STRUCTURAL ANALYSIS OF HUMAN FANCL, THE E3 LIGASE IN THE FANCONI ANEMIA PATHWAY

Charlotte Hodson¹, Ambrose R.Cole¹², Laurence P.C.Lewis¹³, Jennifer A.Miles¹ Andrew Purkiss-Trew¹ and Helen Walden¹*

¹Protein Structure and Function Laboratory, Lincoln's Inn Fields Laboratories of the London Research Institute, Cancer Research UK, 44 Lincoln's Inn Fields, London WC2A 3LY.
²Present address: Department of Biological Sciences, Institute of Structural and Molecular Biology, Crystallography, Birkbeck College, Malet Street, London WC1E 7HX.
³Present address: Sensory Neurogenetics Research Group, Max Planck Institute of Neurobiology, Am Klopferspitz 18, D-82152 Martinsried, Germany.
*Corresponding author Helen.Walden@cancer.org.uk, tel: +44(0) 20 72693106, fax: +44(0) 20 72693258

Running title: Structural analysis of human FANCL
Summary
The Fanconi Anemia (FA) pathway is essential for the repair of DNA interstrand cross-links. At the heart of this pathway is the monoubiquitination of the FANCI/FANCD2 (ID) complex by the multiprotein 'core complex' containing the E3 ubiquitin ligase FANCL. Vertebrate organisms have the 8-protein core complex, while invertebrates apparently do not. We report here the structure of the central domain of human FANCL in comparison with the recently solved Drosophila melanogaster FANCL. Our data represent the first structural detail into the catalytic core of the human system, and reveal that the central fold of FANCL is conserved between species. However, there are macromolecular differences between the FANCL proteins that may account for the apparent distinctions in core complex requirements between the vertebrate and invertebrate FA pathways. In addition we characterize the binding of human FANCL with its partners, Ube2t, FANCD2 and FANCI. Mutational analysis reveals, which residues are required for substrate binding, and we also show the domain required for E2 binding.

Introduction
Fanconi Anemia (FA) is an X-linked or autosomal recessive disorder with a range of clinical presentations including bone marrow failure and high incidence of cancer (1). Mutations in one of at least 14 genes associated with FA lead to dysfunction in DNA interstrand crosslink repair, rendering patients susceptible to DNA-damaging agents (2). Of the FA genes, 8 of the gene products form the FA core complex (FANC-A, -B, -C, -E, -F, -G, -L and -M) (3-6). The role of the complex, in conjunction with the E2 ubiquitin conjugating (UBC) enzyme Ube2T (7), is to monoubiquitinate each of the components of the FANC1-FANCD2 (ID) complex (8-11). Mutation in any of the genes coding for the core complex results in the loss of this monoubiquitination event. The FA pathway is at least partially present in all eukaryal organisms, with yeast having only a FANCm homologue (12), but all higher organisms possessing at least FANCD2. However, in the case of the core complex, vertebrates have the full complement, conserved from zebrafish to man (13), while the invertebrates vary in their FA composition – Caenorhabditis elegans have no obvious core complex (14), the slime mold Dictyostelium discoideum has a FANCE homologue (15) while Drosophila melanogaster has FANCm and FANCl homologues (16).

FANCL is the E3 ubiquitin ligase of the core complex (17). It was originally predicted from primary sequence to be a WD40-propeller fold, with a RING domain at the C-terminus. A recent study also suggested the presence in human FANCL of an N-terminal RWD domain (18). The RWD classification is a submember of the UBC-superfamily (19,20). The UBC superfamily has a core 4-stranded $\beta$-meander, flanked by an N-terminal helix and 1 or 2 C-terminal helices. In the catalytic E2s, this meander has a C-terminal $\beta$-flap containing the catalytic cysteine (Fig.1A). Most E2 folds contain a YPXXXP motif between strands 3 and 4 that form a triple turn motif also predicted to be characteristic of the RWD-like fold (18,20). In the RWD fold, the $\beta$-flap assumes a helical conformation (19,20) (Fig.1A).

