



Molecular basis for arginine C-terminal degron recognition by Cul2^{FEM1} E3 ligase

Xinyan Chen^{1,3}, Shanhui Liao^{1,3}, Yaara Makaros^{2,3}, Qiong Guo¹, Zhongliang Zhu¹, Rina Krizelman², Karin Dahan², Xiaoming Tu¹, Xuebiao Yao¹, Itay Koren^{1,2}✉ and Chao Xu¹✉

Degrans are elements within protein substrates that mediate the interaction with specific degradation machineries to control proteolysis. Recently, a few classes of C-terminal degrons (C-degrons) that are recognized by dedicated cullin-RING ligases (CRLs) have been identified. Specifically, CRL2 using the related substrate adapters FEM1A/B/C was found to recognize C degrons ending with arginine (Arg/C-degron). Here, we uncover the molecular mechanism of Arg/C-degron recognition by solving a subset of structures of FEM1 proteins in complex with Arg/C-degron-bearing substrates. Our structural research, complemented by binding assays and global protein stability (GPS) analyses, demonstrates that FEM1A/C and FEM1B selectively target distinct classes of Arg/C-degrons. Overall, our study not only sheds light on the molecular mechanism underlying Arg/C-degron recognition for precise control of substrate turnover, but also provides valuable information for development of chemical probes for selectively regulating proteostasis.

Ubiquitin-dependent protein degradation is orchestrated by ubiquitin proteasome system (UPS), in which the ubiquitin is transferred to the substrate protein via the E1-E2-E3 cascade^{1,2}. E3 ligase (E3) mediates protein ubiquitination and subsequently directs the target protein to the proteasome through recognizing a specific sequence within it, termed a degron^{3,4}. Within the cullin-RING E3 ligases (CRL) complex, cullins function as scaffolds and recruit substrates through adapter proteins^{5,6}. Adapter protein-degron interaction determines the specificity of CRLs, thereby precisely regulating a broad range of cellular events, including cell cycle⁷, DNA replication⁸, signal transduction⁹ and more⁶. Perturbation of proteolysis results in many severe diseases, such as cancer, neurodegenerative and autoimmune diseases^{10–13}.

The first characterized degron was the N-degron that hallmarks the destabilizing residues at the N terminus of short-lived proteins^{3,14–16}. Recently, the C-terminus of full-length proteins, protein fragments generated after proteolytic cleavage and nascent polypeptides from premature termination were also found to serve as destabilizing residues, termed C-degrons^{17–20}. Pioneering work at the Elledge and Yen laboratories identified a series of C-degrons that are recognized by distinct adapter proteins in CRL complexes and regulate the stability of a wide variety of proteins *in vivo*^{17,18,20,21}.

Using the global protein stability (GPS) method, all three mammalian FEM1 proteins, FEM1A, FEM1B and FEM1C, were demonstrated to serve as CRL2 adapters recognizing the Arg/C-degron^{17,18}. Although all FEM1 proteins contain an N-terminal ankyrin domain followed by tetratricopeptide repeats (TPRs), a C-terminal ankyrin domain²² and a von Hippel–Lindau box for Cul2 binding^{23,24} (Fig. 1a), FEM1A/C and FEM1B were found to selectively target C-degrons of nucleotide exchange factor SIL1^{25,26}, and Cyclin-dependent kinase 5 activator 1 (CDK5R1, also known as p35), respectively¹⁷. The molecular mechanism underlying distinct recognition of Arg/C-degrons of SIL1 and CDK5R1 by FEM1 proteins was not addressed.

Using an isothermal titration calorimetric (ITC) assay, we found that FEM1A/C binds to C-degrons ending with -K/R-X_{1–2}-R (X denotes any residue), including those of Olfactory receptor 51B2 (OR51B2)²⁷ and SIL1, whereas FEM1B selectively recognizes the C-degron of CDK5R1 ending with -G-L-D-R. By solving several C-degron-bound structures of FEM1C, we unraveled a bipartite mechanism for recognition of the -K/R-X_{1–2}-R C-degron. To uncover the distinct substrate binding properties of FEM1A/C and FEM1B, we further determined the structure of CDK5R1-bound FEM1B. Furthermore, by using an *in vivo* protein stability assay, we show the interactions between FEM1s and the Arg/C-degron efficiently control the turnover of substrates via the UPS pathway. Overall, our structural approach, complemented by biochemical, biophysical and cell biology experiments, provides structural insights into the selectivity of Cul2^{FEM1A/B/C} E3 machinery toward the Arg/C-degron.

Results

FEM1A/B/C directly interact with Arg/C-degron. To study the binding properties of FEM1 proteins, we first cloned, expressed and purified the recombinant proteins FEM1A_{1–442}, FEM1C_{1–390} and FEM1B_{1–356} spanning their N-terminal ankyrin repeats and TPRs (Fig. 1a). We then synthesized several peptides bearing Arg/C-degron recognized by FEM1 proteins^{17,18}, including SIL1_{452–461} (YSVNSLLKELR), OR51B2_{303–312} (YRLLSKHRFSR), NS11 (YQERGPTWDKKNLR), Clone13 (PTQGRAR)¹⁷ and CDK5R1_{298–307} (YKKRLLGLDR), and examined their FEM1-binding properties by ITC binding assay.

We noticed that FEM1A_{1–442} and FEM1C_{1–390}, but not FEM1B_{1–356}, bind to peptides ending in -K/R-X-X-R or K/R-X-R, including SIL1_{452–461}, OR51B2_{303–312}, NS11 and Clone13 with apparent dissociation constants (K_d) in the range of 6.2–13 and 3.5–20 μ M, respectively. The binding of FEM1A or FEM1C to SIL1 is abolished by the terminal Arg to Ala mutation (R-1A), underscoring

¹MOE Key Laboratory for Membraneless Organelles and Cellular Dynamics, Hefei National Laboratory for Physical Sciences at the Microscale, School of Life Sciences, University of Science and Technology of China, Hefei, China. ²The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat-Gan, Israel. ³These authors contributed equally: Xinyan Chen, Shanhui Liao, Yaara Makaros. ✉e-mail: itay.koren@biu.ac.il; xuchaor@ustc.edu.cn

the essential role of terminal arginine in FEM1A/C binding (Fig. 1b–e and Supplementary Table 1). FEM1B_{1–356} selectively binds to a peptide derived from the CDK5R1 C-degron with a K_d of 6.0 μM , roughly 4–6-fold stronger than those of FEM1A and FEM1C ($K_d = 25\text{--}36\ \mu\text{M}$) (Supplementary Table 1). We also expressed and purified recombinant full-length FEM1C and FEM1B proteins, and found that full-length FEM1C binds to the Clone13 peptide with a K_d comparable with that of FEM1C_{1–390} (K_d of 4.9 versus 3.5 μM), and similarly, full-length FEM1B and FEM1B_{1–356} display comparable binding affinities toward the CDK5R1 peptide (K_d of 10 versus 6.0 μM) (Supplementary Table 1). In summary, our binding data unambiguously demonstrate that although all FEM1 proteins directly interact with the Arg/C-degron, FEM1A/C and FEM1B exhibit distinct binding properties.

Structure of FEM1C bound to SIL1 Arg/C-degron peptide. To uncover the molecular basis for C-degron recognition by FEM1C, we sought to solve the structure of SIL1-bound FEM1C. Although the cocrystallization of FEM1C with the SIL1 peptide failed, we succeeded in the crystallization of a fusion construct by fusing FEM1C_{1–390} and the SIL1 C-degron (¹⁰SVNSLLKELR⁻¹) with a (GGGS)₄ linker and solved its crystal structure at a 2.9 Å resolution (Supplementary Table 2). There are two molecules in one asymmetric unit, which are almost identical to a root-mean-square deviation (r.m.s.d.) of 0.43 Å over 370 C α atoms.

Overall, FEM1C_{1–390} adopts a crescent-shape architecture, which consists of seven ankyrin repeats (ANK1–7) immediately followed by three TPRs (TPR1–3) (Fig. 1f). Most of residues in FEM1C_{1–390} are visible except residues at the N or C terminus, or in flexible loop regions. The seven ankyrin repeats spanning residues 1–240 of FEM1C, are well aligned and spaced uniformly, with each repeat folding into a helix–loop–helix structural motif (Fig. 1f and Extended Data Fig. 1a). However, some unique structure features exist within the FEM1C ankyrin repeats. First, there is a helix insertion ($\alpha 3$) between ANK1 and ANK2; second, ANK2 is connected with ANK3 by antiparallel beta strands ($\beta 1\text{--}\beta 2$) rather than a hairpin loop, with $\beta 1\text{--}\beta 2$ packing with the first helix of ANK2 and ANK3 (Fig. 1f and Extended Data Fig. 1a). Three TPRs are connected with ankyrin repeats via the short linker between ANK7 and TPR1. TPR1 and TPR3, but not TPR2, adopt a canonical TPR fold, given that the first helix of TPR2 ($\alpha 18$) is much shorter. The TPRs stabilize the ankyrin repeats primarily via the contact between TPR1 and ANK7 (Fig. 1f and Extended Data Fig. 1a).

Bipartite recognition of the SIL1 C-degron by FEM1C. The last seven residues of the fusion protein, ⁷SLLKELR⁻¹, corresponding to the C terminus of SIL1, are visible in the electron density map (Extended Data Fig. 2a), which folds into a 3₁₀ helix and is positioned into a concave formed by $\beta 1$, $\beta 2$ and ANK3–6 of FEM1C (Figs. 1f and 2a). The C-terminal Arg (R⁻¹) is inserted into a negatively charged pocket of FEM1C, named p1 pocket hereafter, comprising F76, D77, W88, S117, R121, F125 and D126 (Fig. 2b). The guanidino group of R⁻¹, which forms hydrogen bonds with the side chains of D77 and D126, is sandwiched between F76 and F125, stacking with the ring plane of F125 via cation– π interaction (Fig. 2b,c). R⁻¹ also forms two main chain hydrogen bonds with the side chains of S117 and R121, respectively, and makes hydrophobic contacts with W88 via its aliphatic chain (Fig. 2b,c). Taken together, the cation– π and hydrogen-bonding interactions constitute the R⁻¹ specific recognition.

We found that the K⁻⁴ of SIL1 is accommodated into another negatively charged pocket of FEM1C, named p2 pocket hereafter, with its N ϵ group forming three hydrogen bonds with the side chains of N183, D188 and E191, respectively. The aliphatic chain of K⁻⁴ stacks with the side chains of H150 and Y158 (Fig. 2b,c). In addition to R⁻¹ and K⁻⁴, the L⁻² of SIL1 also makes hydrophobic

contact with T115 and H148, with its main chain carbonyl group hydrogen bonded to R121; the E⁻³ forms salt bridge with K159, while the L⁻⁵ makes hydrophobic contact with H148 (Fig. 2b,c). In contrast to ⁻⁵LKELR⁻¹, neither L⁻⁶ nor S⁻⁷ directly interacts with FEM1C residues, and this is supported by ITC binding data showing that peptides spanning the last 5–10 residues of SIL1 bind to FEM1C with comparable affinities ($K_d = 7\text{--}13\ \mu\text{M}$) (Supplementary Tables 1 and 3).

To pinpoint key residues within the SIL1 C-degron that are essential for FEM1C binding, we generated several single mutants of the SIL1 peptide, including R-1K, R-1A, K-4R and K-4A, and tested their FEM1C-binding affinities by ITC. ITC data showed that R-1K, R-1A and K-4A disrupt the binding, whereas K-4R demonstrated slightly reduced FEM1C-binding affinity ($K_d = 19\ \mu\text{M}$) (Supplementary Table 3). Given that the FEM1C p1 pocket accommodates R⁻¹ rather than K⁻¹, it is the nature of the R⁻¹ guanidino group rather than its positive charge that governs the Arg/C-degron recognition by FEM1C (Fig. 2b). Moreover, the p2 pocket of FEM1C favors a basic residue at the –4 position, such as K⁻⁴ or R⁻⁴.

Next, to define the FEM1C residues involved in SIL1 recognition, we made several mutants for FEM1C_{1–390}, including R121A, F125A, H148A, the double mutant D77A/D126A and the triple mutant N183A/D188A/E191A. By ITC binding assay, we found that H148A weakened the FEM1C-binding affinity by roughly six-fold (K_d of 68 versus 11 μM), whereas the other mutants abolished the binding (Supplementary Table 3). The mutagenesis and ITC binding assay further confirmed the bipartite recognition of SIL1 by FEM1C.

FEM1C–C-degron complexes define a -K/R-X-X-R consensus.

To determine whether the bipartite recognition mode also applies for other FEM1C-binding C-degrons, we further solved two more structures of FEM1C_{1–390} fused with NS11 and OR51B2 C-degrons, at resolutions of 2.38 and 2.51 Å, respectively (Supplementary Table 2). Overall, the two structures are similar to that of SIL1-bound FEM1C, with r.m.s.d.s in the range of 0.4–0.9 Å (Extended Data Fig. 3a,b). The last six residues of NS11 (⁻⁶WDKNLR⁻¹) and the last five residues of OR51B2 (⁻⁵HRFSR⁻¹) are visible in the two structures, respectively (Extended Data Fig. 2b,c).

