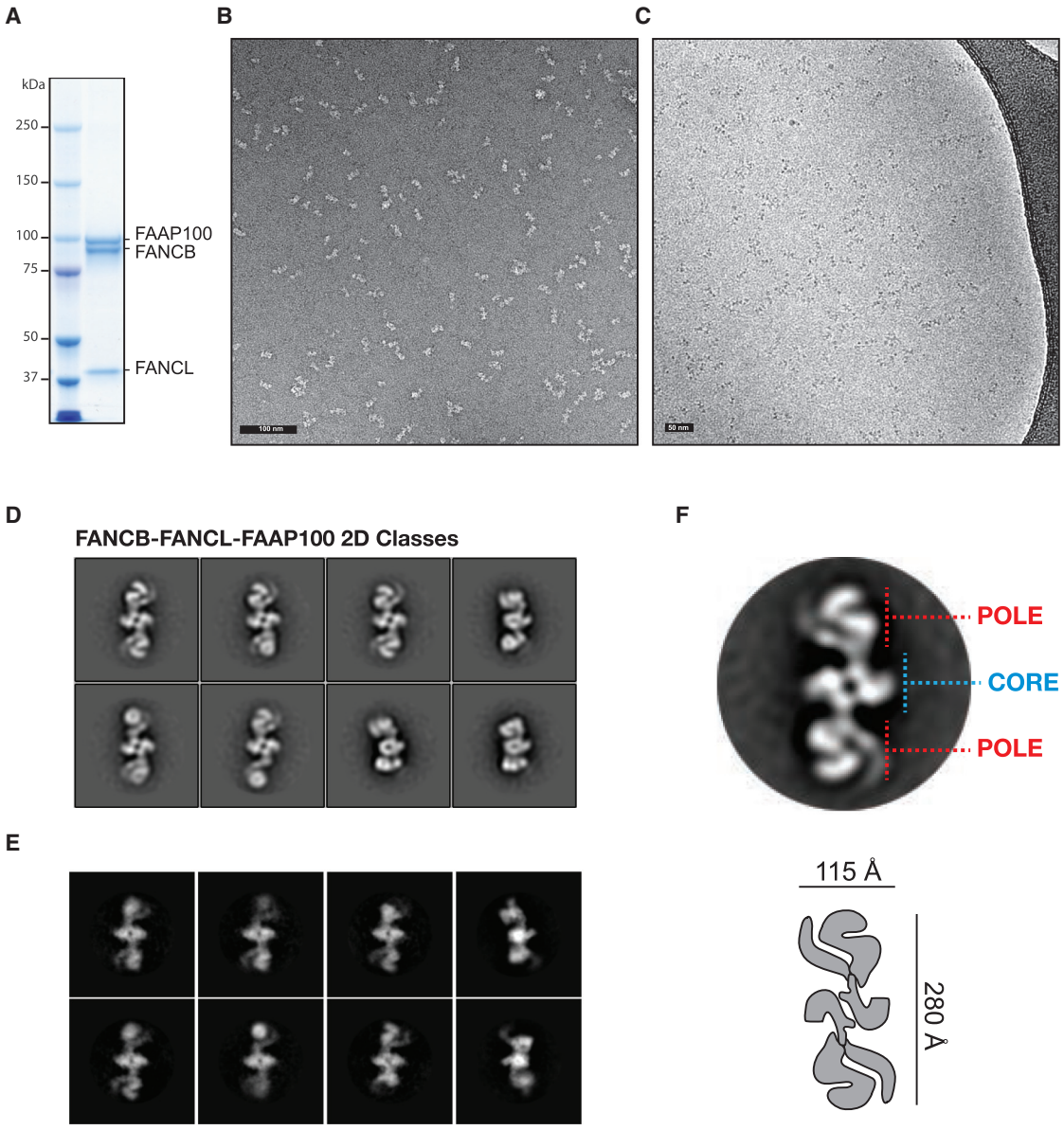


Figure 1. FANCB-FANCL-FAAP100 under the Electron Microscope

(A) Coomassie-stained SDS-PAGE gel of the purified BL100 complex.

(B and C) Negative-stain (B) and cryo-EM (C) of the purified BL100 complex.

(D and E) Representative negative-stain (D) and cryo-EM (E) two-dimensional class averages. Box size is 458 Å × 458 Å.

(F) Top: BL100 forms an elongated, apparently 2-fold symmetric particle containing a dimerization core (blue) and two globular poles (red). Bottom: a cartoon representation of the assembly with overall particle dimensions.

See also Figure S1.

and S3). Comparison of FAAP100 with the B100 two-dimensional class averages, centered on the flexible globular pole component, allows us to recognize FAAP100 as the element mapping at the extremities of the B100 particle. No resemblance was detected between the isolated FAAP100 and the dimerization core in the assembly (Figures 2H and S3). From these data, we conclude that FANCB and FAAP100 provide a dimeric scaffold to position FANCL on opposite poles in the protein assembly (Figure 2I). This suggests that efficient substrate ubiquitination requires a 2-fold symmetric arrangement in the FA core complex. Furthermore, our subunit dropout experiments are incompatible with a homo-dimerization role for either FANCL or FAAP100, leaving the FANCB subunit as the sole candidate for the dimerization element of the BL100 complex.

XL/MS Analysis of the BL100 Assembly

Analytical biochemistry combined with single-particle EM analysis allowed us to propose a model for the molecular

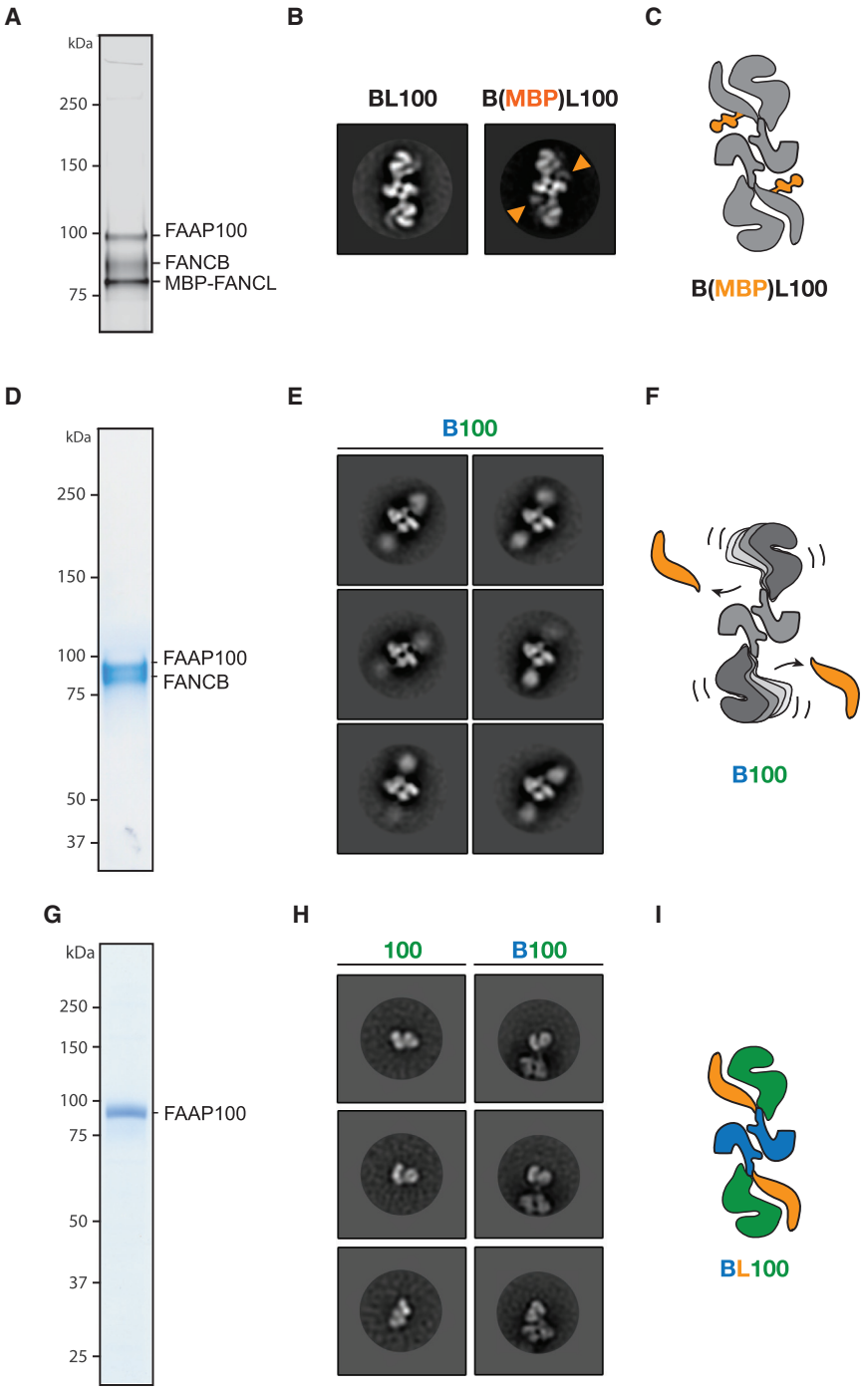


Figure 2. Subunit Architecture of the FANCB-FANCL-FAAP100 Assembly Studied by Single-Particle Analysis

(A) Silver-stained SDS-PAGE gel of the assembly containing an MBP N-terminally fused to FANCL.

(B) Comparison between wild-type and MBP fused BL100 (negative-stain two-dimensional classes). Two MBP density features are marked with an orange arrowhead. Box size is 458 Å × 458 Å.

(C) Cartoon representation of the BL100 complex with MBP in orange.

(D) Coomassie-stained SDS-PAGE gel of the purified B100 subcomplex.

(E) Negative-stain two-dimensional classes of B100. Box size is 458 Å × 458 Å.

(F) Cartoon model of B100. In the absence of FANCL, the dimerization core remains unperturbed. The poles in the particle are more flexible.

(G) Coomassie-stained SDS-PAGE gel of the isolated FAAP100 subunit.

(H) Comparison between FAAP100 two-dimensional classes and two-dimensional classes of B100, focused on one pole in the particle. Box size is 458 Å × 458 Å.

(I) Cartoon of the BL100 architecture derived from single-particle EM combined with subunit dropout experiments.

See also Figures S2 and S3.

architecture of a dimeric BL100 assembly. Difference mapping suggests that FANCB is located at the homo-dimerization interface; however, we failed in our attempts to directly show that the isolated FANCB can self-associate. Therefore, we turned to XL/MS analysis, a powerful tool to study the architecture of macromolecular assemblies, especially when combined with structural EM (Bui et al., 2013) (Figure 3A; Table S1). We have used an amino-reactive bi-functional crosslinker (BS3) to covalently bridge between surface-exposed primary amines that map within a distance range of 26–30 Å between Cα of lysine residues. Samples were crosslinked and trypsinized, enriched for larger crosslinked peptides through gel filtration, and analyzed by nanoLC-MS (nano-liquid chromatography-mass spectrometry) using an LTQ Orbitrap Velos. We first validated the XL/MS data by comparing our results with the available crystal structure of the human FANCL central domain. All detected crosslinks in FANCL exist within solvent-exposed lysines that are spaced within a range compatible with the dimensions of the crosslinker spacing arm (Figure 3B). This control provided us with confidence in the interaction network built for the rest of the protein assembly. Remarkably, FANCB contains, by far, the highest frequency of intra-molecular crosslinks, followed by FANCL (Figures 3A and 3C). Only one FAAP100 intra-molecular crosslink was observed. Possible explanations for this deficiency include the unusually low number of lysines in this subunit (1.9% of the sequence, with 5.7% being the average in human proteins), their location in large peptides lacking trypsin-cleavage sites, or the absence of proximal lysines in the tertiary structure of FAAP100. Instead, the majority of crosslinked lysines observed for FAAP100 were involved in inter-molecular crosslinks with FANCB. Altogether, these data provide

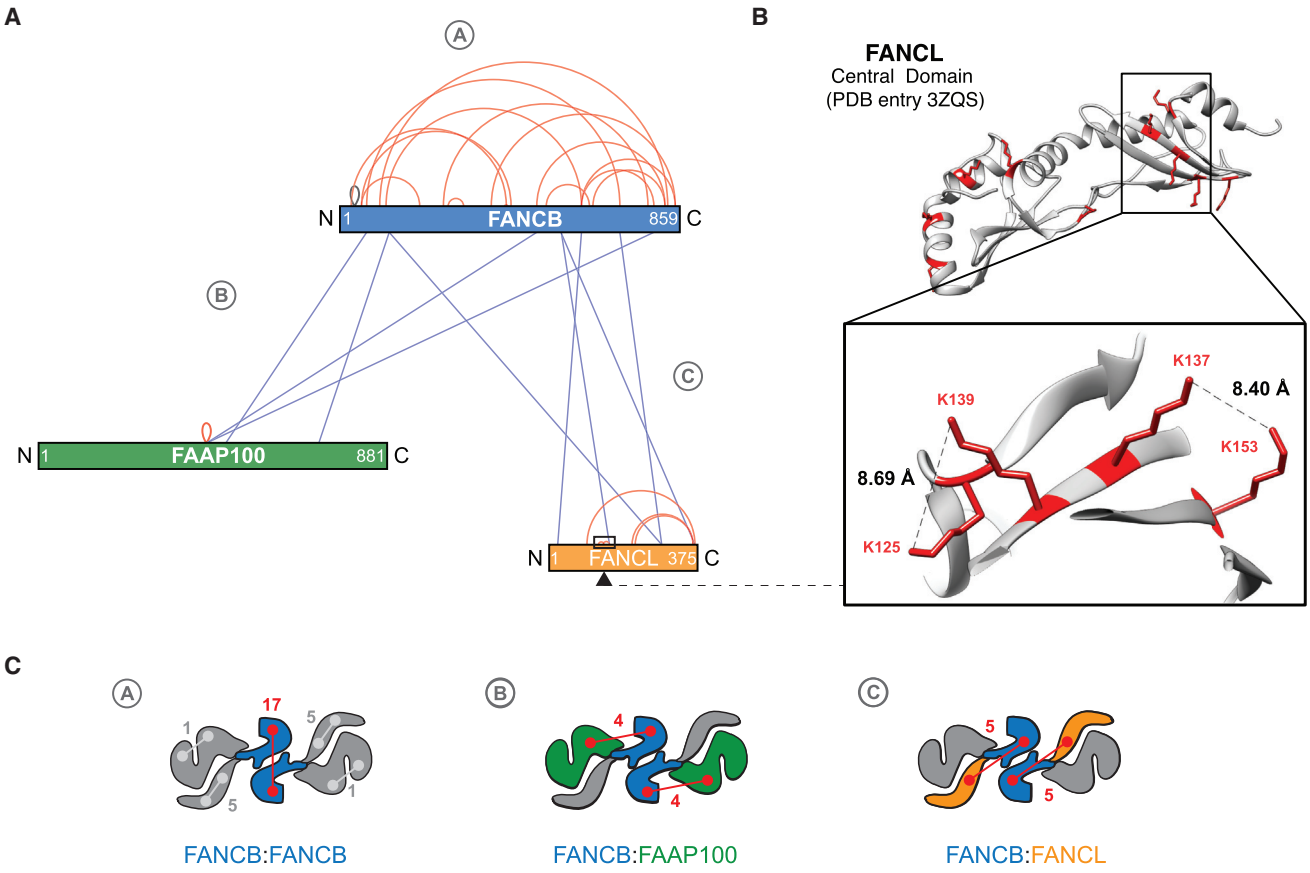


Figure 3. Subunit Architecture of the FANCB-FANCL-FAAP100 Assembly Studied by XL/MS

(A) Inter-subunit crosslinks between the six-protein BL100 complex. The lengths of the subunits correspond to the lengths of the colored rectangles that they represent: blue, FANCB; green, FAAP100; orange, FANCL. Red lines represent intra-molecular crosslinks. Blue lines represent inter-molecular crosslinks.