We recently reported the structure of full length FANCL from Drosophila melanogaster (21). DmFANCL comprises 3 domains, an N-terminal E2-like fold (ELF), a double RWD (DRWD) domain and a C-terminal RING domain. The DmFANCL DRWD domain harbours no YPXXXP motifs, yet each lobe contains a helix in place of the $\beta$-flap, giving rise to the DRWD terminology.

Conservation in the core between the Drosophila and vertebrate FANCLs suggests that they would be broadly similar in structure. However, the sequence identity between human and DmFANCL is low, at 21% (21). Of the 3 domains, the DRWD domain, particularly the N-terminal half, has the least sequence identity, at 9%. Given that the requirements for a core complex are apparently different between species, and in humans, the loss of any member of the complex results in failure to ubiquitinate the ID complex, we set out to solve the structure of the human FANCL protein. We report here the structure of the central domain of human FANCL, representing ~50% of the protein. We show that it is globally similar to the DmDRWD domain, but that there are macromolecular differences between the proteins that may account for the biological differences in
the systems. In addition, we demonstrate that the central human FANCL domain is dispensable for interaction with Ube2T but is essential to bind substrates FANCD2 and FANCI.

Experimental Procedures

Protein expression and purification
Previous efforts to structurally characterise human FANCL had proved unsuccessful, but the understanding of the domain architecture of DmFANCL allowed us to design new constructs. We synthesised a codon-optimised version of the human FANCL coding sequence, using GeneArt. We then subcloned the central domain (residues 109-294) and the RING domain (residues 289-375) into a vector containing an N-terminal His-Smt3 tag using restriction-free cloning (22). Human Ube2T cDNA was purchased as an IMAGE clone (Gene service) and cloned into the same vector as described above. Proteins were expressed in E.coli BL21 cells (Invitrogen) in lysogeny broth medium supplemented with antibiotics, and 0.5mM ZnCl$_2$ in the case of the RING domain. Cells were cultured at 37°C to an OD$_{600}$nm of 0.6, induced with 0.25mM IPTG (0.5mM IPTG in the case of Ube2T), and cultured overnight at 16°C. Cells were harvested by centrifugation, and lysed by sonication on ice (10s bursts, followed by 30s on ice, repeated 4 times) in 100mM Tris pH 8.0, 500mM NaCl, 20mM Imidazole, 250µM tris(carboxyethyl)phosphine (TCEP) (10µM ZnCl$_2$ added when purifying the RING domain). Sonicates were subjected to high-speed centrifugation at 4°C to remove cell debris. Clarified supernatants were applied to a Ni$^{2+}$-NTA affinity resin (Invitrogen) equilibrated with sonication buffer. Fusion proteins were cleaved overnight at 4°C with Smt3 protease, Ulp1 at a weight:weight ratio of 1:30 Ulp1:His-Smt3-protein. Flow-through was collected, concentrated and applied to a SuperDex 200 or SuperDex 75 column for size-exclusion chromatography. Purified proteins were concentrated to 20mg/mL (central domain), 8mg/mL (RING), 15mg/ml (Ube2T) and flash-frozen in liquid nitrogen and stored at -80°C. Human central domain mutants were generated using the stratagene site-directed mutagenesis method with Pfu Turbo DNA polymerase (Strategene) and verified by sequencing. The central domain mutants were prepped as the wild type (WT). Human His-FANCD2 was a kind gift from M.R.Hodkinson and K.J.Patel, and prepared as described (23). Xenopus laevis FLAG-FANCI was a kind gift from P.Knipscheer and J.C.Walter and prepared as described (24).