In both structures, the R⁻¹ is inserted into p1 via cation– π and hydrogen-bonding interactions, and the K⁻⁴ (NS11) or R⁻⁴ (OR51B2) is positioned into p2 via hydrogen-bonding interactions (Fig. 3 and Extended Data Figs. 4a,b and 5a,b), demonstrating a similar bipartite recognition mechanism to that observed in the FEM1C–SIL1 structure. In the structure of FEM1C–OR51B2, while the R⁻⁴ of OR51B2 maintains the electrostatic interactions with the p2 pocket of FEM1C, the F⁻³ of OR51B2 partially occupied p2, with its aromatic ring sandwiched between side chains of Y158 and R⁻⁴, thereby forming $\pi\text{--}\pi$ and cation– π interactions with Y158 and R⁻⁴, respectively (Extended Data Figs. 5a,b and 6a). However, the alanine substitution of F⁻³ only slightly reduces the FEM1C-binding affinity (K_d of 11 versus 7.4 μM , Supplementary Table 3), suggesting that the F⁻³ is dispensable for the FEM1C binding. In contrast, p1 pocket mutants of FEM1C, including R121A, F125A and D77A/D126A bind to the OR51B2 peptide roughly 12–31-fold weaker than the wild-type protein (K_d of 7.4 versus 91–230 μM), underscoring the critical role of the p1 residues in R⁻¹ recognition (Supplementary Table 3). Overall, the two structures further define the bipartite recognition of the -K/R-X-X-R consensus by FEM1C (Extended Data Fig. 1b).

Unveiling the -R-X-R binding mode by FEM1C. In above solved structures, R⁻¹ and K/R⁻⁴ binding pockets are spaced by two residues. To study the impact of the special distance on the FEM1C-binding affinity, we synthesized a SIL1 peptide variant ending with KAR (YSVNSLLKAR⁻¹) and examined its FEM1C-binding affinity by ITC. The binding data show that the variant binds to FEM1C

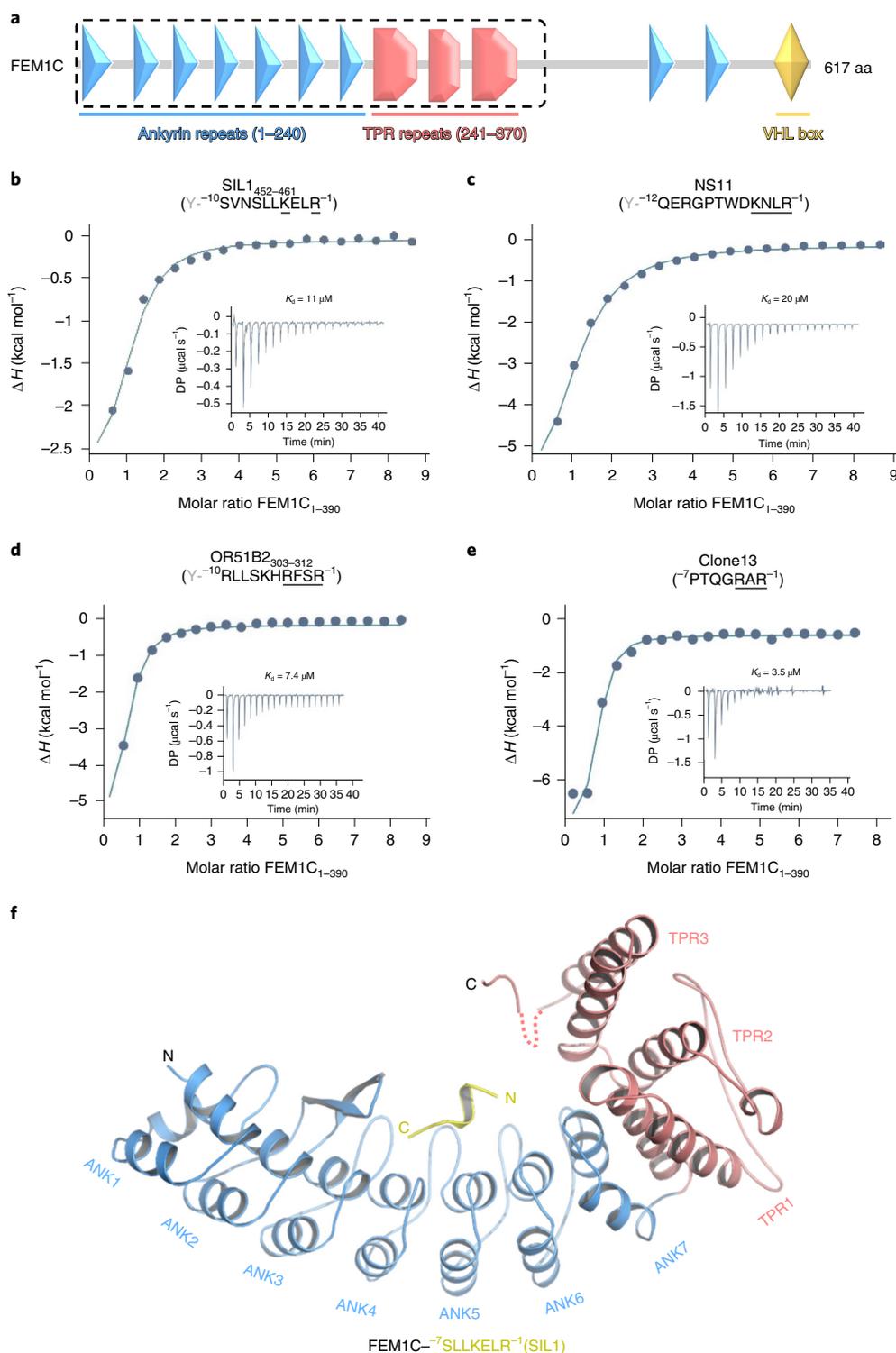


Fig. 1 | FEM1C₁₋₃₉₀ specifically recognizes peptides containing a C-terminal -K/R-X₁₋₂-R motif. a, Domain architecture of FEM1C₁₋₃₉₀, with the ankyrin repeats, the TPRs and the von Hippel–Lindau box shown in blue, red and orange, respectively. aa, amino acid. **b–e**, ITC binding curves for FEM1C₁₋₃₉₀ binding to peptides. **b**, SIL1₄₅₂₋₄₆₁. **c**, NS11. **d**, OR51B2₃₀₃₋₃₁₂. **e**, Clone13. **f**, Overall structure of FEM1C₁₋₃₉₀ in fusion with the C-degron of SIL1. The ankyrin repeats (ANK1–7) and TPRs (TPR1–3) of FEM1C and the SIL1 peptide are shown in blue, red and yellow, respectively. The invisible loop region is indicated with red dashes.

roughly 4.7-fold stronger than the wild-type SIL1 (K_d of 2.3 versus 11 μM) (Supplementary Table 3). Consistently, the Clone13 peptide ending with RAR binds to FEM1C roughly threefold stronger than the SIL1 peptide (K_d of 3.5 versus 11 μM) (Supplementary Table 3).

In sum, our findings suggest that FEM1C favors -K/R-X-R over -K/R-X-X-R.

To better understand an enhanced binding affinity between -K/R-X-R C-degron and FEM1C, we determined the structure

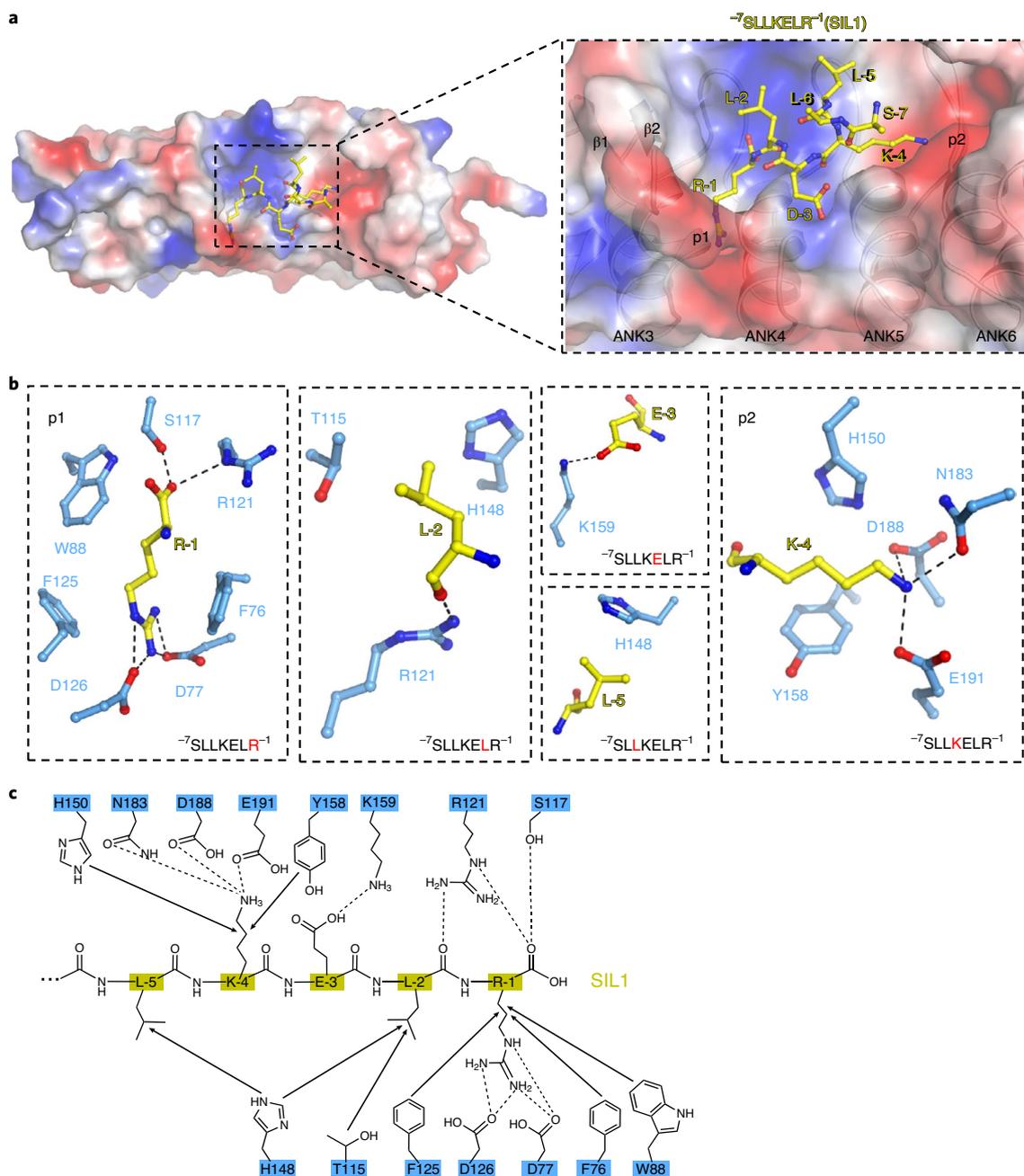


Fig. 2 | Structure of the FEM1C₁₋₃₉₀ with SIL1 C-degron ($^{-10}\text{SVNSLLKELR}^{-1}$). **a**, Left, the electrostatic surface of the FEM1C ankyrin repeats (1-240) bound with SIL1; right, SIL1 interacts with the residues of ankyrin repeats 3-6 in FEM1C, with its Arg⁻¹ and Lys⁻⁴ positioned into two negatively charged pockets. The peptide residues are shown in yellow sticks and labeled. **b**, Detailed interactions of Arg⁻¹, Leu⁻², Glu⁻³, Lys⁻⁴ and Leu⁻⁵ with FEM1C residues; peptide residues and FEM1C residues are shown as yellow and blue sticks, respectively. The intermolecular hydrogen bonds are indicated with black dashes. **c**, Interaction diagram between FEM1C and the SIL1 peptide, with peptide residues and FEM1C residues labeled in yellow and blue, respectively. Intermolecular hydrogen bonds and hydrophobic interactions are indicated with black dashes and black arrows, respectively.

of FEM1C fusion protein ending with the Clone13 peptide (Supplementary Table 3 and Extended Data Fig. 3c). The last six residues, $^{-6}\text{TQGRAR}^{-1}$ are visible in the solved structure (Extended Data Fig. 2d). The R⁻¹ and R⁻³ are recognized by the p1 and p2 pockets of FEM1C, respectively, indicating that the bipartite recognition also applies for the RAR⁻¹ motif (Fig. 3 and Extended Data Figs. 4c and 5c,d). Superposition of the FEM1C-Clone13 structure with that of SIL1-bound one indicates that the R⁻³ of Clone13 is closer to the aromatic ring of Y158, demonstrating stronger cation- π interaction (Extended Data Figs. 5d and 6b). Presumably, the replacement of

R⁻³ with K reinforces cation- π and the hydrogen-bonding interactions, accounting for the preference of FEM1C for peptides ending with -K/R-X-R.

