

# Structure of the Fanconi anaemia monoubiquitin ligase complex

<https://doi.org/10.1038/s41586-019-1703-4>

Received: 26 April 2019

Accepted: 18 September 2019

Published online: 30 October 2019

Shabih Shakeel<sup>1,5</sup>, Eeson Rajendra<sup>1,5</sup>, Pablo Alcón<sup>1</sup>, Francis O'Reilly<sup>2</sup>, Dror S. Chorev<sup>3</sup>, Sarah Maslen<sup>1</sup>, Gianluca Degliesposti<sup>1</sup>, Christopher J. Russo<sup>1</sup>, Shaoda He<sup>1</sup>, Chris H. Hill<sup>1</sup>, J. Mark Skehel<sup>1</sup>, Sjors H. W. Scheres<sup>1</sup>, Ketan J. Patel<sup>1</sup>, Juri Rappsilber<sup>2,4</sup>, Carol V. Robinson<sup>3</sup> & Lori A. Passmore<sup>1\*</sup>

The Fanconi anaemia (FA) pathway repairs DNA damage caused by endogenous and chemotherapy-induced DNA crosslinks, and responds to replication stress<sup>1,2</sup>. Genetic inactivation of this pathway by mutation of genes encoding FA complementation group (FANC) proteins impairs development, prevents blood production and promotes cancer<sup>1,3</sup>. The key molecular step in the FA pathway is the monoubiquitination of a pseudosymmetric heterodimer of FANCD2–FANCI<sup>4,5</sup> by the FA core complex—a megadalton multiprotein E3 ubiquitin ligase<sup>6,7</sup>. Monoubiquitinated FANCD2 then recruits additional protein factors to remove the DNA crosslink or to stabilize the stalled replication fork. A molecular structure of the FA core complex would explain how it acts to maintain genome stability. Here we reconstituted an active, recombinant FA core complex, and used cryo-electron microscopy and mass spectrometry to determine its structure. The FA core complex comprises two central dimers of the FANCB and FA-associated protein of 100 kDa (FAAP100) subunits, flanked by two copies of the RING finger subunit, FANCL. These two heterotrimers act as a scaffold to assemble the remaining five subunits, resulting in an extended asymmetric structure. Destabilization of the scaffold would disrupt the entire complex, resulting in a non-functional FA pathway. Thus, the structure provides a mechanistic basis for the low numbers of patients with mutations in FANCB, FANCL and FAAP100. Despite a lack of sequence homology, FANCB and FAAP100 adopt similar structures. The two FANCL subunits are in different conformations at opposite ends of the complex, suggesting that each FANCL has a distinct role. This structural and functional asymmetry of dimeric RING finger domains may be a general feature of E3 ligases. The cryo-electron microscopy structure of the FA core complex provides a foundation for a detailed understanding of its E3 ubiquitin ligase activity and DNA interstrand crosslink repair.

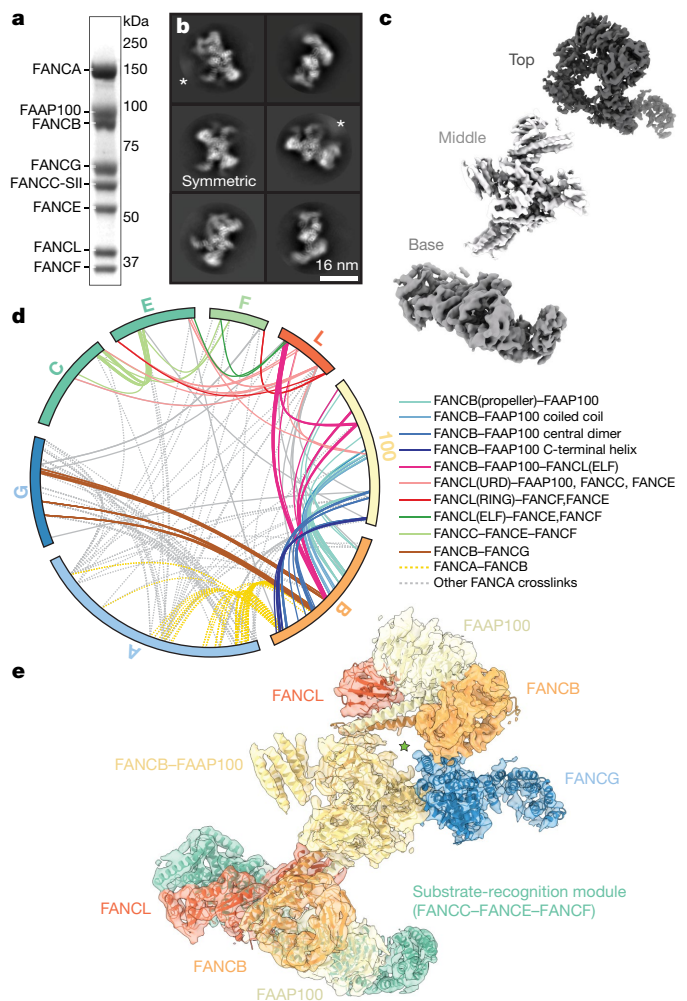
The FA core complex is composed of eight stably-associated subunits: FANCA, FANCB, FANCC, FANCE, FANCF, FANCG, FANCL and FAAP100<sup>6</sup>. FANCL contains a RING-finger domain which acts as the E3 ubiquitin ligase. It associates with FANCB and FAAP100 to form a catalytic module<sup>6,8</sup>; a low-resolution negative-stain electron microscopy (EM) study suggested that, in the absence of the other subunits, this is a symmetric dimer of FANCB–FANCL–FAAP100 heterotrimers<sup>9</sup>. FANCA and FANCG are proposed to act as a chromatin-targeting module, whereas FANCC, FANCE and FANCF form a substrate-recognition module<sup>6,8,10,11</sup>. Despite the central role of the FA core complex in DNA repair, we lack a molecular understanding of how FANCL incorporates into the complex to perform site-specific monoubiquitination of the FANCD2–FANCI substrate and how mutation disrupts the function of the complex<sup>12</sup>.

To determine the structure of the FA core complex, we overexpressed all eight subunits from *Gallus gallus* (chicken) on a single baculovirus

in insect cells, which enabled us to purify an intact, recombinant complex (Fig. 1a). The purified complex specifically monoubiquitinated FANCD2 but not FANCI in vitro (Extended Data Fig. 1), similar to the native chicken complex<sup>6</sup>.

To investigate the molecular basis of subunit association, we imaged this recombinant FA core complex using cryo-EM. This revealed an elongated structure, about 25 nm in length (Fig. 1b, Extended Data Fig. 2a). We determined a 3D reconstruction of the FA core complex at an overall resolution of 4.2 Å (Extended Data Fig. 2b–e, Extended Data Table 1). The peripheral regions were less well resolved than the central core, possibly owing to conformational flexibility. In agreement with conformational heterogeneity in the complex, we detected structural variations using multi-body refinement<sup>13</sup>, including a continuum of conformational movement of the top and base regions (Extended Data Fig. 2f, Supplementary

<sup>1</sup>MRC Laboratory of Molecular Biology, Cambridge, UK. <sup>2</sup>Bioanalytics, Institute of Biotechnology, Technische Universität Berlin, Berlin, Germany. <sup>3</sup>Physical and Theoretical Chemistry Laboratory, University of Oxford, Oxford, UK. <sup>4</sup>Wellcome Centre for Cell Biology, University of Edinburgh, Edinburgh, UK. <sup>5</sup>These authors contributed equally: Shabih Shakeel, Eeson Rajendra. \*e-mail: [passmore@mrclmb.cam.ac.uk](mailto:passmore@mrclmb.cam.ac.uk)



**Fig. 1 | Overall structure of the FA core complex.** **a**, SDS-PAGE analysis of purified FA core complex with subunits and molecular weight markers indicated. FANCC carries a 2× Strep II tag on its C terminus (FANCC-SII). This purification was repeated more than three times with similar results. For gel source data, see Supplementary Fig. 1. **b**, Selected 2D reference-free class averages of the FA core complex. One class appears to be symmetric (labelled). Asterisks mark disordered density extending from the side of the complex that does not align well. **c**, Focused classification and refinement on the top and base regions, and multibody refinement on the middle region resulted in three independent cryo-EM maps that are shown separately, in three different shades of grey. **d**, Crosslinking mass spectrometry revealed 834 crosslinks (1% false discovery rate) between residues that are in close proximity. Intermolecular crosslinks are shown, coloured by interacting regions. **e**, Model of FA core complex (cartoon representation) fitted into the EM density (isosurface representation with transparency). Map and model are coloured by assigned subunits. The green star marks a channel with diameter approximately 23 Å.

Videos 1, 2). Particle subtraction followed by focused classification and refinement<sup>14</sup> generated separate, improved reconstructions of the top and base regions (Fig. 1c, Extended Data Fig. 2g, h and Supplementary Video 3).

Our map of the complete FA core complex was of sufficient resolution to dock existing structures and to resolve secondary structure elements (Extended Data Fig. 2i). We fit two previously determined high-resolution structures (FANCL<sup>15</sup> and part of FANCF<sup>16</sup>) into the map, accounting for about 12% of the entire mass of the complex. Most subunits do not have substantial homology to proteins of known structure (Extended Data Table 2), so we modelled  $\alpha$ -helices and  $\beta$ -strands into the remainder of the map. To determine which subunits these  $\alpha$ -helices and  $\beta$ -strands belong to, we required additional data.

Next, we purified subcomplexes and imaged them using cryo-EM. By comparing the 2D class averages of subcomplexes to the complete FA core complex, we identified regions corresponding to specific subunits (Extended Data Fig. 3a–c). Removal of FANCA did not substantially change the class averages. This suggested that FANCA may be conformationally heterogeneous and blurred out in reconstructions, or it may dissociate or denature during cryo-EM specimen preparation. The base was absent in a complex of FANCA, FANCG, FANCB, FANCL and FAAP100, suggesting that the base probably contains the substrate-recognition module (FANCC, FANCE and FANCF). The partially disordered arm that extends from the central part of the complex is probably FANCG because this density is lost when FANCG is removed from a FANCB–FANCL–FAAP100 complex. Finally, 2D classes of the catalytic module (FANCB, FANCL and FAAP100) resemble the middle region of the FA core complex.

We also studied the structure of the FA core complex using non-covalent native mass spectrometry (Extended Data Fig. 3d, e). During ionization, the FANCE subunit tends to dissociate and subcomplexes are formed, providing information on subunit stoichiometry and protein–protein interactions. The largest complex (808 kDa) that we detected in native mass spectrometry contains seven of the eight different subunits, including two copies of FANCB, FANCL and FAAP100, and a single copy of each of the remaining subunits. This agrees with the subunit stoichiometries of a purified native FA core complex<sup>6</sup>. FANCB, FANCL and FAAP100 were present in most of the subcomplexes identified by native mass spectrometry (Extended Data Fig. 3d), suggesting that they form a central core.

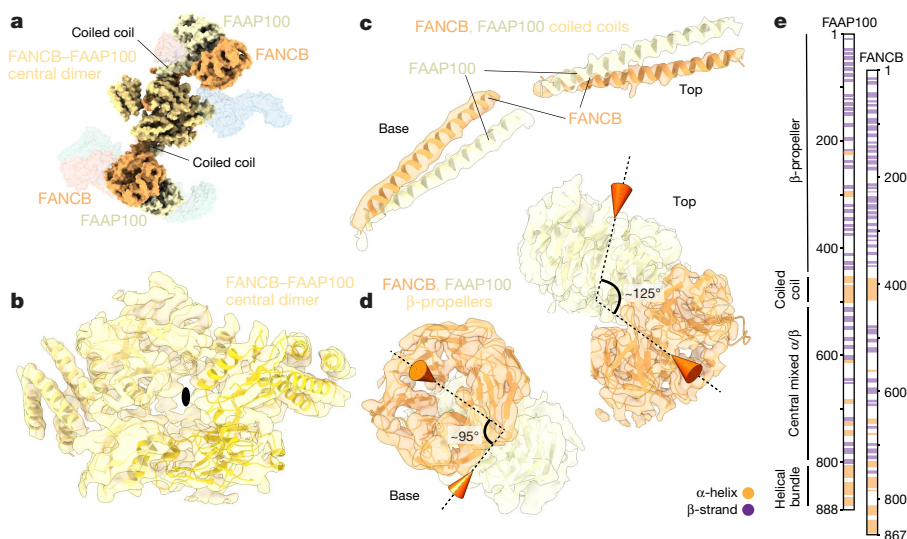
To identify residues in close proximity, we performed crosslinking mass spectrometry (Fig. 1d, Extended Data Fig. 4). This revealed 834 inter- and intramolecular crosslinks, with 40% of these located in FANCB, FANCL and FAAP100. This is consistent with these three subunits forming an intimate complex. By combining the crosslinking mass spectrometry data showing which residues are in close proximity with the subunit assignment from subcomplexes, the subunit stoichiometry from native mass spectrometry, homology modelling and secondary structure predictions, we generated models for all FA core complex subunits except FANCA (Fig. 1e, Extended Data Fig. 5a–d, Extended Data Table 1, 2, Supplementary Video 4, Methods).

A dimer of FANCB–FAAP100 heterodimers is located in the middle region of the structure (Fig. 2). Two pairs of long  $\alpha$ -helices (coiled coils) connect central  $\beta$ -strands and helical bundles with peripheral densities (Fig. 2b, c). Crosslinking mass spectrometry and modelling showed that each coiled coil is probably composed of  $\alpha$ -helices from FANCB and FAAP100 (Extended Data Fig. 5e, f). At the peripheral ends of the coiled coils, we identified pairs of  $\beta$ -propellers, each containing a  $\beta$ -propeller from the N-terminal region of FANCB or FAAP100 (Fig. 2d). We could differentiate the two  $\beta$ -propellers on the basis of crosslinks: the FANCB  $\beta$ -propeller is near the coiled coil, whereas the FAAP100  $\beta$ -propeller is close to the ELF domain of FANCL. Unexpectedly, despite the lack of sequence homology, FANCB and FAAP100 share markedly similar overall structures and domain organizations (Fig. 2e).

A homology model of FANCL, including the ELF, URD and RING finger domains<sup>15</sup>, fits into the base of the complex (Fig. 3a, b) but the relative orientations of the individual domains are different compared with the crystal structure (Extended Data Fig. 4b). By contrast, only the ELF domain could be placed into the second copy of FANCL at the top of the complex (Fig. 3c). Hydrogen–deuterium exchange mass spectrometry (HDX-MS) confirmed that FANCL interacts with the FANCB–FAAP100 coiled coil (Extended Data Fig. 6).

FANCG contains tetratricopeptide repeats (TPRs) (Fig. 3d, e), crosslinks to the central FANCB–FAAP100 dimer (Fig. 1d), and is required for FANCA association with the FA core complex (Extended Data Fig. 3b). Of note, there is a channel between FANCG and the catalytic module (Fig. 1e). FANCA, which is absent in the maps, is probably peripheral to FANCG, and possibly located in the blurred density visible in 2D class averages (indicated by asterisks in Fig. 1b; Extended Data Fig. 5d).