Crystallisation and structure determination
Human FANCL central domain crystals were grown by hanging drop vapour diffusion at 4°C, in spacegroup C2, with cell dimensions $a = 147.6$ Å, $b = 102.5$ Å, $c = 65.8$ Å, $\alpha = 90$, $\beta = 94.09$, $\gamma = 90$. Solvent-content estimates suggested 2 monomers per asymmetric unit, with a solvent content of 79%. Crystallisation conditions were: 0.2M L-proline, 0.1M sodium tri citrate, pH5.5, 2% PEG 3,350. Crystals were harvested and cryocooled after protection with 30% glycerol. Data were collected at the DIAMOND synchrotron light source beamline I03, at 0.98Å wavelength. Both a low and high-resolution pass was collected (3.5Å and 1.75Å respectively). The 2 datasets were processed separately using Mosflm (25), and then merged using XDS and scaled using XSCALE (26,27). The last 55 images from the high-resolution data set were omitted from data processing due to radiation damage to the crystal, which accounts for 75% completeness in the outer shell (2.05-2.0Å, overall completeness 79%). The data were phased using the DmFANCL DRWD domain, residues 109-292, as a search model (pdb code 3K1L, (21)), with the program PHASER (28). The model was rebuilt manually through iterative cycles of refinement using COOT (29) for building and PHENIX (30) for refinement. Omit maps were generated across regions of the protein to assess model bias. The structure was refined to 2.0Å resolution. The asymmetric unit contains 2 molecules, A and B. A is complete and in a single conformation throughout, and is used for the analyses reported. Copy B however, has a region of disorder where residues Cys174-Ser182 in loop 6 appear in the electron density maps in multiple conformations. To reflect this movement/disorder, these atoms have been assigned varying occupancies in the PDB file as judged manually in the final 2Fo-Fc and Fo-Fc maps. Along with the high resolution of the data and that chain B has crystal contacts the NCS was not used for some of the loop regions (a total of 80 residues left out of the NCS). The final model possesses excellent stereochemistry with all residues in the favoured
regions of the Ramachandran diagram (98.7%) on the basis of an analysis with MolProbity (31). Table 1 summarises the data collection and refinement statistics.

**Structural Analyses**
All structural analyses were carried out on Chain A of the model. Figures were produced using PyMol (32), and structural based alignments were produced using MegaAlign and manual adjustments. To produce the electrostatic surfaces programs pdb2pqr (33) and apbs (34) were used in conjunction with PyMol.

**Thioester Charge Assay**
2µM of E2, His-UbcH7 (Boston Biochem), or human central domain was charged with an excess (50µM) of HA-Ub (Caltag Medsystems) using 150nM of E1 (Boston Biochem). A total reaction volume of 10µl was made up with assay buffer containing 50mM Tris pH8, 250mM NaCl, 2mM dithiothreitol (DTT), 5mM MgCl₂ and 5mM ATP. Reactions were left for 30min at room temperature. Reactions were then terminated by addition of 40µl of 2x non-reducing SDS buffer. 20µl of each reaction then had 1µl of 1M β-Mercaptoethanol added to reduced the thioester bond. 4µl of each sample were loaded on SDS-PAGE gels and visualized using western blots. Antibodies used were anti-His (GE Healthcare), anti-HA, and anti-FANCL raised against recombinant human FANCL central domain (Pettingill Technology Ltd).

**In vitro Pull Down Assays**
In a total reaction volume of 1ml containing assay buffer: 0.5M NaCl, 0.1M Tris pH8 and 250µM TCEP, human His-FANCD2 (225nM) or *Xenopus* laevis FLAG-FANCI (225nM) was added with an excess of human FANCL central domain or mutants at 454nM. Reactions were left to bind for 1hr at room temperature. 100µl Ni-NTA agarose (Qiagen) or 100µl anti-FLAG affinity gel (Sigma) were equilibrated in assay buffer and added to the 1ml reaction and left on a roller at 4°C for at least an hour. Samples were washed with 10mls of assay buffer. Agarose/affinity gel was then resuspended in 100µl of assay buffer and 50µl of 2x SDS buffer was added and the sample boiled. 9µl of the samples was loaded onto SDS page gel and subjected to Western blotting and probed with appropriate antibodies. Anti-FLAG antibody was purchased from abcam.

**Analytical Gel-Filtration**
Initial establishment of interactions between the human central domain (109-294), and human RING domain (289-375), with Ube2T was carried out by analytical gel-filtration. Interaction buffer contained 100mM Tris pH8, 250µM TCEP, 10µM ZnCl₂ with 100mM NaCl. An increased amount of NaCl to 500mM was used for central domain interactions, as the protein is unstable in lower amounts of NaCl. Samples were incubated on ice for 1hr with a total volume of 500µl. Samples were then loaded onto a Superdex 75 column (equilibrated with the sample buffer) and 0.5ml fractions were collected for analysis on 12% SDS-PAGE gels.

