



Published in final edited form as:

*Science*. 2011 July 15; 333(6040): 312–316. doi:10.1126/science.1205805.

## Structure of the FANCI-FANCD2 Complex: Insights into the Fanconi Anemia DNA Repair Pathway

Woo Joo<sup>1,\*</sup>, Guozhou Xu<sup>2,\*</sup>, Nicole S. Persky<sup>1,3,\*</sup>, Agata Smogorzewska<sup>4,5</sup>, Derek Rudge<sup>1,3</sup>, Olga Buzovetsky<sup>1,3</sup>, Stephen J. Elledge<sup>3,4</sup>, and Nikola P. Pavletich<sup>1,3,‡</sup>

<sup>1</sup>Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, NY 10069, USA

<sup>2</sup>Sloan-Kettering Division, Joan and Sanford I. Weill Graduate School of Medical Sciences, Cornell University, New York, NY 10069, USA

<sup>3</sup>Howard Hughes Medical Institute

<sup>4</sup>Department of Genetics, Harvard Medical School, Division of Genetics, Brigham and Women's Hospital, Boston, MA 02115, USA

<sup>5</sup>Laboratory of Genome Maintenance, The Rockefeller University, New York, NY 10069, USA

### Abstract

Fanconi Anemia is a cancer predisposition syndrome caused by defects in the repair of DNA interstrand crosslinks (ICL). Central to this pathway is the FANCI-FANCD2 (ID) complex, which is activated by DNA damage-induced phosphorylation and monoubiquitination. The 3.4 Å crystal structure of the ~300 kDa ID complex reveals that monoubiquitination and regulatory phosphorylation sites map to the I-D interface, suggesting that they occur on monomeric proteins or an opened-up complex, and that they may serve to stabilize I-D hetero-dimerization. The 7.8 Å electron density map of FANCI-DNA crystals and *in vitro* data show that each protein has binding sites for both single- and double-stranded DNA, suggesting that the ID complex recognizes DNA structures that result from the encounter of replication forks with an ICL.

---

Fanconi Anemia (FA) is a recessive cancer predisposition and developmental syndrome characterized by hypersensitivity to DNA inter-strand crosslinking (ICL) agents (1). The proteins mutated in the thirteen FA complementation groups act in a common pathway that results in DNA-repair by homologous recombination (HR) (2, 3). The FA pathway is activated when a DNA replication fork encounters an ICL, although FA proteins can be recruited to ICLs in the absence of DNA-replication as well (4-7).

Central to the FA pathway is the ID complex formed by the Fanconi Anemia D2 (FANCD2) and I (FANCI) proteins (8-11). Activation of the pathway culminates in FANCD2 and FANCI phosphorylation by ATR (1, 12), their mono-ubiquitination by the FA Core Complex ubiquitin ligase (1, 2, 13), and their localization to chromatin foci considered sites repair (9, 14, 15).

Replication-dependent ICL repair involves nucleolytic incisions flanking the ICL on one strand, translesion DNA synthesis across the unhooked ICL, removal of the ICL by additional incisions, and homologous recombination (3-5). Depletion of FANCD2 or mutation of its ubiquitination site arrests the process prior to the initial incision step (5).

---

<sup>‡</sup>To whom correspondence should be addressed..

\*These authors contributed equally to this work.

Among the nucleases implicated in these incisions, FAN1 and SLX4-SLX1 contain a ubiquitin-binding UBZ domain required for their association with mono-ubiquitinated ID (3, 16).

FANCI and FANCD2 consist of 1328 and 1451 amino acids, respectively. They share a ~150 amino acid region of homology around their mono-ubiquitination sites, and they have been proposed to be paralogs (17). Neither protein has any recognizable sequence motifs. They have been shown to bind to DNA, although the nature of their specific DNA substrates has not been clear (18-20).

Here, we present the 3.4 Å crystal structure of the 296 kDa ID complex, the 3.3 Å structure of individual FANCI, as well as the 7.8 Å crystallographic electron density map of a FANCI bound to a splayed Y DNA.

## Overall structure of the ID complex

Crystals were grown using mouse FANCI (residues 1 to 1301) and FANCD2 (residues 33-1415) proteins that were truncated to remove regions highly susceptible to proteolysis (figs. S1 and S2).