By using ITC binding assay, we found that Y158A mutant of FEM1C severely weakened the binding to SIL1 and Clone13 by roughly 13-fold (K_d of 150 versus 11 μM) and 68-fold (K_d of 240 versus 3.5 μM), respectively (Supplementary Table 3), confirming the critical role of Y158 in maintaining the higher binding affinity with K/R⁻³. Further superposition of all four FEM1C structures show that the R⁻¹ recognition by p1 pocket is absolutely conserved,

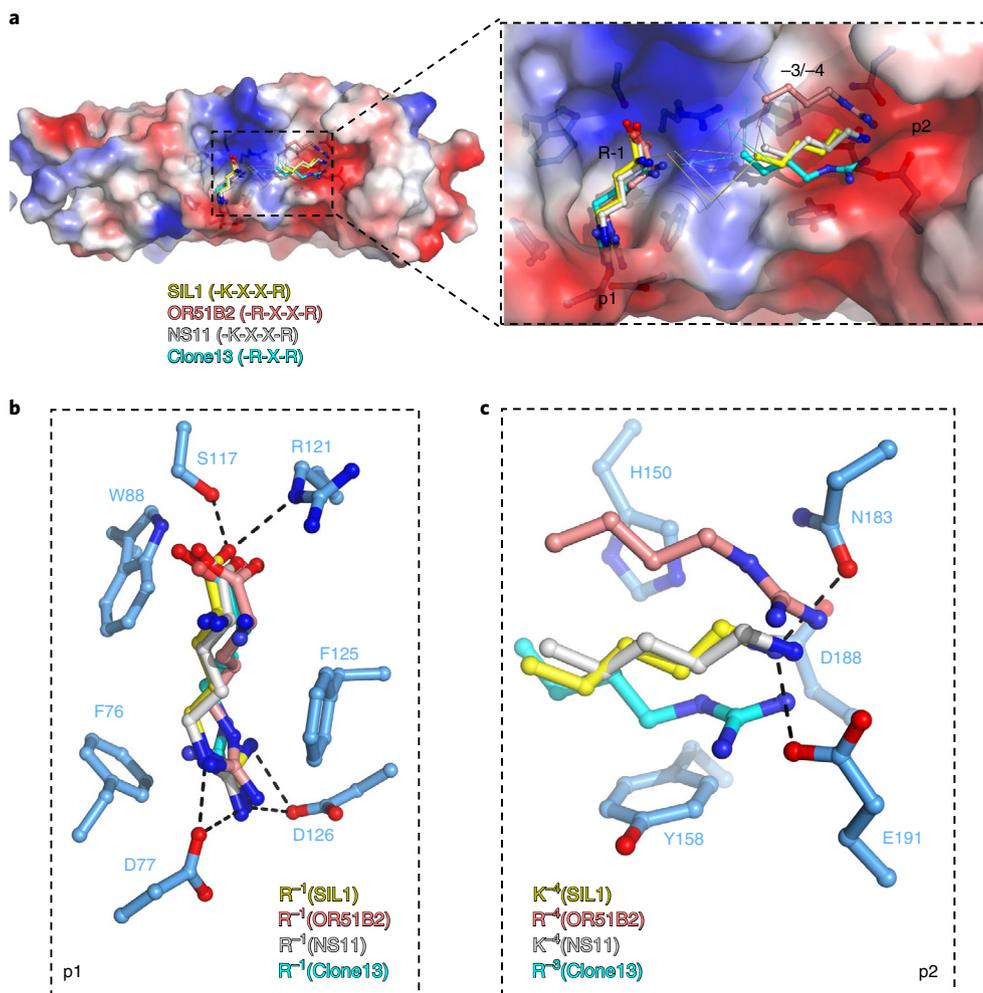


Fig. 3 | The consensus recognition mode of -K/R-X_{1,2}-R by FEM1C. **a**, The structures of NS11-bound FEM1C, OR51B2-bound FEM1C and Clone13-bound FEM1C are superimposed with that of SIL1-bound FEM1C. The electrostatic surface of the FEM1C ankyrin repeats (1–240) is shown. SIL1, OR51B2, NS11 and Clone13 are shown as yellow, red, gray and cyan ribbons, respectively. R⁻¹ of all peptides, K⁻⁴ of SIL1 and NS11, R⁻⁴ of OR51B2, as well as R⁻³ of Clone13, are shown as sticks. **b**, The four degron-bound FEM1C structures are superimposed to indicate the conserved Arg-1 binding mode. Arg⁻¹ of degrons are shown in a way similar to that shown in **a**, and Arg⁻¹ binding residues of FEM1C are shown as blue sticks. **c**, The four degron-bound FEM1C structures are superimposed to indicate that the p2 pocket preferentially accommodates a basic residue (Lys or Arg) at the –3 or –4 position, as shown in a way similar to that of **b**.

whereas the recognition of a basic residue at the –3 or –4 position by p2 reflects the structural plasticity for the bipartite binding (Fig. 3b,c). Collectively, the p1 and p2 pockets of FEM1C coordinately confer the preference for -K/R-X_{1,2}-R (Extended Data Fig. 1b). It is noteworthy that both FEM1C pockets are conserved in FEM1A (Extended Data Fig. 1a), suggesting that the bipartite binding mode also applies for FEM1A.

On the basis of the uncovered bipartite recognition mode, we identified >20 proteins that contain a C-terminal -K/R-X_{1,2}-R as potential substrates for Cul2^{FEM1C} based on our previous work¹⁷ (Supplementary Table 4). Two olfactory receptors, OR13J1 (-RASR⁻¹) and OR2T33 (-RSR⁻¹)²⁷, were also found to contain a -R-X_{1,2}-R motif, suggesting that they might be regulated by Cul2^{FEM1C} (ref. 17). Future work is needed to validate that full-length proteins corresponding to the C-terminal peptides that scored as CRL substrates are indeed FEM1C substrates.

Tolerance of FEM1C for residues downstream to C-Arg. By analyzing all degron-bound FEM1C structures, we found that although one of terminal carboxyl group of R⁻¹ is hydrogen bonded to the

FEM1C S117, R⁻¹ might still use the other carboxyl group to connect a downstream residue (Fig. 3b). Consistently, the ITC binding data show that FEM1C binds weakly to another SIL1 peptide variant ending with RG⁻¹ (⁻¹¹SVNSLLKELRG⁻¹) (K_d = 30 μM) (Supplementary Table 3).

To understand how FEM1C tolerates an extra residue downstream of Arg/C-degron, we further solved the structure of a FEM1C fusion proteins ending with ⁻¹¹SVNSLLKELRG⁻¹ (Supplementary Table 3 and Extended Data Fig. 3d). In the structure, the last seven residues of the variant are visible in the density map (Extended Data Fig. 2e), with its R⁻² and K⁻⁵ positioned into the p1 and p2 pockets, respectively. Comparing the RG-bound structure with the SIL1-bound one, we found that besides R⁻² and K⁻⁵, G⁻⁴ and L⁻⁶ also maintain their conformations (Extended Data Fig. 5e,f). However, the RG-bound structure is different from the SIL1-bound one in two aspects. First, G⁻¹, the extra C-terminal residue, forms two hydrogen bonds via its terminal carboxyl groups with the side chains of T115 and N146, respectively (Extended Data Figs. 4d and 5f). Second, L⁻³ alters its conformation to avoid potential steric clash with G⁻¹, resulting in not only the disruption of main

chain hydrogen bonds with R121, but also the loss of the hydrophobic contacts with the T115 and H148 of FEM1C (Extended Data Figs. 5f and 6c). The unfavorable conformational change of L⁻³ of the SIL1 variant might account for its weaker FEM1C-binding affinity (Supplementary Table 3).

Consistently, we found that addition of Ala at C terminus of SIL1 decreases the FEM1C-binding affinity by roughly 4.5-fold (K_d of 52 versus 11 μM), whereas addition of Val or Lys at its C terminus abolishes the binding. An Arg downstream of the SIL1 reduced the binding affinity by roughly eightfold (K_d of 89 versus 11 μM) by converting the C-degron to -K-X-X-X-R (Supplementary Table 3). Additionally, the addition of diglycine at the SIL1 C terminus further weakened the FEM1C-binding affinity by more than eightfold (K_d of 94 versus 11 μM) (Supplementary Table 3). Taken together, our structure analysis, supported by mutagenesis and biochemical experiments, is consistent with the work from Lin et al.¹⁸, and demonstrates that although FEM1C tolerates residues downstream of Arg/C-degron, the degree of tolerance is negatively correlated with the number and the size of the amino acid added.

Recognition of CDK5R1 Arg/C-degron by FEM1B and FEM1C.

ITC binding data show that FEM1B selectively binds to the CDK5R1 C-degron sixfold stronger than FEM1C (K_d of 6.0 versus 36 μM); however, in contrast with FEM1A/C, it does not bind to other Arg/C-degrons tested here (Extended Data Fig. 7a,b and Supplementary Table 1). To unveil the recognition mechanism of CDK5R1 by FEM1B, we first solved a fragment of FEM1B (1–356) at a resolution of 3.25 Å (Supplementary Table 2 and Supplementary Fig. 1a). In this structure, FEM1B_{1–337}, spanning N-terminal ankyrin repeats (ANK1–7) and TPRs (TPR1–2), represents the minimal structured region. The last six residues of FEM1B_{1–356}, ³⁵¹GAVYAD³⁵⁶, are found to occupy a concave pocket that represents the counterpart of the p1 pocket of FEM1C, probably due to the crystal packing (Supplementary Fig. 1b).

Inspired by the successful crystallization of FEM1C fusion proteins, we fused FEM1B_{1–337} and the CDK5R1 C-degron (⁻¹⁰KKRLLGLDR⁻¹) with a (GGGS)₂ linker, and solved the structure of the fusion protein at a 3.50 Å resolution (Fig. 4a and Supplementary Table 2). In the solved structure, the last eight residues of CDK5R1 are visible and exhibit an extended conformation, contacting with the ANK3–7 of FEM1B (Fig. 4b and Extended Data Fig. 2f). Similar to the bipartite recognition of Arg/C-degron by FEM1C, the R⁻¹ of CDK5R1 interacts with the p1 pocket of FEM1B mainly through cation- π and hydrogen-bonding interactions, while the L⁻³ is positioned into a hydrophobic pocket of FEM1B composed of Y163, F193 and L⁻⁶ (Fig. 4c). The CDK5R1 residues upstream of ⁻³LDR⁻¹ also interact with FEM1B. Especially the G⁻⁴, whose main chain is hydrogen bonded to N151, makes van der Waals interaction with the Y153 of FEM1B. The side chains of L⁻⁵ and L⁻⁷, which are invisible due to lower resolution of the structure, likely make hydrophobic contacts with Y153, and C186 and H218, respectively; R⁻⁸ forms a salt bridge with the side chain of E196 (Fig. 4c and Extended Data Fig. 4e).

To validate the FEM1B–CDK5R1 interface, we generated several CDK5R1 mutants and examined their FEM1B binding affinities by ITC. The binding data show that R-1K, R-1A and L-3A abolished or reduced the FEM1B binding, underscoring the important roles of R⁻¹ and L⁻³. While the G-4A mutant decreased the binding affinity by 11-fold (K_d of 66 versus 6.0 μM), G-4V disrupt the binding (Supplementary Table 3), suggesting that any bulky residue at the -4 position is disfavored because of potential steric clash with the FEM1B Y153 (Fig. 4c). Additionally, L-6A, L-7A and the double mutant L-5A/R-8A decrease the FEM1B binding affinity by roughly 3–7.5-fold (K_d of 18–45 versus 6.0 μM) (Supplementary Table 3), confirming the contribution of the residues upstream of ⁻⁴GLDR⁻¹ in FEM1B binding. Next, we generated

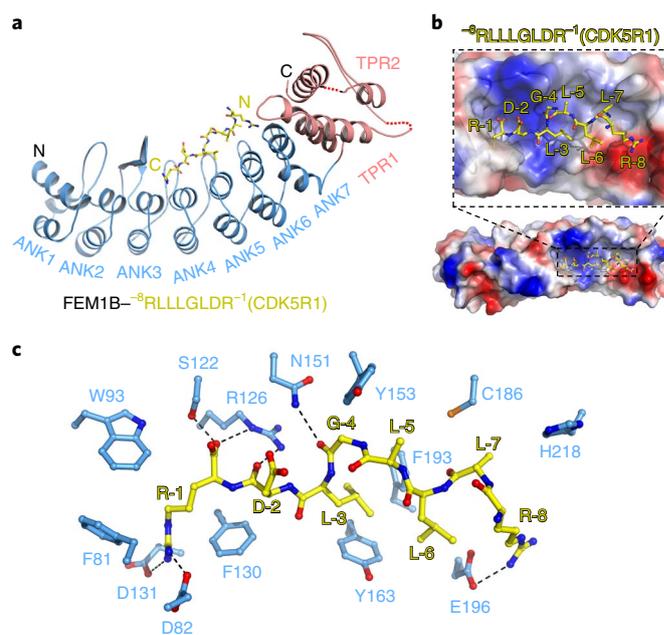


Fig. 4 | FEM1B selectively binds to the C-degron of CDK5R1. **a**, Overall structure of FEM1B_{1–356} fused with a CDK5R1 C-degron, with the ankyrin repeats and TPRs shown in blue and red cartoon representations, respectively. Invisible loop regions are indicated with red dashes. CDK5R1 are shown in yellow sticks. **b**, The electrostatic surface of the FEM1B ankyrin repeats (1–243) bound with CDK5R1, with CDK5R1 shown in yellow sticks. **c**, Detailed interactions between FEM1B and CDK5R1. CDK5R1 and CDK5R1 binding residues of FEM1B, are shown as blue and yellow sticks, respectively.

several FEM1B mutants and examine their CDK5R1 binding affinities by ITC. F130A and the double mutant D82A/D131A abolished the binding, whereas another double mutant Y163A/F193A reduced the binding affinity by >50-fold (K_d > 300 versus 6.0 μM) (Supplementary Table 3), further validating the FEM1B–CDK5R1 interface. Collectively, our data indicate that FEM1B prefers the Arg/C-degron ending with -G-L-X-R and interaction requires an additional upstream sequence.

