

# Insights into Fanconi Anaemia from the structure of human FANCE

Ravi K. Nookala, Shobbir Hussain and Luca Pellegrini\*

Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1GA, UK

Received November 2, 2006; Revised December 20, 2006; Accepted January 5, 2007

## ABSTRACT

**Fanconi Anaemia (FA) is a cancer predisposition disorder characterized by spontaneous chromosome breakage and high cellular sensitivity to genotoxic agents. In response to DNA damage, a multi-subunit assembly of FA proteins, the FA core complex, monoubiquitinates the downstream FANCD2 protein. The FANCE protein plays an essential role in the FA process of DNA repair as the FANCD2-binding component of the FA core complex. Here we report a crystallographic and biological study of human FANCE. The first structure of a FA protein reveals the presence of a repeated helical motif that provides a template for the structural rationalization of other proteins defective in Fanconi Anaemia. The portion of FANCE defined by our crystallographic analysis is sufficient for interaction with FANCD2, yielding structural information into the mode of FANCD2 recruitment to the FA core complex. Disease-associated mutations disrupt the FANCE–FANCD2 interaction, providing structural insight into the molecular mechanisms of FA pathogenesis.**

## INTRODUCTION

A group of rare genetic conditions collectively defined as chromosome instability syndromes has received much attention in recent years, as their study continues to provide important insight into the molecular mechanisms responsible for the integrity of our genome. One such condition is Fanconi Anaemia (FA), a genetically heterogeneous disorder characterized by congenital abnormalities, aplastic anaemia and predisposition to cancer, especially acute myeloid leukemia and squamous cell carcinomas (1–3). A conspicuous cellular feature of FA is chromosomal fragility and hypersensitivity to DNA cross-linking agents such as mitomycin C, diepoxybutane and cisplatin. Sensitivity to genotoxic agents suggests that the pathogenic effects of FA are due to defects in the

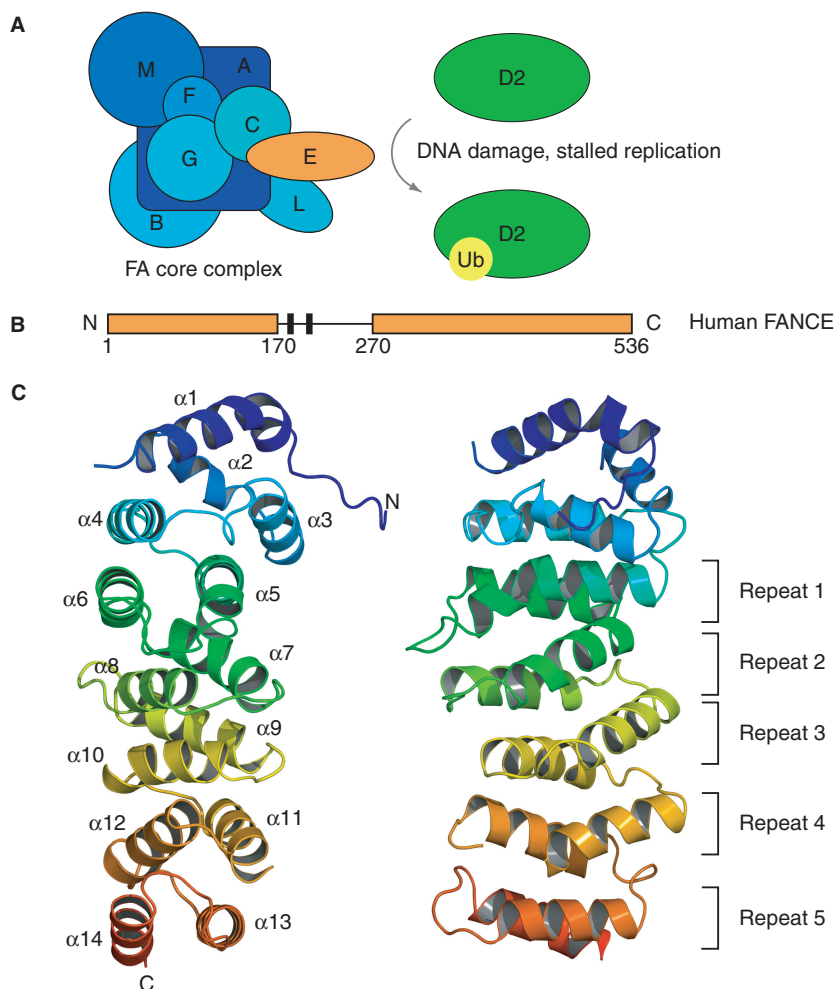
molecular mechanisms of DNA damage signalling and repair (4–6).

Twelve different FA subtypes (A, B, C, D1, D2, E, F, G, I, J, L, M) have been isolated, and the genes for all but type I have been cloned (7). The majority of FA proteins do not possess clear functional motifs, and only a subset of them have been associated with an enzymatic activity, including a E3 ubiquitin ligase, FANCL (8,9), and two helicases, FANCI (10,11) and FANCD1 (12,13). A nuclear multi-subunit complex of at least eight FA proteins (FANCA, FANCB, FANCC, FANCE, FANCF, FANCG, FANCL and FANCM), the FA core complex (14), adds a single ubiquitin chain to FANCD2 following DNA damage or replicative stress (15) (Figure 1A). Monoubiquitination acts as a signal for FANCD2 recruitment to nuclear foci where it colocalizes with cell-cycle checkpoint regulation and DNA repair proteins such as BRCA1, BRCA2 and RAD51 (15–17). Within the FA core complex, individual constituents engage in multiple interactions with each other, giving rise to functional subcomplexes (18,19). Recent evidence also points to additional roles of the FA core complex besides FANCD2 ubiquitination (20). Although much has been learned about the role of the FA proteins in maintenance of genome stability, our understanding of the molecular mechanisms underlying their function remains largely incomplete.

FANCE is essential for FANCC accumulation in the nucleus and assembly of the FA core complex (21,22). Moreover, FANCE localizes to constitutive nuclear foci (21) and becomes associated with ubiquitinated FANCD2 and BRCA2 in a chromatin complex (23). FANCE is the only member of the FA core complex for which a direct association with FANCD2 has been demonstrated (21). Indeed, it has been proposed that FANCE represents the essential link between the FA core complex and FANCD2 (19,21,22).

Here we describe the identification and crystallographic analysis of a large, evolutionarily conserved region of human FANCE. The first structure of a FA protein reveals the presence of a repeated helical motif, which was not apparent from the analysis of its amino acid sequence and represents a structural template for other proteins

\*To whom correspondence should be addressed. Tel: 00441223333667; Fax: 00441223766002; Email: luca@cryst.bioc.cam.ac.uk



**Figure 1.** Structure of human FANCE. (A) In response to DNA damage or replicative stress, the FA core complex monoubiquitinates FANCD2. The eight identified subunits of the FA complex are shown. (B) Schematic representation of the domain structure of the FANCE protein. Orange boxes represent regions of the protein that are predicted to constitute independently folded domains. The two nuclear localization sequences in the middle section of the protein are drawn in black. (C) Cartoon representation of the crystal structure of amino acids 273 to 536 of human FANCE. The protein chain is shown as a ribbon, rainbow-coloured from blue at the N-terminal end to red at the C-terminal end. Two views are shown, differing by a 90° rotation around an axis aligned to the long dimension of the molecule. The alpha-helical segments in the structure are labelled,  $\alpha 1$  to  $\alpha 14$ . The positions of the five helical repeats identified in FANCE are indicated next to the structure.

defective in Fanconi Anaemia. We demonstrate that the FANCE region defined by the structure is sufficient for interaction with FANCD2 and identify an epitope on the FANCE surface that is critical for FANCD2 binding. Disease-associated mutations in FANCE and FANCD2 disrupt the FANCE–FANCD2 interaction, providing a structural rationale for their pathological effect in FA patients.