The structures of FANCI and FANCD2 have overall similar shapes and share extensive local similarity essentially throughout their length (Fig. 1A fig. S3). They consist predominantly of  $\alpha$  helices, the majority of which are arranged in antiparallel pairs that form  $\alpha$ - $\alpha$  superhelical structural segments, commonly referred to as  $\alpha$  solenoids. Each structure has four distinct  $\alpha$  solenoid segments that range in size from 156 to 310 residues (Solenoids 1 to 4; Fig. 1A and fig. S4). Two other, mostly helical segments intervene between Solenoids 1 and 2 (~80 residue helical domain HD1), and between Solenoids 2 and 3 (~240 residue HD2; fig. S4).

FANCI and FANCD2 fold into a saxophone-like structure consisting of a long N-terminal neck (Solenoid 1 – HD1 – Solenoid 2), a U-shaped middle bow (HD2), and a bulbous C-terminal bell (Solenoid 3 – Solenoid 4; Fig. 1A and fig. S4). The ID complex has a trough-like shape with partially open ends (Fig. 1, B and C). The interior of the trough, which is ~105 Å long, ~70 Å wide and ~40 Å deep, is marked by multiple grooves that have a highly positive electrostatic potential. Two of these grooves contain electron density for double-stranded and single-stranded DNA in the 7.8 Å map of Y DNA-FANCI crystals (fig. S3C). The FANCD2 K559 and FANCI K522 mono-ubiquitination sites are within the ID interface, in solvent-accessible tunnels located at the bottom of the trough-like structure (Fig. 1A).

The individual Solenoid and HD segments of FANCI and FANCD2 share extensive structural homology, although the entire structures cannot be superimposed because of the cumulative effects of differences in individual segments and in inter-segment packing arrangements. The most homologous are the Solenoid 2 segments, which contain the monoubiquitination sites and the reported FANCI-FANCD2 sequence homology. They can be superimposed with a 1.7 Å C $\alpha$  root mean square deviation (rmsd) for ~82 % of their residues (table S2 and fig. S4B). For the remainder of the segments, the fraction of the superimposing residues ranges from ~66 % for Solenoid 4 to ~60 % for HD2 (table S2). These local and global structural similarities indicate that FANCI and FANCD2 evolved from a common ancestor [fig. S4 and supporting online material (SOM) text].

Ten of the eleven FA-I and FA-D2 missense mutations (21) map to buried residues with clear structure-stabilizing roles, indicating that their pathogenic effects result from structural defects in the mutant proteins (fig. S5).

## I-D interface

The two proteins interact along the long dimension of their saxophone-like shape (Solenoid 1 – HD1 – Solenoid 2 and part of HD2) in an anti-parallel, two-fold pseudo-symmetric manner (Fig. 2A and fig. S4A). The interface extends along a ~560 residue-region on each protein, and buries a total of ~7,100 Å<sup>2</sup> of solvent accessible area. It is nearly continuous, except for gaps at the mono-ubiquitination site of each protein, and an additional gap at the center of pseudo-symmetry (between opposing HD1 segments).

The interface has comparable numbers of hydrophobic (42 side chains) and polar (44 side chain and 11 backbone groups) residues within intermolecular-contact distance (Fig. 2A; contacts marked on fig. S1). Most of these contacts are clustered near the ends of the extended interface, although with substantial asymmetry in the density of contacts at the two ends. The end where the FANCD2 Solenoid 1 packs with the FANCI Solenoid 2-HD2 segments has 60 residues involved in contacts (Fig. 2B) compared to only 30 at the reciprocal end (Fig. 2C). This asymmetry results primarily from two structurally non-homologous regions. One of these involves the N-terminal portions of Solenoid 1 segments (thereafter “caps”; fig. S4), with the FANCD2 cap having 11 contact residues (Fig. 2B) compared to only 2 for the FANCI cap at the reciprocal end (Fig. 2C). The other involves a FANCI-specific HD2 insertion with 9 side chain and backbone groups that contact FANCD2 (α26b-α26c; Fig. 2B bottom). The asymmetry in these elements and their intermolecular contacts might have evolved to drive the duplicated ancestral protein to form heterodimers instead of homodimers.

The remaining intermolecular contacts are scattered in small groups throughout the porous middle portion of the interface, except around the ubiquitination site lysine residues, which make no contacts (Fig. 2A and fig. S6).

## Mono-ubiquitination sites

Although the FANCI K522 and FANCD2 K559 side chains are embedded in the I-D interface (Fig. 1A), the absence of I-D contacts next to each results in two tunnels, one for each lysine, that allow free access to bulk solvent from either side of the trough wall. In the monomeric proteins, both lysine side chains are fully solvent exposed.