We also solved the structure of FEM1C fused with the C-degron of CDK5R1 at a 2.35 Å resolution (Supplementary Table 2 and Extended Data Figs. 2g and 3e). In the structure, CDK5R1 binds to FEM1C in a way similar to that in the FEM1B–CDK5R1 structure, with its R⁻¹ and Leu⁻³ positioned into the p1 and p2 pockets, respectively (Extended Data Fig. 5g,h). Mutagenesis and ITC assay also confirmed the essential role of R⁻¹ in FEM1C binding (Supplementary Table 3).

p2 pocket governs the distinct FEM1-binding properties.

Superposition of two CDK5R1-bound FEM1 structures indicates that despite the conserved p1 pockets (Fig. 5a), FEM1B and FEM1C possess distinct p2 pocket residues. While the Y158 and E191 of FEM1C are conserved in FEM1B, H150, N183 and D188 of FEM1C are replaced in FEM1B by N155, A188 and F193, respectively (Fig. 5b and Extended Data Fig. 1a). The N183 and D188, and H150 of FEM1C make hydrogen-bonding and van der Waals interactions with the K/R at the -3 or -4 position, respectively (Fig. 3c). The corresponding FEM1B replacements would not only disrupt the binding, but also cause the steric clash with the aliphatic chain of K/R (Fig. 5b), well accounting for the nondetectable FEM1B binding affinity to -K/R-X_{1,2}-R degrons (Supplementary Table 3).

As for CDK5R1, the last four residues (⁻⁴GLDR⁻¹) were superimposed very well in the two structures, whereas conformations of

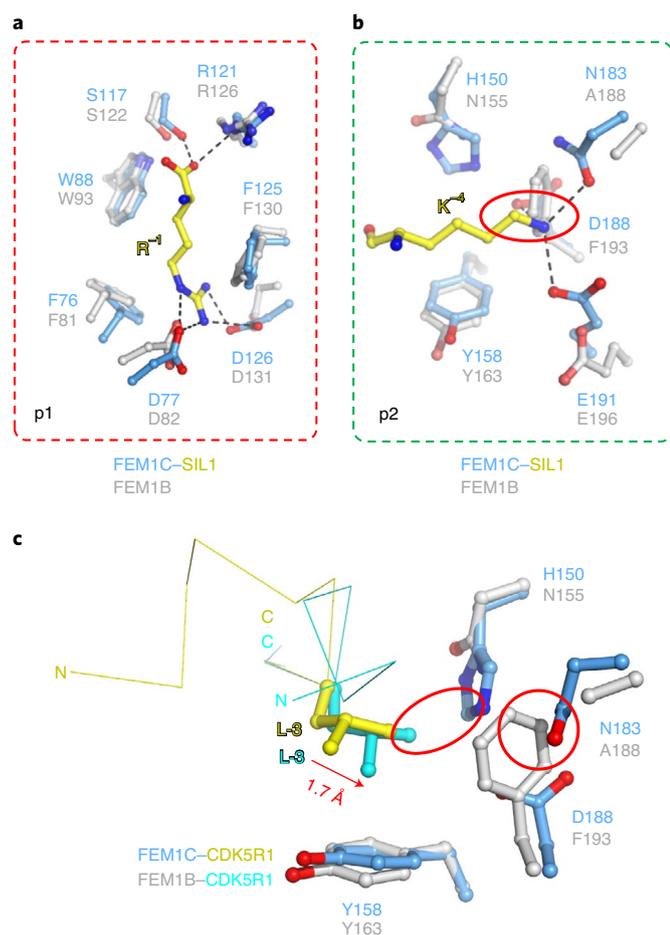


Fig. 5 | Key residues govern the different binding properties of FEM1B and FEM1C. **a**, The structure of FEM1B is superimposed with that of SIL1-bound FEM1C to indicate the conserved p1 pocket. R⁻¹ is shown in yellow sticks. R⁻¹ binding residues of FEM1C and FEM1B, are shown in blue and gray, respectively. **b**, p2 pockets of FEM1B and FEM1C are superimposed, with nonconserved residues highlighted. The K⁻⁴ of SIL1 is shown in yellow sticks. K⁻⁴ binding residues of FEM1C and their counterparts in FEM1B are shown in blue and gray, respectively. The red circle indicates the potential steric clash between the SIL1 K⁻⁴ and the F193 of FEM1B. **c**, The FEM1B-CDK5R1 structure is superimposed with that of FEM1C-CDK5R1 to indicate the interaction between the FEM1B p2 pocket and L⁻³. In both structures, CDK5R1 are shown in yellow and cyan ribbon, respectively, with its L⁻³ shown in sticks. L⁻³ binding residues of FEM1C and FEM1B, are shown in blue and gray sticks, respectively. Three nonconserved residues, H150, N183 and D188 of FEM1C, are highlighted.

⁻⁸RLLL⁻⁵ deviate (Fig. 5c). In the FEM1B-CDK5R1 structure, the L⁻³ side chain slides roughly 1.7 Å deeper into p2, resulting in an increased hydrophobic interaction with F193 and Y163. In contrast, the access of L⁻³ to p2 is restricted by the side chain of H150 in the FEM1C-CDK5R1 structure (Fig. 5c).

To convert the p2 of FEM1C into its FEM1B counterpart, we generated a FEM1C triple mutant, H150N/N183A/D188F, and examined its C-degron-binding affinities. ITC binding data show that the FEM1C mutant binds to SIL1 roughly 12.5-fold weaker than wild type (Fig. 1b and Extended Data Fig. 7c), and binds to CDK5R1 roughly 6.5-fold stronger than wild type (Extended Data Fig. 7b,d), highly reminiscent of the C-degron-binding properties of FEM1B (Extended Data Fig. 7a,d). Collectively, we confirmed that the distinct binding properties of FEM1C and FEM1B result

from their p2 pockets by altering the preference of FEM1C from the C-degron of SIL1 (-K-X-X-R) to that of CDK5R1 (-G-L-X-R) with a triple mutant of FEM1C(H150N/N183A/D188F) (Extended Data Fig. 7c,d).

Comparison of FEM1 structures with relevant complexes.

Previously, the structures of the p62 ZZ domain with N-Arg^{28,29}, and the SND1 extended tudor domain with arginine were reported^{30,31}, enabling us to compare their arginine recognition modes with that by FEM1s. Like FEM1C, the SND1 recognizes arginine via both cation- π and hydrogen-bonding interactions (Extended Data Fig. 8a,b), whereas the ZZ domain of p62 recognizes N-Arg via main chain hydrogen bonds and side chain salt bridges (Extended Data Fig. 8c).

Recently, the solved structure of KLHDC2 with Gly/C-degron has uncovered the recognition of diglycine by KLHDC2 (refs. 17,18,32). By comparing the FEM1C-SIL1 structures with that of diglycine-bound KLHDC2 we found that the two C-degron recognition modes are different in several aspects. (1) FEM1C and KLHDC2 bind to degrons via ankyrin repeats and Kelch domain, respectively. (2) FEM1C recognizes the Arg/C-degron via a bipartite mechanism (Extended Data Fig. 9a), whereas KLHDC2 recognizes GG⁻¹ via a diglycine-specific cavity, which is too small to accommodate any bulky residues (Extended Data Fig. 9b). (3) FEM1C tolerates residues downstream of Arg/C-degron, albeit with weaker binding affinities, because only one terminal carbonyl group of R⁻¹ is hydrogen bonded to FEM1C residues (Extended Data Fig. 9a). In the KLHDC2 complex structures, both terminal carbonyl groups of G⁻¹ are locked by intermolecular hydrogen bonds, indicating the intolerance of extra C-terminal residues (Extended Data Fig. 9b).

GPS assay with FEM1s and their degron-binding mutants.

To validate the functional roles of the C-degron-binding residues in FEM1s, we constructed a series of FEM1 mutants in lentiviral expression vector and tested their activities in cell-based system, called GPS^{17,18,33,34}. GPS is a fluorescent-based reporter system that examine changes in protein stability in live cells. This system is based on a lentiviral expression vector that encodes for DsRed, serving as internal expression reference for the lentiviral cassette and a green-fluorescent protein- (GFP-) fusion that is translated from an internal ribosome entry site. In this method, the GFP to DsRed ratio serves to readout the effect of the fusion partner on the stability of GFP¹⁷.

To this end, we engineered a few representative GPS reporters encoding 10mer C-degron peptides of various motifs. GPS-SIL1 (-R-X-X-R), GPS-OR51B2 (-K-X-X-R) and GPS-Clone13 (-R-X-R) were chosen as FEM1C representative substrates, while GPS-CDK5R1 (-G-L-X-R) was chosen as a FEM1B substrate. For each C-degron, we constructed both wild type and mutant peptides, in which the critical C-degron-binding residues identified by the structural analysis and ITC binding assay were substituted to alanine.

To generate cells stably overexpressing FEM1C and FEM1B, wild-type or degron-binding-pocket mutants, we transduced human embryonic kidney 293T (HEK293T) cells with lentiviral constructs encoding FEM1s (Supplementary Fig. 2a and 3). The expression levels of various exogenously expressed FEM1s were comparable judged by immunoblots (Supplementary Fig. 2b). Then, we individually introduced with lentiviruses the GPS peptides into FEM1s cells (Supplementary Fig. 2a). In each case, we compared the stability of the GFP-fused C-degron between wild-type and mutant FEM1s, with APPBP2 serving as a negative control CRL2 adapter.

In agreement with the solved structure and ITC experiments, for FEM1C, among the residues directly interacting with the C-degron in the p1 pocket, single alanine mutation of R121 and double alanine mutation of D77/D126 as well as the triple mutation of N183/D188/E191 in the p2 pocket, all failed to induce the degradation of all three substrates tested (Fig. 6a, upper panel). For FEM1B, the p1 pocket mutant F130A and the p2 pocket mutant Y163A/F193A

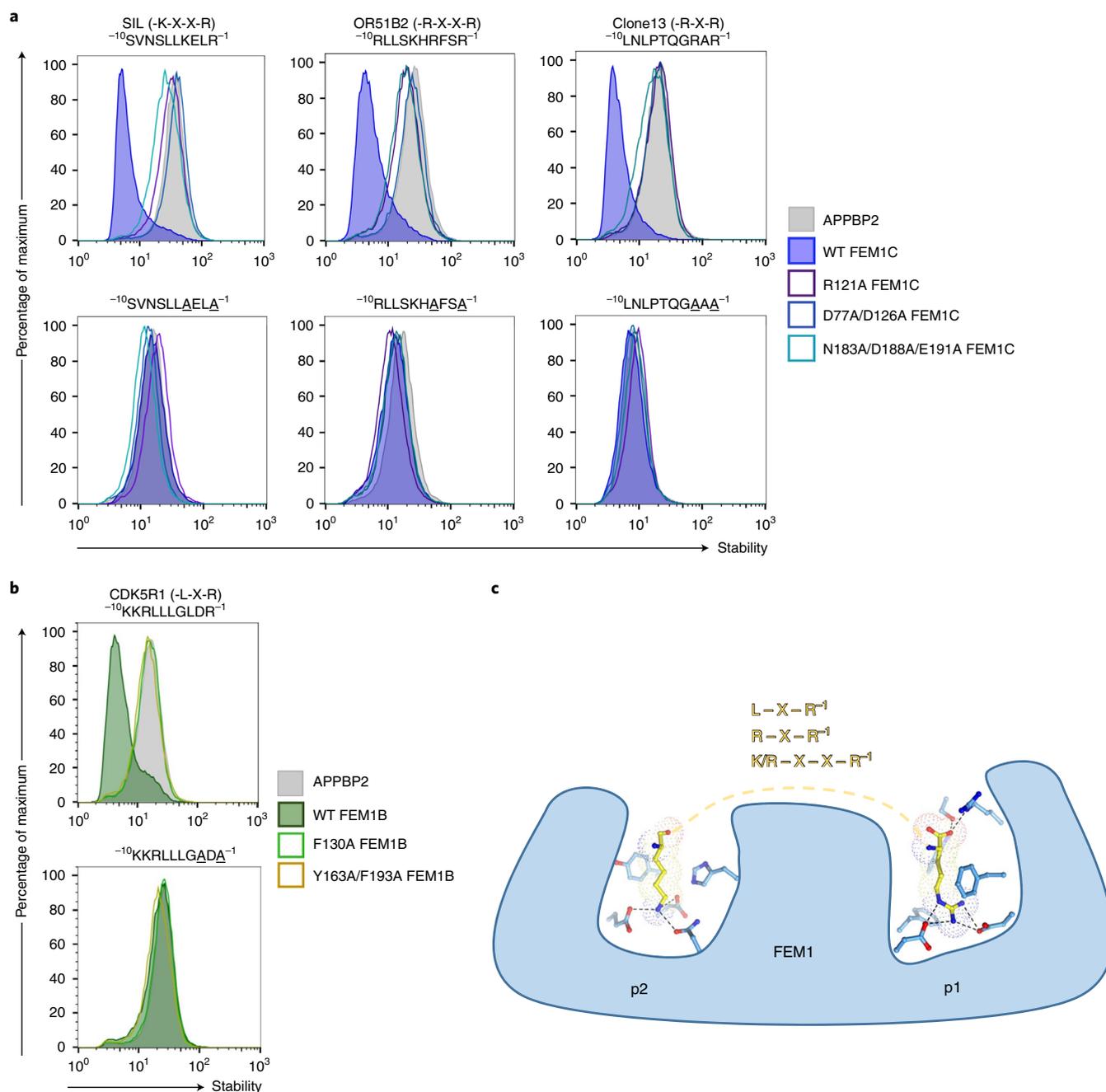


Fig. 6 | GPS assay to dissect critical residues in FEM1 protein degron-binding pocket. a, b, GPS assay to monitor stability of GFP-fused C-end degrons introduced into FEM1 stable cells. The GFP/DsRed ratio is used to indicate the stability of GFP-fused C-end degron and was analyzed by flow cytometry. Flow cytometry data were repeated independently in triplicates with similar results. In each case, the GFP-fused degron is destabilized in cells overexpressing the wild-type (WT) but not the mutant FEM1 proteins. Stability analysis of the FEM1C 10mer peptide substrates, GFP-fused SIL1, OR51B2 and Clone13 (**a**) or the FEM1B substrate, GFP-fused CDK5R1 (**b**). **c,** Proposed model for bipartite recognition of Arg/C-degrons by Cul2^{FEM1}.

lost their function in degrading CDK5R1 (Fig. 6b, upper panel). Another p1 pocket mutant of FEM1B, R126A, was only partially silent (Supplementary Fig. 2c) indicating that R126 is important, but not essential for FEM1B–CDK5R1 interaction. This is also in agreement with ITC results that F130A disrupted the binding and Y163A/F193A double mutant severely impaired the binding, whereas R126A only diminished the interaction (Supplementary Table 3).