## MATERIALS AND METHODS

### Purification and crystallization

A C-terminal segment of the human FANCE protein spanning amino acids 273 to 536 (natural C-end) was cloned into the pET28a plasmid vector and over-expressed in the *E. coli* BL21(DE3) strain as a 6xHis-tagged protein. The recombinant protein was purified by standard  $\text{Ni}^{2+}$ -affinity chromatography over a HisSelect™ Sepharose

(Sigma, UK) column. The histidine tag was cleaved with thrombin protease and the digested sample was passed again over the HisSelect column in order to remove the cleaved tag. The FANCE protein was further purified by size exclusion chromatography using a Superdex™ 16/60 Global 200 column (GE Healthcare, UK).

Mutagenesis was performed with the Quick Change™ II mutagenesis kit according to the manufacturer's instruction (Stratagene). The C391A FANCE mutant used for structure determination was purified like the wild-type protein, and crystallized by vapor diffusion against buffer solution 22 of the Crystal Screen Cryo (Hampton Research, USA), consisting of 0.085 M Tris pH 8.0, 0.170 M sodium acetate, 25.5% w/v PEG 4000, 15% glycerol.

### Phasing and refinement

The X-ray crystal structure of FANCE was solved using the anomalous signal of the selenomethionine-substituted

protein. Native diffraction data to a resolution of 2.0 Å and multiple anomalous dispersion data to 2.8 Å were collected at beamline ID29 of the European Synchrotron Radiation Facility (ESRF), Grenoble, France. The diffraction intensities were measured in MOSFLM (24) and merged in SCALA (25). The positions of the selenium atoms in the asymmetric unit were determined in Shake-and-Bake (26), and used as input for phasing in SHARP (27). The solvent-modified map calculated by SHARP was readily interpreted by ARP/wARP (28), which produced an almost complete model of the structure. Refinement of the structure was carried out in REFMAC5 (29), together with minor manual rebuilding in COOT (30). The final crystallographic model comprises 249 amino acids and 223 water molecules (2067 atoms), and includes amino acids 275 to 535 of the human FANCE sequence. Residues 301 to 307 in the loop linking helix 1 and helix 2, and residues 479 to 483 in the intra-helical loop of repeat FANC4 are not visible in the electron density map and are not included in the final model. The conformation of amino acids 484, 485 and 518 to 521 in the inter-helical loops of repeats FANC4 and FANC5 must be considered tentative as the quality of the electron density is poor for these residues. 97.3% of residues in the crystallographic model are in the most favoured regions of the Ramachandran plot, 2.7% in the allowed regions and none in disallowed regions of the plot. Figures were prepared with PyMOL (<http://pymol.sourceforge.net>).

### Yeast two-hybrid analysis

The MATCHMAKER Two-Hybrid System 3 (Clontech) was used for yeast two-hybrid analysis according to the manufacturer's instructions and as described earlier (31). Briefly, GAL4-activation domain and GAL4-binding domain constructs were either sequentially transformed into AH109 yeast cells and subjected to selection on -trp/-leu/-his/-ade medium, or transformed separately into AH109 and Y187 yeast strains and mating cultures plated onto selection medium. Transformations were performed using a PEG/ssDNA/lithium acetate procedure. Colonies that grew on -trp/-leu/-his/-ade selection media were transferred onto filters and tested for  $\beta$ -galactosidase expression with X-gal. The activation of the three reporters in this system: *His3*, *Ade2* and *LacZ*, was assayed for in each single experiment. Yeast colony growth therefore represents simultaneous activation of the *His3* and *Ade2* reporters and blue colouring of colonies following X-gal treatment represents activation of the *LacZ* reporter. Experiments were performed at least in triplicate.

## RESULTS

### Domain mapping and structural analysis of FANCE

The human FANCE gene encodes a protein of 536 amino acids. Examination of its sequence in DISOPRED (32) reveals the presence of a disordered region between residues 170 and 270, linking N- and C-terminal domains with high predicted secondary structure content

(Figure 1B). In addition, amino acids in the C-terminal domain display a higher degree of evolutionary conservation relative to the rest of the protein (Figure 2). Thus, FANCE sequence analysis suggests that its C-terminal region might represent a distinct protein domain capable of autonomous folding and suitable for biophysical investigation. We expressed and purified a region of human FANCE spanning residues 273 to 536 (264 amino acids). Initial crystallization experiments were unsuccessful. SDS-PAGE analysis showed a marked tendency of the recombinant protein to multimerize through disulphide-mediated cross-linking. Systematic replacement of cysteine residues with alanine showed that the C391A mutation removed the tendency of the protein to form covalent aggregates (Supplementary Figure 1) and promoted the growth of crystals suitable for high-resolution X-ray analysis. The crystal structure was solved to a resolution of 2.0 Å by the multiple anomalous dispersion method using selenomethionine-substituted protein (Table 1). In the rest of the article, we will refer to the FANCE region studied here simply as FANCE.

### General features of the structure

The crystal structure reveals that FANCE consists predominantly of helices (thirteen  $\alpha$ -helices, one  $3_{10}$ -helix) and no  $\beta$ -strand (Figure 1C). The molecule adopts an elongated, non-globular shape, with a size of 70 Å in its longest dimension, a width of 30 Å and thickness of 20 Å. The polypeptide folds in a continuous, right-handed solenoidal pattern from the N- to the C-terminal end of the chain. Beginning with helix  $\alpha 5$ , the loops of the solenoid become more regular, and it is possible to identify five copies of a helical motif repeating to the C-end of the protein. Thus, the most outstanding feature of the FANCE structure is the presence of a repeated motif, which was not detected by the inspection of the amino acid sequence.

The repeats vary between 30 and 40 amino acids in length and fold in an antiparallel helical hairpin (Figures 2 and 3). The two helices (H1 and H2) of the repeat are of similar size, spanning between three and four turns, and cross with an angle varying between 21° and 35°. The helices in the repeat are straight or display only a minimal degree of bending. The only exception to these characteristics is seen in helix H1 of repeat 1, which is longer at five turns and shows a pronounced kink between the second and third helical turn. In each repeat, helices H1 and H2 make extensive contacts with each other and with the helices of neighbouring repeats, thus generating a continuous hydrophobic core extending throughout the molecule. The loops connecting the helices within a repeat, as well as adjacent repeats, vary significantly in length and conformation.