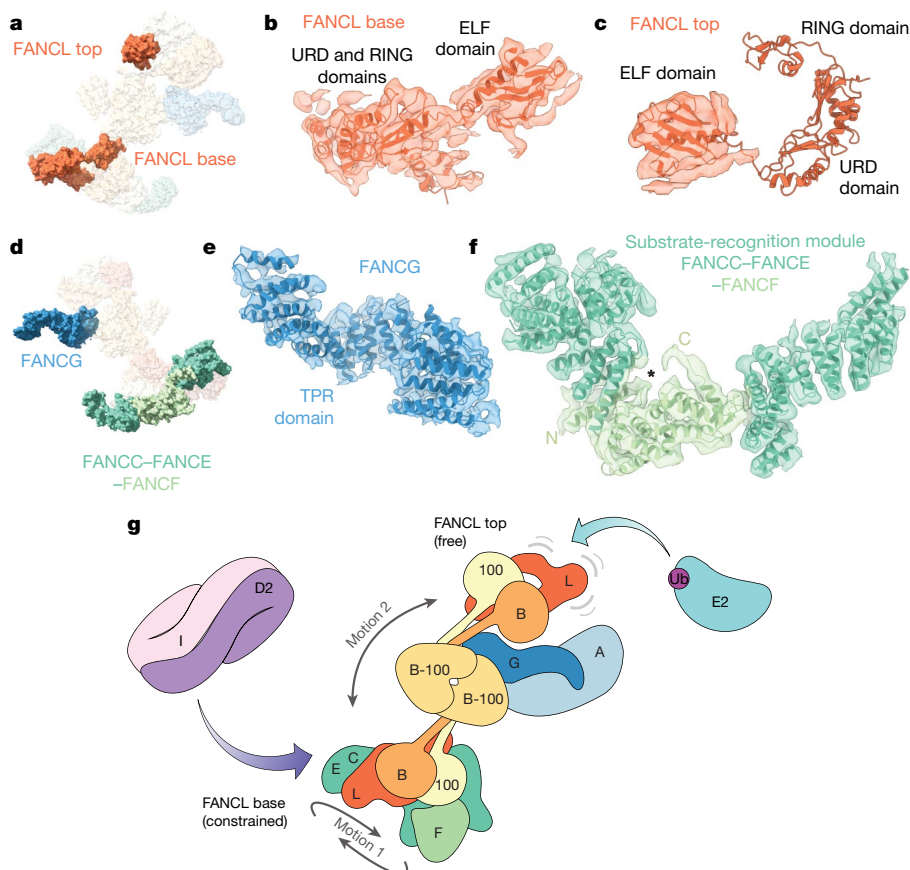


**Fig. 2 | The molecular scaffold of the FA core complex includes a dimer of FANCB-FAAP100 heterodimers.** **a**, Surface representation of the FA core complex model, highlighting FANCB and FAAP100. FANCB, orange; FAAP100, yellow; regions where we are unable to distinguish FANCB and FAAP100, yellow-orange. **b-d**, Models of FANCB and FAAP100 subunits in cartoon representation placed into the cryo-EM map. In **b**, a black oval marks the pseudo two-fold symmetry axis. There are substantial differences between the two symmetry-related copies, which are shown in different shades of yellow in the model. In **d**, cones indicate the orientations of the central pores of the  $\beta$ -propellers and the angles between them are shown. **e**, Proposed domain organization of FANCB and FAAP100, showing their structural similarity.

The large number of crosslinks between FANCA and FANCG are consistent with their proximity (Fig. 1d).

The substrate-recognition module (FANCC, FANCE and FANCF), located in the base, comprises an arc of  $\alpha$ -helices (Fig. 3f). FANCF occupies a central position within the arc. Crosslinking mass spectrometry showed that the FANCL RING finger and ELF domain contact FANCE and FANCF, the FANCL URD domain contacts FANCC and FANCE, and the FANCC C-terminal region and FANCE are in close proximity. Native mass spectrometry also showed direct contact between FANCE and FANCF (Extended Data Fig. 3d).

We also determined a structure of a subcomplex, present at a lower abundance in our sample, at an overall resolution of 4.6 Å (Extended Data Fig. 7a-d). This subcomplex was symmetric but the map did not improve on application of C2 symmetry. We therefore implemented a local symmetry algorithm in Relion for averaging the two halves of the subcomplex map (Methods, Extended Data Fig. 7e, f). This symmetric structure revealed an assembly comprising two copies of each of FANCB, FANCL, FAAP100 and FANCG (Extended Data Fig. 8a-c, Supplementary Video 5). It is unclear whether this subcomplex has a functional role in vivo or whether it is an assembly intermediate. In both copies



**Fig. 3 | Asymmetric dimerization in the FA core complex.** **a-c**, Models of FANCL in the FA core complex. **a**, Surface representation of the FA core complex model, highlighting the two copies of FANCL. **b, c**, Models of FANCL<sub>base</sub> and FANCL<sub>top</sub> subunits in cartoon representation are shown fitted in the cryo-EM map. Density for the URD and RING domains is not well defined in the top copy (**c**). **d-f**, Models of FANCG and the substrate-recognition module. **d**, Surface representation of the FA core complex model highlighting FANCG and the substrate-recognition module (FANCC-FANCE-FANCF). **e, f**, Models of FANCG TPR domain (**e**) and the substrate-recognition module (**f**) are shown fitted into the cryo-EM map. The crystal structure of FANCF could be assigned and the N and C termini of the model are indicated. The first helix (N to \*) was not present in the crystal structure. Since all three subunits of the substrate-recognition module are substantially helical, it was not possible to assign the remaining helices of the base to individual subunits. **g**, Model for monoubiquitination of FANCD2 by the FA core complex. The major motions detected in multi-body refinement are indicated with grey arrows.

of FANCL, the URD and RING finger domains have weak density or are not visible, similar to the top FANCL (FANCL<sub>top</sub>) in the full, asymmetric FA core complex.

Comparison of subcomplex and FA core complex structures suggests that the substrate-recognition module alters the relative orientations of the  $\beta$ -propellers and coiled coil in the base (Fig. 2d, Supplementary Video 6). This disrupts the symmetry of the catalytic module in the complete FA core complex, provides a binding site in the base for the URD and RING finger domains of FANCL and probably disrupts docking of a second (symmetric) copy of FANCG onto the middle region owing to steric clashes (Extended Data Fig. 8d). These structural alterations may be transmitted to the top region to prevent the binding of a second substrate-recognition module, consistent with allosteric coupling proposed previously<sup>9</sup>. In agreement with this, a purified substrate-recognition module did not readily associate with the symmetric subcomplex in vitro to form the asymmetric FA core complex (Extended Data Fig. 8e, f), suggesting that in vivo assembly is required.

Dimerization is required for the activity of other E3 ligases, including the Rad18, RNF8 and CHIP homodimers<sup>17–19</sup> and the BRCA1–BARD1 and Ring1b–Bmi1 heterodimers<sup>20,21</sup>. These contain two RING finger–U-box domains arranged in an asymmetric manner with only a single functional E2 binding site. Notably, the activity of multi-subunit cullin–RING finger E3 ligases is stimulated by dimerization<sup>22,23</sup>. Thus, structural and functional asymmetry appear to be a common feature of E3 ligases but the spatial separation of the FANCL RING fingers in the FA core complex is unusual (Extended Data Fig. 9a, b). Like other dimeric RING finger E3s, one RING finger (in the FANCL subunit at the base (FANCL<sub>base</sub>)) of the FA core complex may have a structural role in promoting substrate binding along with FANCE<sup>10,11,24,25</sup>, whereas the other (FANCL<sub>top</sub>) may be the active E3 that promotes ubiquitin transfer<sup>26</sup> to FANCD2 (Fig. 3g) within the structural state we observe. The FANCD2–FANCI substrate is also an asymmetric dimer. Thus, each of the two FANCL RING finger domains could monoubiquitinate one of the substrate proteins. Nevertheless, this purified complex does not monoubiquitinate FANCI, so an additional activation step might be required. Substrate binding is not required to activate the E3 ligase activity (Extended Data Fig. 9c, d).

The majority of FA-complex mutations detected in patients with FA result in protein truncation and are found in the structural periphery of the FA core complex<sup>27</sup> (Extended Data Fig. 9e). Residual monoubiquitin ligase activity is still present after deletion of peripheral subunits in cells<sup>6,8</sup> and in vitro (Extended Data Fig. 9f), indicating a partially functioning core complex. By contrast, deletion of FANCB, FANCL or FAAP100, which comprise the catalytic module and the structural scaffold for the complex, eliminates this residual activity<sup>6,8</sup>. Patients with FA who carry mutations in FANCB or FANCL are severely afflicted, and we predict that FANCB missense mutations (L43S, P230S, L329P and L676P) disrupt stability of the catalytic module (indicated by asterisks in Extended Data Fig. 4a). Together, these data provide genetic and clinical support that mutations in the catalytic module result in disruption of the FA core complex structure. By contrast, mutations in the periphery do not prevent complex formation and may be better tolerated.

In summary, our data provide a structural model for the FA core complex, enabling an interpretation of the molecular pathophysiology of FA. The reconstituted system we describe will also enable further mechanistic questions to be addressed, including precisely how this large complex functions as a DNA damage-inducible monoubiquitin ligase.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-019-1703-4>.

1. Crossan, G. P. & Patel, K. J. The Fanconi anaemia pathway orchestrates incisions at sites of crosslinked DNA. *J. Pathol.* **226**, 326–337 (2012).
2. Schlacher, K., Wu, H. & Jasin, M. A distinct replication fork protection pathway connects Fanconi anemia tumor suppressors to RAD51–BRCA1/2. *Cancer Cell* **22**, 106–116 (2012).
3. Nalepa, G. & Clapp, D. W. Fanconi anaemia and cancer: an intricate relationship. *Nat. Rev. Cancer* **18**, 168–185 (2018).
4. Knipscheer, P. et al. The Fanconi anemia pathway promotes replication-dependent DNA interstrand cross-link repair. *Science* **326**, 1698–1701 (2009).
5. Smogorzewska, A. et al. Identification of the FANCI protein, a monoubiquitinated FANCD2 paralog required for DNA repair. *Cell* **129**, 289–301 (2007).
6. Rajendra, E. et al. The genetic and biochemical basis of FANCD2 monoubiquitination. *Mol. Cell* **54**, 858–869 (2014).
7. Garcia-Higuera, I. et al. Interaction of the Fanconi anemia proteins and BRCA1 in a common pathway. *Mol. Cell* **7**, 249–262 (2001).
8. Huang, Y. et al. Modularized functions of the Fanconi anemia core complex. *Cell Rep.* **7**, 1849–1857 (2014).
9. Swuec, P. et al. The FA core complex contains a homo-dimeric catalytic module for the symmetric mono-ubiquitination of FANCI–FANCD2. *Cell Rep.* **18**, 611–623 (2017).
10. Pace, P. et al. FANCE: the link between Fanconi anaemia complex assembly and activity. *EMBO J.* **21**, 3414–3423 (2002).
11. van Twest, S. et al. Mechanism of ubiquitination and deubiquitination in the Fanconi anemia pathway. *Mol. Cell* **65**, 247–259 (2017).
12. Walden, H. & Deans, A. J. The Fanconi anemia DNA repair pathway: structural and functional insights into a complex disorder. *Annu. Rev. Biophys.* **43**, 257–278 (2014).
13. Nakane, T., Kimanius, D., Lindahl, E. & Scheres, S. H. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *eLife* **7**, e36861 (2018).
14. Bai, X. C., Rajendra, E., Yang, G., Shi, Y. & Scheres, S. H. Sampling the conformational space of the catalytic subunit of human  $\gamma$ -secretase. *eLife* **4**, e11182 (2015).
15. Cole, A. R., Lewis, L. P. C. & Walden, H. The structure of the catalytic subunit FANCL of the Fanconi anemia core complex. *Nat. Struct. Mol. Biol.* **17**, 294–298 (2010).
16. Kowal, P., Gurtan, A. M., Stuckert, P., D'Andrea, A. D. & Ellenberger, T. Structural determinants of human FANCF protein that function in the assembly of a DNA damage signaling complex. *J. Biol. Chem.* **282**, 2047–2055 (2007).
17. Huang, A. et al. Symmetry and asymmetry of the RING–RING dimer of Rad18. *J. Mol. Biol.* **410**, 424–435 (2011).
18. Mattioli, F. et al. RNF168 ubiquitinates K13–15 on H2A/H2AX to drive DNA damage signaling. *Cell* **150**, 1182–1195 (2012).
19. Zhang, M. et al. Chaperoned ubiquitylation—crystal structures of the CHIP U box E3 ubiquitin ligase and a CHIP–Ubc13–Uev1a complex. *Mol. Cell* **20**, 525–538 (2005).
20. Brzovic, P. S., Rajagopal, P., Hoyt, D. W., King, M. C. & Kleit, R. E. Structure of a BRCA1–BARD1 heterodimeric RING–RING complex. *Nat. Struct. Biol.* **8**, 833–837 (2001).
21. Buchwald, G. et al. Structure and E3-ligase activity of the Ring–Ring complex of polycomb proteins Bmi1 and Ring1b. *EMBO J.* **25**, 2465–2474 (2006).
22. Passmore, L. A. et al. Structural analysis of the anaphase-promoting complex reveals multiple active sites and insights into polyubiquitylation. *Mol. Cell* **20**, 855–866 (2005).
23. Tang, X. et al. Suprafacial orientation of the SCFCdc4 dimer accommodates multiple geometries for substrate ubiquitination. *Cell* **129**, 1165–1176 (2007).
24. Gordon, S. M., Alon, N. & Buchwald, M. 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 (2005).
25. Polito, D. et al. The carboxyl terminus of FANCE recruits FANCD2 to the Fanconi Anemia (FA) E3 ligase complex to promote the FA DNA repair pathway. *J. Biol. Chem.* **289**, 7003–7010 (2014).
26. Zimmerman, E. S., Schulman, B. A. & Zheng, N. Structural assembly of cullin–RING ubiquitin ligase complexes. *Curr. Opin. Struct. Biol.* **20**, 714–721 (2010).
27. Neveling, K., Endt, D., Hoehn, H. & Schindler, D. Genotype–phenotype correlations in Fanconi anemia. *Mutat. Res.* **668**, 73–91 (2009).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019

# Article

## Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

### Cloning, expression and purification

cDNAs encoding full length *G. gallus* (Gg) FANCA, FANCB, FANCC, FANCE, FANCF, FANCG, FANCL and FAAP100 were synthesized (GeneArt). FANCC contained a C-terminal extension with a 3C protease site and double Strep II tag. For protein expression, all genes were cloned into the MultiBac expression system and constructs were generated as previously described<sup>28</sup>. In brief, to make gene expression cassettes, FANCA and FANCG were subcloned via BamHI and XbaI into pACEBac1; FANCC, FANCE, FANCF and FAAP100 were subcloned via XhoI and KpnI into pIDS; and FANCB and FANCL were subcloned via BamHI and XbaI into pIDC. Gene cassettes were then sequentially subcloned as BstXI–I-CeuI or BstXI–PI-SceI fragments into I-CeuI or PI-SceI sites to generate pACEBac1-FANCA-FANCG, pIDS-FANCC-3C-2×StrepII-FANCE-FANCF, and pIDS-FAAP100-FANCB-FANCL. The spectinomycin antibiotic resistance cassette of pIDS-FAAP100-FANCB-FANCL was substituted with the kanamycin antibiotic resistance cassette of pIDK as a SnaBI-PI-SceI fragment to generate pIDK-FAAP100-FANCB-FANCL.

pACEBac1-FANCA-FANCG, pIDK-FAAP100-FANCB-FANCL and pIDS-FANCC-3C-2×StrepII-FANCE-FANCF were fused using Cre recombinase (NEB) to generate a single vector (gentamicin, spectinomycin and kanamycin resistant) containing a single copy of each gene (FA core complex) used to make protein for the initial data collection. All constructs were confirmed by restriction digest analysis, PCR and sequencing.