(B) Crystal structure of the human FANCL central domain with detected lysine residues highlighted in red (PDB: 3ZQS). All detected crosslinks in the FANCL region covered by the crystal structure map involve lysines that are surface exposed and are spaced within a range compatible with the spacer arm in the BS3 crosslinking agent.

(C) Cartoon representation integrating the crosslinking/mass spectrometry and single-particle analysis characterization of the BL100 architecture. The total number of detected inter-\intra-molecular crosslinks is indicated for each cartoon.

See also Table S1.

additional support to the notion that FANCB maps at the homo-dimerization interface in the core complex (Figures 2 and 3C). We note that one of the intra-FANCB crosslinks detected involves surface-exposed lysines (in position 35 and 39), which flank a known FA-associated mutation (L37S). Given the immediate proximity of Lys35 and Lys39, Leu37 is likely solvent exposed and could serve a role in interacting with other FA proteins. Further confidence in our XL/MS results is derived from the striking match between the inter-molecular crosslinks and our biochemical/EM analysis of the BL100 assembly. For example, four crosslinks connect FAAP100 with FANCB, in accord with our finding that FANCB-FAAP100 can exist as an isolated assembly (Figures 3A and 3C). FANCL and FANCB are connected by five crosslinks, which agrees with the observation that the MBP N-terminally fused to FANCL maps in close proximity to the central FANCB homo-dimer (Figure 2B). Conversely, FANCL and FAAP100 are not crosslinked, in agreement with our findings that (1) co-expression of FANCL and FAAP100 does not yield a L100 hetero-dimer (data not shown) and that (2) the two subunits only appear loosely associated in the globular pole in the BL100 particle (Figure 2I). In conclusion, XL/MS analysis, combined with our single-particle EM, provides a compelling architectural description of the dimeric BL100 catalytic assembly of the FA core complex.

Two Copies of CEF Can Associate with the BL100 Dimer

We have, so far, established that the BL100 catalytic assembly of the FA core complex forms a homo-dimer. Whether this arrangement extends to the rest of the core complex, or whether the full assembly contains a symmetry mismatch element, remains to be established. Therefore, we decided to purify the CEF assembly and image it in isolation (as an MBP-FANCC fusion derivative to increase the overall molecular mass of the complex) (Figures 4A and S4). (MBP)CEF appears as an electron-dense, tri-lobed entity with no detectable 2-fold symmetry component (Figures 4B and S4). We have then reconstituted a BL100-(MBP)CEF

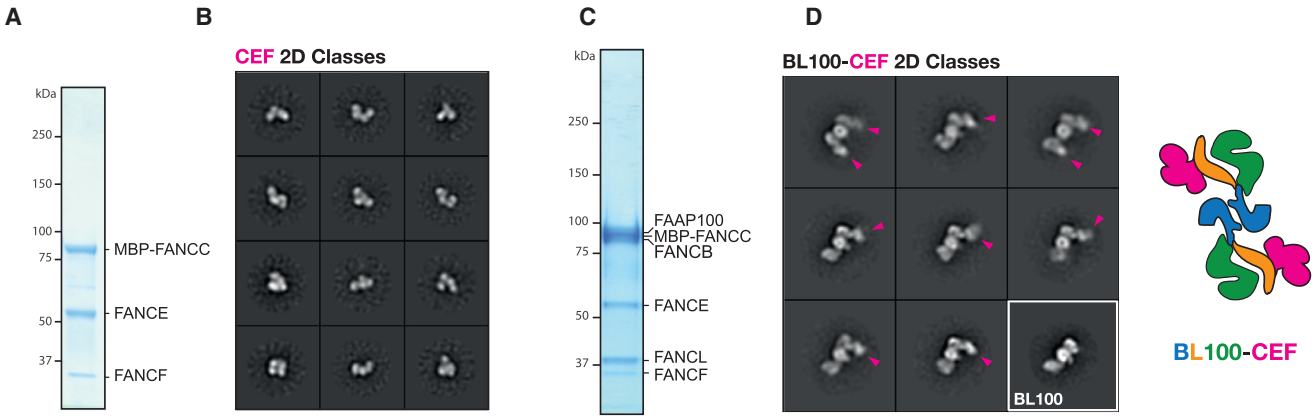


Figure 4. The FANCC-FANCE-FANCF Complex and Its Interaction with FANCB-FANCL-FAAP100

(A) Coomassie-stained SDS-PAGE gel of the CEF complex, containing an MBP N-terminally fused to FANCC.

(B) Negative-stain two-dimensional classes of the of the MBP-CEF complex. Box size is 441 Å × 441 Å.

(C) Coomassie-stained SDS-PAGE gel of the reconstituted and purified MBP-CEF/BL100 assembly.

(D) Negative-stain two-dimensional classes of the reconstituted and purified (MBP)CEF-BL100 complex. Purple arrowheads point at the MBP-CEF density. Due to the flexible nature of the BL100-(MBP)CEF interaction, the MBP-CEF complex appears slightly smaller when BL100 is incorporated. A view of the isolated BL100 is shown, framed in white, for comparison. Box size is 580 Å × 580 Å.

(E) Cartoon representation of the BL100-CEF architecture.

See also Figure S4.

co-assembly (Figures 4C and S4) and found that particles maintain the characteristic dimeric architecture of BL100 and are decorated with either one or two CEF additional densities, connected to the globular poles in the bow-tie structure (Figures 4D and S4). Therefore, binding of two copies of CEF to the BL100 complex is physically possible, indicating that the BL100 dimer acts as an organizing center for other core complex components. Whether two copies of AG20 are found in the full FA core complex remains to be established and will be the focus of future studies.

CEF-Co-expressed ID2 Particles Are Compact and Symmetric

In the BL100-CEF complex, CEF projects outward from the BL100 dimerization center and appears perfectly poised to latch onto the outer perimeter of the ID2 mono-ubiquitination substrate (Figure 4D). A direct contact between FANCE and FANCD2 has been previously reported (Gordon et al., 2005; Pace et al., 2002; Polito et al., 2014), and in an accompanying study, we show that human CEF can bind both FANCI and FANCD2 directly (van Twest et al., 2017). Reconstituting a CEF-ID2 assembly could, therefore, provide key insights into the mechanism of the FANCI-FANCD2 mono-ubiquitination by the FA core complex.

Xenopus laevis ID2 can be expressed and purified to homogeneity and efficiently processed in the mono-ubiquitination reaction by the human FA core complex (van Twest et al., 2017). To gain mechanistic insights into ID2 mono-ubiquitination by the human FA core complex, we chose to produce *Xenopus* ID2 for single-particle EM studies (Figure S5). In agreement with recently published work (Liang et al., 2016), we observed that ID2 forms an asymmetric, flexible, dimeric complex, which resists averaging and high-resolution structure determination (Figure S5). We hypothesized that CEF binding might stabilize the ID2 assembly, and we co-expressed the five proteins in the attempt to purify an ID2-CEF complex by affinity purification via a FLAG-tagged FANCI subunit. This purification strategy, including stringent ion exchange and size exclusion chromatography (SEC) steps, yielded an apparently stoichiometric ID2 assembly, according to Coomassie-stained SDS-PAGE gel analysis, but only trace amounts of the CEF assembly (Figure S6). As expected from inspection of the SDS-PAGE gel, the vast majority of ID2 particles show no sign of CEF co-complex, and no clear evidence for CEF association was detected by inspection of the two-dimensional classes. To our surprise, however, ID2 particles from the CEF co-expression experiments yielded neatly averaging two-dimensional classes (Figure S6). A pseudo-2-fold symmetric end-on view can be recognized in the assembly and perfectly matches the two-dimensional projections of an electron density map generated from the PDB coordinates of the mouse ID2 crystal structure (PDB: 3S4W) (Joo et al., 2011) (Figure 6A). To confirm this unexpected result, we repeated the co-expression experiment using a different CEF derivative, containing a Strep II tag on FANCC (Figure 5A). Single-particle analysis on the reconstituted complex confirms our initial observations (Figures 5B–5D and S6). To obtain a three-dimensional structure of the stabilized ID2 assembly, we generated an initial model using EMAN2, followed by three-dimensional classification, and auto-refined the most populated three-dimensional class to 18.3 Å resolution using RELION (Figure 6B; Figure S7). Rigid-body docking of the mouse atomic structure into the *Xenopus* ID2 map results in an excellent fit, with a striking match at the hetero-dimerization interface (Figure 6B). Conversely, minor adjustments would be needed in the C-terminal region of the two protomers to perfectly fit the atomic structure into the EM density (although we chose not to perform these local adjustments,

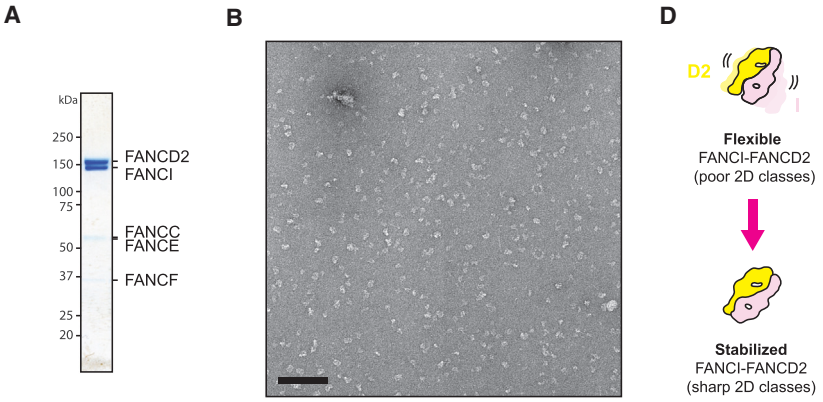


Figure 5. CEF Co-expression Stabilizes the FANCI-FANCD2 Interface

(A) Coomassie-stained SDS-PAGE gel of the ID2 complex, with trace amounts of CEF complex.

(B) Representative negative-stain micrograph of the ID2 assembly. Scale bar, 100 nm.

(C) Two-dimensional class averages of the CEF-stabilized ID2 assembly. Box size is 441 Å × 441 Å.

(D) CEF has a structural role in reshaping the ID2 interface, summarized in a cartoon.

See also Figures S5 and S6.

given the limited resolution of our three-dimensional EM data). Coherent with a more flexible nature of the FANCI and FANCD2 carboxy-terminal modules, analysis of complex dynamics based on two-dimensional class averages of end-on views indicates that the peripheral lobes in the ID2 dimer are flexible (Movie S3).

In summary, our characterization of the ID2-CEF interaction reveals that the two protein assemblies do not form a stoichiometric complex, at least in the experimental conditions tested. Remarkably, however, CEF co-expression alters the configuration of the ID2 dimer, stabilizing the dimerization interface and locking the assembly in a pseudo-2-fold symmetric structure (Figure 5D). This unexpected observation has key implications for the mechanism of ID2 mono-ubiquitination and suggests a role for the CEF assembly of the FA core complex.

DISCUSSION

Because the two strands of DNA are antiparallel (Watson and Crick, 1953), the factors that spatially organize, cut, or open the double helix are often arranged as symmetric dimers. For example, a dyad element can be found in the protein core of the nucleosome (the building block of chromatin in eukaryotic cells) (Richmond et al., 1984), in DNA integrases (which concertedly cut the two DNA strands to insert foreign DNA) (Engelman and Cherepanov, 2014), or in DNA helicases that unwind the double helix (Li et al., 2015). This 2-fold symmetric architecture can then be propagated to a subsequent set of proteins that physically interact with the DNA-binding dimer. A paradigm for this concept is found in the eukaryotic origin of DNA replication, where two replisomes are built symmetrically around a dimer of ring-shaped helicases (Araki, 2016; Costa et al., 2014; Deegan et al., 2016) by a set of regulatory and chaperone proteins that are, in turn, symmetric dimers (Itou et al., 2015).

In our study, we report one additional example of how a 2-fold symmetric enzyme modifies a dimeric DNA interactor. In this case, a homo-dimeric regulator protein acts symmetrically on a hetero-dimeric assembly. In the FA/BRCA ICL repair pathway, the FANCI-FANCD2 dimer binds the double helix at the damaged site and acts as a landing platform that recruits a set of DNA-repair proteins for ICL processing (Ceccaldi et al., 2016). FANCI and FANCD2 are structural homologs that can arrange in a pseudo-2-fold-symmetric assembly (Joo et al., 2011). Recruitment of the DNA-repair machinery requires mono-ubiquitination of both factors (Smogorzewska et al., 2007). This function is provided by the FA core complex, an enzyme assembly whose architecture was previously unknown (Ceccaldi et al., 2016; Walden and Deans, 2014). Here, we showed that the trimeric center of the FA core complex (BL100) exists as a symmetric homo-dimer (Figure 2C). FANCB and FAAP100 play an architectural role, orienting the two E3 ligases in an ideal position to symmetrically target the FANCI and FANCD2 ubiquitination sites (Figure 2I). In particular, FANCB plays a key structural function, being the dimerization factor. We note that FANCB is the only protomer in the BL100 catalytic module that contains disease-associated missense mutations, emphasizing the key role of this factor in the assembly of a functional FA core complex (Chandrasekharappa et al., 2013) (http://databases.lovd.nl/shared/variants/FANCB/unique?search_var_status=%3D%22Marked%22|%3D%22Public%22). Inspection of EM two-dimensional class averages reveals that the BL100 particle is dynamic, with the two catalytic poles of the complex free to tilt and stretch with respect to the fixed FANCB dimerization center (Figure S1; Movie S1). Remarkably, this movement occurs symmetrically across the particle (when the upper lobe moves to the left, the lower lobe moves to the right, and vice versa). If the catalytic subunit FANCL is omitted, however, the two external lobes in the particle move independently of each other (Figure S3;