Each repeat lies within an ideal plane that is broadly perpendicular to the long axis of the molecule, and stacking of the repeats in a consecutive array generates a double layer of helices. Within the structure, adjacent repeats are related by a rotation and a tilt along an axis parallel to the long dimension of the structure,

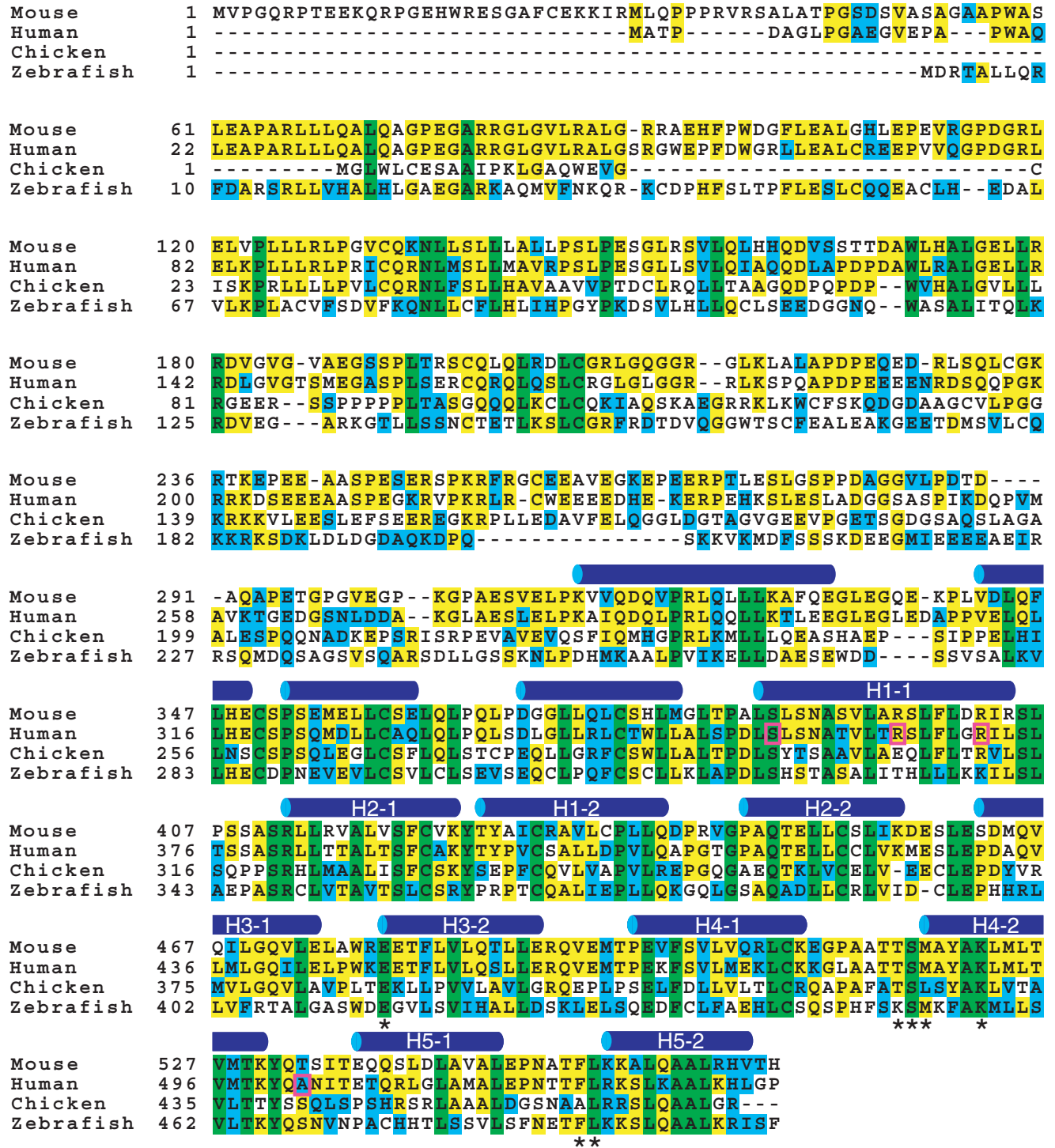


Figure 2. Multiple sequence alignment of FANCE orthologues. The amino acid sequences of mouse, human, chicken and zebrafish FANCE were aligned in the program ClustalW. Absolutely conserved residues are highlighted in green, identical residues in yellow and conserved residues in cyan. The secondary structure elements identified in the FANCE structure are marked above the alignment. FANCE residues involved in FANCE binding by bioinformatic and yeast two-hybrid analysis are marked by an asterisk below the alignment. Residues mutated in Fanconi Anemia are highlighted by a red box.

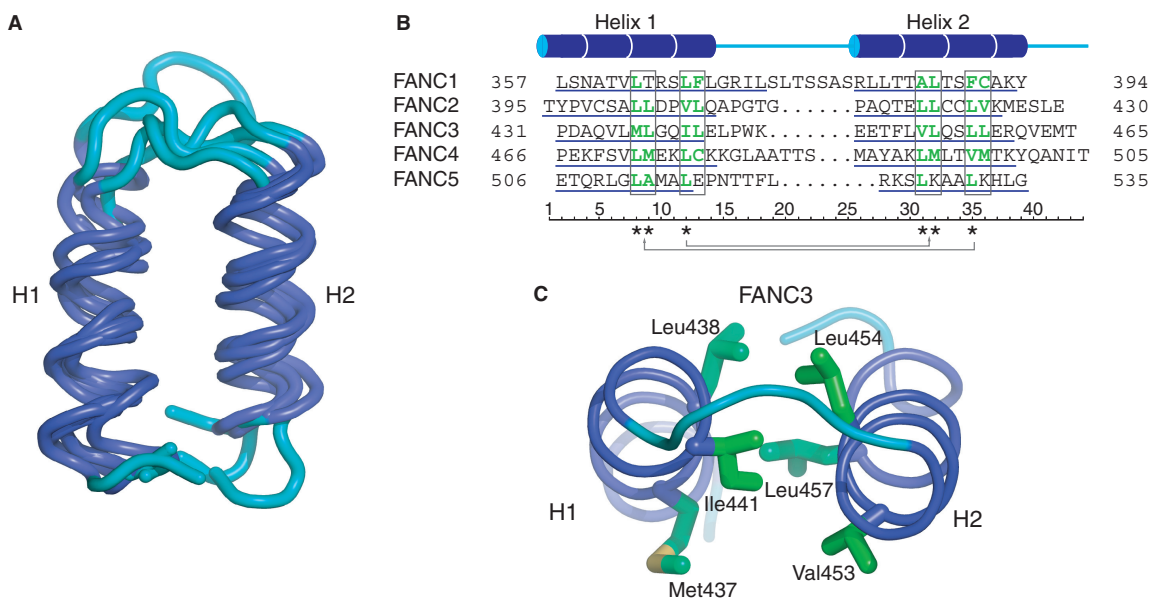
which together generate a super-helical curvature. The rotation angle between repeats 1 and 2 and repeats 3 and 4 spans  $\sim 35^\circ$ , whilst the rotation angle between repeats 2 and 3 and 4 and 5 is  $\sim 15^\circ$ . The presence of

super-helical curvature does not confer an overall curved shape to the structure, since the direction of rotation changes as one moves from the N- to the C-terminus of the chain: repeats 2 and 3 show a left-handed rotation,

**Table 1.** X-ray diffraction data and refinement statistics

		Se-Met			Native
Data collection <sup>a</sup>					
Space group	P4 <sub>1</sub> 2 <sub>1</sub> 2				
Cell dimensions (Å):					
a					59.3
b					59.3
c					140.2
		Peak	Remote	Inflection	
Energy (keV)		12.6609	12.700	12.6590	12.6588
Resolution (Å)		2.8	2.8	2.8	2.0
R <sub>sym</sub> (%)		5.8 (26.8)	6.8 (47.4)	9.9 (88.0)	5.0 (19.5)
Multiplicity		12.9 (13.8)	13.0 (13.8)	13.0 (13.8)	8.6 (8.9)
I/σI		10.0 (2.8)	9.2 (1.6)	6.8 (0.9)	8.1 (3.9)
Completeness (%)		98.3 (98.3)	98.6 (98.3)	98.4 (98.6)	98.9 (98.7)
Refinement					
Resolution (Å)	2.0				
No. of reflections	17665				
R <sub>work</sub> /R <sub>free</sub>	0.194/0.244				
No. non-H atoms	2067				
Mean B value (Å <sup>2</sup> )	32.9				
Matthews coefficient	2.13				
Solvent content (%)	41.8				
R.m.s deviations					
Bond lengths (Å)	0.018				
Bond angles (°)	1.7				

<sup>a</sup>Highest resolution shell is shown in parenthesis.