For subsequent protein preparations, A-G-B-L-100-C-E-F, A-G-B-L-100, G-B-L-100 and C-E-F complexes (letters indicate the FANCL family members) were prepared using a modified BiGbac system as described previously<sup>29,30</sup>. The individual genes were PCR amplified for cloning into pBIG vectors from pACEBac1, pIDC or pIDS vectors. Sequences encoding 2×Strep II tag and 3C protease site were included on FANCC. If FANCC was not present, the tag sequence was added to FANCB. The combined vector carrying the FA core complex or subcomplex was transformed into EMBacY cells to generate a bacmid. Bacmid DNA was transfected into Sf9 cells and virus was passaged twice in the same cell line before large-scale infection in Sf9 cells. Infected cells were collected when cell growth arrested. Sf9 cells were obtained from Oxford Expression Technologies, catalogue no. 600100 (negative for mycoplasma, identity not independently authenticated by us).

Cells were lysed by sonication in lysis buffer (100 mM HEPES pH 8.0, 300 mM NaCl, 1 mM TCEP, 5% glycerol, EDTA-free protease inhibitor, 5 mM benzamidinium hydrochloride and 100 U ml<sup>-1</sup> benzonase). Clarified cell lysate was incubated with StrepTactin resin (GE Healthcare) for 1 h followed by wash with lysis buffer. Proteins were eluted in elution buffer (100 mM HEPES pH 8.0, 300 mM NaCl, 1 mM TCEP, 5% glycerol and 8 mM desthiobiotin). Further purification was performed by HiTrap Heparin HP affinity column (GE Healthcare) using a linear gradient of NaCl from concentration of 150 mM to 1 M in 50 mM HEPES pH 8.0, 1 mM TCEP, over 22 column volumes. For FA core complex, this was followed by anion-exchange chromatography (MonoQ, GE Healthcare) in the same buffer using a linear gradient of NaCl from concentration of 150 mM to 1 M over 20 column volumes. The final buffer for purified FA core complex and subcomplexes was 50 mM HEPES pH 8.0, ~500 mM NaCl, 1 mM TCEP.

### Ubiquitination assay

Ubiquitination assays were performed as described previously<sup>6,31</sup>. In brief, a reaction volume of 20 µl contained 75 nM E1 (Boston Biochem), 1 µM E2 (GgUBE2T), 0.25 µM E3 (FA core complex), 1 µM substrate (His-GgFANCI, GgFANCI<sup>K525R</sup>, His-GgFANCD2, GgFANCD2<sup>K563R</sup>), 50 µM 5'-flapped DNA and 20 µM haemagglutinin (HA)-ubiquitin (Boston Biochem). For ubiquitin discharge assays, concentrations of 125 nM

E1 (Boston Biochem), 5 µM E2 (GgUBE2T), 1.5 µM E3 (FA core complex) and 50 mM free lysine (instead of substrate) were used. The 0.25 µM E3 enzyme concentration is based on the FANCL subunit, estimated by comparing the amount of FANCL in FA core complex and subcomplexes against purified FANCL of known concentration on SDS-PAGE. The reaction buffer was 50 mM HEPES pH 8.0, 64 mM NaCl, 4% glycerol, 5 mM MgCl<sub>2</sub>, 2 mM ATP and 0.5 mM DTT. The reaction was incubated at 30 °C for 90 min, stopped by adding NuPAGE LDS sample buffer (Thermo Fisher) and run on an SDS polyacrylamide gel (3–8% NuPAGE Tris-acetate). Samples were analysed by Coomassie staining or by western blot using a HA antibody (Santa Cruz Biotechnology). All assays were performed independently three times.

### Electron microscopy

Protein complexes were vitrified by applying 3–3.5 µl purified protein (~1 µM) to UltraAuFoil RL2/1.3 grids (Quantifoil)<sup>32</sup> with a thin continuous carbon support layer (for initial FA core complex dataset) or in unsupported ice (for final FA core complex dataset and subcomplexes) that had been made hydrophilic using an argon:oxygen plasma, blotting for 2.5, 3 or 4.5 s at 4 °C with relative humidity of 100% and plunging into liquid ethane using a Vitrobot Mark IV (FEI).

Cryo-EM data were collected on a FEI Titan Krios transmission electron microscope operated at 300 keV acceleration voltage using EPU automated data collection software. An initial dataset was collected on a Falcon II detector at 47,000× nominal magnification with a pixel size of 1.774 Å per pixel. The final data for FA core complex (Extended Data Table 1) were collected on a Falcon III detector in counting mode at 75,000× nominal magnification, and pixel size of 1.04 Å (MRC LMB) or 1.085 Å (eBIC). The subcomplexes G-B-L-100-C-E-F, A-G-B-L-100 and G-B-L-100 were imaged at 59,000× nominal magnification on a Falcon III detector in integrating mode. B-L-100 data were collected at 47,000× on a Falcon II detector in integrating mode.

### Cryo-EM image processing

Image processing was performed in Relion (v.2 and v.3.0-beta)<sup>33–35</sup>, and Relion wrappers were used for external programs except for EMAN2<sup>36</sup>. For all datasets, whole-frame alignment was performed using MotionCorr<sup>37</sup>, and contrast transfer function parameters were estimated using gCTF<sup>38</sup>. 3D maps were post-processed to automatically estimate and apply the B-factor and to determine the resolution by Fourier shell correlation (FSC) between two independent half datasets using 0.143 criterion<sup>39</sup>. Local resolution was estimated using ResMap<sup>40</sup>.

### Initial model

The initial FA core complex dataset was processed in Relion v.2. Particles were picked manually from a few micrographs, and used for 2D class averaging with a box size of 390 pixels. The resulting 2D class averages were used to pick particles from all micrographs using template-matching in Relion's autopicker. After 2D classification of the auto-picked particles, selected classes were used to make an initial 3D model in EMAN2<sup>36</sup>. This initial model was used for subsequent 3D classification and refinement in Relion.

### Refinement

The map generated from the Falcon II dataset was used during the first round of 3D classification with particles from the Falcon III dataset of FA core complex. The datasets from MRC LMB and eBIC were initially processed separately, to generate separate 3D reconstructions. The pixel size of eBIC data was determined using the 2.35 Å spacing from the gold foil images collected under the same imaging conditions as the sample data. The eBIC micrographs were then rescaled to the pixel size of LMB data and CTF was re-estimated, and all datasets were merged.

The FA core complex map was divided into three bodies: body 1 (middle region), body 2 (base region) and body 3 (top region). Bodies 2 and 3 were rotated relative to 1 because body 1 (middle) region appears to be

the most rigid (Extended Data Fig. 2f). Sigma angles of 10 and a sigma offset of 2 were used during refinement. Multibody refinement<sup>13</sup> was continued from the last iteration of consensus refinement. The middle region was best resolved with an overall resolution of 4.4 Å whereas the top and base were ~7 Å resolution. To estimate the flexibility in the FA core complex we performed principal component analysis on the optimal orientations of all the bodies for all particle images in the dataset using *relion\_flex\_analyse*<sup>13</sup>. We rendered videos for principal components 1 and 2 as these described ~30% of the variance in the rotations and translations.

Per-particle CTF refinement and beam tilt estimation were performed by dividing the datasets into individual data collection sessions and processing in a bigger box of 586 pixels, followed by further refinement<sup>35</sup>. The resolution of the maps for the top and base regions of the complete complex were further improved by performing focused classification with signal subtraction<sup>14</sup> followed by refinement (Extended Data Fig. 2g, h). The overall resolution is probably limited by heterogeneity. A regularization T-value of 5 was used during 3D refinement to boost the contribution of higher spatial frequencies<sup>41</sup>, to improve the quality of the map to aid in model building.

The subcomplexes were processed up to 2D classification as there were not enough different views to generate a 3D map.

### Local symmetry averaging

Cryo-EM structure determination by single-particle analysis relies on the reduction of noise through averaging over multiple copies of extremely noisy projection images of individual macromolecular complexes. For symmetric complexes, for example, for homo-multimers or for icosahedral virus capsids, additional averaging can be performed by imposing point-group symmetry on the reconstruction. Because each projection image of a symmetrical object provides multiple views of the asymmetric unit, compared to asymmetric complexes, the same number of images will yield a better reconstruction, or fewer images are needed to obtain a reconstruction of the same quality. Therefore, point-group symmetry averaging is commonly employed in single-particle analysis refinement programs.

The FA subcomplex described here does not obey overall point-group symmetry, but still contains multiple, potentially identical, subcomplexes. We call this local symmetry. Averaging over locally symmetric subcomplexes in cryo-EM single-particle analysis has previously been performed to improve reconstructions after completion of the refinement process, for example on the subunits of triangulation number  $T > 1$  virus capsids<sup>42,43</sup>. However, local symmetry averaging has the potential to improve the reconstruction at every stage of the iterative refinement process. Better reconstructions during refinement will lead to better alignments, and hence a better final reconstruction. Therefore, we implemented a local symmetry averaging approach inside the *relion\_refine* program. This approach has conceptual similarities to imposing non-crystallographic symmetry (NCS) in X-ray crystallography<sup>44</sup>.

This new implementation within Relion3.0 allows the user to define an arbitrary number of groups, each with an arbitrary number of assumed identical subcomplexes. For each group, the user provides a mask around one member of the group, as well as 3D transformation matrices (expressed as three Euler angles and three translations in  $x$ ,  $y$  and  $z$ ) in real space to superimpose that member onto each of the other members in the group. This information is expressed in a Relion-style STAR file, which is passed using the *local\_symmetry* command line option to the *relion\_refine* program. A helper program to find and optimize the 3D transformation matrices, *relion\_localsym*, was also implemented. To minimize artefacts in Fourier space, the edges of the masks should be kept soft, that is, they should gradually change from zero to one over multiple real-space pixels.

Our new implementation allows local symmetry averaging in both 3D classifications and 3D refinements, and works with 2D projection images as well as 3D images, such as sub-tomograms. At every step of

the expectation-maximization algorithm, local symmetry is applied according to the masks and transformations defined in the STAR file. This symmetrization is performed after the maximization step in real space. As such, the signal-to-noise gain that arises from the additional averaging is not considered when calculating the FSC between the two independent half-reconstructions in 3D auto-refinement, or when calculating the signal from the power of the reconstruction in 3D classification. Therefore, in cases of local symmetry, it may be advantageous to increase the empirical regularization T-value,  $-\tau_{2\_fudge}$ , to account for the expected gain in signal during the refinement. When doing so, care should be taken not to overfit the data by using a T-value that is too high. Calculation of a final FSC curve using Relion's post-processing program will lead to a realistic resolution estimate. However, care should be taken to use soft-edged masks for the local symmetrization, as the real-space mask operations may lead to artefacts in the Fourier-based resolution assessment.

Another note of caution considers the assumption that all the members of each of the local symmetry groups are identical. Since, by definition, the complex does not obey point group symmetry, this assumption can never be entirely true. At the very least, some of the subcomplex interfaces will be chemically different, whereas in worst-case scenarios biologically relevant conformational differences may exist within the members of each group. Imposing (local) symmetry on objects that are not identical will lead to a false impression of similarity in the output reconstruction. Besides minimizing artefacts in Fourier space, the soft edges of the local symmetry masks are also relevant here. Local symmetry is applied for all real-space pixels where the mask has values  $m > 0$ , but the symmetrized map will be calculated as  $(1 - m)$  times the original reconstruction plus  $m$  times the average of the corresponding voxel in all members of the group. Thereby, mask values of  $m < 1$  can be used to impose local symmetry only partially.

For the subcomplex reconstruction, masks around each of the symmetric regions were created and filtered to 25 Å with mask extension to 6 pixels and a soft-edge of 10 pixels (Extended Data Fig. 7e). The transformation operator for the three Euler angles and the three translations that relate one symmetric region with the other were calculated using *relion\_localsym\_mpi* in real space, followed by 3D refinement using a regularization T-value of 100. This improved the overall appearance of the map (Extended Data Fig. 7f).

### Model building

Model building was performed using Coot<sup>45,46</sup>. All models described were built as polyaniline chains except FANCL (see below).

A homology model for chicken FANCL (residues 1–373) was generated with I-TASSER<sup>47</sup> using *Drosophila* FANCL (PDB 3KIL)<sup>15</sup> as a template. The FANCL homology model was rigidly fitted into the EM map in Chimera using 'fit in map' tool<sup>48</sup> and then flexibly fitted using Jiggle fit in Coot. Densities corresponding to the ELF, URD and RING domains of FANCL were identified in the focused map of the base region, however, we were not able to orient the RING domain unambiguously in the density. The density for the ELF domain was well-defined in the focused map for the top region. There was only weak density for the URD domain and no density for the RING domain in FANCL<sub>top</sub> (Extended Data Figs. 2g, 5d). The ELF domain is next to the coiled-coil helix of FANCB, in agreement with FANCL crosslinks to the region C-terminal of the coiled-coil helix of FANCB (residues 439, 441, 454 and 460) and to the FAAP100  $\beta$ -propeller (residues 25, 180, 188, 262, 267 and 274). There are several crosslinks between FANCL URD-RING and the substrate-recognition module (FANCC, FANCE and FANCF) consistent with its placement within the base of the complex.

The 'multi-body refinement map' for the middle region, the map from  $T = 5$  regularization, and the map for symmetric subcomplex were used to build the central region of FANCB-FAAP100 de novo by placing idealized helices and strands in Coot, and refining their fits with the real space refine zone tool. In agreement with their role as a scaffold, FANCB



and FAAP100 have 64 intermolecular crosslinks and they crosslink to four other subunits.

One of the two  $\beta$ -propellers in the top region was built de novo using the 'focused map' for the top region and this model was used to identify a structural homologue in DALI<sup>49</sup>. The  $\beta$ -propeller of Bardet–Biedl syndrome 1 protein (PDB 4VON)<sup>50</sup> was the top hit, and this was also a good fit for the other  $\beta$ -propeller. The sequence of this model was changed to polyalanine and used to build all four  $\beta$ -propellers after rigid fitting into the 'focused map' for the top or base region.