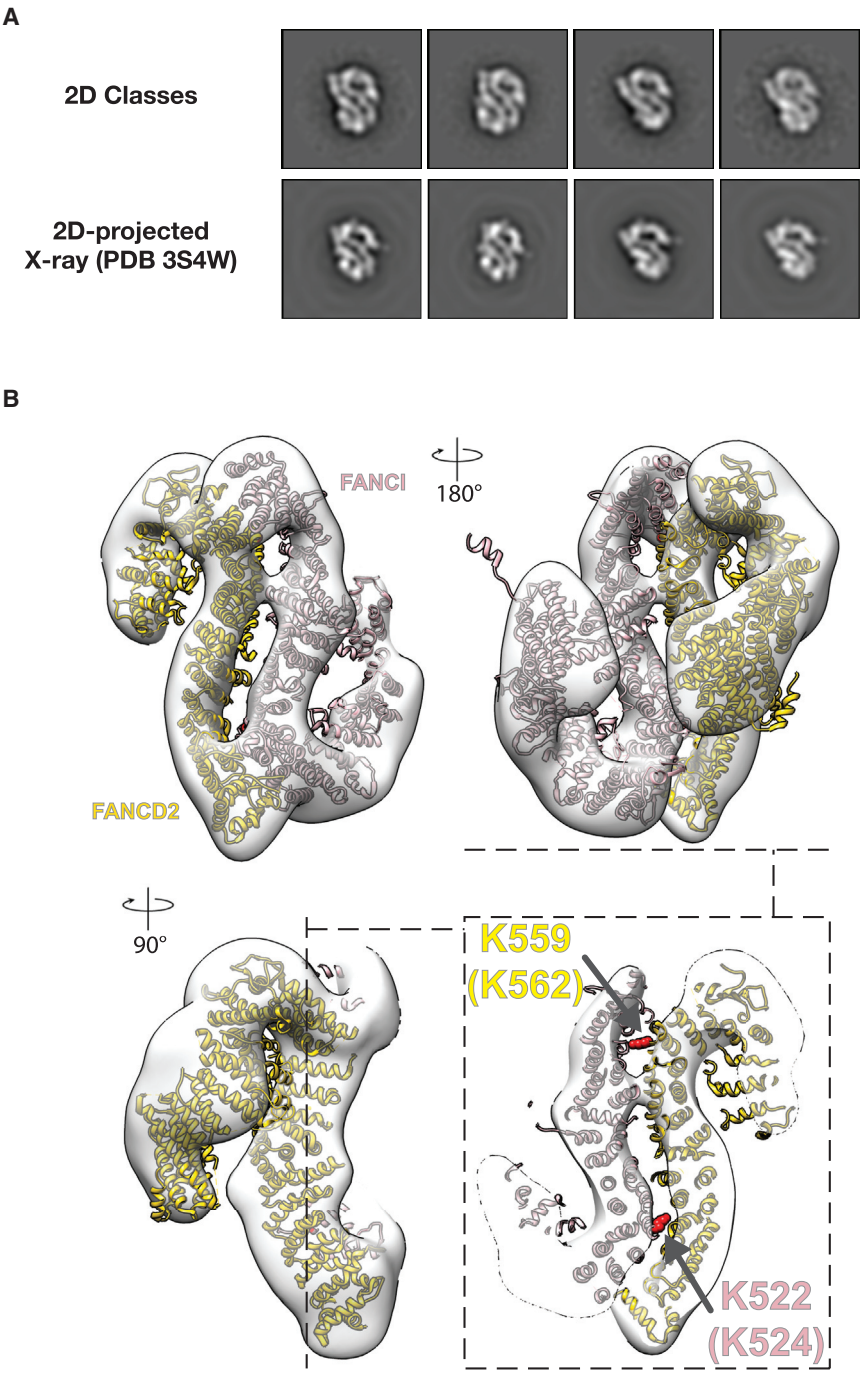


Figure 6. Three-Dimensional Reconstruction of the CEF-Co-expressed FANCI-FANCD2 Assembly

(A) Representative two-dimensional classes of the CEF-co-expressed ID2 complex matched to the two-dimensional projections of the electron density map generated from the ID2 crystallographic atomic coordinates (PDB: 3S4W). Box size is 350 Å × 350 Å.

(B) Three orientations and a cut-through view of the *X. laevis* ID2 EM structure with docked mouse ID2 crystal structure. The fit is excellent, in particular, at the dimerization interface. Lysines targeted for mono-ubiquitination are highlighted in red (residue numbering refers to the mouse proteins, and the corresponding lysines in the *X. laevis* ortholog are indicated in parentheses).

See also Figure S7.

Movie S2). Although the functional significance of this concerted movement is unclear, it is evident that each of the two catalytic subunits in the assembly must be able to sense the state of the symmetry mate in the assembly, and this allosteric communication must be mediated by the FANCB dimerization center. In summary, our study provides the first insights into the mechanics of this dimeric E3 ligase super-assembly and starts to explain how the FANCL mono-ubiquitination activity can be stimulated by the FANCB and FAAP100 architectural factors (Rajendra et al., 2014).

The BL100 assembly also functions as the organizing center of the FA core complex, providing a dimerization scaffold for the recruitment of a set of two CEF trimeric assemblies (Figures 4D and S4). The FANCE subunit of CEF has been previously reported to interact with the ID2 assembly (Gordon et al., 2005; Pace et al., 2002; Polito et al., 2014), and our structural data suggest that CEF might act as a molecular matchmaker to bridge between the ID2 mono-ubiquitination substrate and the BL100 E3 ligase machinery. Previous structural studies suggest that the ID2 dimer might exist in different configurations during the ubiquitination reaction. For example, crystallization of the murine complex captured a symmetric, compact state of the ID2 assembly, with the ubiquitination sites mapping at the dimerization interface, probably impairing access to the ubiquitination machinery (Joo et al., 2011). Conversely, a cryo-EM study of the human ID2 shows that the complex is more flexible (poorly averaged) in solution (Liang et al., 2016). In agreement with these recent findings, we show that the intact *Xenopus* ID2 complex expressed in isolation cannot be captured in discrete conformational states (two-dimensional images do not average) (Figure S5). In contrast, the co-expression of CEF with ID2 appears to stabilize a symmetric state of ID2, which is highly reminiscent of the form captured in the crystal structure (Figures 5C and 6; Movie S3). This hints at a possible role for CEF in manipulating the ID2 substrate to facilitate symmetric mono-ubiquitination of the two related protomers (Figure 7). Biochemical analysis of the

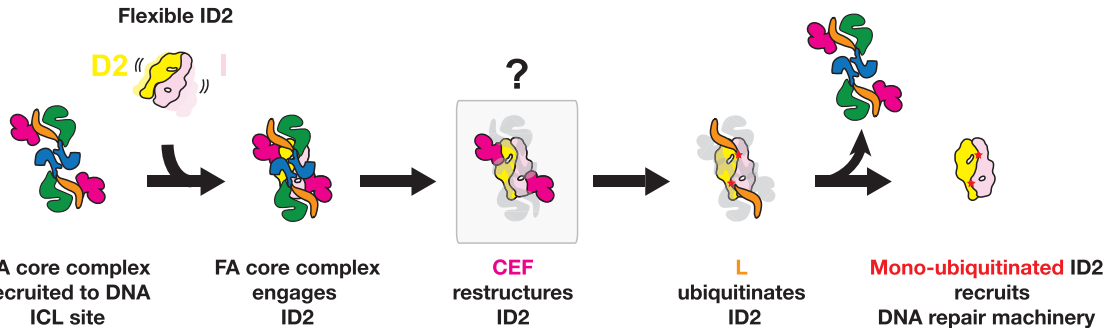


Figure 7. Molecular Model for the Mechanism of FANCI-FANCD2 Mono-ubiquitination by the FA Core Complex

reconstituted ID2 mono-ubiquitination reaction by the FA core complex supports our suggestion that the CEF assembly acts as a nexus between the BL100 enzyme and the ID2 substrate while also reconfiguring the ID2 structure (van Twest et al., 2017). Whereas the isolated BL100 assembly solely ubiquitinates the FANCD2 subunit, the presence of CEF significantly stimulates FANCI mono-ubiquitination, necessary for triggering initiation of the FA/BRCA ICL repair pathway. How BL100-CEF and ID2 interact and influence one another remains to be explained at a molecular level. We note that a disease-associated mutation on FANCE (R371W), predicted to destabilize the FANCE fold, abrogates the interaction with FANCD2 in a yeast two-hybrid study. The same Arg-to-Trp point mutation on FANCD2 (R302W), mapping within the outer perimeter of the N-terminal ID2 dimerization module, likewise abrogates the FANCE-D2 interaction (Joo et al., 2011; Nookala et al., 2007; Timmers et al., 2001). Such elements likely play an important role in the CEF-mediated stabilization of the ID2 substrate during the mono-ubiquitination reaction catalyzed by the homo-dimeric BL100 catalytic module. Notably, in addition to interacting with FANCE, the same R302 element in FANCD2 has been previously implicated in mediating a histone H3 contact (Sato et al., 2012). How the ID2 mono-ubiquitination reaction occurs in the context of chromatinized DNA is an exciting frontier topic.

Conclusions

Several outstanding questions will need to be addressed to understand the mechanism of activation of the FA/BRCA ICL repair pathway. For example: how do disease mutations affect the structure and function of the FA core complex? What is the molecular mechanism for the formation of the FA core complex in cells? How does DNA influence ID2 mono-ubiquitination? Our discovery that the FA core complex is a symmetric dimer changes the way we think about these problems.

EXPERIMENTAL PROCEDURES

Construction of Baculoviruses

Baculoviruses were constructed as detailed in the Supplemental Information.

Protein Expression and Purification

For the expression of the FANCB-FANCL-FAAP100 complex and the isolated FAAP100, 1 L of Sf9 cells at a titer of 2.0 × 10^6 cells per milliliter were suspended in 4 L spinner flasks (Corning) and infected with baculovirus at an MOI of 2 for 72 hr at 27°C. Cells were collected post-infection by centrifugation (5,000 rpm, 4°C, 15 min), washed in PBS buffer supplemented with 5 mM $MgCl_2$, and resuspended in lysis buffer (25 mM HEPES NaOH [pH 7.6], 15 mM NaCl, 0.02% Tween 20, 10% glycerol, 1 mM EDTA, 1 mM EGTA, 2 mM $MgCl_2$, 2 mM 2-BME, 0.4 mM PMSF, and Roche protease inhibitor cocktail). The suspension was lysed by Dounce homogenization (40 strokes, tight pestle), and NaCl was added to a final concentration of 250 mM. The lysate was cleared by centrifugation (15000 rpm, 4°C, 35 min), and the supernatant was incubated (3 hr, 4°C) with anti-FLAG M2 affinity resin (Sigma Aldrich) pre-equilibrated with FLAG washing buffer (25 mM HEPES NaOH [pH 8.0], 200 mM NaCl, 0.02% Tween 20, 10% glycerol, 1 mM EDTA, 1 mM EGTA, and Roche protease inhibitor cocktail). After wash, the sample was eluted with FLAG elution buffer (25 mM HEPES NaOH [pH 8.0], 100 mM NaCl, 0.02% Tween 20, 5% glycerol, 1 mM EDTA, 1 mM EGTA, 0.3 mg/mL FLAG peptide, and Roche protease inhibitor cocktail). Elution fractions were combined and loaded onto a Mono Q 5/50 GL for ion-exchange chromatography (GE Healthcare). The column was washed with Q-Wash buffer (25 mM HEPES NaOH [pH 8.0], 150 mM NaCl, 5% glycerol, 1 mM EDTA, 1 mM EGTA, 0.5 mM TCEP, and Roche protease inhibitor cocktail) and eluted over a 150-mM to 650-mM NaCl linear gradient in the same buffer. The peak fraction from cation-exchange chromatography was loaded onto a Superose 6 PC 3.2/30 column (GE Healthcare) operating on an ÄKTAmicro HPLC (high-performance liquid chromatography) system (GE Healthcare) and was equilibrated in 25 mM HEPES NaOH (pH 8.0), 250 mM NaCl, and 1 mM TCEP. The proteins were separated by SDS-PAGE, using 4%–12% Bis-Tris gels (Invitrogen) running for 50 min at 150 V in MOPS buffer or using 3%–8% Tris-acetate gels (Invitrogen) running for 60 min at 150 V in Tris-acetate buffer. Protein identities were confirmed by MS. Protein concentration was determined by measuring absorbance at 280 nm (A_{280nm}) using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies).

Expression of the FANCB-FAAP100 and FANCB-(MBP)FANCL-FAAP100 complexes were performed using the FANCB-FANCL-FAAP100 protocol; however, infection was performed using an MOI of 1 for 72 hr at 27°C. Protein extraction, purification, and identification were performed as described earlier. Freshly purified samples (BL100 at 1 mg/mL, B100 at 0.25 mg/mL, and FAAP100 at 1.5 mg/mL) were immediately used for negative-stain EM or cryo-EM experiments.

Expression and purification of (MBP)FANCC-FANCE-FANCF was performed as follows. 1 L of Sf9 cells at a titer of 1.0 × 10^6 cells per milliliter was suspended in 4-L spinner flasks (Corning) and infected with baculovirus at an MOI of 1 for 72 hr at 27°C. Cells were collected post-infection by centrifugation (5,000 rpm, 4°C, 15 min), washed in PBS buffer supplemented with 5 mM $MgCl_2$, and resuspended in lysis buffer (25 mM HEPES NaOH [pH 7.6], 15 mM NaCl, 0.02% Tween 20, 10% glycerol, 1 mM EDTA, 1 mM EGTA, 2 mM $MgCl_2$, 15 mM 2-BME, 0.6 mM PMSF, and Roche protease inhibitor cocktail). The suspension was lysed by Dounce homogenization (40 strokes, tight pestle), and NaCl was added to a final concentration of 200 mM. The lysate was cleared by centrifugation (15,000 rpm, 4°C, 35 min), and the supernatant was incubated (2 hr, 4°C) with Amylose resin (New England Biolabs) pre-equilibrated with Amylose wash buffer (25 mM

HEPES NaOH [pH 8.0], 150 mM NaCl, 0.02% Tween 20, 10% glycerol, 1 mM EDTA, 1 mM EGTA, 0.25 mM TCEP, and Roche protease inhibitor cocktail). After extensive wash, the sample was eluted by ligand competition using the Amylose wash buffer supplemented with 20 mM maltose (Sigma). Peak fractions were pooled and loaded onto a Mono Q 5/50 GL for ion-exchange chromatography (GE Healthcare). The column was washed with buffer Q-A (50 mM HEPES NaOH [pH 8.0], 150 mM NaCl, 5% glycerol, 1 mM EDTA, and 0.5 mM TCEP) and eluted over a linear salt gradient with buffer Q-B (50 mM HEPES NaOH [pH 8.0], 1 M NaCl, 5% glycerol, 1 mM EDTA, and 0.5 mM TCEP). Peak fraction was successively loaded onto a Superose 6 PC 3.2/30 column (GE Healthcare) operating on an ÄKTAmicro HPLC system (GE Healthcare) and separated by SEC in 25 mM HEPES NaOH (pH 8.0), 100 mM NaCl, and 2 mM DTT. Peak fraction was analyzed by SDS-PAGE. Instant Blue Safe-Stain-stained (Expedeon) bands were identified by MS. Sample concentration was measured by A_{280nm} using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies) and the (MBP)CEF theoretical extinction coefficient. Freshly purified (MBP)CEF sample at a final concentration of 0.5 mg/mL was immediately used for preparation of negative-stain EM grids or reconstitution with freshly purified BL100 sample.