As expected, mutation of the identified critical residues in all tested C-degrons greatly impairs or abolishes the Cul2^{FEM1}-mediated degradation (Fig. 6a,b, lower panel), underscoring the bipartite

recognition of Arg/C-degron by Cul2^{FEM1} E3 machinery (Fig. 6c). Overall, the *in vivo* cell-based assays validate the functional roles of Arg/C-degron-binding residues of FEM1 proteins that were identified in structural experiments and *in vitro* ITC assays.

Discussion

In higher eukaryotes, UPS serves as the main route to guide specific proteolysis to maintain proteostasis^{40,35}. The specificity of the UPS system is dictated by the interactions between E3s or adapter proteins within the E3 complexes and the degrons within

substrates^{16,36–38}. Previous work identified FEM1A/B/C, KLHDC2/3/10 and APPBP2 as the adapter proteins in different Cul2 complexes that recognize distinct substrates via targeting specific C-degrons^{17,18}. Our structure-guided analyses delineate the binding modes and sequence selectivity underlying recognition of Arg/C-degrons by FEM1 ankyrin repeats, which extends early findings on recognition of ankyrin via its concave inner surface^{39–42}.

Our structural work, complemented by biochemical experiments and GPS assays, demonstrates that FEM1A/C and FEM1B prefer Arg/C-degrons ending with -K/R-X_{1,2}-R and -G-L-X-R, respectively. In contrast with the recognition of -K/R-X_{1,2}-R by FEM1C, the FEM1B–CDK5R1 interaction requires upstream residues of GLDR⁻¹, suggesting that it would be more challenging to identify potential substrates for Cul2^{FEM1B}. However, the role of FEM1B in mediating the turnover of CDK5R1 and other unknown regulators of cell fate might facilitate the understanding of the FEM1B function in stimulating apoptosis⁴³. In the future, it would be important to delineate how Cul2^{FEM1B} controls proteolysis in space and time during cell division and cell cycle⁴⁴.

In sum, our study provides structure insights underlying the Arg/C-degron recognition and the substrate selectivity accounting for the C-degron pathways. The selectivity of FEM1A/B/C revealed in this study not only provides an insight into the classic view that selective degradation orchestrates spatiotemporal regulation in different contexts, but also sheds light on the future design of chemically antagonizing molecules for Cul2^{FEM1A/B/C} to advance our understanding of their relevance in health and disease.

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/s41589-020-00704-3>.

Received: 2 April 2020; Accepted: 30 October 2020;

Published online: 04 January 2021

References

- Dikic, I. Proteasomal and autophagic degradation systems. *Annu. Rev. Biochem.* **86**, 193–224 (2017).
- Hershko, A., Ciechanover, A. & Varshavsky, A. Basic Medical Research Award. The ubiquitin system. *Nat. Med.* **6**, 1073–1081 (2000).
- Varshavsky, A. N-degron and C-degron pathways of protein degradation. *Proc. Natl Acad. Sci. USA* **116**, 358–366 (2019).
- Zheng, N. & Shabek, N. Ubiquitin ligases: structure, function, and regulation. *Annu. Rev. Biochem.* **86**, 129–157 (2017).
- Sarikas, A., Hartmann, T. & Pan, Z. Q. The cullin protein family. *Genome Biol.* **12**, 220 (2011).
- Petroski, M. D. & Deshaies, R. J. Function and regulation of cullin-RING ubiquitin ligases. *Nat. Rev. Mol. Cell Biol.* **6**, 9–20 (2005).
- Bosu, D. R. & Kipreos, E. T. Cullin-RING ubiquitin ligases: global regulation and activation cycles. *Cell Div.* **3**, 7 (2008).
- Jang, S. M. et al. The replication initiation determinant protein (RepID) modulates replication by recruiting CUL4 to chromatin. *Nat. Commun.* **9**, 2782 (2018).
- Li, L. et al. The CRL3(BTBD9) E3 ubiquitin ligase complex targets TNFAIP1 for degradation to suppress cancer cell migration. *Signal Transduct. Target Ther.* **5**, 42 (2020).
- Rape, M. Ubiquitylation at the crossroads of development and disease. *Nat. Rev. Mol. Cell Biol.* **19**, 59–70 (2018).
- Tank, E. M. & True, H. L. Disease-associated mutant ubiquitin causes proteasomal impairment and enhances the toxicity of protein aggregates. *PLoS Genet.* **5**, e1000382 (2009).
- Dey, A. et al. Loss of the tumor suppressor BAP1 causes myeloid transformation. *Science* **337**, 1541–1546 (2012).
- Zhao, Y. & Sun, Y. Cullin-RING ligases as attractive anti-cancer targets. *Curr. Pharm. Des.* **19**, 3215–3225 (2013).
- Bachmair, A., Finley, D. & Varshavsky, A. In vivo half-life of a protein is a function of its amino-terminal residue. *Science* **234**, 179–186 (1986).
- Timms, R. T. et al. A glycine-specific N-degron pathway mediates the quality control of protein N-myristoylation. *Science* **365**, eaaw4912 (2019).
- Chen, S. J., Wu, X., Wadas, B., Oh, J. H. & Varshavsky, A. An N-end rule pathway that recognizes proline and destroys gluconeogenic enzymes. *Science* **355**, eaal3655 (2017).
- Koren, I. et al. The eukaryotic proteome is shaped by E3 ubiquitin ligases targeting C-terminal degrons. *Cell* **173**, 1622–1635 e14 (2018).
- Lin, H. C. et al. C-terminal end-directed protein elimination by CRL2 ubiquitin ligases. *Mol. Cell* **70**, 602–613 e3 (2018).
- Ravalin, M. et al. Specificity for latent C termini links the E3 ubiquitin ligase CHIP to caspases. *Nat. Chem. Biol.* **15**, 786–794 (2019).
- Timms, R. T. & Koren, I. Tying up loose ends: the N-degron and C-degron pathways of protein degradation. *Biochem. Soc. Trans.* **48**, 1557–1567.
- Chatr-Aryamontri, A., van der Sloot, A. & Tyers, M. At long last, a C-terminal bookend for the ubiquitin code. *Mol. Cell* **70**, 568–571 (2018).
- Mosavi, L. K., Cammett, T. J., Desrosiers, D. C. & Peng, Z. Y. The ankyrin repeat as molecular architecture for protein recognition. *Protein Sci.* **13**, 1435–1448 (2004).
- Mahrour, N. et al. Characterization of cullin-box sequences that direct recruitment of Cul2-Rbx1 and Cul5-Rbx2 modules to Elongin BC-based ubiquitin ligases. *J. Biol. Chem.* **283**, 8005–8013 (2008).
- Dankert, J. F., Pagan, J. K., Starostina, N. G., Kipreos, E. T. & Pagano, M. FEM1 proteins are ancient regulators of SLBP degradation. *Cell Cycle* **16**, 556–564 (2017).
- Senderek, J. et al. Mutations in SIL1 cause Marinesco-Sjogren syndrome, a cerebellar ataxia with cataract and myopathy. *Nat. Genet.* **37**, 1312–1314 (2005).
- Anttonen, A. K. et al. The gene disrupted in Marinesco-Sjogren syndrome encodes SIL1, an HSPA5 cochaperone. *Nat. Genet.* **37**, 1309–1311 (2005).
- Trimmer, C. et al. Genetic variation across the human olfactory receptor repertoire alters odor perception. *Proc. Natl Acad. Sci. USA* **116**, 9475–9480 (2019).
- Kwon, D. H. et al. Insights into degradation mechanism of N-end rule substrates by p62/SQSTM1 autophagy adapter. *Nat. Commun.* **9**, 3291 (2018).
- Zhang, Y. et al. ZZ-dependent regulation of p62/SQSTM1 in autophagy. *Nat. Commun.* **9**, 4373 (2018).
- Liu, K. et al. Structural basis for recognition of arginine methylated Piwi proteins by the extended Tudor domain. *Proc. Natl Acad. Sci. USA* **107**, 18398–18403 (2010).
- Liu, H. et al. Structural basis for methylarginine-dependent recognition of Aubergine by Tudor. *Genes Dev.* **24**, 1876–1881 (2010).
- Rusnac, D. V. et al. Recognition of the diglycine C-end degron by CRL2(KLHDC2) ubiquitin ligase. *Mol. Cell* **72**, 813–822 e4 (2018).
- Emanuele, M. J. et al. Global identification of modular cullin-RING ligase substrates. *Cell* **147**, 459–474 (2011).
- Yen, H. C., Xu, Q., Chou, D. M., Zhao, Z. & Elledge, S. J. Global protein stability profiling in mammalian cells. *Science* **322**, 918–923 (2008).
- Komander, D. & Rape, M. The ubiquitin code. *Annu. Rev. Biochem.* **81**, 203–229 (2012).
- Tasaki, T., Sriram, S. M., Park, K. S. & Kwon, Y. T. The N-end rule pathway. *Annu. Rev. Biochem.* **81**, 261–289 (2012).
- Tasaki, T. et al. The substrate recognition domains of the N-end rule pathway. *J. Biol. Chem.* **284**, 1884–1895 (2009).
- Dougan, D. A., Micevski, D. & Truscott, K. N. The N-end rule pathway: from recognition by N-recognins, to destruction by AAA+proteases. *Biochim. Biophys. Acta* **1823**, 83–91 (2012).
- Xu, C. et al. Sequence-specific recognition of a PxLPxI/L motif by an ankyrin repeat tumbler lock. *Sci. Signal* **5**, ra39 (2012).
- Wang, C. et al. Structural basis of diverse membrane target recognitions by ankyrins. *eLife* **3**, e04353 (2014).
- Collins, R. E. et al. The ankyrin repeats of G9a and GLP histone methyltransferases are mono- and dimethyllysine binding modules. *Nat. Struct. Mol. Biol.* **15**, 245–250 (2008).
- Nie, J. et al. Ankyrin repeats of ANKRA2 recognize a PxLPxL motif on the 3M syndrome protein CCDC8. *Structure* **23**, 700–712 (2015).
- Subauste, M. C. et al. RACK1 downregulates levels of the pro-apoptotic protein Fem1b in apoptosis-resistant colon cancer cells. *Cancer Biol. Ther.* **8**, 2297–2305 (2009).
- Liu, X. et al. Phase separation drives decision making in cell division. *J. Biol. Chem.* **295**, 13419–13431 (2020).

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 America, Inc. 2021

Methods

Cloning, mutations, protein expression and purification. Gene encoding FEM1B_{1–356} was amplified by PCR from a complementary DNA library and cloned into a modified pET28a vector fused with the open reading frame of yeast *SMT3* gene. FEM1C_{1–403} and FEM1A_{1–442} were amplified by PCR from cDNA library and cloned into pET28a-MHL vector (NCBI txid433626)⁴⁶. Plasmids for expressing fusion proteins, including FEM1C_{1–390}-(Gly-Gly-Gly-Ser)₁-peptide and FEM1B_{1–337}-(Gly-Gly-Gly-Ser)₂-CDK5R1_{298–307}, were constructed by overlap extension PCR and were cloned into the pET28a-MHL and pET28a-SUMO vectors, respectively.

All recombinant proteins were over-expressed in *Escherichia coli* BL21 (DE3). Cells were grown in LB medium at 37 °C until the optical density (OD₆₀₀) reached roughly 0.8. Protein expression was induced with 0.2 mM β-D-1-thiogalactopyranoside for 20 h at 16 °C. Cells were gathered by centrifuge at 3,600g for 10 min at 4 °C, and pellets were resuspended in lysis buffer containing 20 mM Tris, 400 mM NaCl, 2 mM imidazole, pH 7.5. Recombinant proteins were purified by Ni-NTA (GE healthcare), columns were then washed by 20 mM Tris, 400 mM NaCl, 50 mM imidazole, pH 7.5 and proteins were eluted by 20 mM Tris, 400 mM NaCl, 500 mM imidazole, pH 7.5. Gel filtration and ion exchange experiments were used for further purification. FEM1B_{1–356} and FEM1C_{1–403} were further purified by size exclusion using a HiLoad 16/600 Superdex 75 column (GE healthcare) in a buffer containing 20 mM Tris, 200 mM NaCl, pH 7.5, the corresponding fractions were then subjected to Hitrap Q HP (1 ml) and Hitrap SP HP (1 ml), respectively. Other peptide-fusion proteins were further purified by size exclusion using a HiLoad 16/600 Superdex 75 column (GE healthcare) in buffer containing 20 mM Tris, 400 mM NaCl, pH 7.5. Seleno-methionine (SeMet) labeled FEM1B_{1–356} was purified in the same way, except that cells were cultured in M9 medium supplied with 50 mg l⁻¹ of SeMet.