On the basis of the crosslinking patterns and secondary structure predictions, we assigned one of each of the pairs of  $\beta$ -propellers, and one helix of each of the coiled coils to FANCB (N-terminal region for  $\beta$ -propeller and residues 390–429 for the helix) and the other to FAAP100 (N-terminal region for  $\beta$ -propeller and residues 461–501 for the helix). These assignments agree with the hydrogen deuterium exchange experiments of B-100 versus B-L-100. The sequences of the predicted long helices of FANCB and FAAP100 were used in MARCOIL<sup>51,52</sup> for assignment of heptad repeats (Extended Data Fig. 5e). Then, a FANCB–FAAP100 coiled-coil model was built in CCbuilder 2.0<sup>53</sup> using the advanced mode (Extended Data Fig. 5f). This model was then fitted into the map and refined in Coot using real space refinement.

A homology model for chicken FANCG (residues 1–648) was generated in I-TASSER using TTC7B–hyccin complex (PDB 5DSE) and O-linked GlcNAc transferase (PDB 1W3B) as the top two templates. The TPR region of the model (residues 153–432) was fitted into the focused map of the top region by rigid fitting in Chimera followed by Jiggle fit in Coot and refinement in Refmac. Additional helices were added towards the C-terminal region of FANCG using Coot.

A homology model for chicken FANCF (residues 117–343) was built in I-TASSER using the crystal structure of human FANCF C-terminal region (PDB 2IQJ)<sup>36</sup> as one of the templates. Residues 117–191 and 215–343 were rigidly fitted into the focused map for the base region followed by Jiggle fit and real space refinement in Coot. In addition to the FANCF model, several short, idealized helices were also placed into the base. A crystal structure for human FANCE C-terminal region (PDB 2ILR)<sup>54</sup> was available but we could not fit it in the map, presumably due to conformational variation. The C terminus of FANCC is near FANCE (residues 155–190; Fig. 1d).

The above models were assembled together into models for top (FANCG,  $\beta$ -propellers and long helices of FANCB and FAAP100), middle and base (FANCL, FANCF,  $\beta$ -propellers and long helices of FANCB, FAAP100, unassigned helices) regions. These models were then further refined in Refmac<sup>55</sup> and Phenix iteratively<sup>56</sup>.

The models for FANCB<sub>top</sub>, FAAP100<sub>top</sub>, FANCB–FAAP100, FANCG and FANCL<sub>top</sub> built in the complete FA core complex map were rigidly fitted into the subcomplex map and refined in Refmac. All models and maps were visualized and rendered in UCSF Chimera<sup>48</sup> or ChimeraX<sup>57</sup>.

### Native mass spectrometry

Native mass spectrometry experiments were carried out on a Q-Exactive Plus UHMR modified to facilitate the transmission of high-energy species and adapted for membrane proteins<sup>58–60</sup>. The FA core complex at concentration of 9.5  $\mu$ M was buffer exchanged using Bio-Spin 6 columns (BioRad). Typically, 2–3  $\mu$ l of sample in 750 mM ammonium acetate were injected, using a 1.2 mm outer diameter, gold-coated borosilicate capillary (Harvard Apparatus). The following parameters were used for protein transmission: capillary voltage 1.2 kV, desolvation voltage –300 V, source fragmentation 0 V, HCD energy 0 V, HCD pressure 4, EMR on, C-trap entrance lens tune offset 2, injection flatopole 8 V, inter flatopole lens 6 V, and bent flatopole 4 V. Threshold was set to 3. Data were analysed using Xcalibur 2.2 (Thermo Fisher), Masslynx 4.2 (Waters) and SUMMIT<sup>61</sup>. The formation of some of the subcomplexes observed may be the result of buffer exchange from 50 mM HEPES pH 8.0, –500 mM NaCl and 1 mM TCEP into 750 mM ammonium acetate, which is often used in native mass spectrometry to improve resolution and generate subcomplexes to aid structure determination.

### HDX-MS

Deuterium exchange reactions were initiated by diluting the complexes in D<sub>2</sub>O (99.8% D<sub>2</sub>O ACROS, Sigma) to give a final D<sub>2</sub>O percentage of ~95%. Deuterium labelling was generally carried out at 23 °C at four time points (3 s, 30 s, 300 s and 3,000 s) in triplicate. The labelling reaction was quenched by adding chilled 2.4% v/v formic acid in 2 M guanidinium hydrochloride and immediately frozen in liquid nitrogen. Samples were stored at –80 °C before analysis.

The quenched samples were rapidly thawed and subjected to proteolytic cleavage using pepsin followed by reverse-phase high performance liquid chromatography separation. The proteins flowed through an Enzymate BEH immobilized pepsin column, 2.1  $\times$  30 mm, 5  $\mu$ m (Waters) at 200  $\mu$ l min<sup>–1</sup> for 2 min, and the resulting peptides were trapped and desalted on a 2.1  $\times$  5 mm C18 trap column (Acquity BEH C18 Van-guard pre-column, 1.7  $\mu$ m, Waters). Trapped peptides were eluted over 11 min using a 3–43% gradient of acetonitrile in 0.1% v/v formic acid at 40  $\mu$ l min<sup>–1</sup> on to a reverse phase analytical column (Acquity UPLC BEH C18 column 1.7  $\mu$ m, 100 mm  $\times$  1 mm (Waters)). The liquid chromatography elute was coupled to a SYNAPT G2-Si HDMS mass spectrometer (Waters) and data were acquired over a *m/z* of 300 to 2,000, using the standard electrospray ionization (ESI) source with lock mass calibration using [Glu1]-fibrinopeptide B (50 fmol  $\mu$ l<sup>–1</sup>). The mass spectrometer was operated in ion mobility mode, at a source temperature of 80 °C and a spray voltage of 2.6 kV. Spectra were collected in positive-ion mode.

Peptide identification was performed with a non-deuterated sample using MS<sup>E</sup> (Waters) to fragment peptides. An identical gradient of increasing acetonitrile in 0.1% v/v formic acid over 11 min was used and the resulting MS<sup>E</sup> data were analysed using Protein Lynx Global Server software (Waters) with an MS tolerance of 5 ppm.

Mass analysis of the peptide centroids was performed using DynamX software (Waters). Only peptides with a score >6.4 were considered. All peptides (deuterated and non-deuterated) were manually verified at every time point for the correct charge state, presence of overlapping peptides and correct retention time. Deuterium incorporation was not corrected for back-exchange and represents relative, rather than absolute changes in deuterium levels. Changes in H/D amide exchange in any peptide may be due to a single amide or several amides within that peptide.

### Crosslinking mass spectrometry of purified FA core complex

The purified FA core complex (50 mM HEPES pH 8.0, –500 mM NaCl and 1 mM TCEP) at a concentration of 7.6  $\mu$ M was crosslinked with 100-fold molar ratio of disulfosuccinimidyl suberate (BS3) for 2 h on ice and the reaction was quenched with 50 mM NH<sub>4</sub>HCO<sub>3</sub> for 30 min at room temperature. The crosslinked samples were cold-acetone precipitated and resuspended in 8 M urea and 100 mM NH<sub>4</sub>HCO<sub>3</sub>. Peptides were reduced with 10 mM DTT and alkylated with 50 mM iodoacetamide. Following alkylation, proteins were digested with Lys-C (Pierce) at an enzyme-to-substrate ratio of 1:100 for 4 h at 22 °C and, after diluting the urea to 1.5 M with 100 mM NH<sub>4</sub>HCO<sub>3</sub> solution, further digestion with trypsin (Pierce) at an enzyme-to-substrate ratio of 1:20.

Digested peptides were eluted from StageTips and split into two, for parallel crosslink enrichment by strong cation-exchange chromatography (SCX) and size exclusion chromatography (SEC), and were dried in a vacuum concentrator (Eppendorf). For SCX, eluted peptides were dissolved in mobile phase A (30% acetonitrile (v/v), 10 mM KH<sub>2</sub>PO<sub>4</sub>, pH 3) before strong cation exchange chromatography (100  $\times$  2.1 mm Poly Sulfoethyl A column; Poly LC). The separation of the digest used a nonlinear gradient<sup>62</sup> into mobile phase B (30% acetonitrile (v/v), 10 mM KH<sub>2</sub>PO<sub>4</sub>, pH 3, 1 M KCl) at a flow rate of 200  $\mu$ l min<sup>–1</sup>. Ten 1-min fractions in the high-salt range were collected and cleaned by StageTips, eluted and dried for subsequent liquid chromatography with tandem mass spectrometry (LC-MS/MS) analysis. For peptideSEC, peptides were fractionated on an ÄKTA Pure system (GE Healthcare) using a Superdex

Peptide 3.2/300 (GE Healthcare) at a flow rate of 10  $\mu\text{L min}^{-1}$  using 30% (v/v) acetonitrile and 0.1% (v/v) trifluoroacetic acid as mobile phase. Five 50- $\mu\text{L}$  fractions were collected and dried for subsequent LC-MS/MS analysis.

Samples for analysis were resuspended in 0.1% v/v formic acid, 1.6% v/v acetonitrile. LC-MS/MS analysis was conducted in duplicate for SEC fractions and triplicate for SCX fractions, performed on an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific) coupled on-line with an Ultimate 3000 RSLCnano system (Dionex, Thermo Fisher Scientific). The sample was separated and ionized by a 50 cm EASY-Spray column (Thermo Fisher Scientific). Mobile phase A consisted of 0.1% (v/v) formic acid and mobile phase B of 80% v/v acetonitrile with 0.1% v/v formic acid. Flow-rate of 0.3  $\mu\text{L min}^{-1}$  using gradients optimized for each chromatographic fraction from offline fractionation ranging from 2% mobile phase B to 45% mobile phase B over 90 min, followed by a linear increase to 55% and 95% mobile phase B in 2.5 min, respectively. The MS data were acquired in data-dependent mode using the top-speed setting with a three second cycle time. For every cycle, the full scan mass spectrum was recorded in the Orbitrap at a resolution of 120,000 in the range of 400 to 1,600  $m/z$ . Ions with a precursor charge state between 3+ and 7+ were isolated and fragmented. Fragmentation by higher-energy collisional dissociation (HCD) employed a decision tree logic with optimized collision energies<sup>63</sup>. The fragmentation spectra were then recorded in the Orbitrap with a resolution of 50,000. Dynamic exclusion was enabled with single repeat count and 60-s exclusion duration.

A recalibration of the precursor  $m/z$  was conducted based on high-confidence (<1% false discovery rate (FDR)) linear peptide identifications<sup>64</sup>. The recalibrated peak lists were searched against the sequences and the reversed sequences (as decoys) of crosslinked peptides using the Xi software suite (v.1.6.746)<sup>65</sup> (<https://github.com/Rappsilber-Laboratory/XiSearch>) for identification. The following parameters were applied for the search: MS1 accuracy = 3 ppm; MS2 accuracy = 10 ppm; enzyme = trypsin (with full tryptic specificity) allowing up to four missed cleavages; crosslinker = BS3 with an assumed reaction specificity for lysine, serine, threonine, tyrosine and protein N termini; fixed modifications = carbamidomethylation on cysteine; variable modifications = oxidation on methionine, hydrolyzed/aminolyzed BS3 from reaction with ammonia or water on a free crosslinker end. The identified candidates were filtered to 1% FDR on link level using XiFDR v.1.1.26.58<sup>66</sup>.

### Pull down assay

The purified subcomplexes A-G-B-L-100 and C-E-F (Strep II tag cleaved by 3C protease) were mixed in a 1:1 molar ratio at concentrations of 1.4  $\mu\text{M}$  each, for 1 h at 4 °C. A 20- $\mu\text{L}$  reaction was incubated with 15  $\mu\text{L}$  of StrepTactin beads (GE Healthcare) equilibrated in 50 mM HEPES pH 8.0, 300 mM NaCl and 1 mM TCEP. The flow through was collected (unbound fraction) and the beads were washed three times with equilibration buffer. The unbound and bound fractions were analysed on SDS-PAGE.

### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

### Data availability

Cryo-EM maps generated during this study have been deposited in the Electron Microscopy Data Bank with accession codes EMD-10290 (FA core complex consensus), EMD-10291 (focused classification top region), EMD-10292 (focused classification middle region), EMD-10293 (focused classification base region) and EMD-10294 (subcomplex). Models generated during this study have been deposited in the Protein Data Bank (PDB) with accession codes 6SRI (FA core complex) and 6SRS (subcomplex). Native mass spectrometry data are available from figshare at <https://doi.org/10.6084/m9.figshare.9692192>. Crosslinking mass spectrometry data have been deposited in the PRIDE database

with accession code PXD014282. All other data are available from the authors upon reasonable request.

28. Sari, D. et al. The MultiBac baculovirus/insect cell expression vector system for producing complex protein biologics. *Adv. Exp. Med. Biol.* **896**, 199–215 (2016).
29. Weissmann, F. et al. biGBac enables rapid gene assembly for the expression of large multisubunit protein complexes. *Proc. Natl Acad. Sci. USA* **113**, E2564–E2569 (2016).
30. Hill, C. H. et al. Activation of the endonuclease that defines mRNA 3' ends requires incorporation into an 8-subunit core cleavage and polyadenylation factor complex. *Mol. Cell* **73**, 1217–1231.e11 (2019).
31. Sato, K., Toda, K., Ishiai, M., Takata, M. & Kurumizaka, H. DNA robustly stimulates FANCD2 monoubiquitylation in the complex with FANCI. *Nucleic Acids Res.* **40**, 4553–4561 (2012).
32. Russo, C. J. & Passmore, L. A. Ultrastable gold substrates for electron cryomicroscopy. *Science* **346**, 1377–1380 (2014).
33. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
34. Fernandez-Leiro, R. & Scheres, S. H. W. A pipeline approach to single-particle processing in RELION. *Acta Crystallogr. D* **73**, 496–502 (2017).
35. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).
36. Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
37. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
38. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
39. Scheres, S. H. A Bayesian view on cryo-EM structure determination. *J. Mol. Biol.* **415**, 406–418 (2012).
40. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).
41. García-Nafria, J., Lee, Y., Bai, X., Carpenter, B. & Tate, C. G. Cryo-EM structure of the adenosine  $A_{2A}$  receptor coupled to an engineered heterotrimeric G protein. *eLife* **7**, e35946 (2018).
42. Stewart, P. L., Burnett, R. M., Cyrklaff, M. & Fuller, S. D. Image reconstruction reveals the complex molecular organization of adenovirus. *Cell* **67**, 145–154 (1991).
43. He, J., Schmid, M. F., Zhou, Z. H., Rixon, F. & Chiu, W. Finding and using local symmetry in identifying lower domain movements in hexon subunits of the herpes simplex virus type 1 B capsid. *J. Mol. Biol.* **309**, 903–914 (2001).
44. Rossmann, M. G. & Blow, D. M. Detection of sub-units within crystallographic asymmetric unit. *Acta Crystallogr. D* **15**, 24–31 (1962).
45. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
46. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
47. Yang, J. et al. The I-TASSER suite: protein structure and function prediction. *Nat. Methods* **12**, 7–8 (2015).
48. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
49. Holm, L. & Sander, C. Dali: a network tool for protein structure comparison. *Trends Biochem. Sci.* **20**, 478–480 (1995).
50. Mourão, A., Nager, A. R., Nachury, M. V. & Lorentzen, E. Structural basis for membrane targeting of the BBSome by ARL6. *Nat. Struct. Mol. Biol.* **21**, 1035–1041 (2014).
51. Delorenzi, M. & Speed, T. An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. *Bioinformatics* **18**, 617–625 (2002).
52. Zimmermann, L. et al. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J. Mol. Biol.* **430**, 2237–2243 (2018).
53. Wood, C. W. & Woolfson, D. N. CCBUILDER 2.0: powerful and accessible coiled-coil modeling. *Protein Sci.* **27**, 103–111 (2018).
54. Nookala, R. K., Hussain, S. & Pellegrini, L. Insights into Fanconi anaemia from the structure of human FANCE. *Nucleic Acids Res.* **35**, 1638–1648 (2007).
55. Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355–367 (2011).
56. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
57. Goddard, T. D. et al. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).
58. van de Waterbeemd, M. et al. High-fidelity mass analysis unveils heterogeneity in intact ribosomal particles. *Nat. Methods* **14**, 283–286 (2017).
59. Rose, R. J., Damoc, E., Denisov, E., Makarov, A. & Heck, A. J. High-sensitivity Orbitrap mass analysis of intact macromolecular assemblies. *Nat. Methods* **9**, 1084–1086 (2012).
60. Gault, J. et al. High-resolution mass spectrometry of small molecules bound to membrane proteins. *Nat. Methods* **13**, 333–336 (2016).
61. Taverner, T. et al. Subunit architecture of intact protein complexes from mass spectrometry and homology modeling. *Acc. Chem. Res.* **41**, 617–627 (2008).
62. Chen, Z. A. et al. Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. *EMBO J.* **29**, 717–726 (2010).
63. Kolbowski, L., Mendes, M. L. & Rappsilber, J. Optimizing the parameters governing the fragmentation of cross-linked peptides in a tribrid mass spectrometer. *Anal. Chem.* **89**, 5311–5318 (2017).
64. Lenz, S., Giese, S. H., Fischer, L. & Rappsilber, J. In-search assignment of monoisotopic peaks improves the identification of cross-linked peptides. *J. Proteome Res.* **17**, 3923–3931 (2018).
65. Giese, S. H., Fischer, L. & Rappsilber, J. A study into the collision-induced dissociation (CID) behavior of cross-linked peptides. *Mol. Cell. Proteomics* **15**, 1094–1104 (2016).
66. Fischer, L. & Rappsilber, J. Quirks of error estimation in cross-linking/mass spectrometry. *Anal. Chem.* **89**, 3829–3833 (2017).