Expression and purification of the FANCI-FANCD2 complex was performed as follows. Briefly, 1 L of Sf9 cells at a titer of 2.0 × 10^6 cells per milliliter was suspended in 4-L spinner flasks (Corning) and co-infected with the two baculoviruses, each at MOI of 1, for 72 hr at 27°C. Following the co-infection, cells were collected by centrifugation (5,000 rpm, 4°C, 15 min) and washed in PBS buffer supplemented with 5 mM $MgCl_2$. Cells were resuspended in lysis buffer (25 mM HEPES NaOH [pH 7.6], 15 mM NaCl, 0.02% Tween 20, 10% glycerol, 1 mM EDTA, 1 mM EGTA, 2 mM $MgCl_2$, 2 mM 2-BME, 0.4 mM PMSF, and Roche protease inhibitor cocktail) and lysed by Dounce homogenization (40 strokes, tight pestle). To preserve complex integrity, the lysate was added with NaCl to a final concentration of 150 mM and cleared by centrifugation (15,000 rpm, 4°C, 35 min). The supernatant was collected and incubated (3 hr, 4°C) with anti-FLAG M2 affinity resin (Sigma-Aldrich) pre-equilibrated in FLAG washing buffer (25 mM HEPES NaOH [pH 8.0], 120 mM NaCl, 0.02% Tween 20, 10% glycerol, 1 mM EDTA, 1 mM EGTA, and Roche protease inhibitor cocktail). After extensive wash, the proteins were eluted by incubation of the resin in FLAG elution buffer (25 mM HEPES NaOH [pH 8.0], 120 mM NaCl, 1 mM EDTA, 0.3 mg/mL FLAG peptide, and Roche protease inhibitor cocktail). Peak fractions were combined and loaded onto a Mono Q 5/50 GL(GE Healthcare) column running in buffer Q-A (25 mM HEPES NaOH [pH 8.0], 100 mM NaCl, 5% glycerol, 1 mM EDTA, and 0.25 mM TCEP). Elution of the sample was performed over a linear salt gradient with buffer Q-B (25 mM HEPES NaOH pH 8.0], 750 mM NaCl, 5% glycerol, 1 mM EDTA, and 0.5 mM TCEP). To increase sample homogeneity, the complex was further purified by pull-down of the Strep-tag II on FANCD2. Peak fraction from ion-exchange chromatography was incubated for 2 hr at 4°C with Strep-tactin Superflow plus resin (QIAGEN). The resin was washed in Strep-wash buffer (25 mM HEPES NaOH [pH 8.0], 120 mM NaCl, 10% glycerol, 1 mM EDTA and 0.25 mM TCEP) and eluted by incubation with Strep-wash buffer supplemented with 25 mM desthiobiotin. Fractions containing stoichiometric I-D2 complex were pooled and concentrated on Amicon Ultra MWCO 10K centrifugal filters (Millipore). Concentrated sample was separated by size exclusion on a Superdex 200 PC 3.2/30 column (GE Healthcare) operating on an ÄKTAmicro HPLC system (GE Healthcare) running in 25 mM HEPES NaOH (pH 8.0), 120 mM NaCl, and 1 mM TCEP. Elution peak was analyzed by SDS-PAGE followed by Instant Blue SafeStain staining (Expedeon). Band identity was confirmed by MS. Freshly purified ID2 sample at a final concentration of 0.25 mg/mL was immediately used for preparation of negative-stain EM grids.

Co-expression of FANCI-FANCD2-FANCC-FANCE-FANCF was obtained by infecting Sf9 cells with baculoviruses encoding for FANCI, FANCD2, and either MBP-tagged or StrepII-tagged CEF. Briefly, 1 L Sf9 cells at a titer of 2.0 × 10^6 cells per milliliter were co-infected with baculoviruses encoding for FLAG-FANCI, StrepII-FANCD2, and (MBP)CEF (or (StrepII)CEF), each at a MOI of 1. Cells were incubated for 72 hr at 27°C. Protein extraction, purification, and identification were carried out as previously described for the ID2 complex. The ID2-(StrepII)CEF sample purified from ion-exchange chromatography and the ID2-(MBP)CEF sample purified from SEC were used for preparation of EM grids by negative stain immediately after purification.

Reconstitution of the BL100-(MBP)CEF Complex

The reconstitution of the FANCB-FANCL-FAAP100-(MBP)FANCC-FANCE-FANCF complex was performed by mixing equal volumes of a freshly purified 1-μM solution of BL100 and a 3-μM solution of CEF (resulting in a molar ratio of 1:3) and loaded on a 30K MWCO Slide-a-Lyzer dialysis cassette (Thermo Scientific). The protein mix was first dialyzed for 1 hr at 4°C in Buffer 500 (25 mM Tris-HCl [pH 8.0], 500 mM NaCl, 5% glycerol, and 1 mM DTT), then for 2 hr in Buffer 350 (25 mM Tris-HCl [pH 8.0], 350 mM NaCl, 5% glycerol, and 1 mM DTT), and finally overnight in Buffer 100 (25 mM Tris-HCl [pH 8.0], 100 mM NaCl, 5% glycerol, and 1 mM DTT). The sample was recovered from the dialysis cassette and loaded onto a Superose 6 PC 3.2/30 column (GE Healthcare) operating on a ÄKTAmicro FPLC (fast protein liquid chromatography) system (GE Healthcare) equilibrated in 25 mM Tris-HCl (pH 8.0), 100 mM NaCl, and 2 mM DTT. In control experiments, isolated dialyzed BL100 and (MBP)CEF complexes were also subject to SEC as described for the reconstituted samples. Samples were analyzed by SDS-PAGE on 3%–8% Criterion XT Tris-Acetate gels (Bio-Rad) followed by Instant Blue SafeStain staining (Expedeon). Freshly purified samples were immediately used for the preparation of EM grids.

Negative-Stain EM Sample Preparation

A 2- to 6-nm-thick carbon layer was evaporated onto freshly cleaved mica sheets (EMS) using a Q150TE carbon coater (Quorum Technologies). Carbon was incubated overnight at 50°C before floating on 600-mesh copper grids (Agar Scientific) or Quantifoil R1.2/1.3 300-mesh copper grids. Prior to sample incubation, the carbon-coated grid was glow discharged for 30 s at 45 mA using a 100× glow discharger (EMS). Serial dilutions of each protein sample were performed to achieve optimal particle concentration on the EM grid (final concentrations ranged from 30 to 5 μg/mL). A 4-μL drop of the freshly purified sample solution was applied onto the glow-discharged grid and incubated for 1 min. Approximately 2 μL of sample solution was absorbed by gentle side blotting, and the grid was immediately stained with four subsequent applications onto 4× 50-μL drops of a 2% (wt/vol) uranyl formate solution with gentle stirring for 8 s. After staining, the grid was blotted dry and stored at room temperature prior to imaging.

Negative-Stain EM Data Collection

Particles of reconstituted BL100-(MBP)CEF, isolated (MBP)CEF, isolated ID2, and ID2 co-expressed with (MBP)CEF were imaged on a Tecnai G2 Spirit LaB6 transmission electron microscope (TEM; FEI) operating at 120 keV (Electron Microscopy Science Technology Platform, The Francis Crick Institute). Images were manually acquired on a UltraScan 1000 2k × 2k charge-coupled device (CCD) camera (Gatan) at a nominal magnification of 30,000× (corresponding to a pixel size of 3.45 Å at the specimen level). Particles of BL100, B100, B-(MBP)L100, isolated FAAP100, and ID2 co-expressed with (StrepII)CEF samples were imaged using a JEM-2100 LaB6 electron microscope (JEOL) operated at 120 kV. Images were recorded on a UltraScan 4k × 4k CCD camera (Gatan) at a nominal magnification of 42,000× corresponding to a pixel size of 2.73 Å at the specimen level. On both microscopes, images were collected in low-dose mode (electron dose of 35 $e^-/Å^2$) and with defocus values in the range of 0.5 μm to 2.0 μm. For the BL100 and FAAP100 samples, the micrographs were automatically recorded using the JEOL Automated Data Acquisition System (JADAS) with same electron dose and defocus value as the manual collection. A total of 567 images were collected for the BL100 sample, 250 images for the B100 sample, 412 images for the B-(MBP)L100 sample, 540 images for the FAAP100 sample, 277 images for the (MBP)CEF sample, 455 images for the BL100-(MBP)CEF sample,712 images for the ID2 sample, 468 images for the ID2-(MBP)CEF sample, and 447 images for the ID2-(StrepII)CEF sample.

Cryo-EM Sample Preparation and Data Collection

Ultrathin carbon-coated Lacey 300-mesh copper grids (Ted Pella) were glow discharged for 30 s at 45 mA using a 100× glow discharger (EMS) and mounted onto a Vitrobot Mark IV (FEI). A 4-μL drop of freshly purified BL100 at a concentration of 20 μg/mL was applied to the glow-discharged grid. The sample was incubated on the grid for 1 min in 100% humidity at 21°C. After incubation, the grid was double-side blotted for 2.0 s and rapidly plunged in liquid ethane. Cryo-EM data for the BL100 sample were collected on a Titan

Krios TEM (FEI) operated at 300 keV and equipped with a back-thinned Falcon II (FEI) direct electron detector at a nominal magnification of 47,000× , resulting in a pixel size of 1.77 Å at the specimen level. A total of 2,798 images were recorded at defocus values in the range of 2.5–4.5 μm and with a cumulative electron dose of 56 e^{-}/Å^2 distributed over 26 frames. Beam-induced motion was corrected on each image using the Motioncorr program (Li et al., 2013). Micrographs showing contamination drift, or out-of-range defocus values were discarded.

Single-Particle Image Processing

For each dataset, particles were semi-automatically picked using the EMAN2 package, version 2.12 (Tang et al., 2007). From the negative-stain EM data collection, a total of 63,543 individual particles were extracted from the BL100 dataset, 63,365 particles from the B100 dataset, 51,941 particles from the B-(MBP)L100 dataset, 177,106 particles from the FAAP100 dataset, 36,768 particles from BL100-(MBP)CEF dataset, 47,898 particles from the (MBP)CEF dataset, 95,553 particles from the ID2 dataset, 27,084 particles from the ID2-(MBP)CEF dataset, and 71,295 particles from the ID2-(StrepII) CEF dataset. From the BL100 cryo-EM dataset, a total of 536,754 individual particles were extracted from selected micrographs. The parameters of the contrast transfer function (CTF) were estimated on each micrograph using CTFFIND3 (Mindell and Grigorieff, 2003) for negative-stain data and CTFFIND4 (Rohou and Grigorieff, 2015) for cryo-EM data. Reference-free two-dimensional alignment and averaging was performed using RELION, version 1.4 (Scheres, 2012). For three-dimensional reconstruction of the ID2 complex co-expressed with (StrepII)CEF, an initial three-dimensional volume was generated using the e2inimodel.py program in the EMAN2 package (Tang et al., 2007). The initial volume was used as a reference model for three-dimensional classification and refinement in RELION 1.4 (Scheres, 2012). The atomic model of the FANCI-FANCD2 hetero-dimer (PDB: 3S4W) (Joo et al., 2011) was used for rigid-body fitting into the EM map using UCSF (University of California, San Francisco) Chimera (Pettersen et al., 2004). The refined EM map of the ID2 hetero-dimer has been deposited in the three-dimensional-EM database (http://www.emdatabank.org) with accession number: EMD-3476.

XL/MS

All chemicals were purchased from Sigma at the highest grade possible, unless otherwise stated. BS3 d0 and BS3 d4 were purchased from ThermoFisher Scientific. nano-LC-MS solvents and additives were purchased as LC-MS grade from Fisher Scientific. The XL/MS workflow was conducted as previously described (Leitner et al., 2014). Briefly, BL100 dimer, at a concentration of 0.5 mg/mL in 25 mM HEPES (pH 8), 300 mM NaCl, 10% glycerol, and 0.25 mM tris(2-carboxyethyl)phosphine (TCEP), was reacted with 5 mM BS3 (using a 50:50 ratio of d0 and d4) for 30 min at 37°C. The reaction was quenched by adding 50 mM ammonium bicarbonate for 20 min at 37°C and then dried to completion using vacuum centrifugation. The sample was resuspended at a concentration of 1 mg/mL using 8 M urea, reduced with 2.5 mM TCEP, and alkylated with 5 mM iodoacetamide. The sample was diluted with 50 mM ammonium bicarbonate to reduce the urea concentration to 1 M, and trypsin (Promega) was added at a 1:50 enzyme:substrate ratio and left to digest overnight at 37°C. The solution was then acidified, and peptides were purified using C18 in house solid-phase elution tips (Empore, 3M).

For size exclusion fractionation, the peptides were resuspended in SEC buffer (70% H_2O, 30% acetonitrile, and 0.1% trifluoroacetic acid), and chromatographically separated using an ÄKTAmicro (GE Healthcare) with a Superdex Peptide PC 3.2/30 column (GE Healthcare) and detected using 215 nm absorbance. The flow rate was set to 50 μL/min, and 100 μL fractions were collected and dried to completion using vacuum centrifugation. For nano-LC-MS acquisition, the SEC fractions were resuspended into 30 μL of 0.1% trifluoroacetic acid in H_2O, and 1 μL was injected and chromatographically resolved using an Ultimate 3000 RSLCnano (Dionex) with an EASY-Spray column (2-μm particles, PepMap C18, 100-Å pore size, 50 cm × 75 μm ID) (Thermo Scientific). A flow rate of 0.25 μL/min was used, starting at 98% mobile A (0.1% formic acid, 5% DMSO, 95% H_2O) and 2% mobile B (0.1% formic acid, 5% DMSO, 80% acetonitrile, 10% H_2O). Over 80 min, Mobile B was then increased to 40% followed by a further increase to 90% over 10 min. Spectra were acquired in real time using an LTQ Orbitrap Velos Mass Spectrometer (Thermo Scientific) with a 400 to 2,000 m/z (mass-to-charge ratio) acquisition window. The top ten most intense ions with a charge state of +3 or greater were then selected for tandem MS (MS/MS) using data-dependent acquisition and CID (collision-induced dissociation) fragmentation (at 35% normalized collision energy). A dynamic exclusion list, with a repeat count of 1, repeat duration of 20 s, exclusion list size of 100, and exclusion duration of 30 s, was used to prevent repeat sampling. Data analysis was carried out using the xQuest/xProphet (v2.1.1) workflow. The data were searched for BS3 mono-links (+156.079 Da for water quench, +155.096 Da for ammonium quench), intra-links, and intra- and inter-protein crosslinks (+138.068 Da) on lysine residues, using isotopic pairing with a mass shift of +4.025 Da for BS3 d4 ions, in addition to a fixed carbamidomethylation modification (+57.021 Da) and variable methionine oxidation (+ 15.994 Da). The data were searched using a 10-ppm mass tolerance for charge states +3 or above. xProphet was then used, and peptides with a false discovery rate (FDR) below 5% were then manually assessed for adequate signal intensity and peak assignment.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, one table, and three movies and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2016.11.013.