Site-specific mutations were performed using two reverse and complement primers containing the mutation codon. Primer sets used for mutations are listed in Supplementary Table 5. All mutants were purified in the same way as purification of wild-type proteins.

To express FEM1s in mammalian cells, plasmid vectors encoding cDNAs for FEM1B, FEM1C and APPBP2 were obtained from the Ultimate Open Reading Frame Clone collection (Thermo Fisher Scientific) and were subcloned into the lentiviral pHAGE-Flag-HA Gateway Destination vector via an LR recombination reaction (Thermo Fisher Scientific). Mutations of key residues in degron-binding pocket of FEM1s were generated by PCR-mediated site directed mutagenesis, using the QuikChange Lightning kit (Agilent) according to the manufacturer's protocol. All mutations were confirmed by Sanger sequencing. Primers sets used for mutations are listed in Supplementary Table 6.

Crystallization data collection and structure determination. All crystals were grown by using sitting drop vapor diffusion method at 18 °C. For crystallization, 1 μl of protein solution was mixed with 1 μl of crystallization buffer. All protein samples used for crystallization were in the same buffer containing 20 mM Tris-HCl, pH 7.5 and 400 mM NaCl and were concentrated in a concentration of 10–15 mg ml⁻¹. Crystals were obtained as follows:

- (1) FEM1C_{1–390}-(GGGS)₄-SIL1_{452–461} was cloned into pET28a-MHL vector and crystallized in a buffer containing 0.3 M lithium sulfate, 0.1 M Tris-HCl, pH 8.5, 5% (v/v) polyethylene glycol 400.
- (2) FEM1C_{1–390}-(GGGS)₄-OR51B2_{303–312} was cloned into pET28a-SUMO vector and crystallized in a buffer containing 0.1 M tris-HCl, pH 7.0, 0.8 M lithium sulfate.
- (3) FEM1C_{1–390}-(GGGS)₄-LNLPTQGRAR was cloned into pET28a-SUMO vector and crystallized in a buffer containing 0.1 M bis-tris propane, pH 6.4, 0.6 M lithium sulfate.
- (4) FEM1C_{1–390}-(GGGS)₄-CDK5R1_{298–307} was cloned into pET28a-SUMO vector and crystallized in a buffer containing 0.1 M tris-HCl, pH 7.9, 0.2 M lithium sulfate.
- (5) FEM1C_{1–390}-(GGGS)₄-QERGPTWTKNLR(NS11) was cloned into pET28a-SUMO vector and crystallized in a buffer containing 0.1 M Tris-HCl, pH 8.2, 0.2 M sodium citrate tribasic dehydrate.
- (6) FEM1C_{1–390}-(GGGS)₄-SIL1_{452–461} (R/RG) was cloned into pET28a-MHL vector and crystallized in a buffer containing 1.0 M sodium phosphate monobasic monohydrate, potassium phosphate dibasic, pH 6.9.
- (7) FEM1C_{1–403} was cloned into pET28a-MHL vector and crystallized in a buffer containing 0.1 M tris-HCl, pH 7.3, 0.2 M lithium sulfate.
- (8) FEM1B_{1–356} was cloned into pET28a-SUMO vector and crystallized in a buffer containing 0.1 M tris-HCl, pH 8.5, 0.1 M guanidine hydrochloride, 0.8 M potassium sodium tartrate tetrahydrate, 0.5% (w/v) polyethylene glycol monomethyl ether 5,000.
- (9) FEM1B_{1–337}-(GGGS)₂-CDK5R1_{298–307} was cloned into pET28a-SUMO vector and crystallized in a buffer containing 0.1 M bis-tris propane, pH 6.7, 1.6 M magnesium sulfate hydrate.

All protein crystals appeared in 4 d. Before flash freezing crystals in liquid nitrogen, all crystals were soaked for a short period in cryo-protectant containing glycerol and precipitant to about 30% (v/v). The diffraction data were collected on beamline BL17U1 and BL18U1 at Shanghai Synchrotron Facility at 0.9791 or

0.9792 Å (ref. ⁴⁷). The statistic details about data collection are summarized in Supplementary Table 1. Data sets were processed using HKL2000/3000 (refs. ^{48,49}) or XDS⁵⁰. The low-resolution model of SeMet FEM1B_{1–356} was solved by CRANK2 (ref. ⁵¹), built manually by COOT (Wincoot v.0.8.9.2)⁵², and refined by Phenix 1.14–3260 (ref. ⁵³). Structure of native FEM1B_{1–356} and FEM1C_{1–403} were solved by Phenix 1.14–3260 (ref. ⁵³) using the SeMet structure as the search model for molecular replacement. Other peptide bound structures were solved by molecular replacement with FEM1C_{1–403} as the search model and were further refined by Phenix 1.14–3260 (ref. ⁵³).

After iterative rounds of refinement, the structure of FEM1C_{1–390}-(GGGS)₄-SIL1_{452–461} was refined to 2.90 Å (favored 96.5%, allowed 3.5%); the structure of FEM1C_{1–390}-(GGGS)₄-OR51B2_{303–312} was refined to 2.51 Å (favored 96.5%, allowed 3.5%); the structure of FEM1C_{1–390}-(GGGS)₄-LNLPTQGRAR was refined to 2.33 Å (favored 96.6%, allowed 3.4%); the structure of FEM1C_{1–390}-(GGGS)₄-CDK5R1_{298–307} was refined to 2.35 Å (favored 98.2%, allowed 1.8%); the structure of FEM1C_{1–390}-(GGGS)₄-QERGPTWTKNLR(NS11) was refined to 2.38 Å (favored 97.1%, allowed 2.9%); the structure of FEM1C_{1–390}-(GGGS)₄-SIL1_{452–461} (R/RG) was refined to 2.39 Å (favored 95.4%, allowed 4.6%); the structure of FEM1C_{1–403} was refined to 2.11 Å (favored 97.8%, allowed 2.2%); and the structure of FEM1B_{1–356} was refined to 3.25 Å (favored 93.5%, allowed 6.5%). The structure of FEM1B_{1–337}-(GGGS)₂-CDK5R1_{298–307} was refined to 3.50 Å (favored 94.0%, allowed 6.0%). Structure figures were made using PyMOL v.1.7 (The PyMOL Molecular Graphics System, Schrödinger).

ITC. All proteins were purified using a Ni-NTA column as described above, and further purified by gel filtration using HiLoad 16/600 Superdex 75 pg column (GE Healthcare). Proteins were dialyzed against ITC buffer (20 mM Tris, 400 mM NaCl, 1 mM EDTA, pH 7.5). Peptides were solved in H₂O to a final concentration of 30 mM, and were diluted in ITC buffer to the required concentration for each ITC trial. ITC assays were carried out on a MicroCal iTC200 calorimeter (GE Healthcare) at 25 °C. ITC experiments were performed by titrating 2 μl of peptides (1–2 mM) into cell containing 30–50 μM proteins, with a spacing time of 120 s and a reference power of 5 μCal s⁻¹. Control experiments were performed by titrating peptides (1–2 mM) into buffer, and were subtracted during analysis, respectively. Binding isotherms were plotted, analyzed and fitted based on one-site binding model by MicroCal PEAQ-ITC Analysis software (Malvern Panalytical). Peptide Sequences used for ITC experiments are listed in Supplementary Table 2. The dissociation constants (*K_d*) and peptide sequences used for ITC experiments were also tabulated in Supplementary Table 2. Representative ITC binding curves are shown in Supplementary Data.

Cell lines. HEK293T (ATCC CRL-3216) cells were maintained at 37 °C and 5% CO₂ in Dulbecco's modified Eagle's medium (Life Technologies) supplemented with 10% fetal bovine serum (HyClone), 100 units ml⁻¹ of penicillin and 0.1 mg ml⁻¹ of streptomycin (Thermo Fisher Scientific).

Transfection and lentivirus production. Lentiviral stocks were generated through the transfection of HEK293T cells with the lentiviral transfer vector plus plasmid encoding Gag-Pol, Rev, Tat and VSV-G using PolyJet in vitro DNA Transfection Reagent (SignaGen Laboratories) as recommended by the manufacturer. Lentiviral supernatants were collected 48 h later, passed through a 0.45-mm filter and applied to target cells in the presence of 8 mg ml⁻¹ of hexadimethrine bromide (Polybrene).

GPS assay. The GPS assay using GFP-fused C-degron reporter cell lines was described previously (Koren et al.¹⁷). Briefly, the last ten amino acids of SIL1, OR51B2, CDK5R1 or Clone13 C-degron peptides (referred as 10mer) were encoded as oligonucleotides (gBlocks Gene Fragments, Integrated DNA Technologies) and cloned into the GPS vector using the BstBI and XhoI sites. To test the activity of FEM1s mutants using the GPS assay, lentivirus vectors encoding wild-type or mutant versions of FEM1s were packaged and introduced into HEK293T, following by selection with puromycin (1 μg ml⁻¹) for 3 d to generate cells stably expressing the FEM1 proteins. FEM1 stable cells were then used in a second round of transduction with the GPS reporters. The GPS reporters were packaged and used to transduce FEM1 protein stable cells followed by Blasticidin (20 μg ml⁻¹) selection for 3 d. Flow cytometry was used to record the fluorescence signal of GFP and DsRed from cells transduced with the GPS reporters. Analysis of the GFP to DsRed ratio was done using FlowJo.

Western blotting analyses. Total cell extracts were made from cells growing in six-well plates. Cells were washed twice in cold PBS and lysed in a cold lysis buffer (50 mM Tris, pH 7.6, 150 mM NaCl, 5 mM EDTA, pH 8.0, 0.5% NP-40) supplemented with a protease inhibitor cocktail (Roche, no. 4693159001), 1 mM phenylmethylsulfonyl fluoride, phosphatase inhibitor cocktail (Sigma-Aldrich nos. P5726, P0044), 10 mM NaF, 20 mM β-glycerophosphate, 1 mM Na₃VO₄ and 20 mM P-nitrophenylphosphate. Protein extracts were incubated on ice for 30 min and the nonsoluble components were pelleted by centrifugation for 45 min at 14,000g. Protein concentration was determined by a standard Bradford assay (Bio-Rad no. 500-0006), a linear bovine serum albumin (BSA) calibration curve and an Epoch microplate spectrophotometer.

SDS-PAGE sample buffer (Laemmli buffer) was then added to 30 µg of lysate from each sample, following denaturing (5–10 min, 95 °C). Proteins were separated on freshly made 8–10% acrylamide gels using a Tris-glycine running buffer. Proteins were then electro-transferred onto a nitrocellulose membrane (Bio-Rad no. 162-0115) using Trans-Blot Turbo transfer system (Bio-Rad). Transfer quality was verified by Ponceau S Solution (Sigma-Aldrich no. 81462). Membranes were washed, blocked (5% skimmed milk in TBS-T) and incubated overnight (4 °C) with antibody solution (2.5% BSA and 0.05% sodium azide in PBS). The following primary antibodies were used: anti-Tubulin 1:1,000 (MS581P1, Fisher Scientific) and anti-Flag 1:1,000 (F3165, Sigma-Aldrich). Following three washes with TBS-T, the membrane was incubated with secondary antibody 1:20,000 (Jackson ImmunoResearch (no. 111-035-144, no. 115-035-003)) for 1 h at room temperature, washed a further three times in TBS-T and incubated with SuperSignal West Femtochemiluminescence substrate (Pierce no. 34095) or an EZ-ECL (Biological Industries, no. 20-500-171) for 5 min. Reactive bands visualized using the ImageQuant TL software v.8.2 on Amersham Imager 680 (Cytiva).

Flow cytometry. HEK293T cells were detached with trypsin, washed once with PBS and then analyzed on a LSR Fortessa instrument (Becton Dickinson). Flow cytometry data were collected using BD FACS Diva software v.8.0.2 (Becton Dickinson) and analyzed using FlowJo software v.10.

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

Data availability

The crystal structures of FEM1C and FEM1B were deposited into Protein Data Bank under accession codes 6LBN, 6L BG, 6LE6, 6LDP, 6LEN, 6LEY, 6LFO, 6LBF and 7CNG.

References

- Zhang, W.-Z. et al. The protein complex crystallography beamline (BL19U1) at the Shanghai Synchrotron Radiation Facility. *Nucl. Sci. Techn.* **30**, 170 (2019).
- Guo, Q. et al. Structural insights into SETD3-mediated histidine methylation on beta-actin. *eLife* **8**, e43676 (2019).
- Wang, Q. S. et al. Upgrade of macromolecular crystallography beamline BL17U1 at SSRF. *Nucl. Sci. Techn.* **29**, <https://doi.org/10.1007/s41365-018-0398-9> (2018).
- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Macromol. Crystallogr. A* **276**, 307–326 (1997).
- Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. HKL-3000: the integration of data reduction and structure solution - from diffraction images to an initial model in minutes. *Acta Crystallogr. D. Struct. Biol.* **62**, 859–866 (2006).
- Kabsch, W. XDSs. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 125–132 (2010).
- Pannu, N. S. et al. Recent advances in the CRANK software suite for experimental phasing. *Acta Crystallogr. D. Biol. Crystallogr.* **67**, 331–337 (2011).
- Emsley, P. & Cowtan, K. COOT: model-building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr.* **60**, 2126–2132 (2004).
- Adams, P. D. et al. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D. Biol. Crystallogr.* **58**, 1948–1954 (2002).