67. Naydenova, K. & Russo, C. J. Measuring the effects of particle orientation to improve the efficiency of electron cryomicroscopy. *Nat. Commun.* **8**, 629 (2017).
68. Buetow, L. & Huang, D. T. Structural insights into the catalysis and regulation of E3 ubiquitin ligases. *Nat. Rev. Mol. Cell Biol.* **17**, 626–642 (2016).
69. Knipscheer, P. & Sixma, T. K. Protein–protein interactions regulate Ubl conjugation. *Curr. Opin. Struct. Biol.* **17**, 665–673 (2007).
70. Metzger, M. B., Pruneda, J. N., Klevit, R. E. & Weissman, A. M. RING-type E3 ligases: master manipulators of E2 ubiquitin-conjugating enzymes and ubiquitination. *Biochim. Biophys. Acta* **1843**, 47–60 (2014).
71. Linares, L. K., Hengstermann, A., Ciechanover, A., Müller, S. & Scheffner, M. HdmX stimulates Hdm2-mediated ubiquitination and degradation of p53. *Proc. Natl Acad. Sci. USA* **100**, 12009–12014 (2003).
72. Alpi, A. F., Pace, P. E., Babu, M. M. & Patel, K. J. Mechanistic insight into site-restricted monoubiquitination of FANCD2 by Ube2t, FANCL, and FANCI. *Mol. Cell* **32**, 767–777 (2008).

**Acknowledgements** We thank T. Nakane, J. Zivanov, C. Lau, A. Carter, P. Emsley, G. Murshudov, D. Malinverni and M. Babu for advice and discussions; K. Naydenova, B. Santhanam, G. Dornan, D. Briant, A. Casañal, A. Kumar, M. Carminati, A. Kelley and members of the Passmore laboratory for assistance; G. Cannone, C. Savva and the LMB EM facility, J. Grimmett and T. Darling (LMB scientific computation) for support and J. Shi for assistance with baculovirus. This work was supported by the Medical Research Council, as part of United Kingdom Research and Innovation, MRC file reference number MC\_U105192715 (L.A.P.); Deutsche

Forschungsgemeinschaft (DFG, 329673113) (J.R.); the Wellcome Trust through a Senior Research Fellowship to J.R. (103139); and the European Research Council grant number 695511-ENABLE (C.V.R.). The Wellcome Centre for Cell Biology is supported by core funding from the Wellcome Trust (203149). We acknowledge Diamond Light Source for access to eBIC (proposals EM18091 and EM17434) funded by the Wellcome Trust, MRC and Biotechnology and Biological Sciences Research Council.

**Author contributions** S.S., E.R. and P.A. designed protein expression and purification schemes, performed ubiquitination assays and performed cryo-EM, 3D reconstruction and modelling; S.S. and D.S.C. performed native mass spectrometry; F.O'R. and G.D. performed crosslinking mass spectrometry; S.M. performed hydrogen–deuterium exchange–mass spectrometry; S.H. and S.H.W.S. developed the local symmetry algorithm; C.H.H., C.J.R. and L.A.P. collected cryo-EM data; L.A.P., C.V.R., J.M.S., J.R., S.H.W.S. and K.J.P. supervised the research; L.A.P. conceived the project; and S.S., P.A. and L.A.P. wrote the paper with contributions from all authors.

**Competing interests** The authors declare no competing interests.

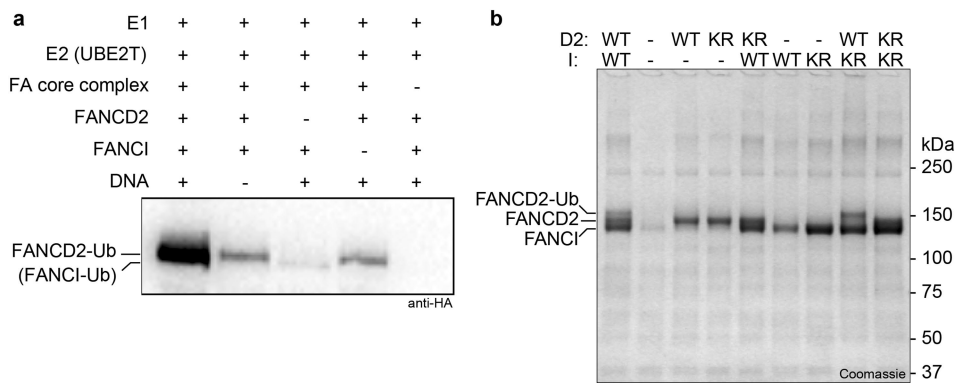
## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-019-1703-4>.

**Correspondence and requests for materials** should be addressed to L.A.P.

**Peer review information** *Nature* thanks Andrew Deans, Xiaodong Zhang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

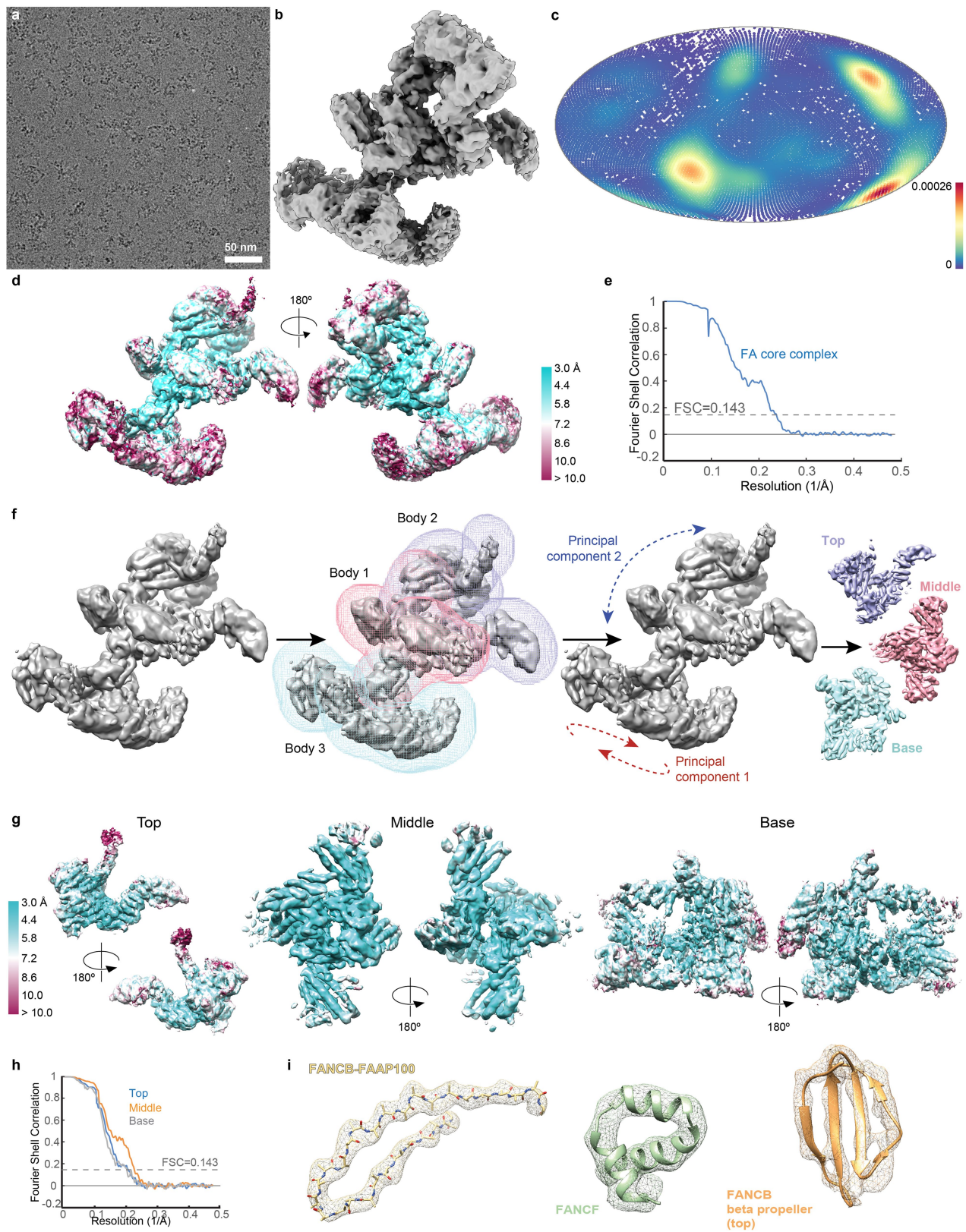


### Extended Data Fig. 1 | Recombinant FA core complex activity.

**a**, Ubiquitination assay analysed by western blot with HA antibody to detect HA-tagged ubiquitin. The migration positions of monoubiquitinated FANCD2 and FANCI are indicated but FANCI is not substantially modified. **b**, Ubiquitination assay analysed by Coomassie-stained SDS-PAGE to show specific monoubiquitination of FANCD2 K563 by recombinant FA core complex. Wild type (WT), FANCD2(K563R) and FANCI(K525R) (KR) were analysed. A native FA core complex purified from chicken DT40 cells monoubiquitinates FANCD2 but does not efficiently monoubiquitinate FANCI<sup>6</sup>.

Therefore, the purified recombinant complex faithfully recapitulates the properties of the native chicken complex. Notably, a purified human complex also did not efficiently monoubiquitinate FANCI, although it did efficiently monoubiquitinate FANCD2<sup>11</sup>. The asymmetry in the FA core complex (see below) reflects this asymmetry in its activity on FANCD2-FANCI. An additional factor or post-translational modification may be required for activation of FANCI monoubiquitination. The ubiquitination assays were repeated at least two times independently with similar results. For gel source data, see Supplementary Fig. 1.

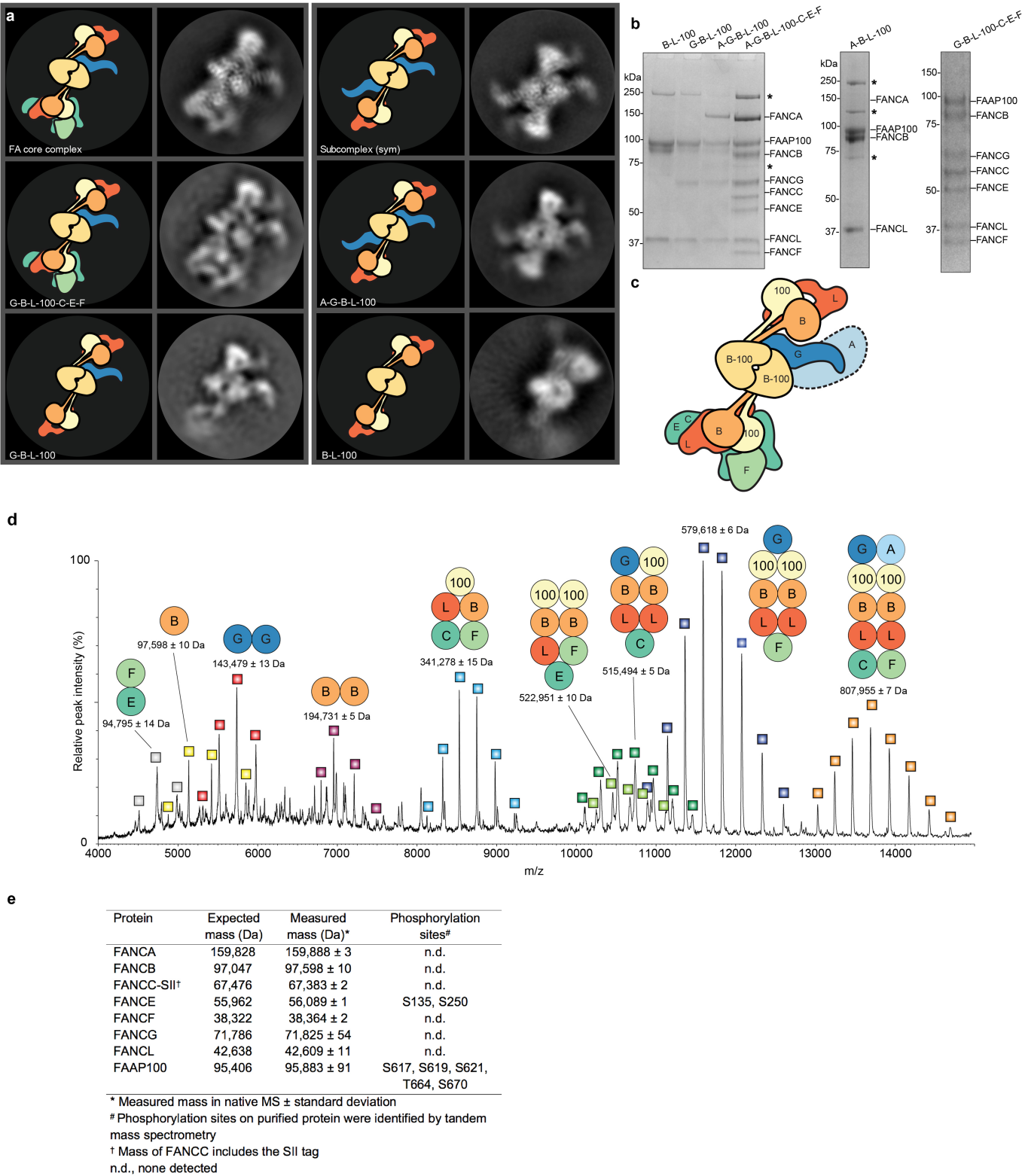




Extended Data Fig. 2 | See next page for caption.