AUTHOR CONTRIBUTIONS

P.S., A.J.D., and A.C. designed the study. A.J.D., V.J.M., S.v.T., and F.S. constructed all baculoviruses and performed preliminary protein preparations. P.S. optimized purification protocols for structural studies, prepared all proteins and EM grids, collected EM data (with L.R.), and performed all image analysis under the supervision of A.C. A.B. performed the XL/MS analysis supervised by B.S. P.S. and A.C. wrote the paper, with input from A.D. and the other authors.

ACKNOWLEDGMENTS

We would like to thank J. Gannon, A. Davies, and Cell Services at The Francis Crick Institute for help with the baculovirus work; and R. Carzaniga, L. Collinson, and K. MacLellan-Gibson for technical support with EM. We acknowledge Diamond for access and support of the cryo-EM facilities at eBIC, proposal EM14266, funded by Wellcome Trust/MRC/BBSRC. We thank A. Siebert, D. Clare, and S. Welsch for help with the Titan Krios. A.J.D. is funded by the National Breast Cancer Foundation (Australia), the Victorian Cancer Agency, the Fanconi Anemia Research Fund, the Victorian Government's OIS Program, and the Cancer Council of Victoria. Research in the A.C. lab is supported by The Francis Crick Institute (which receives its core funding from Cancer Research UK, the UK Medical Research Council, and the Wellcome Trust).

Received: August 2, 2016
Revised: September 22, 2016
Accepted: October 31, 2016
Published: December 13, 2016

REFERENCES

Alpi, A.F., Pace, P.E., Babu, M.M., and Patel, K.J. (2008). Mechanistic insight into site-restricted monoubiquitination of FANCD2 by Ube2t, FANCL, and FANCI. Mol. Cell *32*, 767–777.

Araki, H. (2016). Elucidating the DDK-dependent step in replication initiation. EMBO J. *35*, 907–908.

Brzovic, P.S., Rajagopal, P., Hoyt, D.W., King, M.C., and Klevit, R.E. (2001). Structure of a BRCA1-BARD1 heterodimeric RING-RING complex. Nat. Struct. Biol. *8*, 833–837.

Bui, K.H., von Appen, A., DiGuilio, A.L., Ori, A., Sparks, L., Mackmull, M.T., Bock, T., Hagen, W., Andrés-Pons, A., Glavy, J.S., and Beck, M. (2013). Integrated structural analysis of the human nuclear pore complex scaffold. Cell *155*, 1233–1243.

Ceccaldi, R., Sarangi, P., and D'Andrea, A.D. (2016). The Fanconi anaemia pathway: new players and new functions. Nat. Rev. Mol. Cell Biol. *17*, 337–349.

Chandrasekharappa, S.C., Lach, F.P., Kimble, D.C., Kamat, A., Teer, J.K., Donovan, F.X., Flynn, E., Sen, S.K., Thongthip, S., Sanborn, E., et al.; NISC Comparative Sequencing Program (2013). Massively parallel sequencing, aCGH, and RNA-seq technologies provide a comprehensive molecular diagnosis of Fanconi anemia. Blood *121*, e138–e148.

Chen, Z., Speck, C., Wendel, P., Tang, C., Stillman, B., and Li, H. (2008). The architecture of the DNA replication origin recognition complex in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA *105*, 10326–10331.

Ciccia, A., Ling, C., Coulthard, R., Yan, Z., Xue, Y., Meetei, A.R., Laghmani, 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.

Cole, A.R., Lewis, L.P., and Walden, H. (2010). The structure of the catalytic subunit FANCL of the Fanconi anemia core complex. Nat. Struct. Mol. Biol. *17*, 294–298.

Costa, A., Renault, L., Swuec, P., Petojevic, T., Pesavento, J.J., Ilves, I., MacLellan-Gibson, K., Fleck, R.A., Botchan, M.R., and Berger, J.M. (2014). DNA binding polarity, dimerization, and ATPase ring remodeling in the CMG helicase of the eukaryotic replisome. eLife *3*, e03273.

Coulthard, R., Deans, A.J., Swuec, P., Bowles, M., Costa, A., West, S.C., and McDonald, N.Q. (2013). Architecture and DNA recognition elements of the Fanconi anemia FANCM-FAAP24 complex. Structure *21*, 1648–1658.

Deans, A.J., and West, S.C. (2011). DNA interstrand crosslink repair and cancer. Nat. Rev. Cancer *11*, 467–480.

Deegan, T.D., Yeeles, J.T., and Diffley, J.F. (2016). Phosphopeptide binding by Sld3 links Dbf4-dependent kinase to MCM replicative helicase activation. EMBO J. *35*, 961–973.

Engelman, A., and Cherepanov, P. (2014). Retroviral integrase structure and DNA recombination mechanism. Microbiol. Spectr. *2*, MDNA3-0024-2014. http://dx.doi.org/10.1128/microbiolspec.MDNA3-0024-2014.

Garcia-Higuera, I., Taniguchi, T., Ganesan, S., Meyn, M.S., Timmers, C., Hejna, J., Grompe, M., and D'Andrea, A.D. (2001). Interaction of the Fanconi anemia proteins and BRCA1 in a common pathway. Mol. Cell *7*, 249–262.

Gordon, S.M., Alon, N., and Buchwald, M. (2005). FANCC, FANCE, and FANCD2 form a ternary complex essential to the integrity of the Fanconi anemia DNA damage response pathway. J. Biol. Chem. *280*, 36118–36125.

Hodson, C., Cole, A.R., Lewis, L.P., Miles, J.A., Purkiss, A., and Walden, H. (2011). Structural analysis of human FANCL, the E3 ligase in the Fanconi anemia pathway. J. Biol. Chem. *286*, 32628–32637.

Hodson, C., Purkiss, A., Miles, J.A., and Walden, H. (2014). Structure of the human FANCL RING-Ube2T complex reveals determinants of cognate E3-E2 selection. Structure *22*, 337–344.

Huang, Y., Leung, J.W., Lowery, M., Matsushita, N., Wang, Y., Shen, X., Huong, D., Takata, M., Chen, J., and Li, L. (2014). Modularized functions of the Fanconi anemia core complex. Cell Rep. *7*, 1849–1857.

Itou, H., Shirakihara, Y., and Araki, H. (2015). The quaternary structure of the eukaryotic DNA replication proteins Sld7 and Sld3. Acta Crystallogr. D Biol. Crystallogr. *71*, 1649–1656.

Joo, W., Xu, G., Persky, N.S., Smogorzewska, A., Rudge, D.G., Buzovetsky, O., Elledge, S.J., and Pavletich, N.P. (2011). Structure of the FANCI-FANCD2 complex: insights into the Fanconi anemia DNA repair pathway. Science *333*, 312–316.

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.

Knipscheer, P., Räschle, M., Smogorzewska, A., Enoiu, M., Ho, T.V., Schärer, O.D., Elledge, S.J., and Walter, J.C. (2009). The Fanconi anemia pathway promotes replication-dependent DNA interstrand cross-link repair. Science *326*, 1698–1701.

Kottemann, M.C., and Smogorzewska, A. (2013). Fanconi anaemia and the repair of Watson and Crick DNA crosslinks. Nature *493*, 356–363.

Leitner, A., Walzthoeni, T., and Aebersold, R. (2014). Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software pipeline. Nat. Protoc. *9*, 120–137.

Li, X., Mooney, P., Zheng, S., Booth, C.R., Braunfeld, M.B., Gubbens, S., Agard, D.A., and Cheng, Y. (2013). Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. Nat. Methods *10*, 584–590.

Li, N., Zhai, Y., Zhang, Y., Li, W., Yang, M., Lei, J., Tye, B.K., and Gao, N. (2015). Structure of the eukaryotic MCM complex at 3.8 A. Nature *524*, 186–191.

Liang, C.C., Li, Z., Lopez-Martinez, D., Nicholson, W.V., Vénien-Bryan, C., and Cohn, M.A. (2016). The FANCD2-FANCI complex is recruited to DNA interstrand crosslinks before monoubiquitination of FANCD2. Nat. Commun. *7*, 12124.

Liew, C.W., Sun, H., Hunter, T., and Day, C.L. (2010). RING domain dimerization is essential for RNF4 function. Biochem. J. *431*, 23–29.

Lorick, K.L., Jensen, J.P., Fang, S., Ong, A.M., Hatakeyama, S., and Weissman, A.M. (1999). RING fingers mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination. Proc. Natl. Acad. Sci. USA *96*, 11364–11369.

Machida, Y.J., Machida, Y., Chen, Y., Gurtan, A.M., Kupfer, G.M., D'Andrea, A.D., and Dutta, A. (2006). UBE2T is the E2 in the Fanconi anemia pathway and undergoes negative autoregulation. Mol. Cell *23*, 589–596.

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.

Mindell, J.A., and Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. J. Struct. Biol. *142*, 334–347.

Nookala, R.K., Hussain, S., and Pellegrini, L. (2007). Insights into Fanconi anaemia from the structure of human FANCE. Nucleic Acids Res. *35*, 1638–1648.

Pace, P., Johnson, M., Tan, W.M., Mosedale, G., Sng, C., Hoatlin, M., de Winter, J., Joenje, H., Gergely, F., and Patel, K.J. (2002). FANCE: the link between Fanconi anaemia complex assembly and activity. EMBO J. *21*, 3414–3423.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera–a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

Polito, D., Cukras, S., Wang, X., Spence, P., Moreau, L., D'Andrea, A.D., and Kee, Y. (2014). The carboxyl terminus of FANCE recruits FANCD2 to the Fanconi Anemia (FA) E3 ligase complex to promote the FA DNA repair pathway. J. Biol. Chem. *289*, 7003–7010.

Pruneda, J.N., Stoll, K.E., Bolton, L.J., Brzovic, P.S., and Klevit, R.E. (2011). Ubiquitin in motion: structural studies of the ubiquitin-conjugating enzyme~ubiquitin conjugate. Biochemistry *50*, 1624–1633.

Rajendra, E., Oestergaard, V.H., Langevin, F., Wang, M., Dornan, G.L., Patel, K.J., and Passmore, L.A. (2014). The genetic and biochemical basis of FANCD2 monoubiquitination. Mol. Cell *54*, 858–869.

Richmond, T.J., Finch, J.T., Rushton, B., Rhodes, D., and Klug, A. (1984). Structure of the nucleosome core particle at 7 A resolution. Nature *311*, 532–537.

Rohou, A., and Grigorieff, N. (2015). CTFFIND4: fast and accurate defocus estimation from electron micrographs. J. Struct. Biol. *192*, 216–221.

Sato, K., Ishiai, M., Toda, K., Furukoshi, S., Osakabe, A., Tachiwana, H., Takizawa, Y., Kagawa, W., Kitao, H., Dohmae, N., et al. (2012). Histone chaperone activity of Fanconi anemia proteins, FANCD2 and FANCI, is required for DNA crosslink repair. EMBO J. *31*, 3524–3536.

Scheres, S.H. (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. J. Struct. Biol. *180*, 519–530.

Sims, A.E., Spiteri, E., Sims, R.J., 3rd, Arita, A.G., Lach, F.P., Landers, T., Wurm, M., Freund, M., Neveling, K., Hanenberg, H., et al. (2007). FANCI is a

second monoubiquitinated member of the Fanconi anemia pathway. Nat. Struct. Mol. Biol. *14*, 564–567.

Smogorzewska, A., Matsuoka, S., Vinciguerra, P., McDonald, E.R., 3rd, Hurov, K.E., Luo, J., Ballif, B.A., Gygi, S.P., Hofmann, K., D'Andrea, A.D., and Elledge, S.J. (2007). Identification of the FANCI protein, a monoubiquitinated FANCD2 paralog required for DNA repair. Cell *129*, 289–301.

Tang, G., Peng, L., Baldwin, P.R., Mann, D.S., Jiang, W., Rees, I., and Ludtke, S.J. (2007). EMAN2: an extensible image processing suite for electron microscopy. J. Struct. Biol. *157*, 38–46.

Timmers, C., Taniguchi, T., Hejna, J., Reifsteck, C., Lucas, L., Bruun, D., Thayer, M., Cox, B., Olson, S., D'Andrea, A.D., et al. (2001). Positional cloning of a novel Fanconi anemia gene, FANCD2. Mol. Cell *7*, 241–248.

van Twest, S., Murphy, V.J., Hodson, C., Tan, W., Swuec, P., O'Rourke, J.J., Heierhorst, J., Crismani, W., and Deans, A.J. (2017). Mechanism of ubiquitination and deubiquitination in the Fanconi anemia pathway. Mol. Cell *65*, Published online December 13, 2016. http://dx.doi.org/10.1016/j.molcel.2016.11.005.

Walden, H., and Deans, A.J. (2014). The Fanconi anemia DNA repair pathway: structural and functional insights into a complex disorder. Annu. Rev. Biophys. *43*, 257–278.

Wang, A.T., and Smogorzewska, A. (2015). SnapShot: Fanconi anemia and associated proteins. Cell *160*, 354–354.e1.

Watson, J.D., and Crick, F.H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature *171*, 737–738.

Yan, Z., Delannoy, M., Ling, C., Daee, 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.

Cell Reports, Volume *18*

Supplemental Information

The FA Core Complex Contains a Homo-dimeric Catalytic Module for the Symmetric Mono-ubiquitination of FANCI-FANCD2

Paolo Swuec, Ludovic Renault, Aaron Borg, Fenil Shah, Vincent J. Murphy, Sylvie van Twest, Ambrosius P. Snijders, Andrew J. Deans, and Alessandro Costa

Supplemental Experimental Procedures

Construction of baculoviruses

The cDNA for FANCB gene was codon-optimized for expression in *Spodoptera frugiperda* by GeneArt (Invitrogen). FANCL and FAAP100 genes were expressed as human cDNAs sequences, and ORFs match the consensus CDS sequence available in Genbank. To obtain the FANCB-FANCL-FAAP100 construct, all genes were cloned into the MulitBac expression system and constructs generated as per manufacturers protocols. Briefly, pFL-EGFP was generated by cloning the enhanced GFP gene from pEGFP-C1 into pFL (Bieniossek et al., 2008). N-terminally FLAG-tagged FANCB was cloned into pFL-EGFP. Untagged FAAP100 and FANCL (isoform 2) were cloned into pSPL (Bieniossek et al., 2008). Subsequently, pFL-EGFP-Flag-FANCB and pSPL-FAAP100-FANCL were combined by CRE, selected on Ampicillin and Spectinomycin LB agar plates. To obtain the FANCB-FAAP100 construct, the FANCL coding sequence was deleted from pFL/pSPL-FANCB-FANCL-FAAP100 by restriction digestion and re-ligation, to generate the pFL/pSPL-FANCB-FAAP100 expression vector. To obtain FAAP100-FANCL construct, pUCDM-Flag-FAAP100-FANCL plasmid was generated by cloning FANCL and FAAP100 together with a 5'-Flag tag into pUCDM, followed by CRE recombination with pFL-EGFP. To obtain the FANCB-(MBP)FANCL-FAAP100 construct, FANCL (isoform 2) was cloned downstream of the maltose binding protein CDS in pFL-eForRFP (which contains *Echinopora forskaliana Red chromoprotein,* accession number EU498726, cloned into the second MCS). To obtain the FANCC-E-F construct, FANCC, FANCE and FANCF genes were expressed as human cDNA sequences, and ORFs match the consensus CDS sequence available in Genbank. For generation of the MBP-derivative of the CEF complex, the maltose binding protein was inserted 5' to a PreScission protease cleavage site and the start codon

of FANCC. MBP-FANCC, FANCE and FANCF were cloned into pFL plasmid. For generation of the StrepII-C-E-F construct, a Strep-tag II was inserted directly 5' to the start codon of FANCC. (StrepII)FANCC, FANCE and FANCF were successively cloned into pFL plasmid. All constructs were checked by restriction digestion and sequencing, and successively inserted into the Multibac bacmid by TN7 transposition. The third generation (P3) of the amplified baculovirus was used for expression of recombinant proteins in Sf9 cells. To produce FANCI-FANCD2, pFastbac1-Flag-FANCI and pFastbac1-StrepII-FANCD2 encoding full length *Xenopus laevis* FANCI and FANCD2 were kindly provided by Dr. Puck Knipscheer, Hubrecht Institute-KNAW (Knipscheer et al., 2009) and inserted into the Multibac bacmid by TN7 transposition. The two individual baculoviruses were amplified and the third generation (P3) was used for the expression of recombinant proteins in Sf9 cells.