The solvent-accessible tunnels of both sites are wide enough (~7 to 9 Å) and short enough to accommodate the four amino acid ubiquitin tail, whose C-terminus would be covalently linked to the lysine epsilon amino group (Fig. 3, A and B; fig. S7). The ubiquitin tail is conformationally flexible (22), and this should aid in the tail navigating the tunnels, while the ubiquitin structural domain is positioned outside the interface. It is also possible, however, that ubiquitination induces a conformational change or a rearrangement of the complex that fully exposes the mono-ubiquitination sites.

In principle, the tunnels can accommodate a ubiquitin residing either at the trough interior or exterior. The tunnel entrance at the trough interior (“top” entrance looking down the plane of Fig. 1A) is ~13 Å away from the lysine epsilon amino group. The entrance at the “bottom” is farther away at ~17 Å, although the relatively more open, funnel shape of this entrance could also accommodate part of the ubiquitin structural domain (Fig. 3, A and B; fig. S7 and SOM text).

Irrespective of where the ubiquitin is positioned, the dimensions of the tunnels and entrances suggest that the ubiquitin tail and structural domains may contact both their conjugated protein and its hetero-dimerization partner. As these interactions would be repeated in a reciprocal fashion, the structure raises the possibility that ubiquitination may contribute to

the stability of the ID complex. In this respect, we note that the mouse ID complex has a short half-life, as it dissociates during gel filtration chromatography at concentrations of 1  $\mu$ M (fig. S8).

Although the tunnels can readily accommodate the ubiquitin tail, they are too small for the active site of the ubiquitin-conjugating (E2) enzyme to access the lysine side chains. The structure thus suggests that ubiquitination either occurs on a cellular pool of monomeric FANCI and FANCD2, or it involves a process that opens up the I-D interface.

The sequestration of the lysine-ubiquitin isopeptide bond at the I-D interface would also protect it against de-ubiquitination by the FA-associated USP1 ubiquitin protease (23). Coupled to a possible role of ubiquitination stabilizing I-D association, this may explain why cells lacking either FANCI or FANCD2 exhibit loss or reduction of the ubiquitinated form of the paralog (9).

## Phosphorylation sites

The DNA-damage dependent phosphorylation of FANCI has been suggested to act as a molecular switch that activates the pathway, as multiple alanine substitutions at putative phosphorylation sites impeded the mono-ubiquitination and chromatin association of both FANCI and FANCD2 (12). Three phosphorylation sites have the ATM/ATR kinase consensus and are also conserved in vertebrates (Ser555, Thr558, Thr564 in mouse FANCI; fig. S9).

These three sites map to the 37-amino acid FANCI-specific HD2 insertion at the I-D interface. This region undergoes a complete conformational change on FANCD2 binding. In free FANCI, it forms the  $\beta$ 3- $\beta$ 4 sheet and part of it is disordered. In the ID complex it instead forms two new helices ( $\alpha$ 26b and  $\alpha$ 26c) and a 10-residue extension to  $\alpha$ 27, becoming ordered in its entirety (Fig. 3C). This changes entirely the structural context of the phosphorylation sites. In free FANCI, they map to the disordered segment where they can be readily phosphorylated. In the complex, they end up at the start, middle and immediately after the  $\alpha$ 26c helix. Their side chains are embedded in hydrogen bond networks that anchor  $\alpha$ 26c in the FANCI structure and also contact FANCD2 (Fig. 3D and Fig. 2B bottom panel). The structure suggests that their phosphorylation may augment the intramolecular hydrogen bond networks, stabilizing the FANCD2-bound conformation of this region, and may also result in new FANCD2 contacts (SOM text in fig. S9 legend). These two effects may cooperate to stabilize I-D association. The magnitude of this stabilization could, in principle, be substantial, as the FANCI  $\alpha$ 26b- $\alpha$ 26c- $\alpha$ 27 segment accounts for ~27 % of the total solvent accessible area buried on complex formation (Fig. 2, A and B).

The model of phosphorylation stabilizing I-D association, thereby protecting against deubiquitination, may explain why mutations in FANCI phosphorylation sites reduce levels of the ubiquitinated forms of both FANCI and FANCD2 (12).