Acknowledgements

We thank the staff from the BL17B/BL18U1/BL19U1/BL19U2/BL01B beamline⁴⁵ of National Facility for Protein Science in Shanghai at Shanghai Synchrotron Radiation Facility for assistance during data collection, S.J. Elledge for helpful advice and discussion, D. Wasserman for technical assistance and M. Pagano for kindly providing the plasmids containing cDNAs of human FEM1B and FEM1C. This work is supported by the 'Strategic Priority Research Program' of the Chinese Academy of Sciences (grant no. XDB19000000) and the National Natural Science Foundation of China (grant nos. 92053107, 31770806). C.X. is also supported by the Major/Innovative Program of the Development Foundation of the Hefei Center for Physical Science and Technology (2018CXFX007) and the 'Thousand Young Talent program'. I.K. is supported by Alon fellowship for outstanding young researchers.

Author contributions

X.C. and C.X. conceived the project. X.C. and S.L. performed structural biology and biochemical experiments with assistance from Q.G., Z.Z. and X.T. Y.M., R.K. and K.D. performed the GPS assays. C.X., I.K. and X.Y. analyzed the data and wrote the manuscript. All authors contributed to editing the manuscript. I.K. and C.X. supervised the project.

Competing interests

The authors declare no competing interests.

Additional information

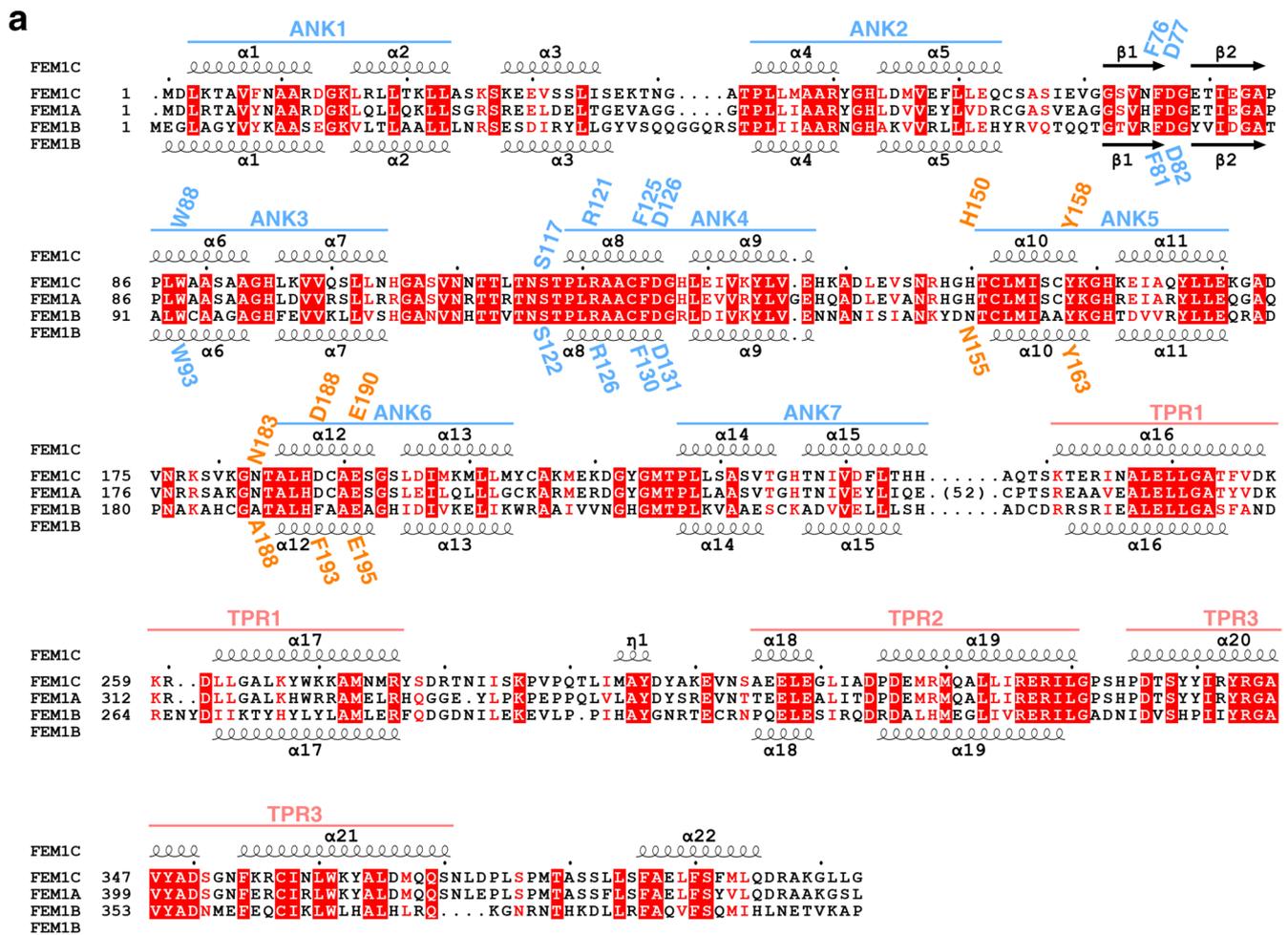
Extended data is available for this paper at <https://doi.org/10.1038/s41589-020-00704-3>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41589-020-00704-3>.

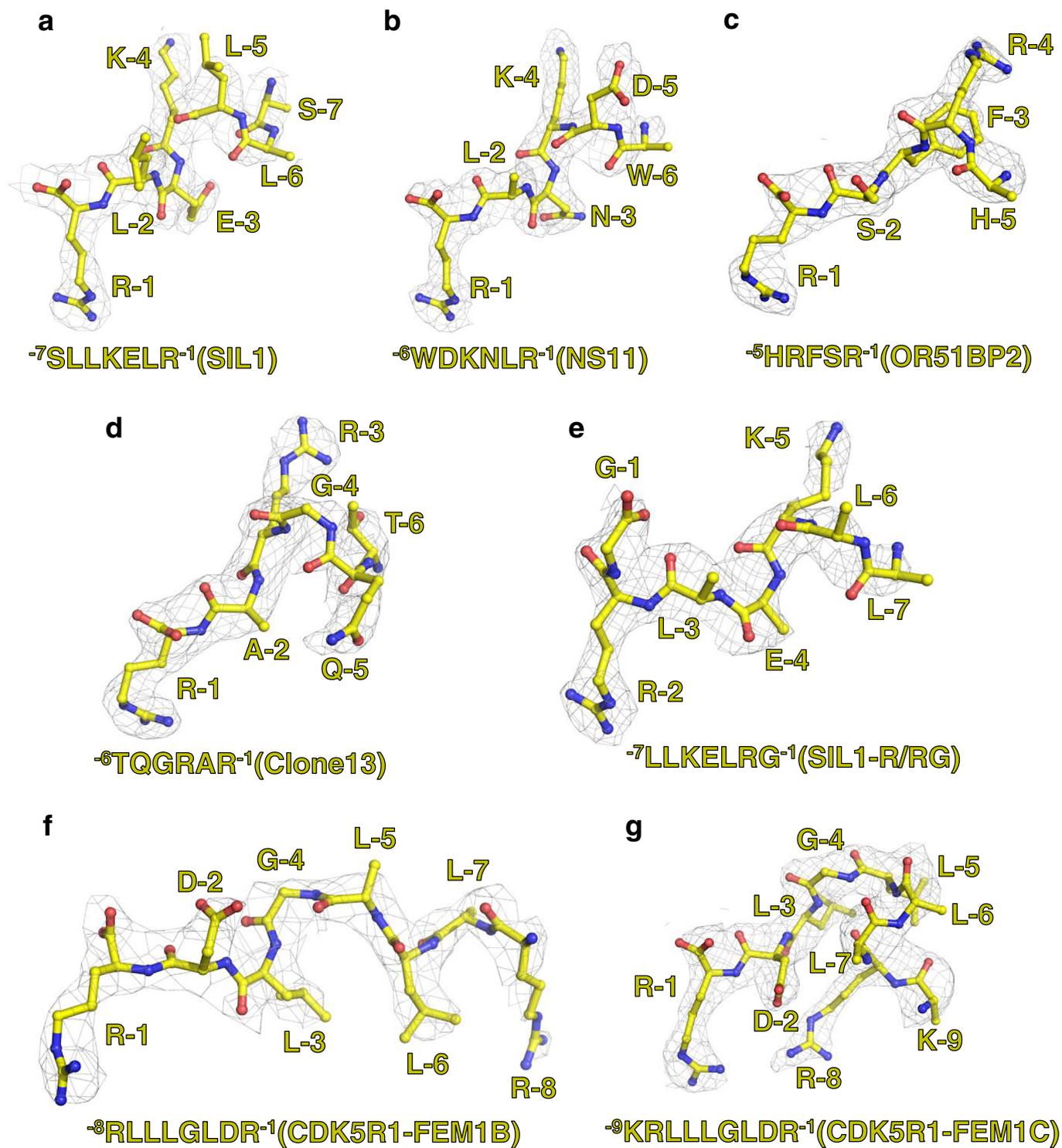
Correspondence and requests for materials should be addressed to I.K. or C.X.

Peer review information *Nature Chemical Biology* thanks David Dougan, Hyun Kyu Song and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

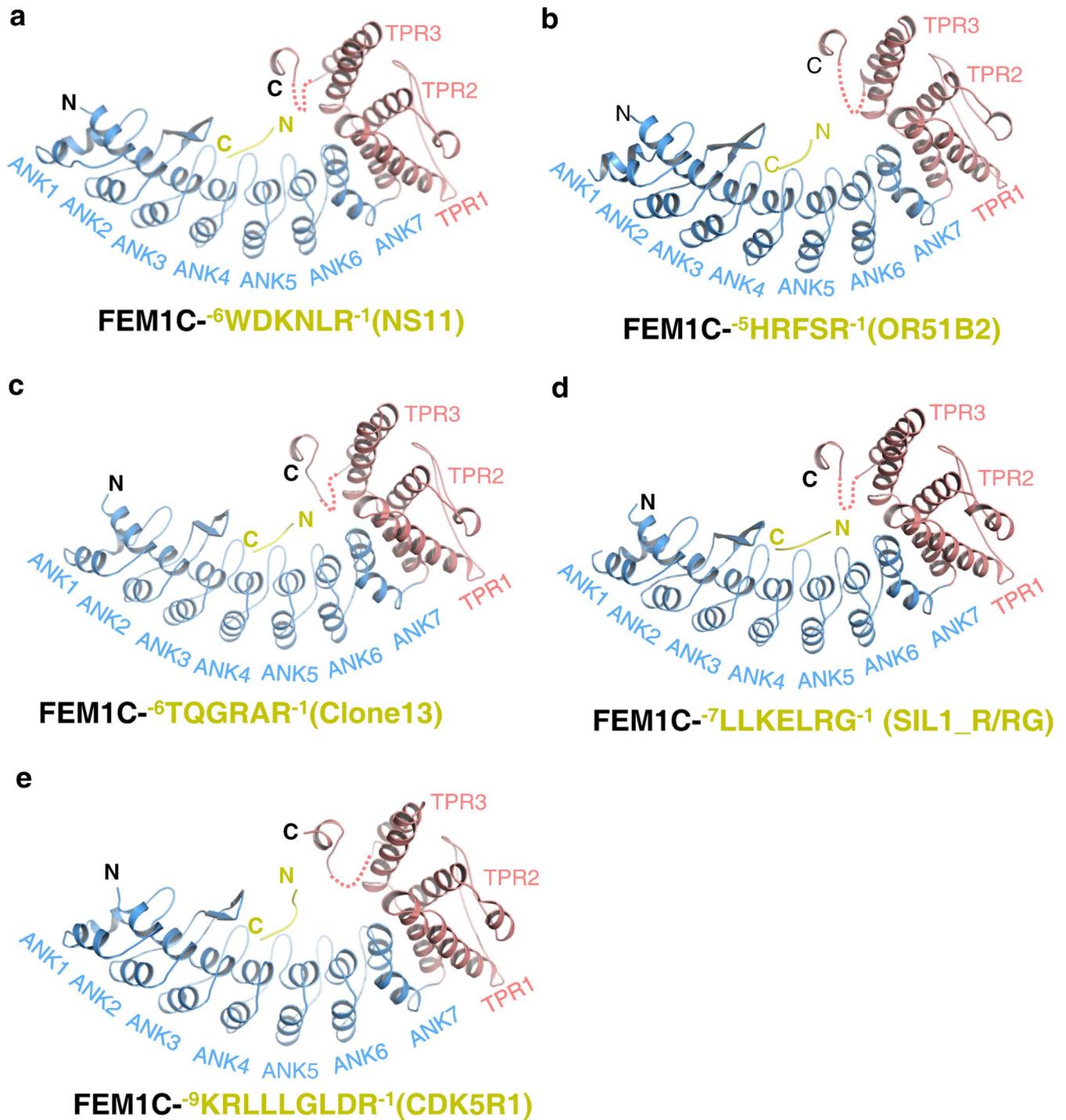
Reprints and permissions information is available at www.nature.com/reprints.