Supplemental Biography

Bieniossek, C., Richmond, T.J., and Berger, I. (2008). MultiBac: multigene baculovirus-based eukaryotic protein complex production. Curr Protoc Protein Sci *Chapter 5*, Unit 5 20.

Knipscheer, P., Raschle, M., Smogorzewska, A., Enoiu, M., Ho, T.V., Scharer, O.D., Elledge, S.J., and Walter, J.C. (2009). The Fanconi anemia pathway promotes replication-dependent DNA interstrand cross-link repair. Science *326*, 1698-1701.

Supplemental Table

Id	Protein1	Protein2	XLType	AbsPos1	AbsPos2	Id-Score	FDR
VAEVQL**K**SDFAAQK-ESGNCP**K**DR	FANCB	FANCB	intra-protein xl	848	611	52.14	0
SY**K**ALINLVQGK-E**K**IISK	FANCB	FANCB	intra-protein xl	424	417	51.05	0
GNFAD**K**EPT**K**TPILHVR	FANCB	FANCB	intralink	35	39	47.15	0
QSFNIIFGECPYCSKPITL**K**MSGR	FANCL		monolink	369		46.85	0
S**K**TVYEGFISAQGR-ESGNCP**K**DR	FANCL	FANCB	inter-protein xl	22	611	46.57	0
AEDASGREHLITLKL**K**	FANCL		monolink	155		43.11	0
HPTMLPECFFLGADHVV**K**PLGIK	FANCL		monolink	261		41.47	0
HPTMLPECFFLGADHVVKPLGI**K**LSR	FANCL		monolink	266		40.82	0
QCPLLLPQNRS**K**TVYEGFISAQGR	FANCL		monolink	22		37.15	0
LLCYNGEVLVFQLS**K**GNFADKEPTK-IIS**K**SYK	FANCB	FANCB	intra-protein xl	29	421	33.03	0
TTSSL**K**LSLNDVTLSLLXDQAHDSR-LLALSA**K**GR	FANCB	FAAP100	inter-protein xl	498	426	32.25	0
STGFFTI**K**EENSHLK	FANCB		monolink	68		30.91	0
KPIEHXEDLFALLAAFHK-**K**MLQTNLK	FANCB	FANCB	intra-protein xl	640	822	30.79	0
LVYADTCFSTI**K**LK-EHLITL**K**LK	FANCL	FANCL	intra-protein xl	137	153	29.6	0
VFLSLEDLSTG**K**YLLTFPK-LL**K**CQNR	FANCB	FANCB	intra-protein xl	631	522	26.42	0
YLLTFP**K**K**K**PIEHMEDLFALLAAFHK	FANCB	FANCB	intralink	638	640	23.44	0
VFVQKSTGFFTI**K**EENSHLK-**K**AVDQR	FANCB	FAAP100	inter-protein xl	68	474	22.53	0
ALINLVQG**K**DDNTSSAEEK-GT**K**VFVQKSTGFFTIK	FANCB	FANCB	intra-protein xl	433	55	21.56	0
TTSSL**K**LSLNDVTLSLLMDQAHDSR-**K**MLQTNLKVSGALYR	FANCB	FANCB	intra-protein xl	498	822	21	0
AEYLPPSVASI**K**VSAELLR-FEMRLSF**K**LGYEMK	FAAP100	FANCB	inter-protein xl	709	125	20.08	0
ELVTLSSLSSAIA**K**HESNFXQR-**K**NQLISFQNGTPK	FANCB	FANCB	intra-protein xl	779	262	19.07	0
KNNVFEYFLLILHSTNK-XLQTNL**K**VSGALYR	FANCB	FANCB	intra-protein xl	101	829	18.79	0
ECCLSEEECTQEPS**K**SDYAIWNTK-GT**K**VFVQKSTGFFTIK	FANCB	FANCB	intra-protein xl	202	55	18.16	0
INYSSEPSDCNEDDLFED**K**QENRYLVVPPLETGLK-VAEVQL**K**SDFAAQK	FANCB	FANCB	intra-protein xl	383	848	16.72	0.263
SAPGDPNALVKILHHLEEPVIFIGAL**K**	FAAP100		monolink	289		16.47	0
THEGGT**K**LLALSA**K**	FAAP100	FAAP100	intralink	419	426	16.42	0
ESGNCP**K**DRYVVCGR-G**K**SSVVAAALSDRR	FANCB	FANCB	intra-protein xl	611	794	15.2	0.263
LSRNIHLWDPENSVLQNL**K**DVLEIDFPAR-LSF**K**LGYEMK	FANCL	FANCB	inter-protein xl	285	125	14.61	0.4
QSFNIIFGECPYCS**K**PITLKMSGRK-AFWDVXDEIDE**K**TWVLEPEKPPR	FANCL	FANCL	intra-protein xl	364	210	14.57	0.263
QSFNIIFGECPYCS**K**PITLKMSGRK-AFWDVXDEIDEKTWVLEPE**K**PPR	FANCL	FANCL	intra-protein xl	364	218	14.55	0.263
EFPEVYFCERPGSFYGTLFTW**K**QR-STGFFTI**K**EENSHLK	FANCB	FANCB	intra-protein xl	708	68	13.48	0.263
NIHLWDPENSVLQNL**K**DVLEIDFPAR-EFPEVYFCERPGSFYGTLFTW**K**QR	FANCL	FANCB	inter-protein xl	285	708	12.71	0.429
QELYALPPPPQFYSSLIEEIGTLGWD**K**LVYADTCFSTIK-L**K**AEDASGR	FANCL	FANCL	intra-protein xl	125	139	12.56	0.263
KEESFVCEHPSKK-EHLITL**K**LKAK	FANCB	FANCL	inter-protein xl	560	153	12.53	0.429
QSFNIIFGECPYCSKPITL**K**XSGRK-RVTLTPDS**K**K	FANCL	FANCB	inter-protein xl	369	559	11.69	0.5
VSAELLRAAL**K**DGHSGVPLCCATLQWLLAENAAVDVVR	FAAP100		monolink	720		10.94	0
GLLTSRQSFNIIFGECPYCSKPITL**K**MSGR-MLLEVAL**K**NR	FANCL	FANCL	intra-protein xl	369	96	10.86	0.263
NNVFEYFLLILHSTN**K**FEMRLSFK-FCCTVLLQIXERESGNCP**K**DR	FANCB	FANCB	intra-protein xl	117	611	10.84	0.263
TTSSL**K**LSLNDVTLSLLMDQAHDSR	FANCB		monolink	498		9.74	0
THEGGTKLLALSA**K**GR-G**K**SSVVAAALSDR	FAAP100	FANCB	inter-protein xl	426	794	9.19	0.667
NQLISFQNGTP**K**NVCQLPFGDPCAVQLMDSGGGNLFFVVSFISNNACAVW**K**	FANCB	FANCB	intralink	274	313	2.8	0.077

Table S1 (related to Figure 3): Inter- and intra-molecular crosslinks detected in the crosslinking and mass spec experiment. The detected peptide sequences (Id) with crosslinked lysines (bold underline), the crosslink type, the absolute crosslink position (AbsPos) for each peptide, the Id-score and the FDR values are reported.

Supplemental Figures

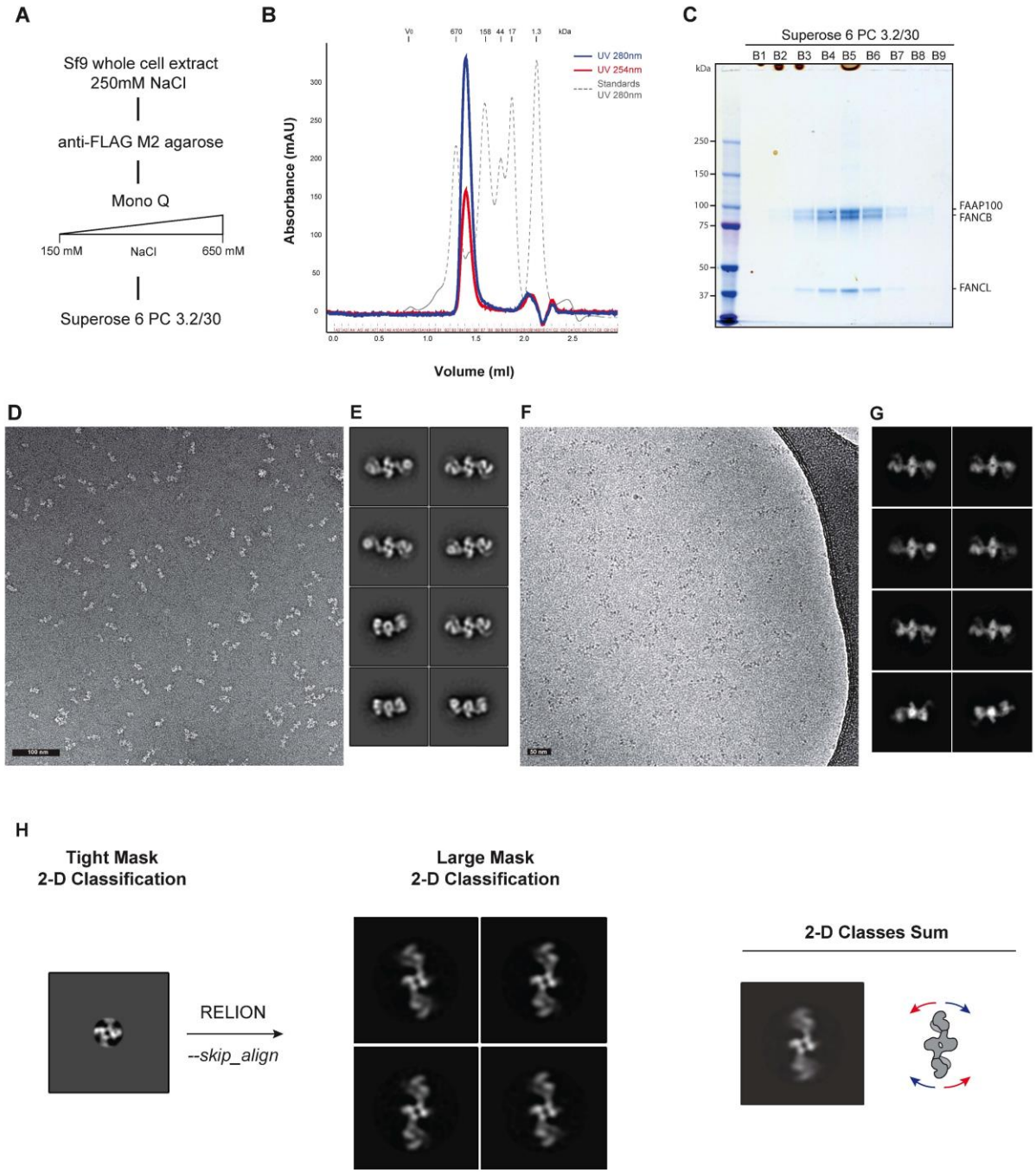


Figure S1 (related to Figure 1): Single-particle analysis of the FANCB-FANCL-FAAP100 complex. **(A)** Purification strategy. **(B)** Superose 6 PC 3.2/30 gel-filtration profile (dashed grey line indicates elution profile of Bio-Rad gel filtration standards: thyroglobulin 670; γ-globulin 158; ovalbumin 44 kDa; myoglobin 17 kDa; vitamin B_{12} 1,3 kDa). **(C)** Coomassie-stained SDS-PAGE gel of the Superose 6 PC 3.2/30 run. **(D)** Representative negative-stain electron micrograph. **(E)** 2D class averages of negatively stained particles. Nicely averaging particles are 21,216 of the original 63,543-particle dataset. Box size is 458 Å × 458 Å. **(F)** Representative cryo-electron micrograph. **(G)** 2D averages of particles embedded in vitrified ice. Nicely averaging

particles are 64,918 of the original 536,754-particle dataset. Box size is 458 Å × 458 Å. **(H)** Conformational variability of the globular extremities in the BL100 particle analyzed by local classification of the dimerization core. A tight mask was applied to classify the dimerization core only. The mask was then relaxed to include the entire particle and a second round of 2D classification was performed skipping particle alignment. Box size is 458 Å × 458 Å.