## Electron density map of the Y DNA-FANCI complex

FANCD2 and FANCI have been shown to bind to double-stranded DNA (dsDNA) with a minor preference for diverse branched DNA structures such as Holliday Junction (HJ), splayed Y, overhang and replication fork DNA (18-20). In co-crystallization experiments with a variety of branched DNA structures, we obtained diffraction-quality crystals of FANCI bound to a splayed Y DNA consisting of a 16-base pair (bp) dsDNA segment and two 20-nucleotide (nt) single-stranded DNA (ssDNA) arms. FANCI binds to this Y DNA with a dissociation constant ( $K_d$ ) of 19 nM, which is ~10-fold tighter than its affinity for either 18-bp dsDNA or for 32-nt ssDNA (fig. S10).

The Y DNA-FANCI crystals diffract to 7.8 Å resolution and contain three complexes in the asymmetric unit (table S1). The electron density map, calculated after three-fold non-crystallographic symmetry (ncs) averaging, shows unambiguously double-helical electron density, which extends throughout the length of a 16 bp ideal B-type DNA model positioned into the map manually (Fig. 4A and fig. S3C).

At one end of the dsDNA, the map shows an undulating tube of continuous electron density that has shape consistent with one ssDNA segment. We do not observe clear electron density for the second ssDNA segment of the Y DNA, and we presume it is disordered. Consistent with FANCI having only one major site for binding to ssDNA, we find that its affinity for Y DNA is indistinguishable from its affinity for the corresponding 3' overhang DNA (19 nM  $K_d$  for both; fig. S10).

The dsDNA lies in a groove that has a semicircular cross-section (Fig. 4B). The groove is built by Solenoid 4 forming one side, and by Solenoid 3 forming the bottom and part of the other side, the remainder of which is completed by portions of HD1, Solenoid 2 and HD2. There are 7 FANCI regions that are within contact distance to 13 bps of dsDNA ( $\alpha$ 19b,  $\alpha$ 20,  $\alpha$ 35,  $\alpha$ 40,  $\alpha$ 42,  $\alpha$ 44,  $\alpha$ 46; Fig. 4, A and B). Most of these regions involve the second helix of individual repeats, and in particular the N-termini of the helices and the loops immediately preceding them. In addition, 4 loops or helix ends that are disordered in the FANCI and ID structures become at least partially ordered near the DNA ( $\alpha$ 19b- $\alpha$ 20,  $\alpha$ 33- $\alpha$ 34,  $\alpha$ 36- $\alpha$ 37 and  $\alpha$ 47- $\alpha$ 48 loops; Fig. 4, A and B). The DNA-proximal FANCI segments are rich in conserved basic residues (marked on fig. S1A).

The transition from double- to single-stranded DNA electron density has  $\alpha$ 19b and  $\alpha$ 20 abutting the last base pair of the duplex (Fig. 4A). The ~55 Å path of the tubular ssDNA electron density could correspond to ~8 to 11 nucleotides, depending on the ssDNA backbone conformation and extent of base-base stacking. The electron density follows a groove formed by the long  $\alpha$ 44,  $\alpha$ 46 and  $\alpha$ 48 helices of Solenoid 4 on one side, and by the  $\alpha$ 15 and  $\alpha$ 17 helices of HD1 on the other (Fig. 4A and fig. S11). This groove is lined with residues containing basic, hydroxyl and amide groups, as well as two conserved phenylalanine residues that are fully solvent exposed (fig. S11).

The polarity of the ssDNA is not clear due to the limited resolution of the map and the proximity of multiple protein elements at the ss-dsDNA junction. In vitro, FANCI exhibits only a minor preference for 3' overhang (19 nM  $K_d$ ) over 5' overhang DNA (40 nM  $K_d$ ; fig. S10).

## Implications for the DNA-binding activity of the ID complex

The overall shapes and highly positive electrostatic potential of the dsDNA/ssDNA binding grooves of FANCI are conserved in FANCD2 (Fig. 4C; fig S12 and SOM text). Consistent with this, FANCD2 binds to dsDNA and ssDNA with affinities comparable to those of FANCI (fig. S10). Taken together, our data suggest that the ID complex has two sets of ds/ssDNA binding sites arranged in a pseudo-2-fold symmetric manner (Fig. 4C and fig. S12).