**Figure 3.** The FANC repeat. (A) Superposition of the five copies of the repeated helical motif revealed by the crystal structure. Each repeat is drawn as narrow tube, with the two helical segments coloured in blue. (B) Structure-based sequence alignment of the five FANC repeats. Conserved hydrophobic residues are highlighted in green. Amino acids that become buried at the intra- and inter-repeat interfaces are boxed. Two sets of three asterisks below the alignment indicate the two triplets of hydrophobic residues that interact in a similar fashion within a repeat. (C) Example of intra-repeat packing of hydrophobic side chains at conserved triplet positions in the FANC3 repeat. The trajectory of the main chain atoms of the repeat is indicated by a narrow tube, whereas the two triplets of interacting side chains are drawn as sticks, coloured in two hues of green.

whereas repeats 4 and 5 show a right-handed rotation. Because of the reversal in the sense of rotation the helices in repeat 5 become parallel again with those of repeat 1.

### The FANC repeat

The repeated motif identified in the structure of human FANCE represents a novel member of the large family

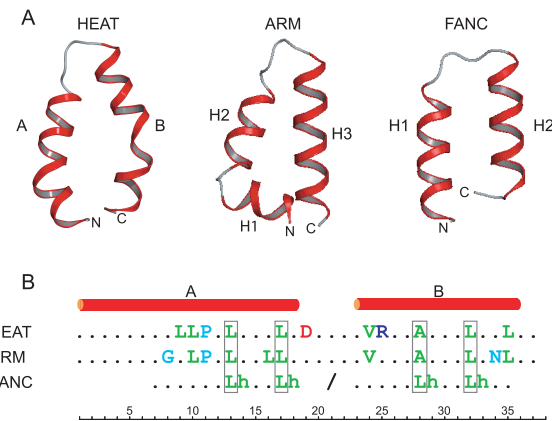
of two- and three-helical motifs that includes the well-characterized HEAT and ARM repeats (33). Here we will refer to the helical motif in human FANCE as the FANC repeat. Although the five FANC repeats clearly share the same architecture, their pairwise superposition gives a root mean square deviation (rmsd) varying between 1.4 and 2.7 Å over 24 alpha carbon positions, thus pointing

to significant differences between repeats (Figure 3A). The three C-terminal repeats are structurally closer to each other, whereas the first and second FANC repeats are more divergent.

A structure-based alignment of the five repeats of human FANCE shows that no amino acid is absolutely conserved across repeats (Figure 3B). However, eight specific positions in the FANC repeat show a strong preference for hydrophobic residues with aliphatic side chains, prevalently leucine but also valine, isoleucine and methionine. No significant conservation of any other residue is observed in the repeat. In keeping with the symmetric architecture of the repeat, the eight conserved positions are distributed equally between the two helical segments. Thus, each helix contains two hydrophobic dipeptides spaced one turn apart, located in the third and fourth turn of helix H1 and in the second and third turn of helix H2, respectively. The only non-hydrophobic residues within the eight conserved positions are Thr364 in FANC1, which is buried at the interface between FANC1 and FANC2, and solvent-exposed Glu517, Lys528 and Lys532 in FANC5.

The FANCE structure provides a rationale for the observed pattern of amino acid conservation in the FANC repeat: residues at conserved positions contribute the majority of side chains concurring to form the hydrophobic core of the protein. Conservation of amino acids at specific positions across repeats reflects the structural conservation in the mode of packing of the two antiparallel helices in a repeat. Thus, within each repeat it is possible to identify two triplets of amino acids, at positions 8, 9, 35 and positions 12, 31, 32, respectively, that make equivalent hydrophobic contacts (Figure 3B). The side chain at position 12 in H1 is inserted between the side chains at positions 31, 32 in H2, whilst the side chain at position 35 in H2 interdigitates in a similar fashion with side chains at positions 8, 9 in H1 (Figure 3C). In the case of the longer FANC1 repeat, an additional triplet is formed by Arg371 and Ile372 in helix H1 and Leu383 in helix H2. Hydrophobic positions 13 in helix H1 and 36 in helix H2 are not involved in triplet-like interactions and engage in inter-repeat contacts with the same helix of the proximal, C-terminal repeat. The conserved hydrophobic side chains further intermesh across neighbouring repeats, giving rise to the compact and uninterrupted hydrophobic core of the molecule. The core of the FANCE structure is completed by aromatic or hydrophobic residues, two or three per repeat, situated in the loops between the helices, that bury their side chains inside the protein.

No other position maintains a significant degree of conservation among FANC repeats. However, fourteen residues outside the eight conserved hydrophobic positions are invariant between human and zebrafish FANCE (Figure 2). Of these, eleven are situated within or near intra- and inter-repeat loops, and their conservation is explained by their involvement in direct or water-mediated polar contacts that are important for maintaining the conformation of the polypeptide chain. Thus, Tyr394, Gln416 and Tyr500 provide sidechain-to-mainchain hydrogen bonds that bridge



**Figure 4.** Comparison of HEAT, ARM and FANC repeats. (A) Different conformations adopted by the HEAT, ARM and FANC repeats. The helices in the different repeats are drawn in red, as narrow ribbons, and are labelled as A, B (HEAT); H1, H2 and H3 (ARM); H1 and H2 (FANC). The pronounced kink in helix A of the HEAT repeat becomes two different helices (H1 and H2) in the ARM repeat. The HEAT and ARM repeats shown in the figure are repeats 10 and 2 of PDB entries 1b3u and 3bct, respectively. (B) Structure-based alignment of consensus sequences for HEAT, ARM and FANC repeats. The helices in the alignment are labelled A and B, following the convention adopted for the HEAT repeat. The numbering below the alignment refers to the HEAT and ARM repeat sequences. Hydrophobic residues are in green, acidic residues in red, basic residues in blue and other residues in cyan. Positions in the FANC repeat marked by a small 'h' have a general preference for hydrophobic residues. Four positions where the chemical nature of the amino acid is conserved across the three repeat types are boxed.

between adjacent repeats, whilst serine residues 356, 380 and 486 act as N-terminal helical caps. Three of the four conserved proline residues are part of a short tetrapeptide sequence of consensus PxL(S/Q) that occurs six times in the structure, spanning inter-helical loops or at the extremities of helical segments.

### Similarity to other structures

Our crystallographic analysis reveals that FANCE belongs to the large family of non-globular protein structures assembled by tandem repetition of helical motifs. Fold-recognition analysis in DALI (34) shows that FANCE is structurally similar to HEAT- and ARM-repeat proteins such as importin- $\beta$  (35),  $\beta$ -catenin (36), the Cand1 protein (37) and the PR65/A subunit of protein phosphatase 2A (38). Indeed, comparative structural analysis reveals that the FANC repeat shares four of the seven conserved positions that form the hydrophobic core in the HEAT and ARM repeats (33) (Figure 4). Residues 8 and 12 in helix H1 of the FANC repeat coincide with residues 13 and 17 in helix A of the HEAT repeat (helix H2 of the ARM repeat); residues 31 and 35 in helix H2 of the FANC repeat correspond to residues 28 and 32 in helix B of the HEAT repeat (helix H3 of the ARM repeat).