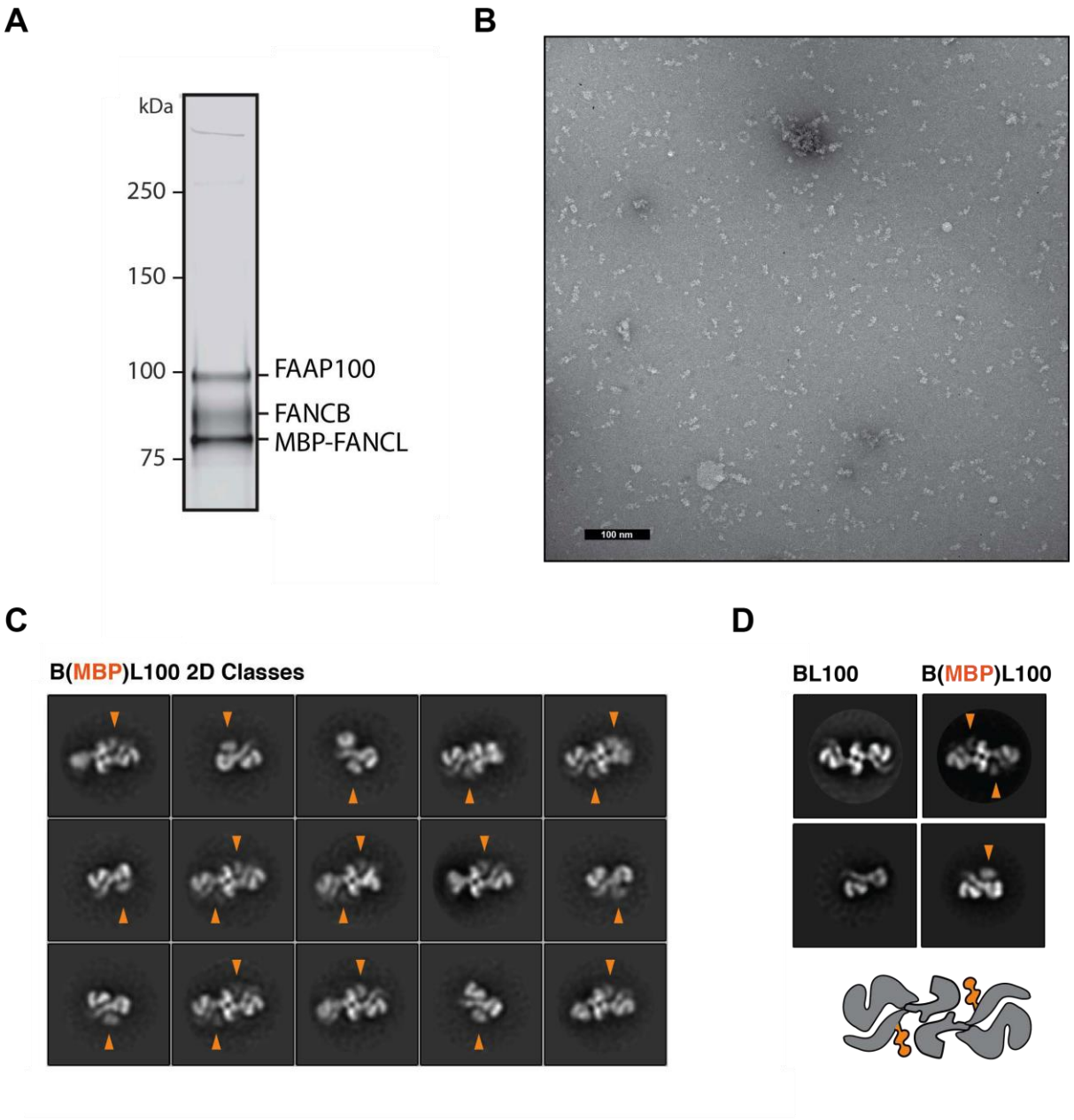
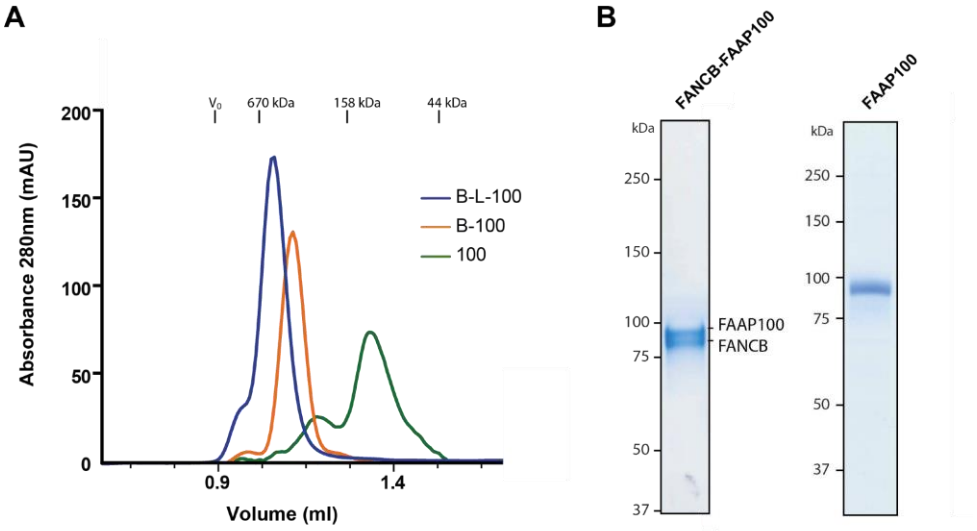
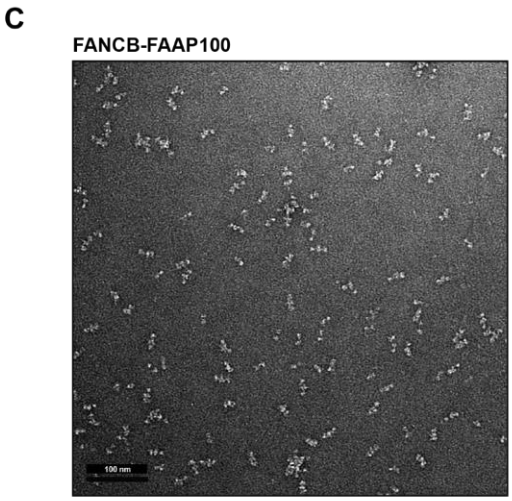
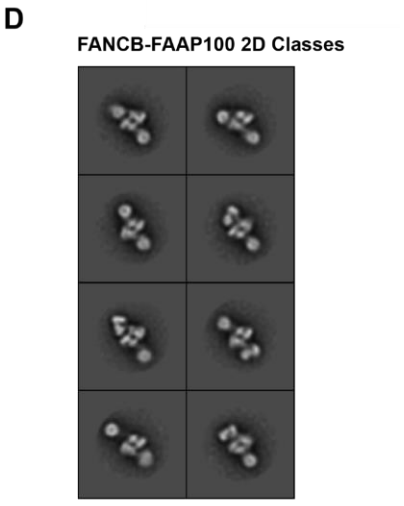
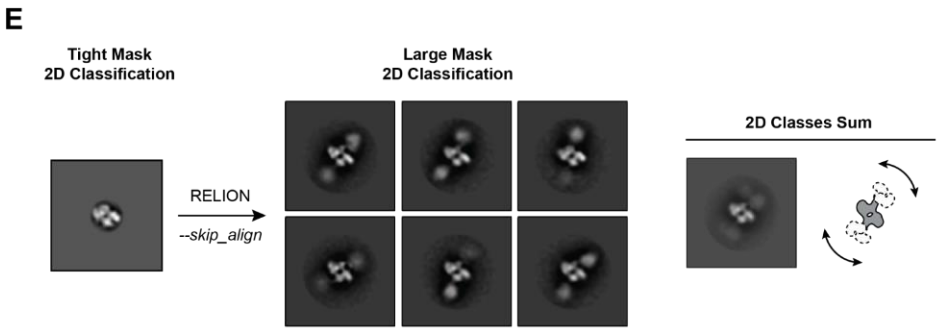


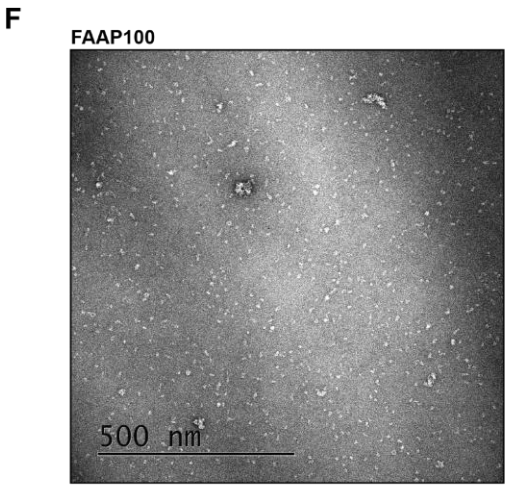
Figure S2 (related to Figure 2): Single-particle analysis of the FANCB-(MBP)FANCL-FAAP100 complex. **(A)** Silver-stained gel of the complex after cation exchange chromatography. **(B)** Representative negative stain electron micrograph. **(C)** 2D class averages with MBP density marked by an orange arrowhead. The presence of MBP partially destabilizes the protein assembly. Nicely averaging particles are 25,421 of the original 51,941-particle dataset. Box size is 458 Å × 458 Å. **(D)** Comparison between the WT and MBP-fused BL100 assemblies. Presence of the MBP allows the interpretation of contaminating monomeric forms of the BL100 complex. Box size is 458 Å × 458 Å.


A
Absorbance 280nm (mAU)
200
150
100
50
0
V0
670 kDa
158 kDa
44 kDa
B-L-100
B-100
100
0.9
1.4
Volume (ml)
B
FANCB-FAAP100
FAAP100
kDa
250
150
100
75
50
37
FAAP100
FANCB


C
FANCB-FAAP100
100 nm


D
FANCB-FAAP100 2D Classes


E
Tight Mask
2D Classification
Large Mask
2D Classification
RELION
--skip_align
2D Classes Sum


F
FAAP100
500 nm

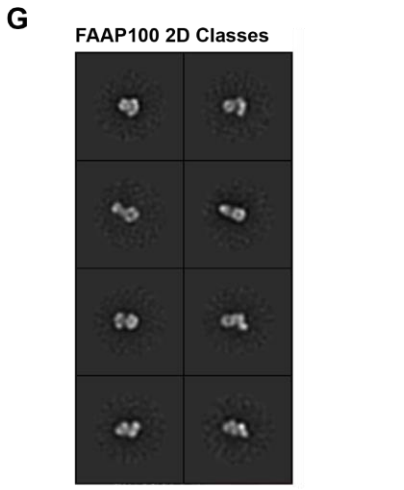

G
FAAP100 2D Classes

Figure S3 (related to Figure 2): Single-particle analysis of the FANCB-FAAP100 complex and the isolated FAAP100 factor. **(A)** Superdex 200 PC 3.2/30 gel-filtration profile of FAAP100 (green) overlaid to the B100 (orange) and BL100 (blue) traces (the molecular mass of Bio-Rad gel-filtration standards elution positions is indicated on top of chromatogram: thyroglobulin 670; γ-globulin 158; ovalbumin 44 kDa). **(B)** Coomassie-stained SDS-PAGE gel of purified B100 and FAAP100. **(C)** Representative negative-stain electron micrograph of B100. **(D)** 2D class averages of negatively stained B100 particles. Nicely averaging particles are 34,747 of the original 63,365-particle dataset. Box size is 458 Å × 458 Å. **(E)** Conformational variability of the globular extremities in the B100 particle analyzed by local classification of the dimerization core. A tight mask was applied to classify the dimerization core only. The mask was then relaxed to include the entire particle and a second round of 2D classification was performed skipping particle alignment. Box size is 458 Å × 458 Å. **(F)** Representative negative-stain electron micrograph of the isolated FAAP100. **(G)** 2D class averages of negatively stained particles. Nicely averaging particles are 52,301 of the original 177,106-particle dataset. Box size is 458 Å × 458 Å.

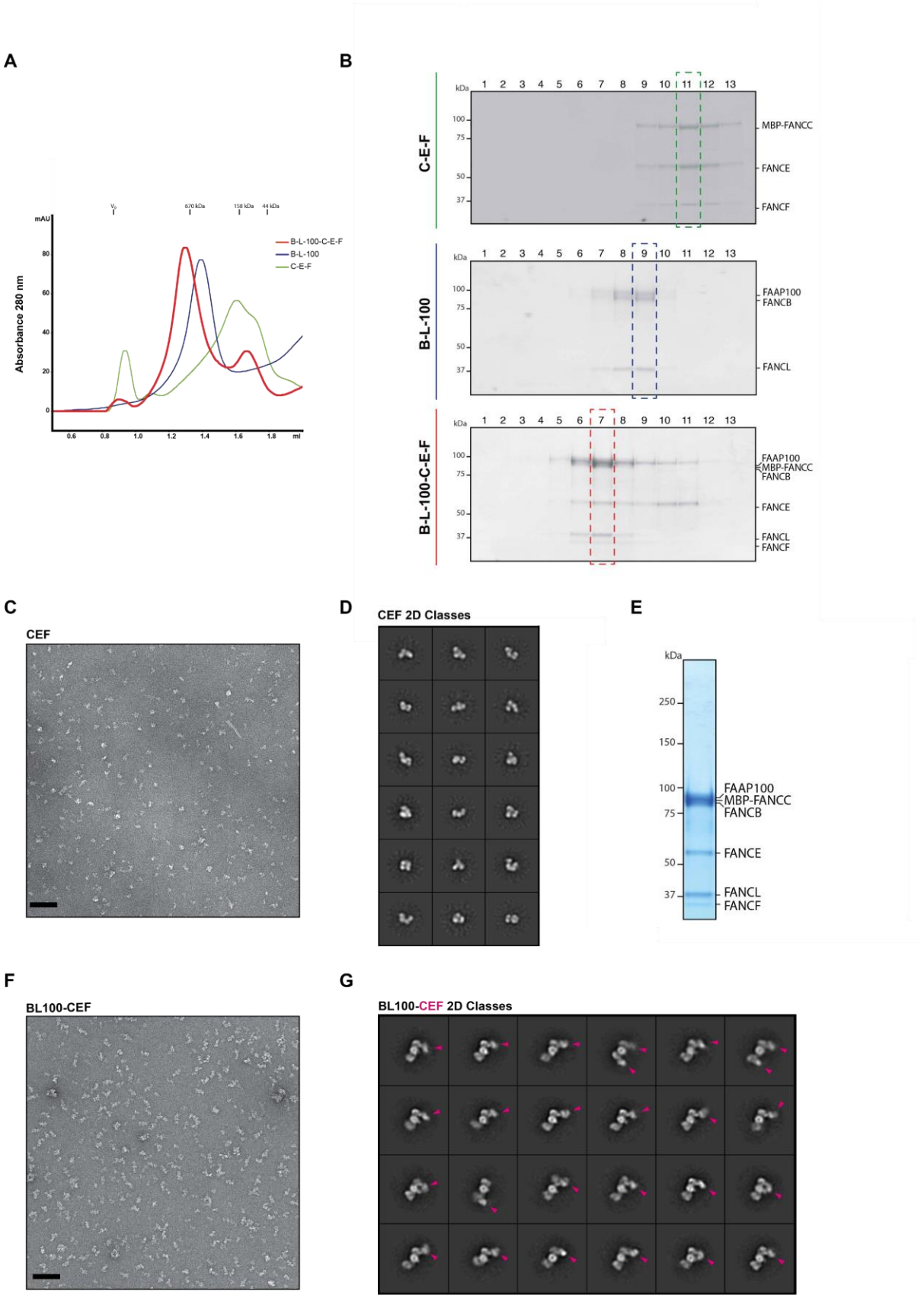
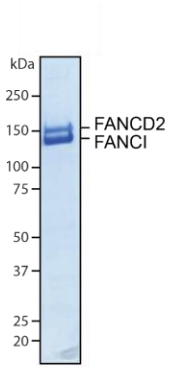


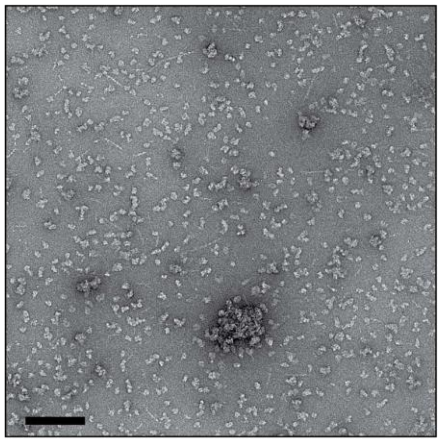
Figure S4 (related to Figure 4): Single-particle analysis of the FANCC-E-F and the FANCB-FANCL-FAAP100-FANCC-FANCE-FANCF complex. **(A)** Superose 6 PC 3.2/30 gel-filtration profile of (MBP)CEF (green) overlaid to the BL100 (blue) and BL100-CEF (red) traces (the molecular mass of Bio-Rad gel-filtration standards elution positions is indicated on top of chromatogram: thyroglobulin 670; γ-globulin 158; ovalbumin 44 kDa). **(B)** Coomassie-stained SDS-PAGE gels of the three Superose 6 gel-filtrations. Dashed-line box for each experiment indicates the fraction

analyzed by EM. **(C)** Representative negative-stain electron micrograph of CEF particles. **(D)** 2D class averages of negatively stained (MBP)CEF particles. Nicely averaging particles are 22,757 of the original 47,898-particle dataset. Box size is 441 Å × 441 Å. **(E)** Coomassie-stained SDS-PAGE gel of purified BL100-(MBP)CEF complex. **(F)** Representative negative-stain electron micrograph of BL100-CEF. **(G)** 2D class averages of negatively stained particles. Nicely averaging particles are 28,498 of the original 36,768-particle dataset. CEF is marked with a purple arrowhead. Box size is 580 Å × 580 Å.