This raises the possibility that the ID complex binds to two ds/ssDNA structures crosslinked by an ICL. The pausing of a replication fork at an ICL has been shown to initially generate a splayed Y DNA structure, with the subsequent advancement of leading strand synthesis to the -1 position of the ICL generating a crosslinked 5' flap structure (4, 5). If two forks converge onto the ICL from opposite directions, this would result in an ICL flanked either by two 5' flaps, or one Y and one 5' flap. Both of these would be compatible with the pseudo-2-fold symmetric arrangement of ds/ssDNA-binding sites in the ID complex (Fig. 4C), with the ICL residing near the center of pseudo-symmetry of the complex. It is also

possible that both single-fork and double-fork events occur in the cell, with the ID complex evolving a certain level of promiscuity in order to recognize an array of single- and double-Y and 5' flap DNA structures.

Our data implicate the ID complex in recognizing DNA structures that result from the encounter of replication forks with an ICL. The complex may function in stabilizing and protecting these DNA structures, and also in providing specificity for the initial incisions around the ICL by directing the FA-associated nuclease to these sites.

## Supplementary Material

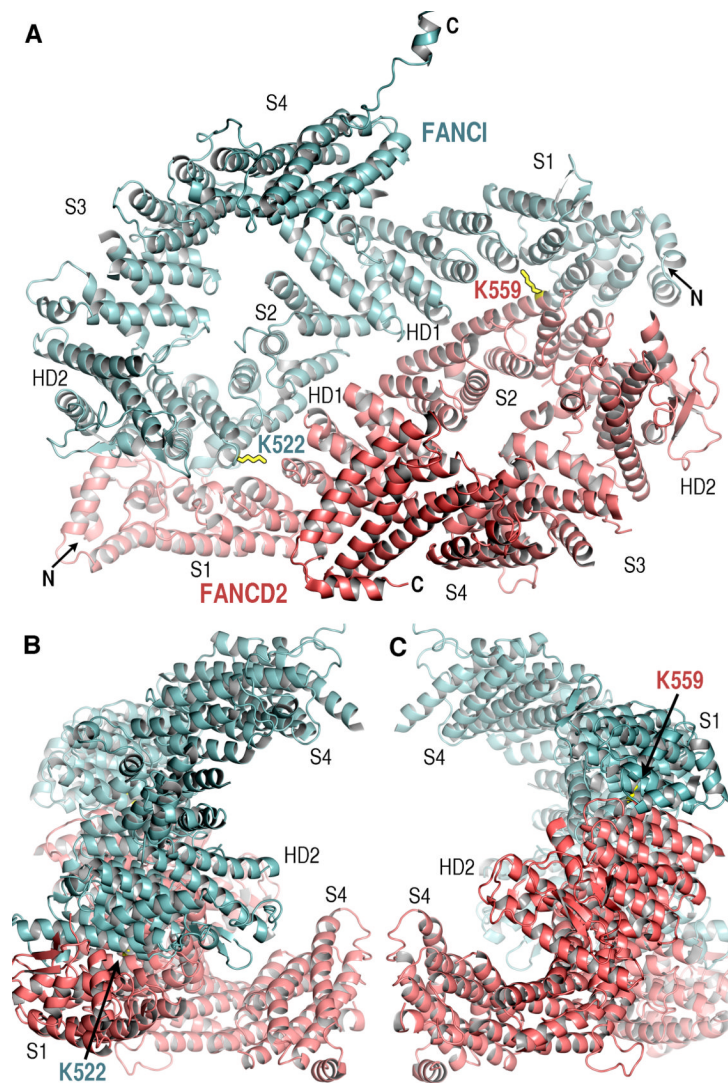
Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We thank David King for mass spectroscopy analysis and the staff of the APS NE-CAT and NSLS X29 beamlines for help with data collection. This work was supported by the NIH and the Howard Hughes Medical Institute. Coordinates and structure factors have been deposited with the PDB (accession codes 3S51 for FANCI, 3S4W for FANCI FANCD2 and 3S4Z for Y DNA-FANCI).