**Isothermal Titration Calorimetry (ITC)**
Kinetic information of Ube2T binding interactions with FANCL domains was established using the iTC200 microcalorimeter (MicroCal, Northampton, MA, USA). Sample buffers were the same as interaction buffers used for analytical gel-filtration. Experiments were carried out at 8°C. The syringe contained Ube2T between 650-750µM that injected a total volume of 40µl by 2.5µl injections. The cell contained a total volume of 205µl of either RING (70µM) or central domain (65µM). Cell concentrations were adjusted for a 1:1 stoichiometric interaction and Microcal Origin software version 7.0 was used to determine dissociation constants (Kd). All measurements were repeated at least twice.

**PDB accession numbers**
The coordinate and structure factor files have been deposited with the Protein Data Bank under the accession code 3ZQS.

**Results**

The structure of human FANCL central domain
We have determined the structure of the predicted human DRWD domain (residues 109-294) to 2.0Å resolution. Overall, the human FANCL central domain has a fused bi-lobal shape with dimensions: 70Å x 25Å x 20Å, consisting of 4 α-helices and 9 β-strands (Fig.2A). Each lobe contains a 4-stranded β-meander, followed by a β-element in the N-terminal lobe, and an α-helix in
the C-terminal lobe. The lobes are fused by a kinked helix that runs from one lobe to the other. In addition there is an α-helix N-terminal to the first β-meander and a C-terminal α-helix (Fig.2B).

Global comparison with DmFANCL
A comparison of known structures using the DALI server unsurprisingly reveals DmFANCL (pdb code 3K1L) as the closest structural homologue. However, a DALI comparison using only the N-terminal lobe (residues 109-194) or C-terminal lobe (residues 195-294) of the human central domain, lists E2 UBC enzymes as the most structurally similar for both, with significant RWD-fold hits only for the C-terminal lobe (Fig.2C). This is due to the single biggest difference between the DmDRWD and the human central domain being the change of secondary structure element in the N-terminal lobe (Fig.2D). In the DmDRWD, both lobes contain a helix after the 4-stranded β-meander, indicative to an RWD-fold, even in the absence of the triple turn YPXXXP motif. The human central domain however, harbours the YPXXXP motif in the N-terminal lobe, and a related HPXXXP motif in the C-terminal lobe (Fig.3A). Yet the human N-terminal lobe has a β-element in the position normally occupied by a helix in RWD-folds. This puts the N-terminal lobe of the human central domain in an almost intermediary position between an E2 fold with a β-flap in the catalytic position, and an α-helix in the non-catalytic RWD-like folds (Fig.2C). Therefore the human central domain is not a double-RWD domain (DRWD), but in fact an UBC-RWD domain. For the remainder of this manuscript the human central FANCL domain is referred to as the UBC-RWD domain (URD). Structure-based sequence alignment with other FANCL proteins reveals that this element swap is likely to be in all the vertebrate FANCLs (Fig.3A). Therefore we refer to this as a vertebrate-specific feature, and it represents a key difference between the vertebrate and invertebrate FANCLs.

Intriguingly, the human β-element in the N-terminal lobe contains a cysteine (Fig.3B). Superposition of the human N-terminal lobe with Ube2T, with an rmsd of 2Å, reveals the cysteine in the URD β-element lies structurally in the same position as Ube2T’s catalytic cysteine (Fig.2C).

We wondered if this cysteine was capable of forming a thioester bond with ubiquitin as in the case of the catalytic E2s. To test whether the cysteine in the URD is catalytic we carried out a thioester charge assay. The thioester assay revealed the URD is unable to form a thioester linkage between its cysteine and the activated C-terminus of ubiquitin (Fig.3B).

The interacting residues between the N and C-terminal lobes of the URD and DmDRWD are not conserved (Fig.3A and 3C). However the residues that make up the hydrophobic core between the two lobes are highly conserved (Fig.3A). A structural alignment of the human and Drosophila domains show an overall conservation of 50%, with only 13% of residues absolutely conserved (Fig.3B). There are also differences in conservation between the two lobes, with the C-terminal lobe being more conserved than the N-terminal lobe (30% and 20% respectively).