**Extended Data Fig. 2 | Cryo-EM reconstruction of FA core complex, multi-body refinement and assessment of 3D reconstructions after focused refinement.** **a**, Representative raw micrograph of FA core complex. **b**, Overall 3D reconstruction of the FA core complex. **c**, Angular distribution density plot of particles used in the 3D reconstruction of the FA core complex. Every point is a particle orientation and the colour scale represents the normalized density of views around this point. The colour scale runs from 0 (low, blue) to 0.00026 (high, red). The efficiency of orientation distribution<sup>67</sup>,  $E_{\text{OD}}$ , was 0.79. **d**, Estimated local resolution map for FA core complex. **e**, FSC plot for gold standard refinement. **f**, Multibody refinement of the FA core complex using

three masks (body 1, body 2 and body 3) shown in pink, purple and cyan, respectively. The motions are shown in Supplementary Videos 1 and 2. **g**, Local resolution maps for reconstructions of top, middle and base regions of FA core complex. The middle region did not substantially change between multi-body refinement and particle subtraction followed by focused classification. The resolution of the base and top regions improved after particle subtraction and focused classification and refinement. **h**, FSC plot for gold standard refinements. **i**, Representative density for  $\beta$ -strand and  $\alpha$ -helical regions. FANCB-FAAP100 is in the middle region which is better defined than more peripheral regions, including FANCF.



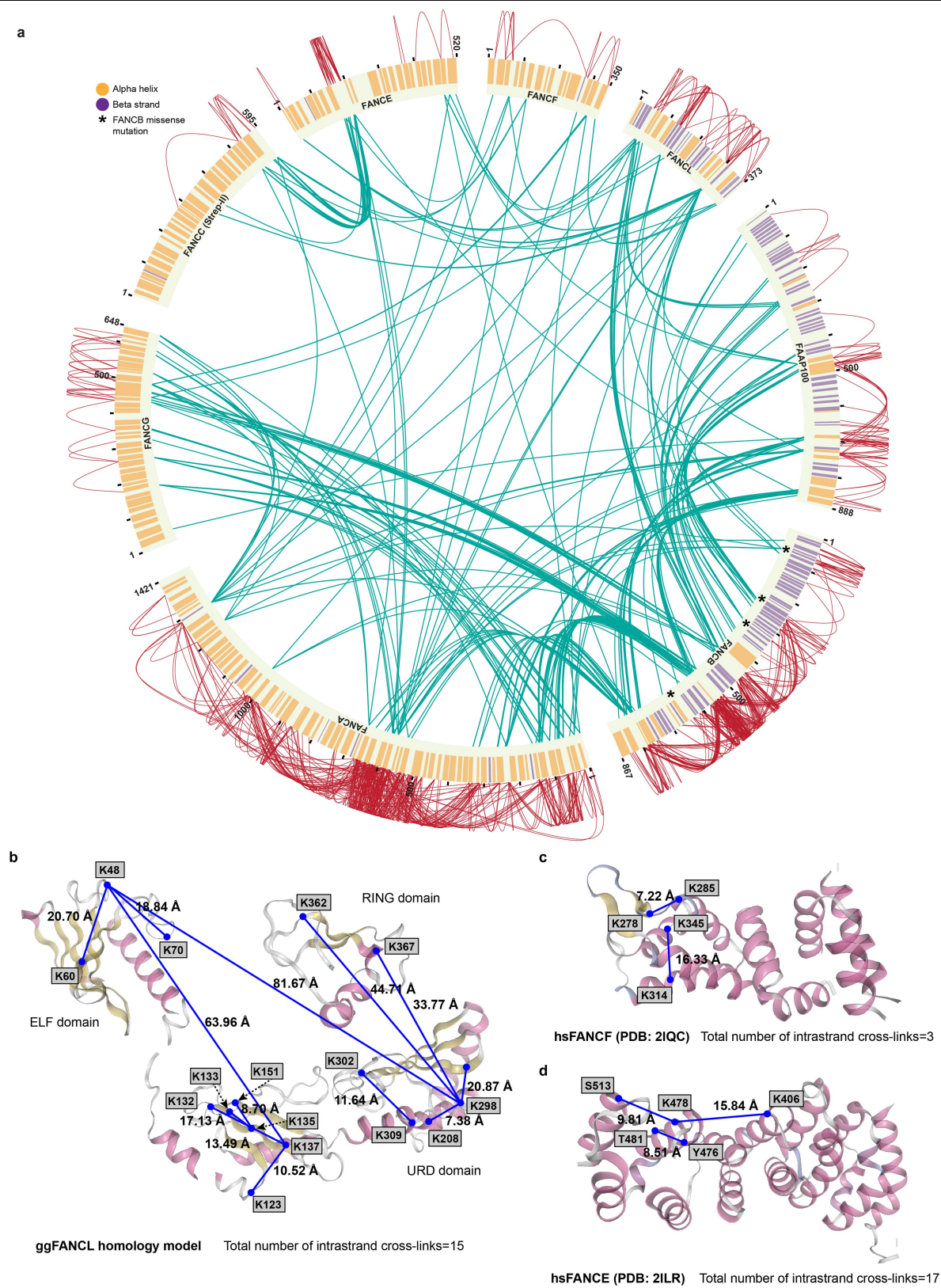
Extended Data Fig. 3 | See next page for caption.

**Extended Data Fig. 3 | Subunit assignment and arrangement in FA core complex.**

**a–c**, Complexes lacking specific subunits were purified and analysed by cryo-EM. **a**, Major 2D class averages identified for subcomplexes, compared with those from FA core complex (A-G-B-L-100-C-E-F). Cartoons are shown to depict the subunits visible in the class averages. The symmetric subcomplex identified in the FA core complex preparation is indicated (sym). The A-G-B-L-100 complex (lacking the substrate-recognition module) has similar symmetric 2D classes. Native mass spectrometry revealed a non-uniform subunit stoichiometry. Thus it is likely that the asymmetric assembly represents the complete FA core complex, while the symmetric structure is a subcomplex that co-purifies with the intact complex. The 2D class average of a complex lacking FANCA (G-B-L-100-C-E-F) appeared similar to the complete FA core complex with no obvious missing density. FANCG is probably the partially disordered arm that extends from the central part of the complex since this was missing when FANCG was not present in the complex. **b**, Coomassie-stained SDS-PAGE analysis of purified subcomplexes. Asterisks indicate contaminant proteins. FANCA did not co-purify with the A-B-L-100 complex but its migration position is indicated on the gel. The purifications were repeated at least two

times independently with similar results. For gel source data, see Supplementary Fig. 1. **c**, Cartoon of FA core complex with subunits labelled. **d**, Native mass spectrum of recombinant FA core complex showing masses and subunit composition of assigned peak series. We dissociated the FA core complex into subcomplexes during ionization and these species were detected by mass spectrometry. Computational analyses then revealed the proteins present in each of the peaks. The standard deviation in fitting the identified peaks to the charge series is given as the  $\pm$  error in the measured mass for a given single measurement. This is the error in the fit and not the error in the mass measurement, which is probably an order of magnitude higher due to, for instance, solvation or adduct effects, or heterogeneous post-translational modifications. Hence, the error gives a rough measure of the accuracy of peak assignment, which is impacted by the broadness, symmetry and signal-to-noise ratio of each peak. **e**, Molecular masses of FA core complex subunits. The expected and measured masses are given, along with phosphorylation sites detected by mass spectrometry. Native mass spectrometry was repeated three times with similar results.

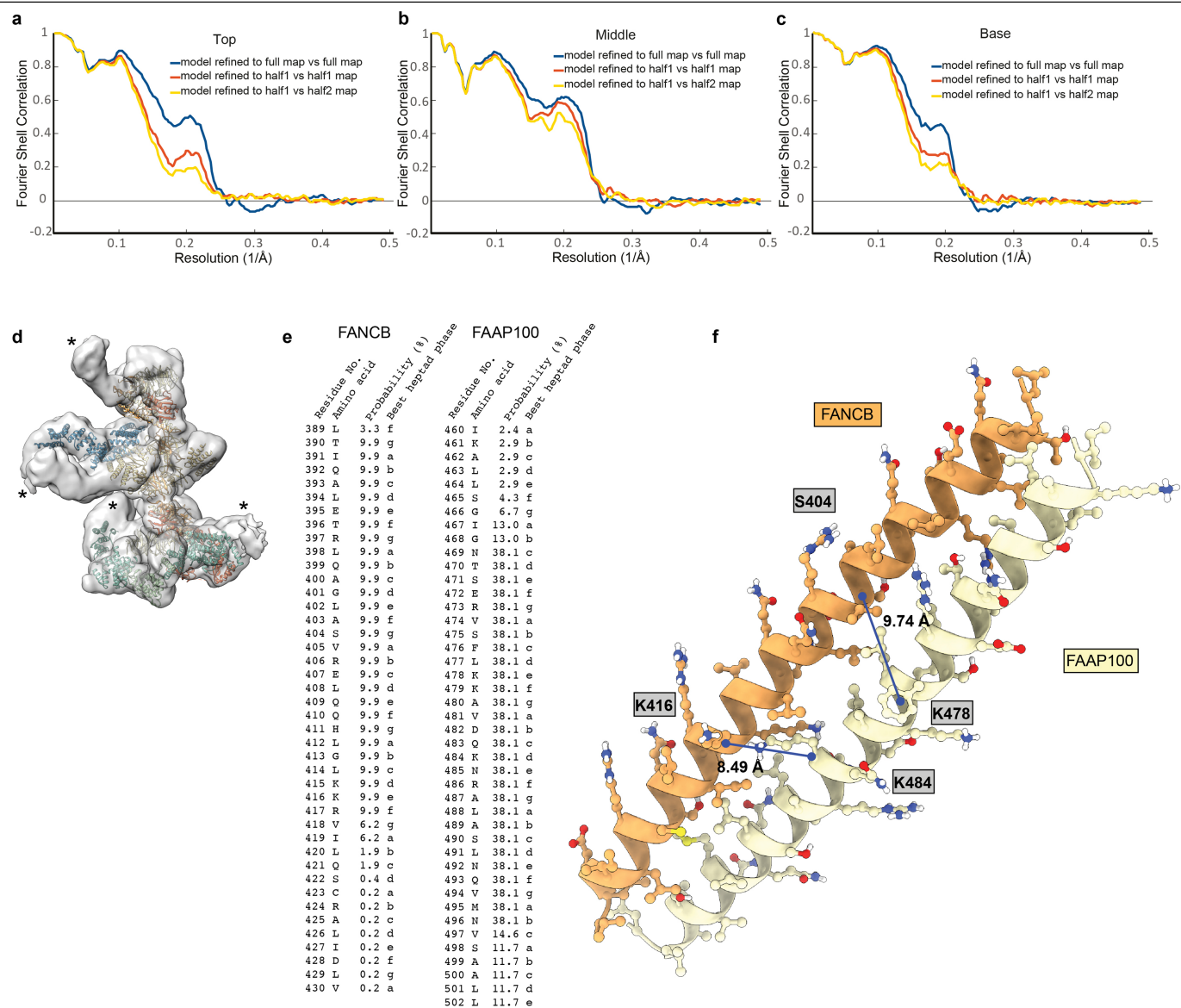




Extended Data Fig. 4 | See next page for caption.

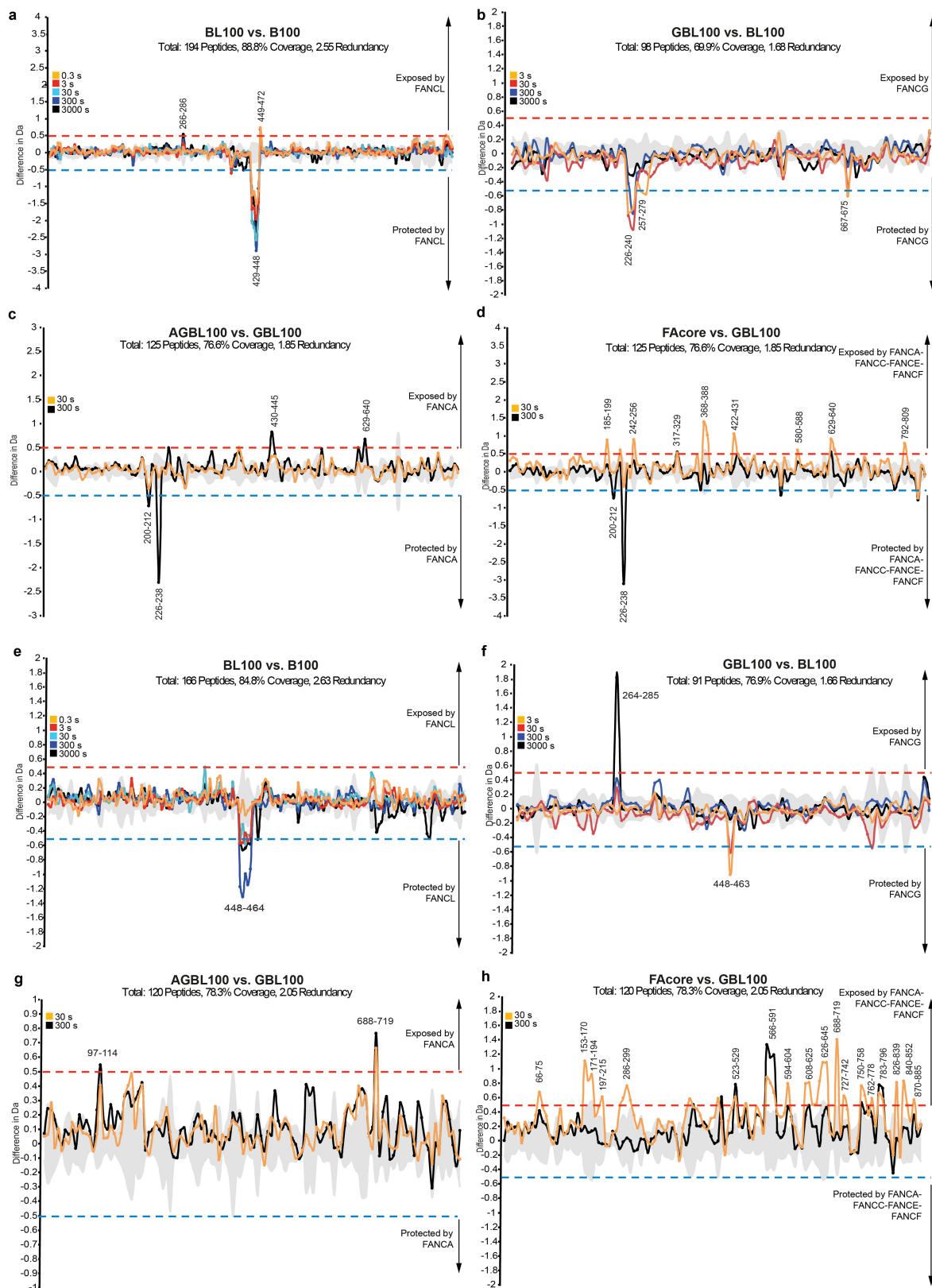
**Extended Data Fig. 4 | Crosslinking mass spectrometry.** **a**, Crosslinking mass spectrometry revealed 834 crosslinks (1% FDR) between residues that are in close proximity. Intermolecular crosslinks are coloured in green; intramolecular crosslinks, red; predicted  $\alpha$ -helices, orange; predicted  $\beta$ -strands, purple. Asterisks on FANCB mark missense mutations from the Fanconi Anaemia Mutation Database (<http://www2.rockefeller.edu/fanconi/>) including L43S in a predicted  $\beta$ -strand, P230S at the N terminus of a predicted helix, and L329P in the middle of a predicted  $\beta$ -strand, all in the  $\beta$ -propeller; and L676P, which is predicted to disrupt an  $\alpha$ -helix in the C-terminal dimerization domain. **b–d**, Validation of crosslinking of FA core complex. Crosslinks were mapped onto a homology model of chicken FANCL (**b**), human FANCF (**c**) and human FANCE (**d**). All intramolecular crosslinks within a domain are consistent with the

maximum crosslinker length (30 Å between the two C $\alpha$ ). Crosslinks between the domains in FANCL are not consistent with the domain arrangement in the crystal structure of FANCL because there is flexibility or changes in the orientation between the domains in the FA core complex. There are two different conformations of FANCL in the structure: FANCL<sub>base</sub> is fully ordered, whereas only the ELF domain is visible in FANCL<sub>top</sub>. Mapping the crosslinks onto FANCL from the base (constrained, with all three domains visible) reveals that the distances for some crosslinks between domains are too large to be consistent with the FANCL conformation in the base. By contrast, for FANCL in the top, where only the ELF domain is ordered, the URD and RING domains are likely to be conformationally flexible. Since the URD and RING domains cannot be modelled for FANCL<sub>top</sub>, it is not possible to validate these interdomain crosslinks.