A

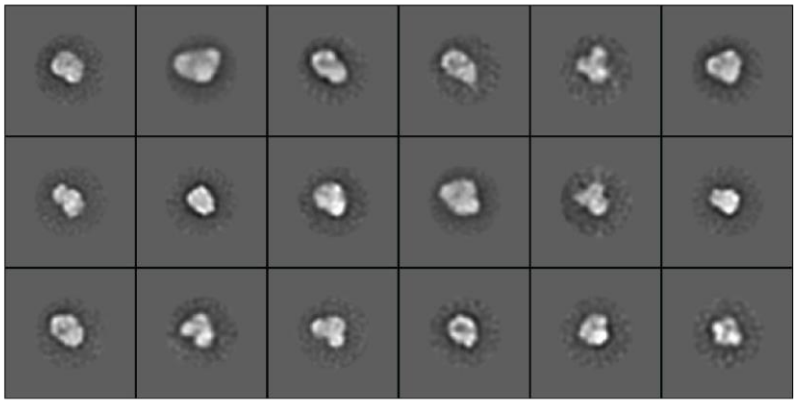


B

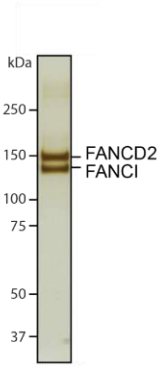


C

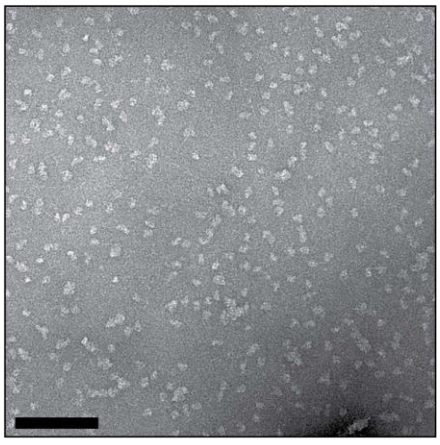
FANCI-FANCD2 from Mono Q 5/50 GL
2D Classes



D



E



F

FANCI-FANCD2 from Superdex 200 PC 3.2/30
2D Classes

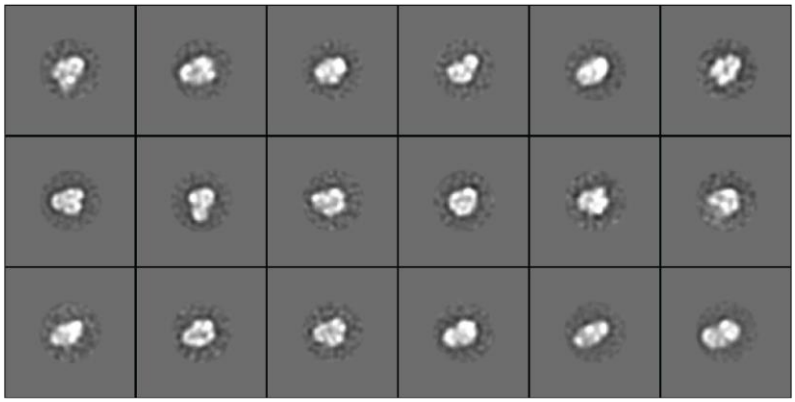


Figure S5 (related to Figure 5): Single-particle analysis of the FANCI-FANCD2 complex. **(A)** Coomassie-stained SDS-PAGE gel of the complex after cation exchange chromatography. **(B)** Representative negative-stain electron micrograph. Particles appear homogeneous and monodisperse. **(C)** 2D class averages of negatively stained particles. ID2 particles average poorly. Box size is 441 Å × 441 Å. **(D)** Silver-stained SDS-PAGE gel of the complex after size-exclusion chromatography. **(E)** Representative negative-stain electron micrograph. Particles appear homogeneous and monodisperse. **(F)** 2D class averages of negatively stained particles. ID2 particles average poorly even after gel filtration. Box size is 441 Å × 441 Å.

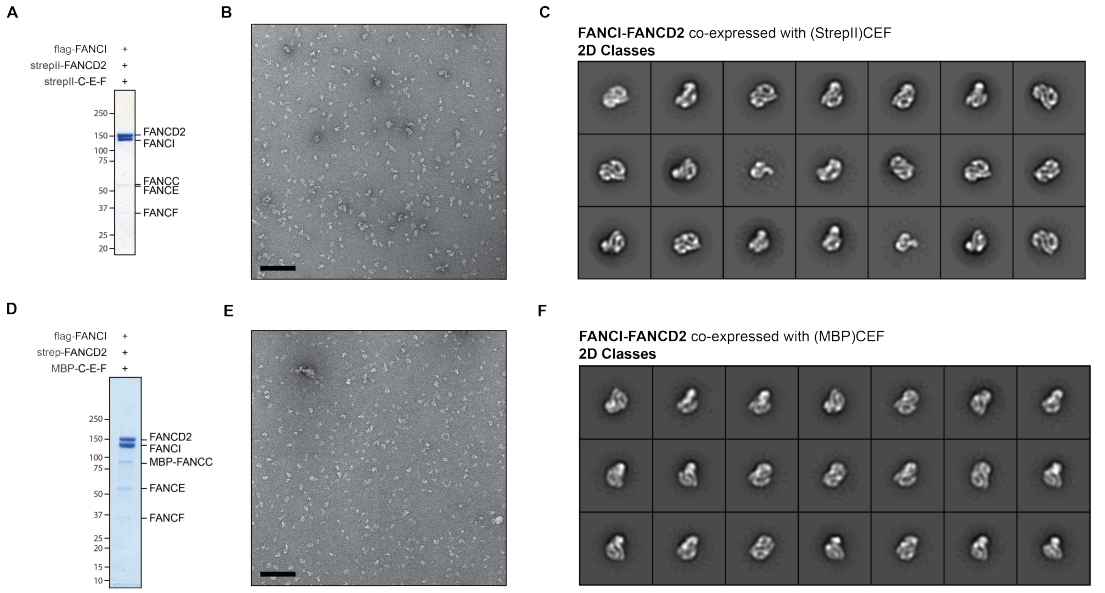
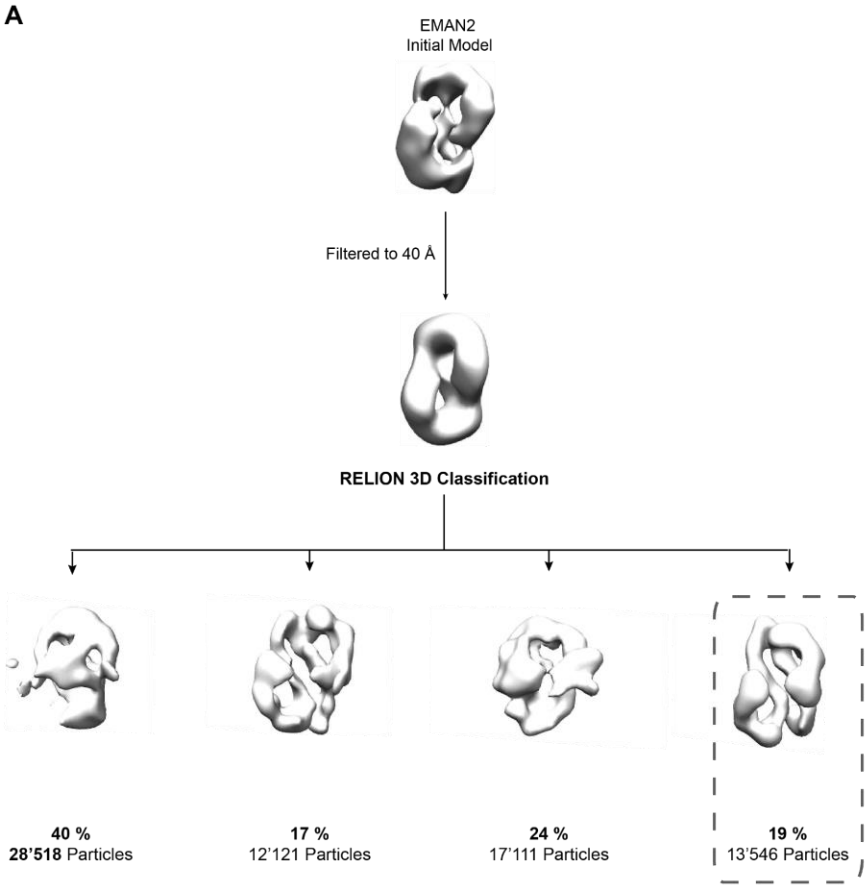
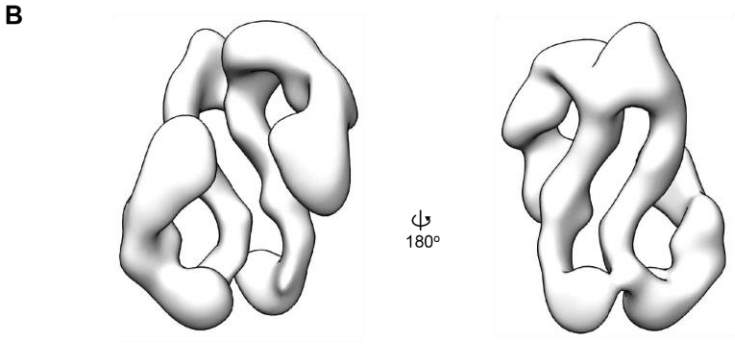
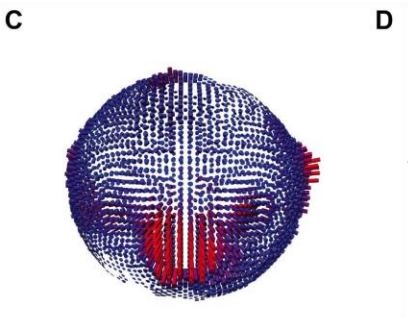
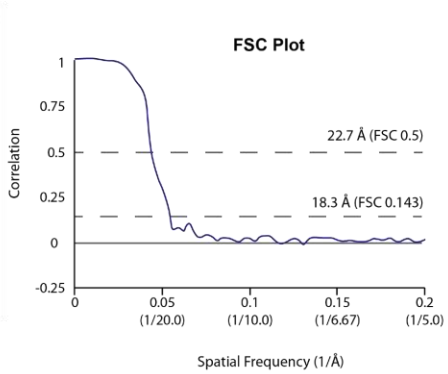


Figure S6: Single-particle analysis of the FANCI-FANCD2 complex, co-expressed with FANCC-E-F. **(A)** Coomassie-stained SDS-PAGE gel of the ID2 complex after size-exclusion chromatography on a Superose 6 PC 3.2/30 column (StrepII-CEF coexpression). StrepII-CEF is present in trace amounts. **(B)** Representative negative stain electron micrograph (StrepII-CEF co-expression). Particles appear homogeneous and monodisperse. **(C)** 2D class averages of negatively stained particles. ID2 co-expressed with StrepII-CEF assembles in stable, pseudo-two-fold symmetric assemblies. Nicely averaging particles are 44,564 of the original 71,295- particle dataset. Box size is 441 Å Å~ 441 Å. **(D)** Coomassie-stained SDS-PAGE gel of the ID2 complex after cation exchange chromatography (MBP-CEF coexpression). MBP-CEF is present in traces amounts **(E)** Representative negativestain electron micrograph. Particles appear homogeneous and monodisperse (similar to isolated ID2, see Figure S8). **(F)** 2D class averages of negatively stained particles. ID2 co-expressed with MBP-CEF assembles in stable, pseudo-two-fold symmetric assemblies. Nicely averaging particles are 14,533 of the original 27,084-particle dataset. Box size is 441 Å Å~ 441 Å.


A
EMAN2
Initial Model
Filtered to 40 Å
RELION 3D Classification
40 %
28'518 Particles
17 %
12'121 Particles
24 %
17'111 Particles
19 %
13'546 Particles


B
180°


C


D
FSC Plot
Correlation
1
0.75
0.5
0.25
0
-0.25
22.7 Å (FSC 0.5)
18.3 Å (FSC 0.143)
0
0.05
(1/20.0)
0.1
(1/10.0)
0.15
(1/6.67)
0.2
(1/5.0)
Spatial Frequency (1/Å)

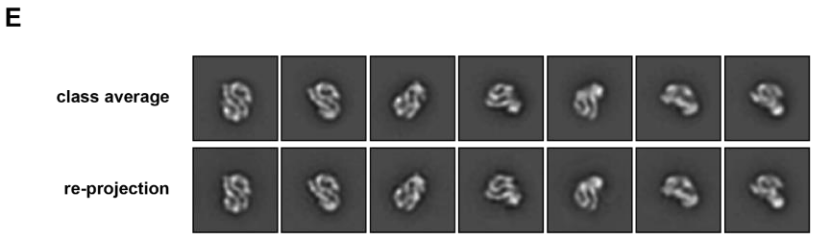

E
class average
re-projection

Figure S7 (related to Figure 6): Three-dimensional reconstruction of the CEF-co-expressed FANCI-FANCD2 complex. **(A)** Overview of the three-dimensional classification. An initial model was obtained using EMAN2, low-pass filtered to 40 Å and used to classify the 2D-cleaned dataset in four 3D classes. The sharpest class (dashed grey box) was selected and used for further three-dimensional auto-refinement in RELION. **(B)** Surface representation of the refined three-dimensional map. **(C)** Angular distribution. **(D)** Fourier-shell correlation plot. **(E)** Reference-free 2D class average matched with 2D-projected ID2 volume. Box size is 350 Å × 350 Å.

Videos:

Video 1 (related to Figure 1): Inherent flexibility of the FANCB-FANCL-FAAP100 assembly.

Video 2 (related to Figure 2): Omitting FANCL from the FANCB-FANCL-FAAP100 assembly increases the flexibility of the peripheral domains in the complex.

Video 3 (related to Figure 5): The dimerization interface in the CEF-treated ID2 assembly remains fixed, while both C-terminal domains in the dimer are free to move.