Surface composition reveals extensive hydrophobic exposure
It is likely that differences in a core complex requirement for the vertebrate system arise from the need to stabilise FANCL. We reasoned that there would likely be differences in the surface conservation and surface chemical composition between DmFANCL and HsFANCL. In order to assess this, we generated surface maps of the URD, and compared it with DmDRWD domain (Fig.4).

The surface conservation between DmDRWD and the URD is approximately 32%. The surface is more conserved on the C-terminal lobe at 20% compared to the N-terminal lobe at 12% (Fig.3B and Fig.4A).

Comparison of solvent-exposed hydrophobics between the URD and the DmDRWD reveal additional surface hydrophobics on the URD. A small hydrophobic patch resides on the N-terminal lobe of the URD, between the last β-strand of the 4-stranded β-meander (β4) and the β-element (residues L149, Y165, F166, V167, Fig.4B). Additionally a patch on the URD lies between α-helix1 and the N-terminal region of the large kinked helix, corresponding to residues: L121, A140, I184, Y187, L191 (Fig.4B). Interestingly, a
A comparison of the electrostatic potential of the surfaces of the URD and the DmDRWD reveal distinct differences between the species (Fig.4C). Interestingly there is a large positive patch on the DmDRWD, which is absent in the URD. Overall the URD has a greater negative charge than the DmDRWD, which wraps around the URD like a ring. However, on the backside of the central FANCL domain in both human and Drosophila, there is a similar overall negative charge.

The human RING domain binds Ube2T
In order to assess whether the URD plays any role in Ube2T binding we carried out size exclusion chromatography and isothermal titration calorimetry (ITC). Initial results from the size exclusion chromatography showed no change in migration that would correspond to a complex formation between the URD and Ube2T (Fig.5A). To further assess the interaction we employed ITC, which again showed no interaction between the URD and Ube2T (Fig.5B). We then tested the purified RING domain to see if it was sufficient for Ube2T binding. We initially observed a 1:1 stoichiometric interaction between the RING and Ube2T as judged by the migration of the complex during size-exclusion chromatography (Fig.5A). To ascertain the affinity of RING for Ube2T we carried out ITC. After titration of Ube2T into RING, the data were fitted with a stoichiometry of 1:1 as seen in the size exclusion chromatography resulting in an equilibrium binding dissociation constant (Kd) of 454nM (Fig.5B). Based on the structure of UbcH7 bound to the RING E3 c-cbl (35), we mutated F63 in Ube2T, predicted to be interacting with the RING, to alanine. No complex shift was observed between Ube2T-F63A and RING using analytical gel-filtration (Fig.5A). Further analysis using ITC revealed mutant Ube2T-F63A was able to bind the RING domain, but with a 11-fold decrease in binding affinity (Kd=5µM, Fig.5B).

The human central domain binds FANCI and FANCD2
In the Drosophila system, pull down studies indicated that the ELF domain was not required for substrate binding (21). Given the role of the RING domain in E2 binding, we hypothesised that the URD domain would be sufficient for substrate binding. In order to test this, we performed in vitro pull down experiments with FANCI, FANCD2 and the URD (Fig.6). As expected the URD was pulled down by both substrates. Our structural analysis reveals several hydrophobic patches exposed on the surface of the URD (Fig. 4B). In order to establish which patch(es) may be required for substrate binding, we generated the following mutants: L149A/F166A; L248A/F252A/L254A/I265A; V127A/Y128A; W212A/L214A (Fig.6A). All mutants were purified to homogeneity as wild type URD, and subjected to pull down arrays with FANCI and FANCD2. Both substrates showed reduced binding with mutants L248A/F252A/L254A/I265A and W212A/L214A as compared to wild type, indicating that the exposed hydrophobic patches on the C-terminal lobe are required for substrate binding. In contrast, mutations in the patches on the N-terminal lobe behaved as wild type (Fig.6B).