The FANC repeat differs from well-characterized helical motifs such as the HEAT and ARM repeats in some important ways. The most evident feature is the

absence of conserved positions beyond the selection of hydrophobic amino acids that become buried at the intra- and inter-repeat interfaces. In the helical regions of the repeat, no conservation of structural amino acids, such as proline at position 11 in the first helix of the HEAT and ARM repeats or glycine at position 8 of the ARM repeat, is observed. A lack of conserved prolines or glycines at specific positions means that both helices in the FANCD2 repeat are straight or minimally bent, at variance with the kinked first helix of the HEAT repeat, which becomes two different helical segments in the ARM repeat. Likewise, no conservation of inter-repeat contacts is present, such as the interaction between the aspartic acid in position 19 of the inter-helical loop of the HEAT repeat and arginine at position 25 of the next repeat (33).

### Functional implications for other FA proteins

A virtually complete complement of FA proteins can be traced back in evolution to the last common ancestor of vertebrate organisms, ~450 million years ago (39). Probing further down the evolutionary tree of individual FA proteins reveals that conservation is limited to the FANCD2 and FANCL proteins (40). Thus, the components of the FA core complex seem to have arisen in order to introduce an additional level of complexity to a primitive FA network of DNA repair. No structural information is presently available about the three-dimensional architecture of the FA core complex, or indeed any of its constituent proteins.

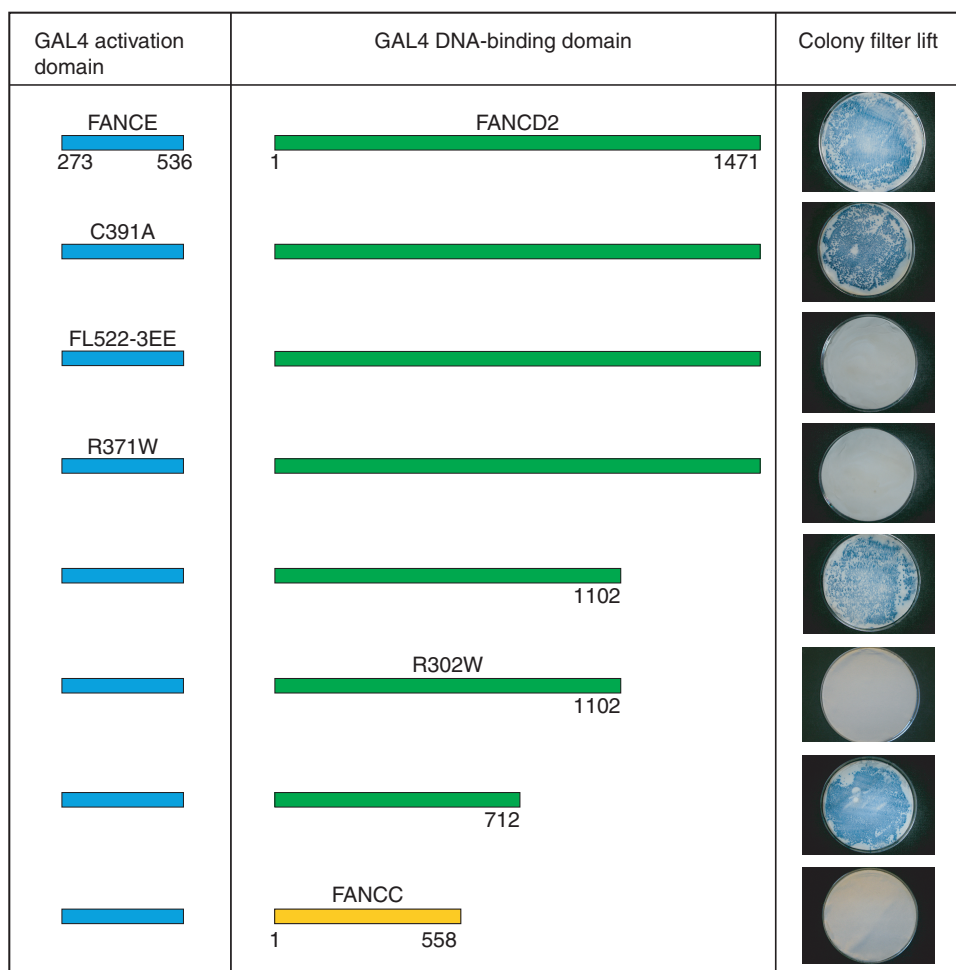
Interestingly, a majority of components of the FA core complex (FANCA, FANCB, FANCC, FANCE, FANCG and FANCF), as well as FANCD2, share an unusually high content of leucine residues (Supplementary Table 1), ranging from 12.1% for FANCB to 19.5% for FANCG, above the average leucine frequency in vertebrate proteins (less than 10%). Furthermore, their predicted content of secondary structure is high, which suggests that the FA proteins of the core complex are independently folded and do not rely on the structural environment provided by the complex for their tertiary structure. An abundance of aliphatic residues, and of leucines in particular, is an essential requirement in order to achieve the kind of packing observed in the hydrophobic core of non-globular, helical repeat proteins. For instance, the leucine content of the protein phosphatase 2A PR65/A subunit and  $\beta$ -catenin is 13.8 and 14.7%, respectively. It is therefore likely that, as shown here for FANCE, other leucine-rich FA proteins adopt a non-globular fold based on arrays of a repeated helical motif. Indeed, seven tetratricopeptide repeats, short helical motifs that are structurally related to the HEAT and ARM repeats, have been convincingly located in the FANCG sequence (41). Based on the FANCE structure, we predict that, in addition to FANCE, leucine-rich FA proteins FANCA, FANCB, FANCC, FANCF and FANCD2 fold partially or entirely in solenoidal structures constituted by arrays of FANCD2-like repeats.

### Interaction with FANCD2

Recruitment of FANCD2 to the FA core complex is essential for its monoubiquitination after DNA damage. Evidence of a direct interaction between FANCD2 and FANCE *in vitro* and *in vivo* has led to the proposal that FANCE represents a physical link between FANCD2 and the FA core complex (21,22). Consequently, we decided to investigate whether the FANCE structure described here included the FANCD2-binding domain. Yeast two-hybrid analysis revealed that a region of FANCE spanning amino acids 273 to 536, coinciding with the recombinant FANCE protein prepared for structural analysis, interacted strongly with full-length FANCD2 (Figure 5). The C391A mutant used for structure determination retained wild-type affinity for FANCD2 in the assay. Thus, the result of the structure-based yeast two-hybrid analysis recapitulates and refines previous findings concerning the FANCE–FANCD2 interaction, by defining the structural domain of FANCE that is sufficient for FANCD2 binding. It has also been reported that the interaction with FANCE was mediated by N-terminal region of FANCD2 (42). In agreement with such findings, we found that two N-terminal truncations FANCD2<sub>1102</sub> and FANCD2<sub>712</sub>, to amino acids 1102 and 712 respectively, maintained FANCE binding in the yeast two-hybrid assay (Figure 5).

Previous experiments have also demonstrated a strong interaction of FANCE with FANCC, another essential component of the FA core complex, and the existence of a ternary FANCC–FANCE–FANCD2 complex has been postulated (19,22). In contrast to the interaction with FANCD2, binding to FANCC was attributed to a central region of FANCE spanning its two NLS sites. In agreement with these observations, the yeast two-hybrid analysis showed that FANCE residues 273 to 536 did not interact with FANCC (Figure 5). Thus, our data support previous reports that different regions of FANCE are responsible for interacting with FANCC and FANCD2.