**Extended Data Fig. 5 | Assessment of model fit in maps and modelling of coiled coils. a–c,** FSC plots of maps versus model for top (a), middle (b) and base (c) regions. **d,** Low-resolution cryo-EM map of FA core complex (transparent surface) with models placed in the map. Asterisks represent density that was not visible at high resolution and was not interpreted. This may

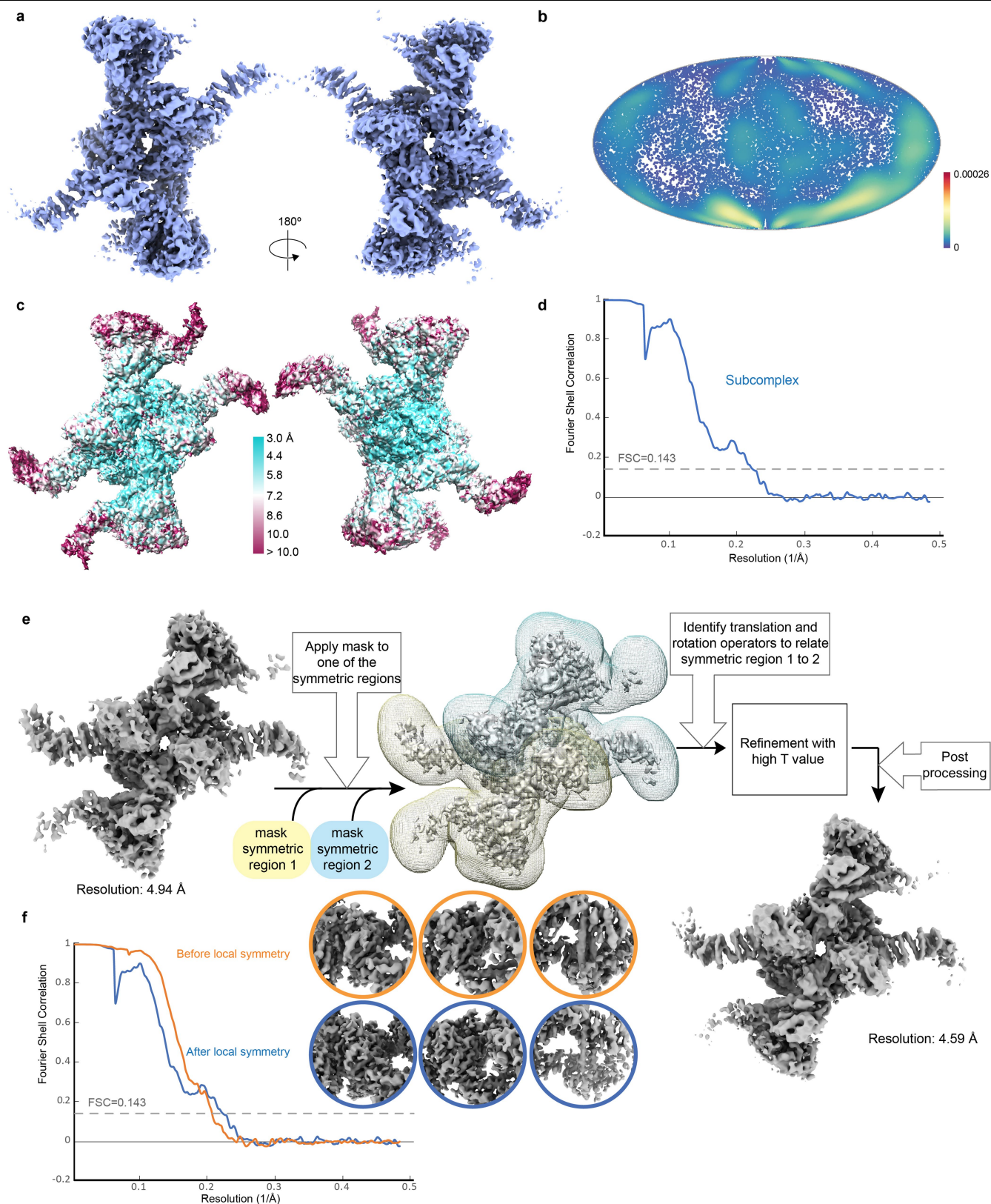
represent FANCA and additional parts of the substrate-recognition module. **e,** MARCOIL<sup>51</sup> prediction for the best heptad phase in the long helices of FANCB and FAAP100. **f,** Predicted coiled-coil model by CCbuilder 2.0<sup>53</sup> for the FANCB and FAAP100 long helices. Crosslinks detected for these helices in the FA core complex crosslinking mass spectrometry are indicated with blue lines.



**Extended Data Fig. 6 | HDX-MS on FANCB and FAAP100. a–d**, Difference plots for FANCB showing peptides that are protected (negative) or exposed (positive) upon binding of additional subunit(s) for B-L-100 vs B-100 (a), G-B-L-100 vs B-L-100 (b), A-G-B-L-100 vs G-B-L-100 (c) and FA core complex vs. G-B-L-100 (d). Exchange of hydrogens in FANCB residues 429–448 was protected after interaction with FANCL, consistent with FANCL being located next to the coiled coil. **e–h**, Difference plots for FAAP100 showing peptides that are protected (negative) or exposed (positive) upon binding of additional

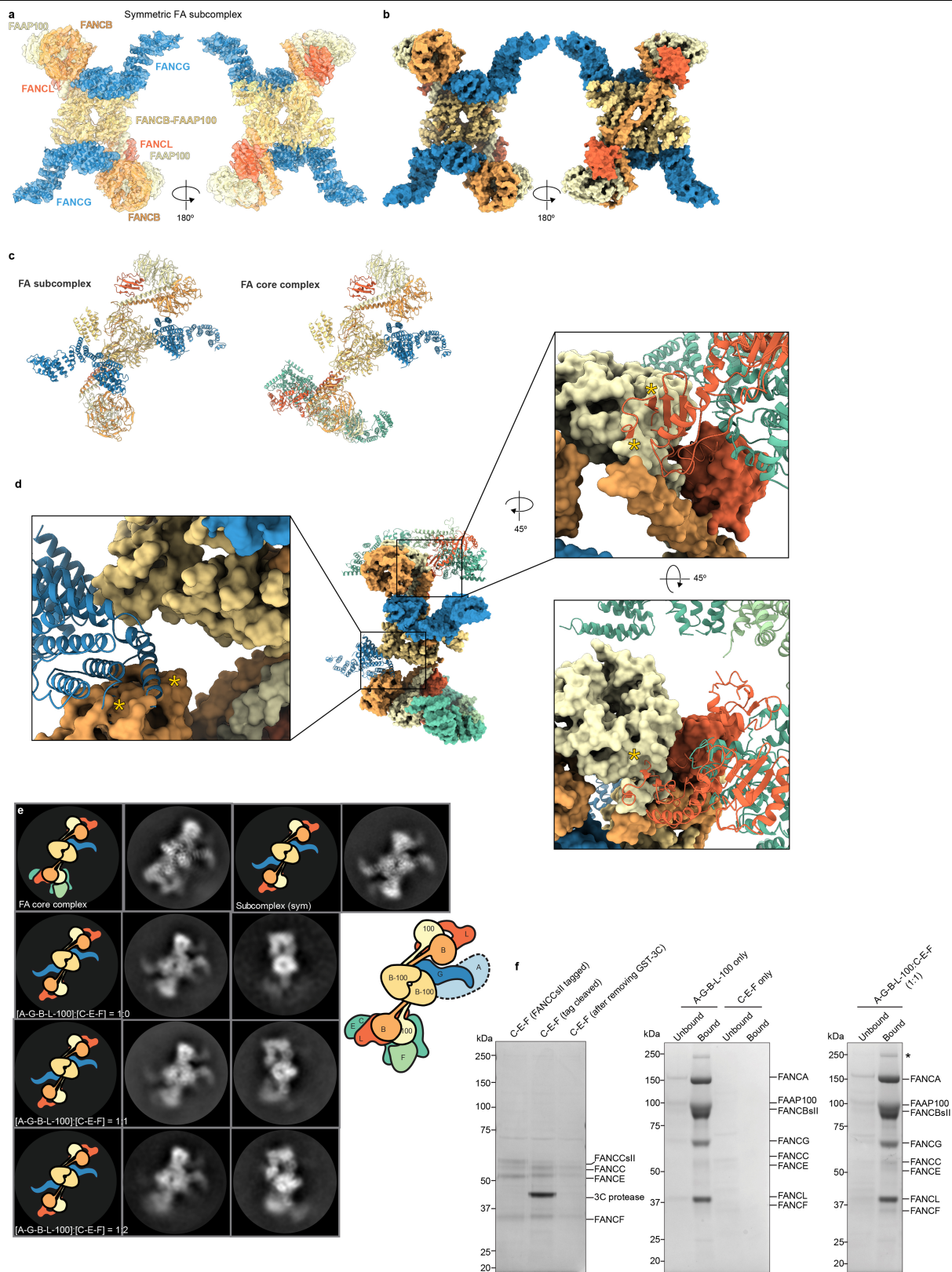
subunit(s) for B-L-100 vs B-100 (e), G-B-L-100 vs B-L-100 (f), A-G-B-L-100 vs. G-B-L-100 (g) and FA core complex vs G-B-L-100 (h). Exchange of hydrogens in FAAP100 residues 448–464 was protected after interaction with FANCL, consistent with FANCL being located next to the coiled coil. For difference plots, triplicate data from four independent colour-coded time points are shown. The significance threshold is indicated by dashed lines. Grey shading indicates the standard deviation of all charge states and replicates per peptide. Sequence coverage is shown in the Supplementary Information.





**Extended Data Fig. 7 | 3D reconstruction of symmetric FA subcomplex and local symmetry refinement.** **a**, Overall 3D reconstruction of the symmetric FA subcomplex. **b**, Angular distribution plot of particles used in the 3D reconstruction of the symmetric FA subcomplex. Every point is a particle orientation and the colour scale represents the normalized density of views around this point. The colour scale runs from 0 (low, blue) to 0.00026 (high, red). The efficiency of orientation distribution<sup>67</sup>,  $E_{OD}$ , was 0.65. **c**, Estimated local

resolution map for symmetric FA subcomplex. **d**, FSC plot for gold standard refinement. **e**, Local symmetry pipeline for reconstruction of the symmetric FA subcomplex (see Methods). This reconstruction could not be improved with C2 symmetry, probably because of local flexibility. **f**, FSC plot for gold standard refinement shown for the subcomplex reconstruction before and after local symmetry refinement. The circular panels show representative densities before and after local symmetry refinement.



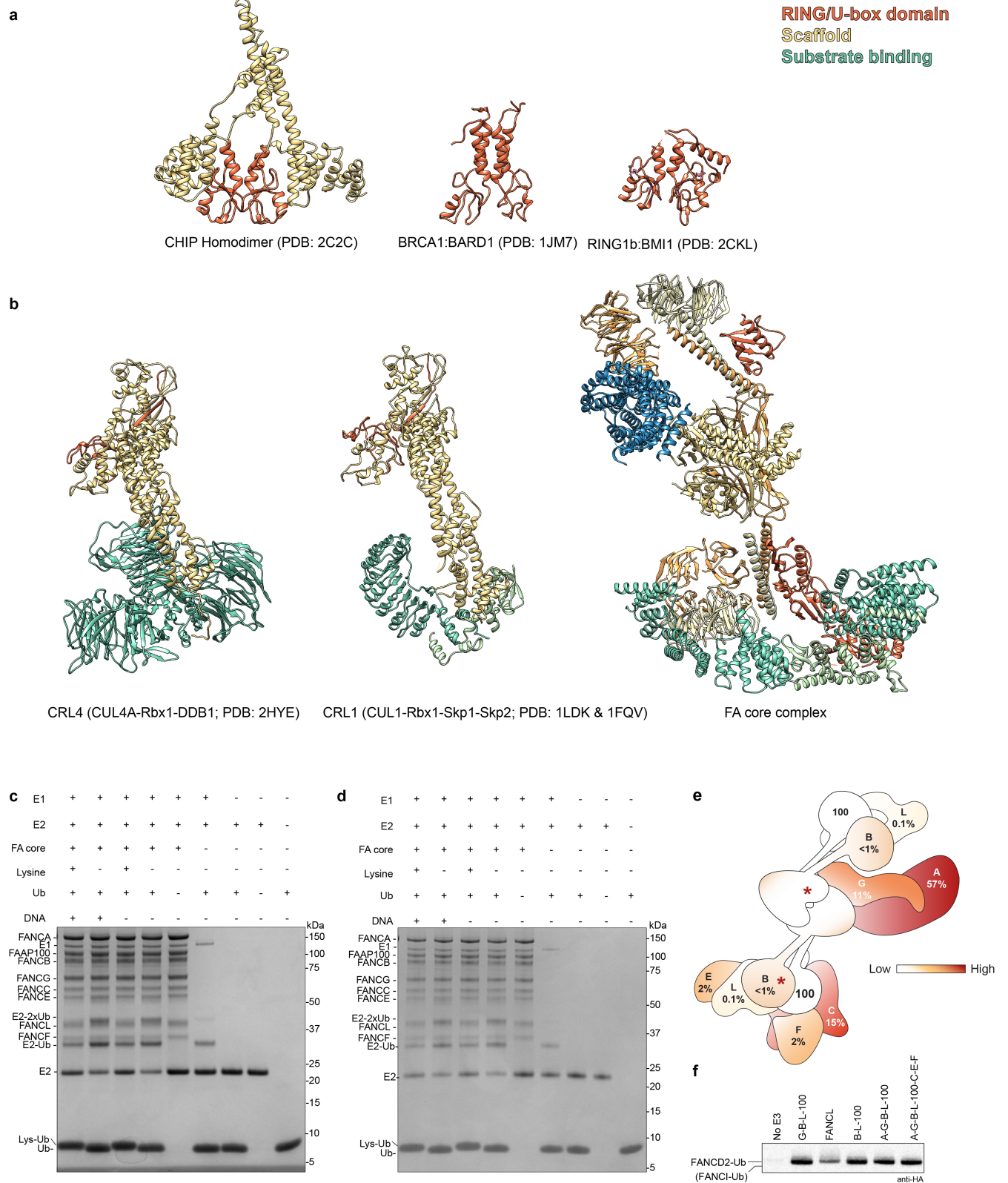
**Extended Data Fig. 8** | See next page for caption.