Discussion
The structure of the human central FANCL domain (URD) adopts the same overall topology as the Drosophila DRWD domain, confirming that the FANCL proteins belong to the UBC and RING superfamilies and are not WD40 proteins as predicted (17). The human central domain does differ to the DmFANCL as it does not have a double RWD domain (DRWD) but contains an UBC-fold fused to an RWD fold (URD). Gene
duplication resulting in fused domains is not a novel occurrence in protein topology. However, the fused consecutive UBC superfamily folds, seen in both the URD and the DmDRWD, are unique and appear to be specific to FANCL.

In the context of the full-length human FANCL, there may be subtle differences in domain associations. Structure-based sequence alignments of vertebrate FANCL proteins to Drosophila FANCL reveal that the loops connecting ELF-URD (or ELF-DRWD) and URD-RING (or DRWD-RING) are divergent between species (21). The ELF-URD loop is conserved in length, but not composition, and in the Drosophila structure, there are no interactions between the ELF and any other part of the protein. One possible explanation for the existence of a β-element in place of the helix in human FANCL URD may be to allow a more intimate association between the URD and ELF domains. The URD-RING loop in human FANCL is 7 residues shorter than in DmFANCL. We show a high affinity of human RING for the E2, Ube2T, which may reflect a different arrangement of the URD and RING domains in the human protein. Further structural analyses of the relationship between the human URD and RING domains and other members of the FANCL family will be required to tease out these differences.

Although we have shown the URD domain is not required for Ube2T binding, the URD domain appears to be essential for the function of FANCL (36). It has been previously shown in the Drosophila system that the ELF domain is not required to bind substrates FANCD2 and FANCI (21). In this analysis we show the human URD domain is sufficient for binding substrates. Additionally, we also reveal the molecular determinants of this binding (Fig.6). In particular, the residues W212, L214, L248 and L254 are conserved across species (Fig.3A). Previous mutational work based on the predicted structure of FANCL indicated that the then-uncharacterised URD domain was essential for core complex assembly (36), with point mutants R226E, W201A and W274A (human numbering, numbered as 275 in (36)) abolishing core complex assembly. Our structure of human FANCL URD provides a molecular rationale for these observations and indicates that loss of core complex assembly is in fact due to structural instability (Fig.7). As predicted from the Drosophila FANCL structure (21), R226 forms buried salt bridges with D208 and the main chain oxygen of A223 (Fig.7). Mutation of this residue to the oppositely-charged glutamate therefore destabilises these salt bridges and consequently the surrounding fold. W201 is right in the middle of the interface connecting the N and C-lobes of the URD domain (Fig.7). A W201A mutation would therefore impact the folding of the lobes as this residue is clearly seen in the structure to be stacking against and making multiple contacts with residues H148, R146 and R221. Therefore, although the hydrophobic nature of the residue is maintained, the contacts in the core are not (Fig.7). Finally, W274 is completely conserved between all the FANCL species, including Drosophila (21). In our URD structure, W274 forms a hydrogen bond between the side-chain nitrogen of the tryptophan and the main-chain oxygen of T246. Mutation of this W274 to alanine removes this hydrogen bond as well as multiple van der Waals contacts with residues M247 and N283 (Fig.7). We can conclude that these particular mutations destabilise the URD domain and we can speculate based on the extent of surface exposed hydrophobic patches on the URD domain, that this domain is very likely to be involved in binding one or more of the core complex proteins. The presence of such patches certainly suggests a number of protein-protein interaction sites. Interestingly, a recent study implicates the URD domain in PCNA binding, providing a potential link between the translesion synthesis and FA pathways (37). Coupled with our observation that the URD binds FANCI and FANCD2, it appears that the URD domain of FANCL is responsible for several interactions. Finally, there has been speculation that phosphorylation of FANCI, shown to trigger the FA pathway in the chicken system, may enhance the ability of FANCL to bind FANCI (38). For this to be the case we would anticipate the existence of a basic patch on FANCL to mediate interaction with a cluster of negatively charged phosphate ions. From our surface electrostatic analysis of the URD domain, there is no such patch, yet intriguingly there is in the Drosophila protein. Further structural studies of FANCL in complex with substrates and other binding partners
will be required to determine the functions of these surfaces.