In order to identify the site of interaction with FANCD2, we performed an *in silico* analysis of the FANCE structure using the optimal docking area (ODA) method (43). Briefly, ODA searches for continuous patches on the surface of a protein with a favourable desolvation energy when buried, as is the case at a protein–protein interface. ODA analysis identified a set of contiguous residues with favourable desolvation energies centred around Phe522, in the exposed intra-helical loop of repeat FANCD5 (Figure 6). Analysis of the ODA epitope reveals that Phe522 is part of a cluster of conserved, hydrophobic residues that includes Met487 and Leu523, which in turn is surrounded by hydrophilic residues Glu448, Ser486 and Lys491. The presence of a hydrophobic centre contoured by polar amino acids poised to make charged interactions is in agreement with known interface features of non-obligate protein complexes. In support of their potential functional role, residues Glu448, Ser486 and Lys491 are conserved between human and zebrafish FANCE.



**Figure 5.** Structure-based yeast two-hybrid analysis of the FANCE–FANCD2 interaction. The FANCE, FANCD2 and FANCC proteins are drawn as light blue, green and orange bars, respectively, of correct relative size. For N- or C-terminally truncated proteins, the number of the first or last residue in the construct are indicated. The result of the assay for each pair of constructs tested is shown in the right-hand column by the  $\beta$ -galactosidase colony filter lift, visualizing activation of the *LacZ* reported gene.

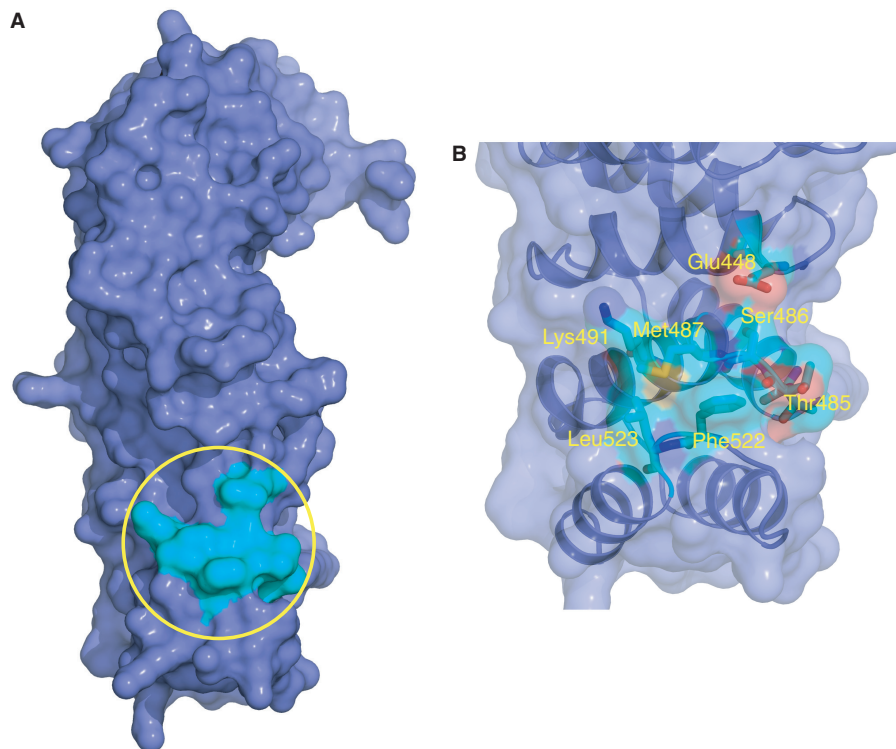
The relevance to FANCD2 binding of the site identified on the FANCE surface was verified experimentally. To this purpose, we designed a double FANCE mutant, where the chemical nature of the solvent-exposed hydrophobic residues Phe522 and Leu523 had been reversed by mutation to the hydrophilic glutamic acid. If the surface patch identified by the ODA analysis is part of the interface in the FANCE–FANCD2 complex, the presence of unpaired charges in the hydrophobic environment of the interface would disrupt or significantly impair the ability of the two proteins to interact. Circular dichroism spectroscopy and gel filtration chromatography experiments confirmed that the mutant FL522-523EE protein retained the native conformation of wild-type FANCE (Supplementary Figure 2). Yeast two-hybrid analysis showed that the simultaneous mutation of Phe522 and Leu523 to glutamic acid abolished the interaction between FANCE and FANCD2. The identification of a single putative protein–protein interaction site implies that the primary function of the C-terminal region of FANCE is to associate with

FANCD2 and does not include multiple, concurrent interactions. The occurrence of conserved Ser486 at the putative protein–protein interface suggests that the FANCE–FANCD2 interaction might be regulated by phosphorylation.

#### Disease-associated mutations

Several missense mutations leading to amino acid substitutions have been identified in the FA genes. However, the molecular basis for the pathogenic effect of these mutations is still largely unknown. Four such mutations have been reported in the human FANCE sequence: S356G, R365K, R371W, A502T (Fanconi Anaemia Mutation Database) (Figure 7A). Mapping the mutations on the structure reveals that three of the four changes (S356G, R365K, R371W) disrupt hydrogen-bond interactions between the side chain of the affected amino acid and neighbouring main-chain atoms, that are important for maintaining the conformation of the polypeptide chain. Thus, the structure predicts that the pathogenic effect of these mutations





**Figure 6.** Protein–protein interaction site in the FANCE structure. **(A)** A bioinformatic analysis of the FANCE structure was performed in order to identify potential protein–protein interaction sites. The solvent-accessible surface of FANCE is shown in blue and the putative protein–protein interaction site is marked by a gold circle. The position of amino acids with favourable desolvation energy is indicated in cyan. **(B)** Close view of the protein–protein interaction site, showing the position of the side chains identified by the bioinformatic analysis. The protein chain is drawn as a ribbon, contained within a semi-transparent representation of the solvent-accessible FANCE surface. The side chains of the relevant amino acids are shown as sticks.

is due to a local destabilization of the ternary structure of the FANCE protein.

The clearest example of such an effect involves Arg371 in the FANCD1 repeat, which is at the centre of a network of direct or water-mediated hydrogen bonds involving the main-chain carbonyl groups of leucines 333, 336 in the loop between helices 3 and 4, and Leu367 of the FANCD1 repeat, as well as the side chain of Asp338 (Figure 7B). Thus, mutation of arginine to tryptophan would cause the loss of several structural hydrogen bonds, leading to disruption of local protein conformation and severe destabilization of the FANCE fold. When tested in the yeast two-hybrid assay, no interaction was observed between the R371W FANCE protein and FANCD2 (Figure 5). Thus, our data provide a structural rationale for the pathological effect of the R371 mutation in FANCE.