## References and Notes

1. Moldovan GL, D'Andrea AD. *Annu Rev Genet.* 2009; 43:223. [PubMed: 19686080]
2. Wang W. *Nat Rev Genet.* Oct.2007 8:735. [PubMed: 17768402]
3. Kee Y, D'Andrea AD. *Genes Dev.* Aug 15.2010 24:1680. [PubMed: 20713514]
4. Raschle M, et al. *Cell.* Sep 19.2008 134:969. [PubMed: 18805090]
5. Knipscheer P, et al. *Science.* Dec 18.2009 326:1698. [PubMed: 19965384]
6. Shen X, et al. *Mol Cell.* Sep 11.2009 35:716. [PubMed: 19748364]
7. Ben-Yehoyada M, et al. *Mol Cell.* Sep 11.2009 35:704. [PubMed: 19748363]
8. Timmers C, et al. *Mol Cell.* Feb.2001 7:241. [PubMed: 11239453]
9. Smogorzewska A, et al. *Cell.* Apr 20.2007 129:289. [PubMed: 17412408]
10. Sims AE, et al. *Nat Struct Mol Biol.* Jun.2007 14:564. [PubMed: 17460694]
11. Dorsman JC, et al. *Cell Oncol.* 2007; 29:211. [PubMed: 17452773]
12. Ishiai M, et al. *Nat Struct Mol Biol.* Nov.2008 15:1138. [PubMed: 18931676]
13. Meetei AR, et al. *Nat Genet.* Oct.2003 35:165. [PubMed: 12973351]
14. Garcia-Higuera I, et al. *Mol Cell.* Feb.2001 7:249. [PubMed: 11239454]
15. Shimamura A, et al. *Blood.* Dec 15.2002 100:4649. [PubMed: 12393398]
16. Yamamoto KN, et al. *Proc Natl Acad Sci U S A.* Apr 19.2011 108:6492. [PubMed: 21464321]
17. Smogorzewska A, et al. *Mol Cell.* Jul 9.2010 39:36. [PubMed: 20603073]
18. Park WH, et al. *J Biol Chem.* Jun 24.2005 280:23593. [PubMed: 15849361]
19. Longeri S, San Filippo J, Liu D, Sung P. *J Biol Chem.* Aug 28.2009 284:23182. [PubMed: 19589784]
20. Yuan F, El Hokayem J, Zhou W, Zhang Y. *J Biol Chem.* Sep 4.2009 284:24443. [PubMed: 19561358]
21. Fanconi Anemia Mutation Database. <http://www.rockefeller.edu/fanconi/mutate/>
22. Winget JM, Mayor T. *Mol Cell.* Jun 11.2010 38:627. [PubMed: 20541996]
23. Cohn MA, et al. *Mol Cell.* Dec 14.2007 28:786. [PubMed: 18082604]
24. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. *Proc Natl Acad Sci U S A.* Aug 28.2001 98:10037. [PubMed: 11517324]
25. Schrodinger, LLC. 2010.

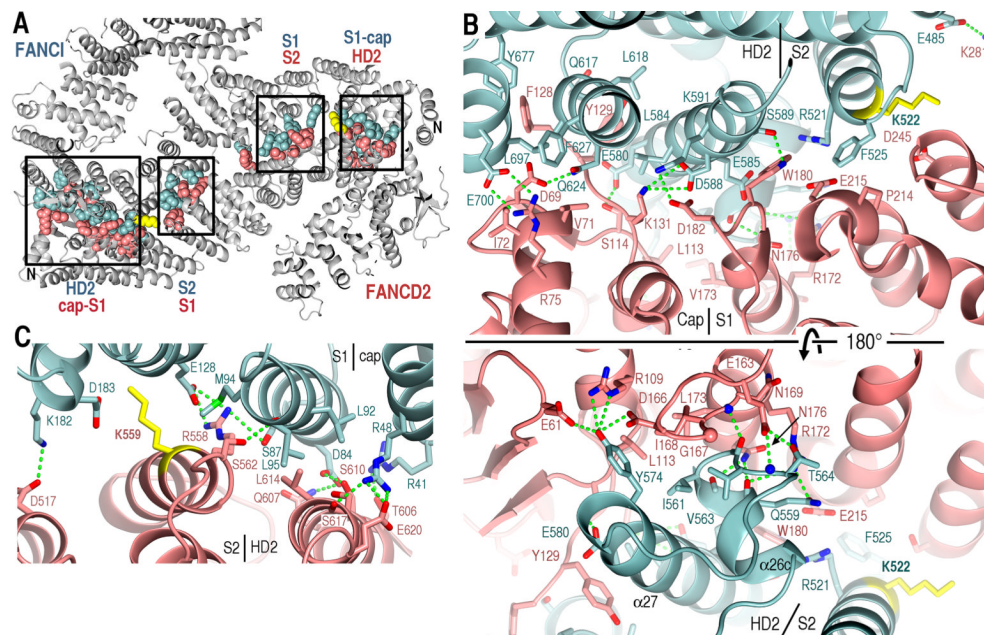


**Fig. 1. The overall structure of the ID complex**

(A) View looking into the interior of the trough-like structure. The FANCI (cyan) and FANCD2 (pink) Solenoid (S1 to S4), HD1 and HD2 domains are labeled near to their center. The ubiquitination site lysine side chains are in black-outlined yellow. The FANCD2 K559 side chain is poorly ordered.