Acknowledgements
We thank Michael Way for discussions and criticisms on the manuscript, and members of the Protein Structure and Function lab for suggestions and support. We also like to thank K.J.Patel for the plasmid for HIS-FANCD2 and P.Knipscheer and J.C.Walter for the FLAG-FANCI plasmid. We would like to thank the protein production facility for generating the viruses for the insect cell protein expression. Cancer Research UK supported this work.
REFERENCES

Figures and Tables

**Figure 1: Structural Evolution of the UBC-family folds**

(A) The UBC fold of a catalytic E2 (Ube2T, pdb code 1YH1) shown in blue, of a non-catalytic E2 (MMS2, pdb code 1J74 (39)) shown in pink, and the RWD fold (RWDD1a, pdb code 2EBM). A gold star shows the catalytic cysteine and a hot pink patch shows the RING domain-binding region. The β-flap indicative to a UBC fold is labelled and the YPXXXP motif.

(B) A structural alignment of UBC folds for both catalytic E2s (Ube2D1 and Ube2T), non-catalytic E2s (Ube2V1 and Ube2V2) and for the RWD folds (RWDD1a and RWDD2a). A scale of red-orange-white shows conservation, with red being 100% conserved and white being 0%. A gold circle labels the catalytic cysteine. The secondary structure is shown above in blue, with the main difference between the UBC fold (a β-strand) and the RWD fold (a2) shown in light green. The most conserved region is the YPXXXP motif.

**Figure 2: Structure of Human FANCL central domain**

(A) Surface representation of the human central domain, with hot pink representing the N-terminal lobe and magenta representing the C-terminal lobe. (B) Secondary structural representation of the human central domain. (C) Structural comparison of the human N-terminal lobe, shown in hot pink overlaid with the UBC fold of Ube2T shown in blue with its catalytic cysteine highlighted by a gold star. A comparison of the human C-terminal lobe shown in magenta overlaid with the RWD fold of GCN2, shown in cyan.
Figure 3: Structural Comparison of Human and Drosophila central domains
(A) A structural alignment of the central domains across species. A scale of red-orange-white shows conservation, with red being 100% conserved and white being 0%. Secondary structure is shown above with hot pink for the human central domain and green for the DmDRWD domain. The YPXXXP and HPXXXP motifs are highlighted in blue. The cysteine found in the differing β-strand region of the human central domain is highlighted in lilac. Residues involved in the lobe interactions have a green circle above and those that are found in the hydrophobic groove of the lobe interacting site have red circles above them. Residues mutated to alanine to determine the region for substrate binding have a square above them, with blue squares representing patch 2 and pink squares representing patch 4. (B) Thioester Charge Assay to determine whether the human β-element cysteine thioester links ubiquitin in a manner analogous to catalytic E2s. Lanes 1 and 2 for each panel correspond to the sample in non-reducing buffer and reducing buffer to determine a thioester linkage, respectively. It is clear that that the URD cannot form a thioester link to ubiquitin. (C) A close up view of the interactions between the human N-terminal (pink) and C-terminal (mauve) central domain lobes. Hydrogen bonds are depicted as black dashed lines.

Figure 4: Surface Comparisons of the Human URD and the Drosophila DRWD domains
(A) A surface conservation comparison between the human URD and the Drosophila DRWD. The surface shown here is of the human URD. (B) A comparison of surface hydrophobics shown in green, for both the human and Drosophila domains. (C) A comparison of surface electrostatics for both the human and Drosophila domains, with blue for basic charge and red for acidic.

Figure 5: RING domain is necessary and sufficient for E2-binding
(A) Size exclusion Chromatograms showing molecular weight shifts if complex formation has occurred. Green lines on the chromatogram represent the proteins allowed to form complexes and red lines show Ube2T or Ube2T-F63A. It is clear that only complex formation occurs between RING and Ube2T from the size exclusion chromatography experiments. The SDS PAGE gel also show that both the RING and Ube2T proteins are present in the shifted peak. (B) Isothermal Titration Calorimetry (ITC) experiments of the interactions between, URD-Ube2T, RING-Ube2T and RING-Ube2T-F63A.