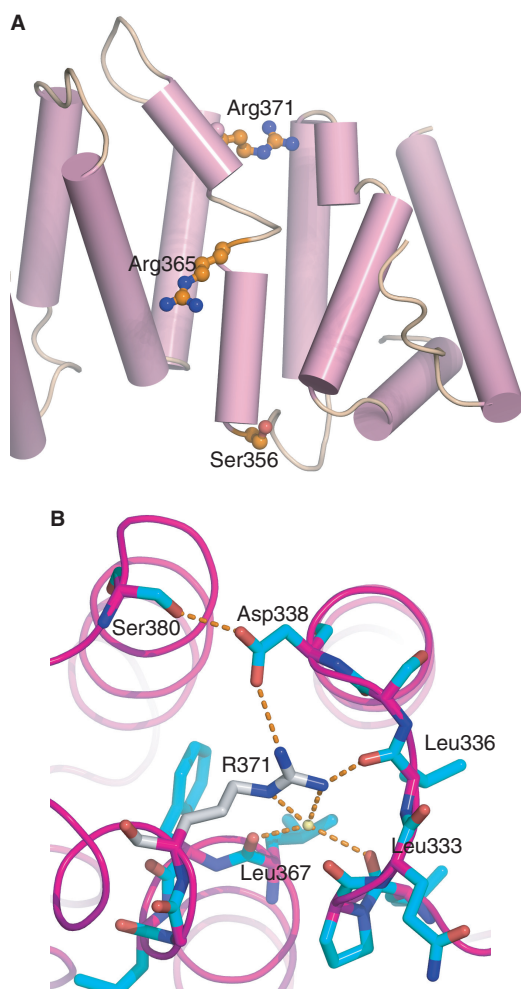
An identical pathogenic mutation, R302W, has also been reported in FANCD2 (Fanconi Anaemia Mutation Database). As the mutation occurs in the part of the FANCD2 sequence responsible for binding FANCE, we determined whether it would disrupt binding to FANCE of the large FANCD2<sub>1102</sub> segment. As observed for the R371W FANCE mutant, no interaction between FANCE and R302W FANCD2<sub>1102</sub> was detected in the yeast two-hybrid assay (Figure 5). It is possible that the R302W mutation in FANCD2 acts in a similar way to the

FANCE R371W mutation, by compromising the stability of the protein structure, and that the observed functional defect is caused by disruption of the FANCE–FANCD2 interaction.

## DISCUSSION

Here we have described the crystal structure of a large, evolutionarily conserved region of FANCE, an essential component of the FA pathway of DNA repair. The structure reveals a non-globular, solenoidal conformation assembled by tandem repeats of a short helical motif. Remarkably, FANCE folding requires only general conservation of amino acid type at a small set of repeat positions, which makes detection of such a motif difficult and explains why its presence had not been noted before. Packing in the hydrophobic core of helical repeat proteins relies on aliphatic side chains of hydrophobic residues, such as leucine. The structure of the FANCE protein therefore provides a rationale for the high leucine content of its sequence and suggests that other FA proteins rich in leucines will adopt a comparable fold.

Multiple arrays of short helical motifs have evolved as a convenient protein architecture in order to mediate protein–protein association. Indeed, helical repeat proteins are usually involved in constitutive or regulated interactions as part of binary or multi-subunit complexes;



**Figure 7.** Disease-associated mutations in the FANCE protein. (A) The position of pathogenic mutations S356G, R365K and R371W identified in the FANCE protein are mapped on the crystal structure. The helices are depicted as pink cylinders, connected by loops drawn as narrow tubes. The side chains of the relevant amino acids are drawn as ball-and-stick models. (B) Close view of the hydrogen-bonding network involving the side chain of Arg371. The disease-associated mutation R371W causes loss of hydrogen-bond interactions with surrounding residues. The protein main chain is drawn as a ribbon, relevant side chains as sticks and a water molecule as a small yellow sphere. Hydrogen bonds are depicted as orange dashed lines.

a cogent example is the regulation of the SCF multi-subunit ubiquitin ligase complex by the HEAT-repeat protein Cand1 (37). A tertiary structure based on arrays of helical repeats, such as that observed in FANCE, would therefore be well suited to the function of the FA proteins, which take part in a complex web of constitutive and regulatory interactions within and outside the FA core complex.

Our data indicate that the region of FANCE defined by the crystallographic analysis is sufficient for interaction with FANCD2. Our results therefore provide insight into the process of FANCD2 association with the FA core complex, a necessary step in the activation of the FA pathway of DNA repair. Taken together with the existing evidence concerning the interaction of FANCE with

FANCC and FANCD2, our finding of a large soluble segment of FANCE able of independent binding to FANCD2, but not to FANCC, suggests a model of FANCE function. In the model, FANCE would be anchored to the FA core complex through a constitutive interaction with FANCC, whereas the C-terminal region of FANCE identified here would be free of contacts with the rest of the core complex and poised to interact with FANCD2 in a regulated way. Furthermore, our bioinformatic and experimental analysis defines an epitope on the FANCE surface that is critical for FANCD2 binding. In principle, the development of small molecules designed to disrupt the FANCE–FANCD2 interface could be useful in tumour therapy based on DNA cross-linking agents.

In summary, the work described here represents an important first step in the structural rationalization of the molecular processes involved in FA. Although the FANCE–FANCD2 interaction is clearly essential, it is one example of a set of concomitant binary interactions that result in FANCD2 monoubiquitination. Future research will investigate the structural basis for the interactions between FANCE, FANCD2 and the other components of the FA core complex that are necessary in order to activate the FA pathway of DNA repair.

**NOTE ADDED IN PROOF** Whilst our manuscript was in production, a structural analysis of the human FANCF protein appeared in print (Kowal, P., Gurtan, A.M., Stuckert, P., D'Andrea, A.D., and Ellenberger, T. (2007) Structural determinants of human FANCF protein that function in the assembly of a DNA damage signalling complex *J. Biol. Chem.*, 282, 2047-2055). The study shows that the C-terminal domain of FANCF folds into a solenoid comprised of repeated helical hairpins, in agreement with our prediction about the widespread adoption of this structural motif among Fanconi Anaemia proteins.

## SUPPLEMENTARY DATA

Supplementary Data is available at NAR Online.

## ACKNOWLEDGEMENTS

We would like to thank Daniel Nietlispach for NMR analysis of wild-type, N<sup>15</sup>-labelled and N<sup>15</sup>- and C<sup>13</sup>-labelled FANCE protein, Gordon Leonard for help with X-ray data collection and Tammy Cheng for the bioinformatic analysis of the FANCE structure. This work was supported by a Wellcome Trust senior research fellowship award to L.P. Atomic coordinates and structure factors have been deposited in the RCS PDB with accession code: 2ILR. Funding to pay the Open Access Publication charge was provided by The Wellcome Trust.

*Conflict of interest statement.* None declared.