(B) View orthogonal to (A) looking down the left end. Only the segments closest to the viewer are labeled.

(C) View orthogonal to (A) looking down the right end.



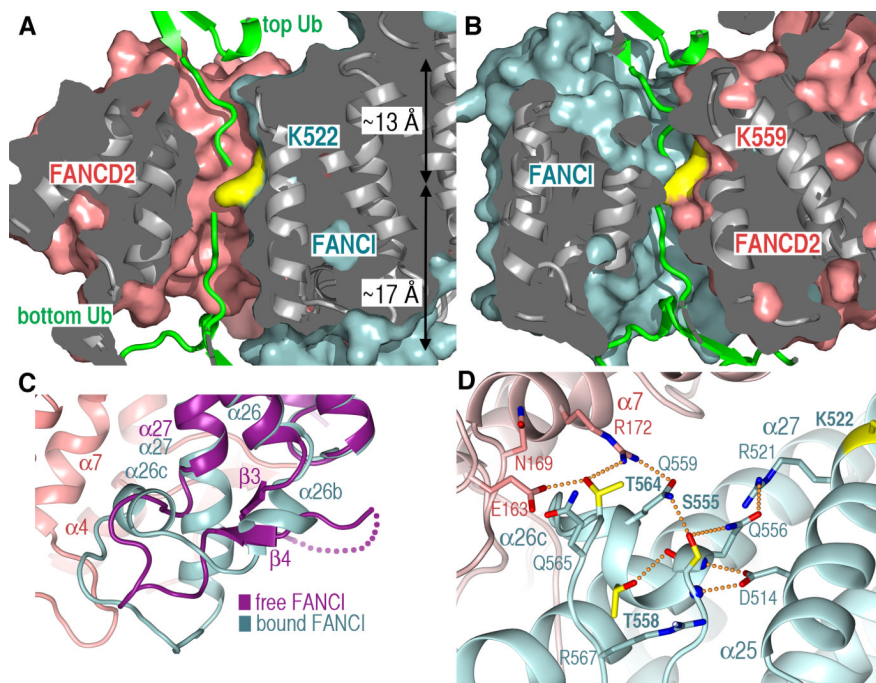
**Fig. 2. The I-D interface**

(A) Portion of the ID structure (gray) showing CPK representations of side chains within intermolecular-contact distance. Boxes indicate clusters of contacts with their respective segments labeled.

(B) Close-up view of the FANCD2 Cap-Solenoid 1 and FANCI Solenoid 2-HD2 segments showing side chains from (A). Green dotted lines indicate hetero-atoms within hydrogen bonding distance and geometry. Top-panel orientation is similar to (A), with bottom panel rotated  $\sim 180^\circ$  about the horizontal axis. The approximate boundaries of segments are labeled.

(C) The reciprocal end of the interface formed by the FANCI Cap-Solenoid 1 and FANCD2 Solenoid 2-HD2 segments.