# Article

**Extended Data Fig. 8 | Model for FA subcomplex.** **a**, Model of FA subcomplex shown as cartoon representations of subunits fit into the cryo-EM map. **b**, Model of FA subcomplex shown as a surface representation of the combined models. Two views are shown down the two-fold symmetric axis. **c**, Comparison of FA subcomplex and complete FA core complex in the same orientations. Both models are shown in cartoon representation. Subunits are coloured as in Fig. 1e. **d**, Modelling of a fully symmetric FA core complex containing two copies of every subunit. Left, the second copy of FANCG (cartoon) from the FA subcomplex was modelled onto the structure of the FA core complex (surface representation). This second FANCG clashes with the FANCB  $\beta$ -propeller in the base (asterisks). Thus, it is likely that upon binding of the substrate-recognition module, rearrangement of the  $\beta$ -propellers of FANCB and FAAP100 prevents binding of a second copy of FANCG. (Right) A second copy of substrate-recognition module (FANCC–FANCE–FANCF; cartoon) is modelled in the top region of the FA core complex by combining the models of FANCC–FANCE–FANCF and FANCL<sub>base</sub> from FA core complex followed by superimposing FANCL<sub>base</sub> on FANCL<sub>top</sub>. There is a clash (asterisks) between the modelled FANCL (cartoon) and FAAP100  $\beta$ -propeller (surface representation). These data suggest that a fully symmetric complex does not readily form. In agreement with this, there was no evidence for any classes containing two copies of C-E-F in any of our EM analyses of the FA core complex. **e**, **f**, The symmetric FA subcomplex

(A-G-B-L-100) does not readily associate with purified substrate-recognition module (C-E-F) to form the asymmetric FA core complex. **e**, The 2D class averages of A-G-B-L-100 mixed with C-E-F compared with complete FA core complex, FA subcomplex and A-G-B-L-100 subcomplex. A-G-B-L-100 was mixed with C-E-F in molar ratios of 1:1 and 1:2 for 1 h at 4 °C before cryo-plunging. Only the symmetric A-G-B-L-100 subcomplex was observed and there was no additional density for C-E-F. Panels for FA core complex, subcomplex and A-G-B-L-100 are replicated from Extended Data Fig. 3a. **f**, Pull-down assay of C-E-F using tagged A-G-B-L-100 (Strep II tagged). Left, the Coomassie Blue-stained gel of purified C-E-F (with Strep II tag, after 3C cleavage of tag and after removal of 3C protease). Tagged A-G-B-L-100 was immobilized on StrepTactin resin and incubated with purified C-E-F at a 1:1 molar ratio. After washing, only a small amount of C-E-F remains bound to the beads. Negative controls (A-G-B-L-100 only and C-E-F only) are shown in the middle panel. Asterisk indicates a contaminant protein. The pulldown experiment was repeated two times independently with similar results. Since C-E-F does not efficiently bind A-G-B-L-100, these experiments suggest that these species are unlikely to be in equilibrium in solution. Previous genetic and biochemical data show that FANCC, FANCE and FANCF are important for monoubiquitination of the FANCD2–FANCL substrate. Together, these data provide evidence that the asymmetric complex is the relevant, functional, physiological entity.





**Extended Data Fig. 9** | See next page for caption.

**Extended Data Fig. 9 | Structural comparison of E3 ligases. a,** There is a strong precedence for dimerization of RING/U-box domain E3 ubiquitin ligases<sup>68–70</sup>. RING/U-box E3s exist both as homo- and heterodimeric complexes, for example, Rad18–Rad18, CHIP–CHIP, RNFS–RNFS, BRCA1–BARD1, RING1b–BMI1 and Hdm2–Hdmx<sup>17–21,71</sup>. Structures of homo- and heterodimeric RING/U-box E3 ligases are shown here with the RING/U-box in orange. Surprisingly, these E3s display functional and structural asymmetry: in all the dimers listed above, only one protomer binds to an E2 enzyme. The homodimeric CHIP E3 ligase has a strikingly asymmetric structure that clearly demonstrates why only one U-box binds E2 enzyme<sup>19</sup>. The FANCL RING subunit is also an asymmetric dimer within the FA core complex and it is possible that only one of these binds E2. However, unlike the smaller E3s, the FANCL RING fingers are not near each other. Together, this suggests that asymmetric dimerization may be a general feature of RING E3s. **b,** Comparison of FA core complex with cullin–RING ubiquitin ligases (CRLs). Many large complexes are predominantly helical suggesting that  $\alpha$ -helices are commonly used as building blocks for complexes. In addition,  $\beta$ -propellers often mediate protein–protein interactions. The CRL complexes and FA core complex are long and extended with substrate-recognition (green), scaffold (yellow) and RING (orange) subunits residing in three different regions of the structure. However, the structural details of these complexes differ. Interestingly, the activities of some multisubunit RING-containing E3 ligases including APC/C and CRL complexes are stimulated by dimerization<sup>22,23</sup>. Thus, dimerization may underpin physiological ubiquitination activity in many E3 ligases. **c, d,** Ubiquitin discharge assay, in

which free lysine is used instead of the FANCD2–FANCI substrate. In these experiments, the FA core complex is incubated with E1, E2, ubiquitin and free lysine. If FANCL is active without a substrate, ubiquitin will be conjugated to lysine, resulting in a shift in its molecular weight; however, if substrate binding is required to activate the E3 ligase activity (for example, through allosteric changes), this will not occur. Coomassie gels of reaction products were run in non-reducing (**c**) and reducing (100 mM DTT) conditions (**d**). Ubiquitin is transferred to free lysine as shown by the increase in molecular weight of ubiquitin as well as a decrease in intensity of the E2–ubiquitin band when compared to the lane containing no free lysine. Thus, substrate binding is not required for activity. Reducing conditions do not eliminate the UBE2T–ubiquitin conjugate, as previously shown<sup>72</sup>. Additionally, DNA is not required for FA core complex E3 ligase activity on free lysines, suggesting that DNA activates the substrate, not the E3. The ubiquitin discharge assays were repeated three times independently with similar results (**c, d**).

**e,** Distributions of patient mutations are indicated on the FA core complex by heat map colouring of subunits and in percentage. **f,** Ubiquitination assay using several subcomplexes (Extended Data Fig. 3b) and the full FA core complex, analysed by western blot with HA antibody to detect HA-tagged ubiquitin. The migration positions of monoubiquitinated FANCD2 and FANCI are indicated but FANCI is not substantially modified, as in Extended Data Fig. 1a. All complexes have similar activities but isolated FANCL is less active. This assay was repeated at least two times independently with similar results. For gel source data, see Supplementary Fig. 1.

**Extended Data Table 1 | Cryo-EM data collection, refinement and validation statistics**

	FA core complex (EMD-10290: consensus, EMD-10291: top, EMD-10292: middle, EMD-10293: base) (PDB 6SRI)	Subcomplex (EMD-10294) (PDB 6SRS)
<b>Data collection and processing</b>		
Magnification	75,000 X	75,000 X
Voltage (keV)	300	300
Electron exposure (e-/Å <sup>2</sup> )	40	40
Defocus range (µm)	-1.8 to -4.0	-1.8 to -4.0
Pixel size (Å)	1.040 (LMB) 1.085 (eBIC)	1.040 (LMB) 1.085 (eBIC)
Symmetry imposed	C1	C1
Initial particle images (no.)	1,947,765	1,947,765
Final particle images (no.)	169,000	49,423
Map resolution (Å)	4.2 (consensus), 4.5 (top), 4.4 (middle), 4.9 (base)	4.6
FSC threshold	0.143	0.143
Map resolution range (Å)	4.2 to > 10	4.6 to > 10
<b>Refinement</b>		
Initial model used	<i>De novo</i> modelling and homology modelling	<i>De novo</i> modelling and homology modelling
Model resolution (Å)	4.5 (top), 4.4 (middle), 4.9 (base)	4.6
FSC threshold	0.143	0.143
Model resolution range (Å)	n/a	n/a
Map sharpening <i>B</i> factor (Å <sup>2</sup> )	-149 (consensus) -198 (top) -190 (middle) -177 (bottom)	-213
Model composition		
Non-hydrogen atoms	15,309	12,424
Protein residues	3,827	3,106
Ligands	0	0
<i>B</i> factors (Å <sup>2</sup> )		
Protein	not estimated	not estimated
Ligand		
R.m.s. deviations		
Bond lengths (Å)	0.23	0.22
Bond angles (°)	0.48	0.50
Validation		
MolProbity score	1.44	1.12
Clashscore	3	3
Poor rotamers (%)	0	0
Ramachandran plot		
Favored (%)	95.95	98.88
Allowed (%)	3.56	1.12
Disallowed (%)	0.49	0



Extended Data Table 2 | Features of individual FA core complex subunits and structural models

Protein	Length (aa)	Sequence features	Models generated in this study	Maps used for modelling	Sequence identity / similarity between <i>Gallus gallus</i> and <i>Homo sapiens</i> (%)
FANCA	1,421	$\alpha$ -helical	N/A	N/A	49.1 / 65.6
FANCB	867	Possible $\beta$ -propeller, plus $\alpha$ -helical	<i>De novo</i> modelling of SSEs <sup>*</sup>	Focused map for top (EMD-10291), middle (EMD-10292), and base regions (EMD-10293), consensus map with T=5 (EMD-10290) and subcomplex map (EMD-10294)	44.1 / 63.2
FANCC	559	$\alpha$ -helical	SSEs placed in maps, not assigned	Focused map for base (EMD-10293)	49.0 / 65.1
FANCE	520	$\alpha$ -helical Crystal structure of human orthologue (C-terminal half; PDB 2ILR)	SSEs placed in maps, not assigned	Focused map for base (EMD-10293)	40.9 / 54.1
FANCF	350	$\alpha$ -helical Crystal structure of human orthologue (C-terminal half; PDB 2IQC)	C-terminal region from homology model based on PDB 2IQC	Focused map for base (EMD-10293)	38.5 / 51.6
FANCG	648	$\alpha$ -helical (TPR)	TPR domain from homology model (I-TASSER)	Focused map for top region (EMD-10291)	37.3 / 52.9
FANCL	373	ELF, URD and RING domains. Crystal structures of human (central domain, PDB 3ZQS; RING domain, PDB 4CCG) and <i>Drosophila</i> orthologues (full-length, PDB 3K1L)	ELF domain in FANCL <sub>top</sub> . ELF, URD and RING domains in FANCL <sub>base</sub> . Homology models based on PDB 3K1L	Focused map for top (EMD-10291) and base (EMD-10293)	69.9 / 82.7
FAAP100	888	Possible $\beta$ -propeller, plus $\alpha$ -helical	<i>De novo</i> modelling of SSEs	Focused map for top (EMD-10291), middle (EMD-10292), and base regions (EMD-10293), consensus map with T=5 (EMD-10290) and subcomplex map (EMD-10294)	55.6 / 68.7

SSE, secondary structure elements

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- |                                     |  |
|-------------------------------------|--|
| n/a                                 | Confirmed  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The exact sample size ( <i>n</i> ) for each experimental group/condition, given as a discrete number and unit of measurement   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of all covariates tested  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i> ) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted<br><i>Give P values as exact values whenever suitable.</i>                                |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i> ), indicating how they were calculated  |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

### Software and code

Policy information about [availability of computer code](#)

Data collection	CryoEM data collected on Titan Krios microscope was performed with EPU (FEI/Thermo Fisher Scientific)
Data analysis	Relion v2, Relion v3.0-beta, Eman2, MotionCor2, gCTF, ResMap, Coot, I-TASSER, Refmac, Marckoil, CCbuilder2.0, Xcalibur 2.2 (Thermo Fisher), Masslynx 4.2 (Waters), SUMMIT, Protein Lynx Global Server software (Waters), DynamX software (Waters), Xi software suite (version 1.6.746) and XIFDR version 1.1.26.58, UCSF Chimera, ChimeraX, Phenix

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

CryoEM maps generated during this study have been deposited in the Electron Microscopy Data Bank (EMDB) with accession codes EMD-10290 (FA core complex consensus), EMD-10291 (focused classification top region), EMD-10292 (focused classification middle region), EMD-10293 (focused classification base region) and EMD-10294 (subcomplex). Models generated during this study have been deposited in the protein databank (PDB) with accession codes 6SRI (FA core complex) and 6SRS (subcomplex). Native MS data is available from figshare with accession code: 10.6084/m9.figshare.9692192. Crosslinking MS data has been deposited in the PRIDE database with accession code PXD014282. All other data are available from the authors upon reasonable request.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes were chosen based on previous experience and published studies to evaluate reproducibility of assays. For cryo-EM, the initial number of particles was ~1,950,000, which was sufficient to obtain the stated resolution after 3D classification .
Data exclusions	No data were excluded.
Replication	All experiments (purifications, ubiquitination assays, pulldowns, nativeMS) were performed at least two or three times (exact number of replicates given in text). All attempts to replicate results were successful.
Randomization	Randomization is not relevant to the experiments performed in this study.
Blinding	Blinding is not relevant to the experiments performed in this study.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Antibodies

Antibodies used	HA-probe (F-7) HRP monoclonal, Santa Cruz, Cat# sc-7392HRP, Lot# H3017, dilution 1:1000
Validation	In this study, the HA antibody was used in Western blots to probe HA-tagged Ubiquitin. Western blot analysis of HA-tagged fusion proteins showing N-terminal HA-tagged JNK2 and JNK1 and C-terminal HA-tagged Daxx was performed by the manufacturer. In addition, we could verify the Western blotting results by Coomassie blue staining.

## Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	Sf9, Oxford Expression Technologies Ltd, Cat No. 600100
Authentication	Cell line was not authenticated.
Mycoplasma contamination	Cell line was negative for mycoplasma.
Commonly misidentified lines (See <a href="#">ICLAC</a> register)	No commonly misidentified cell lines were used.