Figure 6: URD domain binds substrates FANCD2 and FANCI
(A) Structural representation of the surface hydrophobic patch mutants of the URD domain. (B) Pull down of substrates His-FANCD2 and FLAG-FANCI with WT URD and the patch mutants of the URD, with the mutations indicated in (A). URD input and URD incubated with beads alone control are shown using anti-FANCL antibody. The pull down shows either His-FANCD2 with anti-His or FLAG-FANCI with anti-FLAG and the URD species that have been pulled-down with anti-FANCL antibody.

Figure 7: Structural explanation for FANCL mutations
(A) Close-up of the salt-bridge formed between Arg226 and D208. The hydrogen bonding is depicted as black dashed lines. (B) Close-up view of the contribution of Trp201 to the hydrophobic core of the DRWD N/C interface. Van der Waals contacts are represented by grey dotted clouds. (C) Close-up view of the completely conserved Trp274 and its contacts in the URD core. Hydrogen bonding is depicted as black dashed lines, van der Waals contacts as grey dotted clouds.
### Table 1
Data Collection and Refinement Statistics

<table>
<thead>
<tr>
<th><strong>Data collection</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength (Å)</td>
<td>0.98</td>
</tr>
<tr>
<td>Space group</td>
<td>C121</td>
</tr>
<tr>
<td>Cell dimensions (Å)</td>
<td>(a = 147.6, \ b = 102.5, \ c = 65.8)</td>
</tr>
<tr>
<td>Cell dimensions (°)</td>
<td>(\alpha = 90, \ \beta = 94.09 \ \gamma = 90)</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>45-2.0, (2.05-2.0)</td>
</tr>
<tr>
<td>No. observations</td>
<td>153360</td>
</tr>
<tr>
<td>No. unique reflections</td>
<td>103454</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>79%, (75%)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>1.5, (1.5)</td>
</tr>
<tr>
<td>(I / \sigma)</td>
<td>8.91, (2.03)</td>
</tr>
<tr>
<td>(R_{\text{meas}})</td>
<td>7.2%, (59.1%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Refinement</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of reflections used</td>
<td>98143</td>
</tr>
<tr>
<td>Number of protein atoms</td>
<td>3031</td>
</tr>
<tr>
<td>Number of solvent atoms</td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td>413</td>
</tr>
<tr>
<td>Na(^+) ion</td>
<td>2</td>
</tr>
<tr>
<td>PEG</td>
<td>103</td>
</tr>
<tr>
<td>Proline</td>
<td>16</td>
</tr>
<tr>
<td>Mean B values (Å(^2))</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>39.5</td>
</tr>
<tr>
<td>Water</td>
<td>54.5</td>
</tr>
<tr>
<td>Na(^+) ion</td>
<td>35.0</td>
</tr>
<tr>
<td>PEG</td>
<td>62.8</td>
</tr>
<tr>
<td>Proline</td>
<td>81.3</td>
</tr>
<tr>
<td>R.m.s deviations</td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.007</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.0</td>
</tr>
<tr>
<td>Final R(_{\text{work}})</td>
<td>0.183</td>
</tr>
<tr>
<td>Final R(_{\text{free}})</td>
<td>0.207</td>
</tr>
<tr>
<td>Ramachandran</td>
<td></td>
</tr>
<tr>
<td>favoured</td>
<td>98.7%</td>
</tr>
<tr>
<td>outliers</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

Highest resolution shell given in parentheses.
A  Size Exclusion Chromatography

URD-Ube2t  RING-Ube2t  RING-Ube2tF63A

B  Isothermal Titration Calorimetry

Time (min)  Time (min)  Time (min)

μcal/sec  μcal/sec  μcal/sec

KCal/Mole of injectant  KCal/Mole of injectant  KCal/Mole of injectant

Molar Ratio  Molar Ratio  Molar Ratio

N = 0.98 ± 0.01  Kd = 454nM ± 128nM
N = 0.99 ± 0.06  Kd = 5μM ± 2.4μM