## REFERENCES

1. Taniguchi, T. and D'Andrea, A.D. (2006) Molecular pathogenesis of Fanconi anemia: recent progress. *Blood*, **107**, 4223–4233.
2. Bagby, G.C. and Alter, B.P. (2006) Fanconi anemia. *Semin. Hematol.*, **43**, 147–156.
3. Joenje, H. and Patel, K.J. (2001) The emerging genetic and molecular basis of Fanconi anaemia. *Nat. Rev. Genet.*, **2**, 446–457.
4. Levitus, M., Joenje, H. and de Winter, J.P. (2006) The Fanconi anemia pathway of genomic maintenance. *Cell Oncol.*, **28**, 3–29.
5. Mirchandani, K.D. and D'Andrea, A.D. (2006) The Fanconi anemia/BRCA pathway: a coordinator of cross-link repair. *Exp. Cell Res.*, **312**, 2647–2653.
6. Mace, G., Bogliolo, M., Guervilly, J.H., Dugas du Villard, J.A. and Rosselli, F. (2005) 3R coordination by Fanconi anemia proteins. *Biochimie*, **87**, 647–658.
7. Niedernhofer, L.J., Lalai, A.S. and Hoeijmakers, J.H. (2005) Fanconi anemia (cross)linked to DNA repair. *Cell*, **123**, 1191–1198.
8. 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. *et al.* (2003) A novel ubiquitin ligase is deficient in Fanconi anemia. *Nat. Genet.*, **35**, 165–170.
9. Meetei, A.R., Yan, Z. and Wang, W. (2004) FANCL replaces BRCA1 as the likely ubiquitin ligase responsible for FANCD2 monoubiquitination. *Cell Cycle*, **3**, 179–181.
10. Levitus, M., Waisfisz, Q., Godthelp, B.C., de Vries, Y., Hussain, S., Wiegant, W.W., Elghalbzouri-Maghrani, E., Steltenpool, J., Rooimans, M.A. *et al.* (2005) The DNA helicase BRIP1 is defective in Fanconi anemia complementation group. *J. Nat. Genet.*, **37**, 934–935.
11. Litman, R., Peng, M., Jin, Z., Zhang, F., Zhang, J., Powell, S., Andreassen, P.R. and Cantor, S.B. (2005) BACH1 is critical for homologous recombination and appears to be the Fanconi anemia gene product FANCF. *Cancer Cell*, **8**, 255–265.
12. Meetei, A.R., Medhurst, A.L., Ling, C., Xue, Y., Singh, T.R., Bier, P., Steltenpool, J., Stone, S., Dokal, I. *et al.* (2005) A human ortholog of archaeal DNA repair protein Hef is defective in Fanconi anemia complementation group M. *Nat. Genet.*, **37**, 958–963.
13. Mosedale, G., Niedzwiedz, W., Alpi, A., Perrina, F., Pereira-Leal, J.B., Johnson, M., Langevin, F., Pace, P. and Patel, K.J. (2005) The vertebrate Hef ortholog is a component of the Fanconi anemia tumor-suppressor pathway. *Nat. Struct. Mol. Biol.*, **12**, 763–771.
14. Medhurst, A.L., Huber, P.A., Waisfisz, Q., de Winter, J.P. and Mathew, C.G. (2001) Direct interactions of the five known Fanconi anaemia proteins suggest a common functional pathway. *Hum. Mol. Genet.*, **10**, 423–429.
15. 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.
16. Taniguchi, T., Garcia-Higuera, I., Andreassen, P.R., Gregory, R.C., Grompe, M. and D'Andrea, A.D. (2002) S-phase-specific interaction of the Fanconi anemia protein, FANCD2, with BRCA1 and RAD51. *Blood*, **100**, 2414–2420.
17. Hussain, S., Wilson, J.B., Medhurst, A.L., Hejna, J., Witt, E., Ananth, S., Davies, A., Masson, J.Y., Moses, R. *et al.* (2004) Direct interaction of FANCD2 with BRCA2 in DNA damage response pathways. *Hum. Mol. Genet.*, **13**, 1241–1248.
18. Medhurst, A.L., Laghmani el, H., Steltenpool, J., Ferrer, M., Fontaine, C., de Groot, J., Rooimans, M.A., Scheper, R.J., Meetei, A.R. *et al.* (2006) Evidence for subcomplexes in the Fanconi anemia pathway. *Blood*, **108**, 2072–2080.
19. 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.
20. Matsushita, N., Kitao, H., Ishiai, M., Nagashima, N., Hirano, S., Okawa, K., Ohta, T., Yu, D.S., McHugh, P.J. *et al.* (2005) A FANCD2-monoubiquitin fusion reveals hidden functions of Fanconi anemia core complex in DNA repair. *Mol. Cell*, **19**, 841–847.
21. Pace, P., Johnson, M., Tan, W.M., Mosedale, G., Sng, C., Hoatlin, M., de Winter, J., Joenje, H., Gergely, F. *et al.* (2002) FANCE: the link between Fanconi anaemia complex assembly and activity. *EMBO J.*, **21**, 3414–3423.
22. Leveille, F., Ferrer, M., Medhurst, A.L., Laghmani el, H., Rooimans, M.A., Bier, P., Steltenpool, J., Titus, T.A., Postlethwait, J.H. *et al.* (2006) The nuclear accumulation of the Fanconi anemia protein FANCE depends on FANCC. *DNA Repair (Amst.)*, **5**, 556–565.
23. Wang, X., Andreassen, P.R. and D'Andrea, A.D. (2004) Functional interaction of monoubiquitinated FANCD2 and BRCA2/FANCD1 in chromatin. *Mol. Cell. Biol.*, **24**, 5850–5862.
24. Leslie, A.G.W. (1992) Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography, No. 26.
25. Evans, P.R. (1993) Proceedings of the CCP4 Study Weekend. Data Collection and Processing, Daresbury Laboratories, Warrington, UK, pp. 114–122.
26. Weeks, C. and Miller, R. (1999) The design and implementation of SnB version 2.0. *J. Appl. Cryst.*, **32**, 120–124.
27. La Fortelle, E.D. and Bricogne, G. (1997) Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction. *Methods Enzymol.*, **276**, 472–494.
28. Morris, R.J., Perrakis, A. and Lamzin, V.S. (2003) ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol.*, **374**, 229–244.
29. Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.*, **53**, 240–255.
30. Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
31. Hussain, S., Witt, E., Huber, P.A., Medhurst, A.L., Ashworth, A. and Mathew, C.G. (2003) Direct interaction of the Fanconi anaemia protein FANCG with BRCA2/FANCD1. *Hum. Mol. Genet.*, **12**, 2503–2510.
32. Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F. and Jones, D.T. (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.*, **337**, 635–645.
33. Andrade, M.A., Petosa, C., O'Donoghue, S.I., Muller, C.W. and Bork, P. (2001) Comparison of ARM and HEAT protein repeats. *J. Mol. Biol.*, **309**, 1–18.
34. Holm, L. and Sander, C. (1995) Dali: a network tool for protein structure comparison. *Trends Biochem. Sci.*, **20**, 478–480.
35. Cingolani, G., Petosa, C., Weis, K. and Muller, C.W. (1999) Structure of importin-beta bound to the IBB domain of importin-alpha. *Nature*, **399**, 221–229.
36. Huber, A.H., Nelson, W.J. and Weis, W.I. (1997) Three-dimensional structure of the armadillo repeat region of beta-catenin. *Cell*, **90**, 871–882.
37. Goldenberg, S.J., Cascio, T.C., Shumway, S.D., Garbutt, K.C., Liu, J., Xiong, Y. and Zheng, N. (2004) Structure of the Candi-Cull1-Roc1 complex reveals regulatory mechanisms for the assembly of the multisubunit cullin-dependent ubiquitin ligases. *Cell*, **119**, 517–528.
38. Groves, M.R., Hanlon, N., Turowski, P., Hemmings, B.A. and Barford, D. (1999) The structure of the protein phosphatase 2A PR65/A subunit reveals the conformation of its 15 tandemly repeated HEAT motifs. *Cell*, **96**, 99–110.
39. Titus, T.A., Selvig, D.R., Qin, B., Wilson, C., Starks, A.M., Roe, B.A. and Postlethwait, J.H. (2006) The Fanconi anemia gene network is conserved from zebrafish to human. *Gene*, **371**, 211–223.
40. Marek, L.R. and Bale, A.E. (2006) Drosophila homologs of FANCD2 and FANCL function in DNA repair. *DNA Repair (Amst.)*, **5**, 1317–1326.
41. Blom, E., van de Vrugt, H.J., de Vries, Y., de Winter, J.P., Arwert, F. and Joenje, H. (2004) Multiple TPR motifs characterize the Fanconi anemia FANCG protein. *DNA Repair (Amst.)*, **3**, 77–84.
42. Gordon, S.M. and Buchwald, M. (2003) Fanconi Anemia protein complex: mapping protein interactions in the yeast 2- and 3hybrid systems. *Blood*, **102**, 136–141.
43. Fernandez-Recio, J., Totrov, M., Skorodumov, C. and Abagyan, R. (2005) Optimal docking area: a new method for predicting protein-protein interaction sites. *Proteins*, **58**, 134–143.