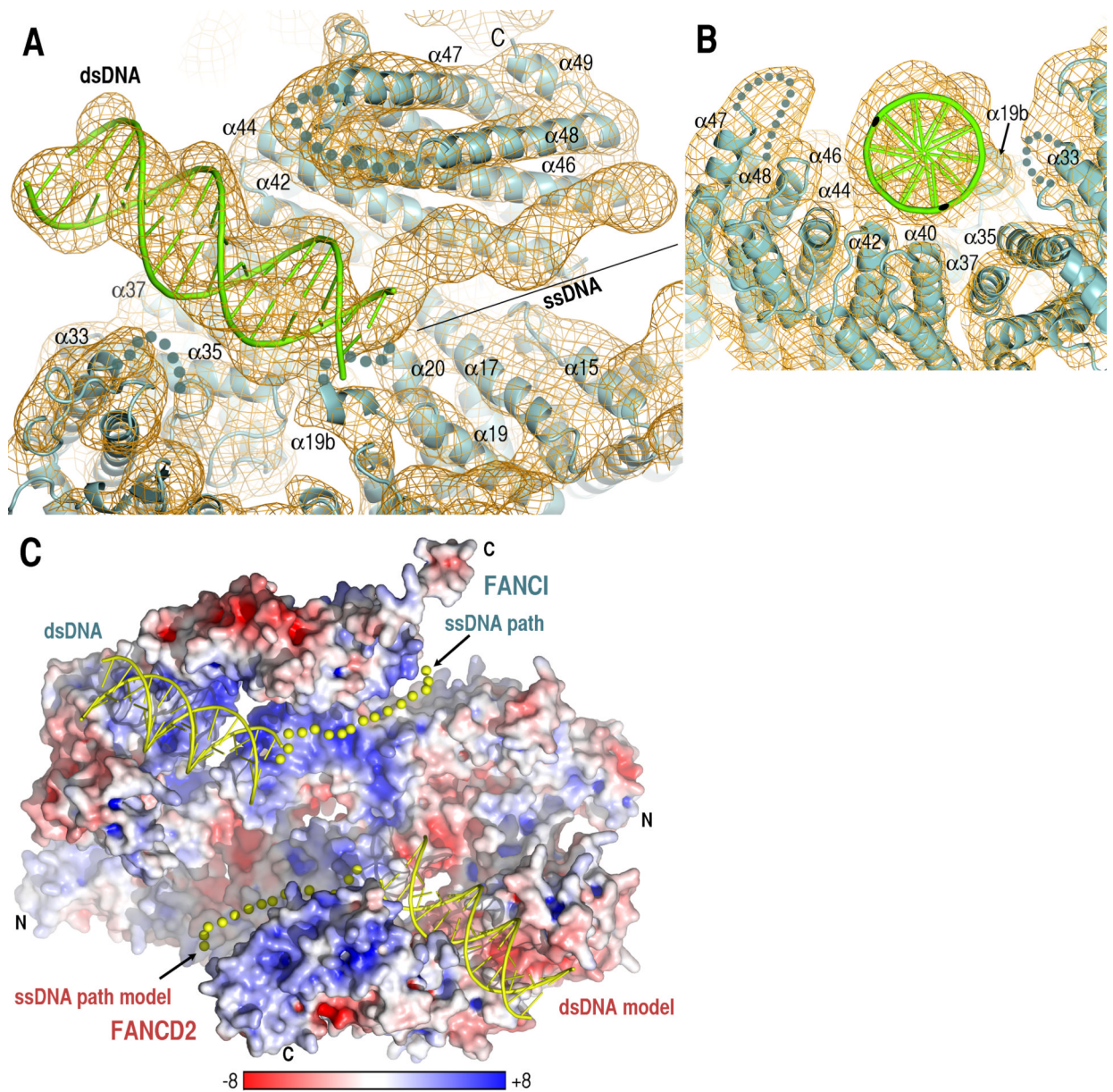
**Fig. 3. Mono-ubiquitination and phosphorylation sites of the ID complex**

(A) Surface representation of the solvent-accessible tunnel harboring the FANCI K522 ubiquitination site, looking from the interface center. The surface portion above the figure plane is clipped, and the surface interior is gray. The approximate distances from the lysine to the tunnel exits are indicated. Free ubiquitin structure (green) was manually docked with its tail approaching from either the “top” or “bottom” exits of the tunnel.

(B) Surface representation of the FANCD2 K559 tunnel. As in (A), two ubiquitin molecules were docked approaching from either tunnel exit.

(C) The FANCI Ser555, Thr564 and Thr566 phosphorylation sites map to an HD2 segment that undergoes a conformational change between free FANCI (purple) and FANCD2-bound FANCI (cyan). In free-FANCI, the three sites map to a disordered loop (dotted purple line). FANCD2 is in pink.

(D) The hydrogen bond network centered on the phosphorylation sites, also showing neighboring residues that may interact with the phosphorylated residues.



**Fig. 4. Electron density map of the Y DNA-FANCI crystals reveals bound double- and single-stranded DNA segments**

(A) 7.8 Å resolution  $F_o(\alpha_{calc})$  map, contoured at 1.2  $\sigma$ , calculated with molecular replacement phases that were improved by 3-fold ncs and multi-crystal averaging. The manually-positioned model of ideal B-type DNA is shown in cartoon representation (green), and the tubular density of ssDNA is marked by a black line. Orientation similar to Figure 1A.

(B) View looking down the blunt-end of the Y DNA.

(C) The electrostatic potential of the ID molecular surface showing the dsDNA cartoon and ssDNA path (yellow spheres) from the Y DNA-FANCI maps, as well as their models on FANCD2, positioned by superimposing Solenoids 3-4 of the two paralogs. The electrostatic potential was calculated with APBS (24) and illustrated (-8 to +8 kT) with Pymol (25).