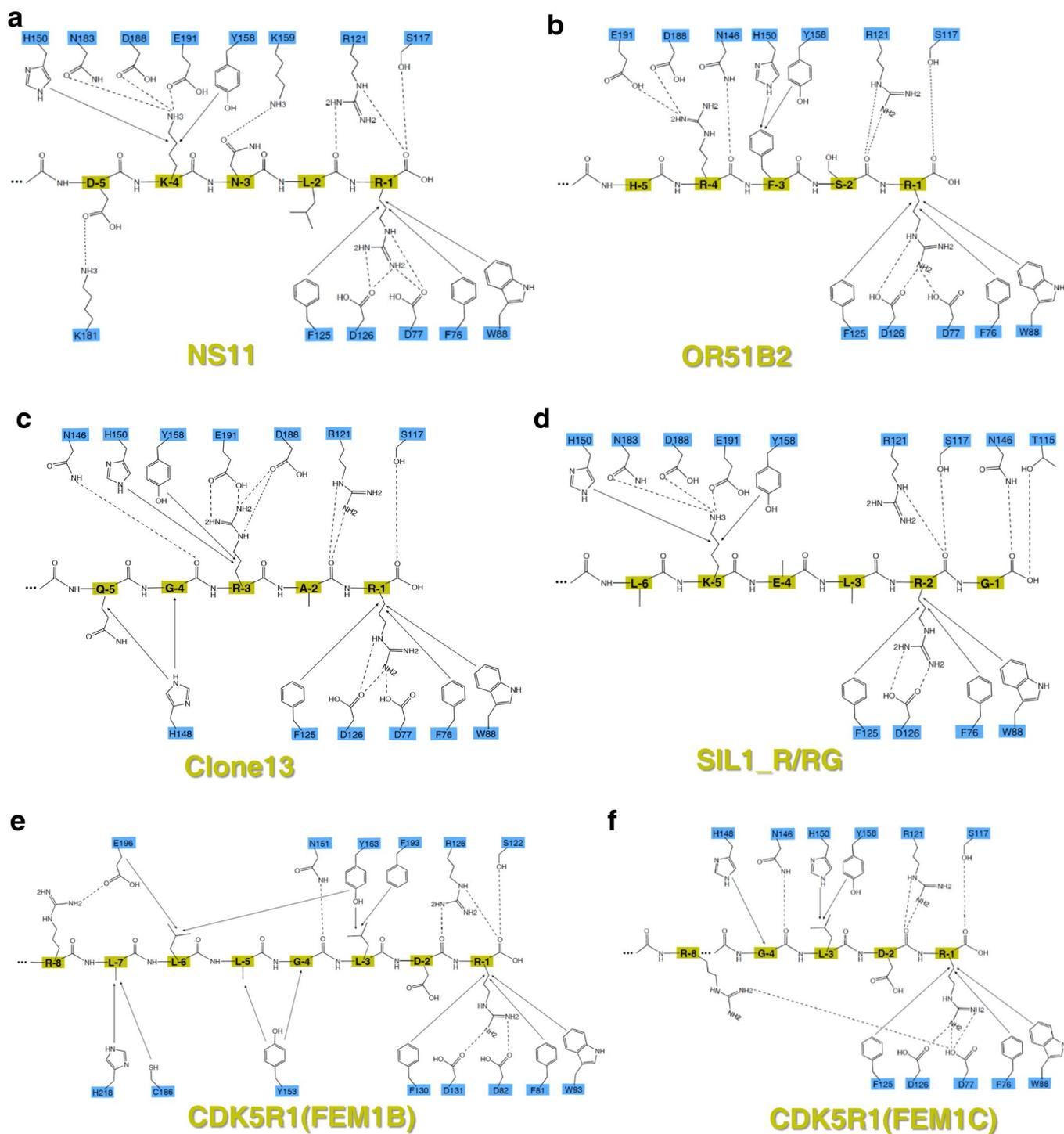
Extended Data Fig. 1 | Sequence alignment of FEM1 proteins and C-degrons, respectively. a, Sequence alignment of human FEM1C (NP_064562.1), FEM1A (NP_061178.1), and FEM1B (NP_056137.1). The secondary structures of FEM1C and FEM1B are labelled at the top and bottom of the sequences, respectively. In the structure of SIL1-bound FEM1C, R-1 and K-4 binding residues of FEM1C, as well as their counterparts in FEM1B, are labelled in blue and orange, respectively. Their counterparts in FEM1B are also labeled accordingly. **b**, Sequence alignment of Arg/C-degrons, including those of SIL1, NS11, OR51B2, Clone13, and CDK5R1. The residues at -1 and -4 positions of SIL1 and their counterparts are marked.



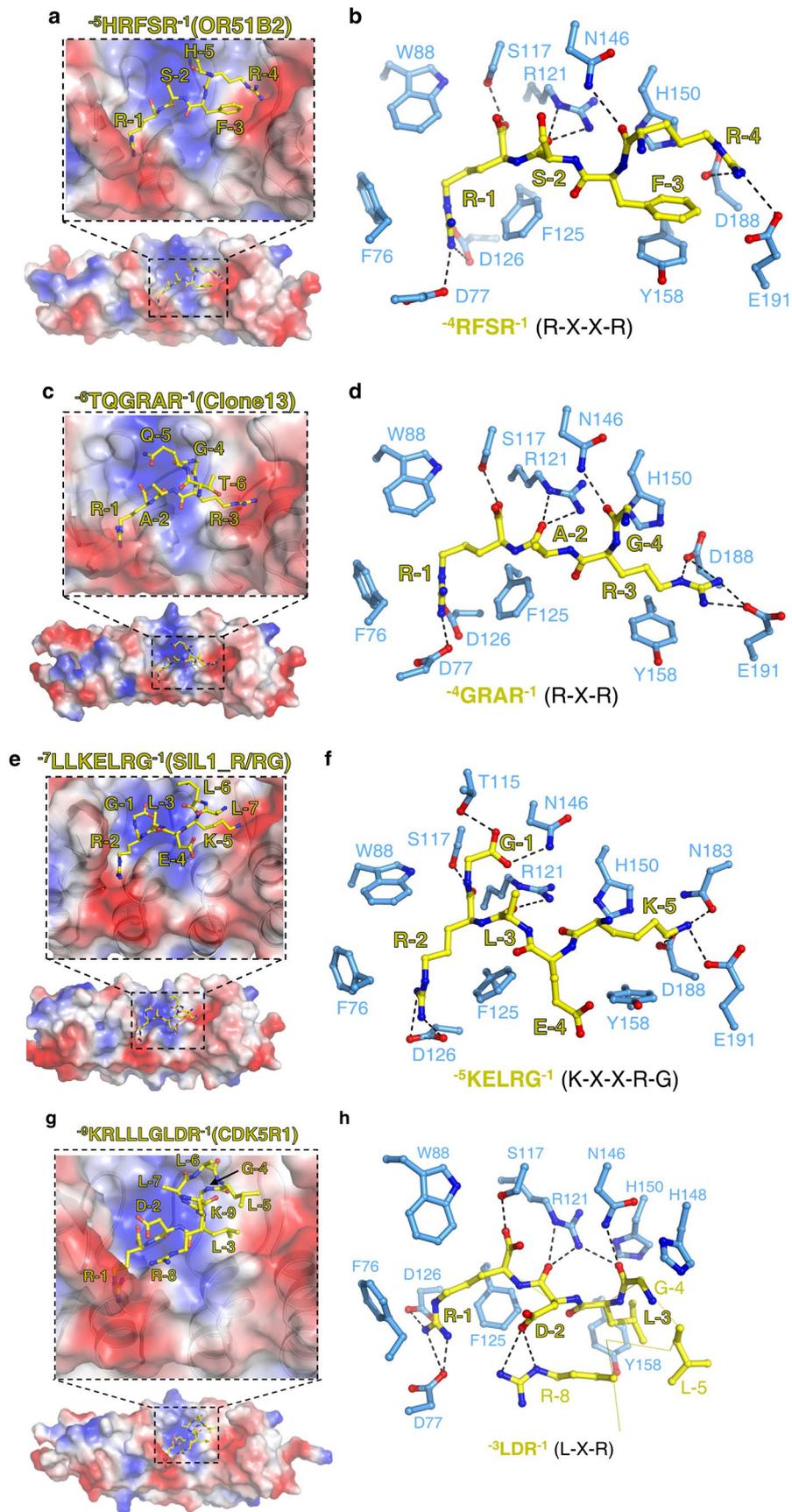
Extended Data Fig. 2 | The $2|F_o|-|F_c|$ omit maps of the fused C-degron peptides in the FEM1C structures contoured at 1.0σ . **a**, SIL1 ($^{-7}\text{SLLKELR}^{-1}$); **b**, NS11 ($^{-6}\text{WDKNLR}^{-1}$); **c**, OR51BP2 ($^{-5}\text{HRFSR}^{-1}$); **d**, Clone13 ($^{-6}\text{TQGRAR}^{-1}$); **e**, SIL1_R/RG ($^{-7}\text{LLKELRG}^{-1}$); **f**, The $2|F_o|-|F_c|$ omit map of CDK5R1 ($^{-8}\text{RLLGLDR}^{-1}$) bound to FEM1B; **g**, The $2|F_o|-|F_c|$ omit map of CDK5R1 ($^{-9}\text{KRLLGLDR}^{-1}$) bound to FEM1C.



Extended Data Fig. 3 | Structures of C-degron-bound FEM1C. **a**, Overall structure of NS11-bound FEM1C; **b**, Overall structure of OR51BP2-bound FEM1C; **c**, Overall structure of Clone13-bound FEM1C; **d**, Overall structure of SIL1_R/RG-bound FEM1C; **e**, Overall structure of CDK5R1-bound FEM1C. The ankyrin repeats and TPR repeats of FEM1C are shown in blue and red cartoon representations, respectively, and the C-degron peptides are shown in yellow cartoon.

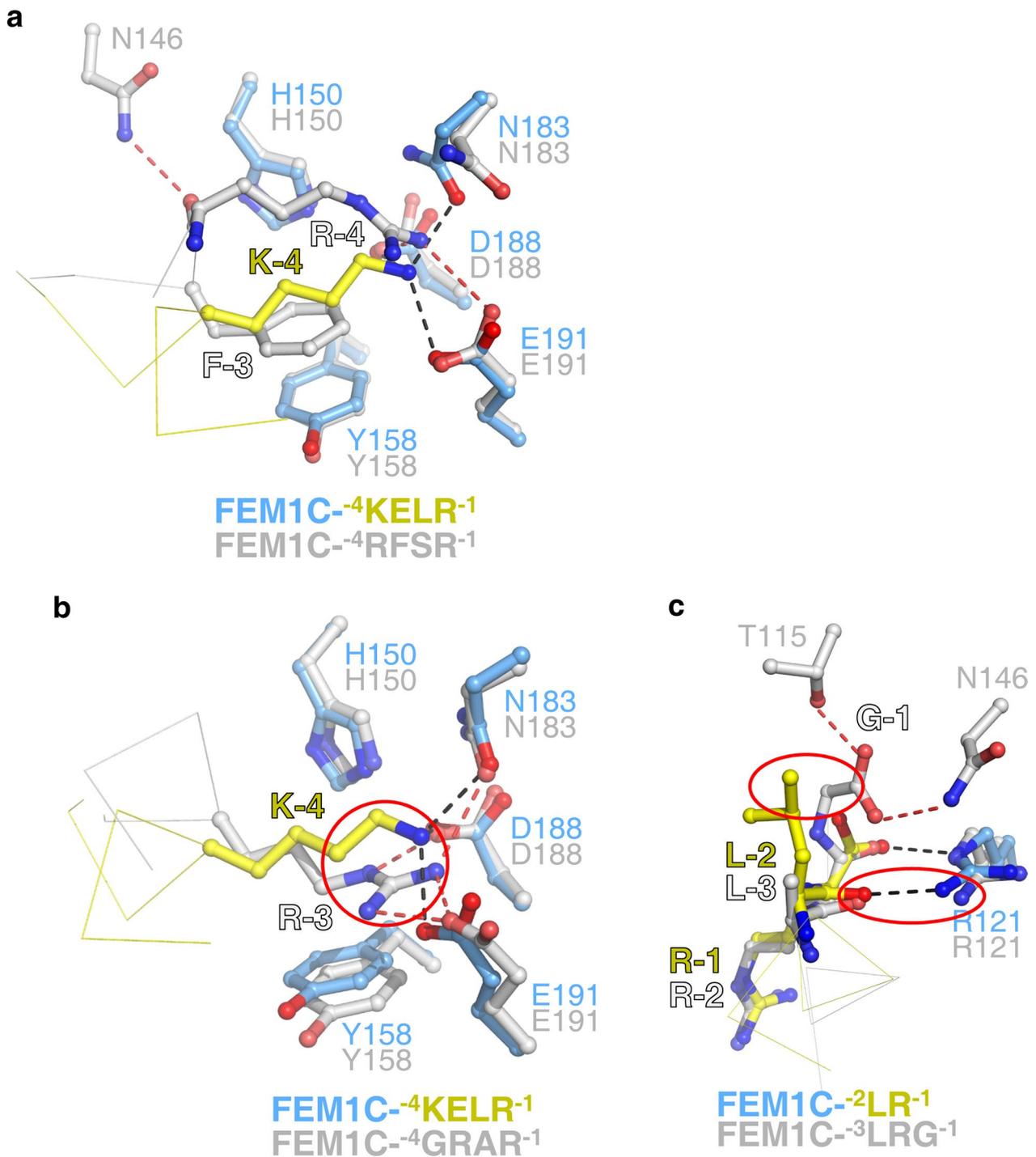


Extended Data Fig. 4 | Schematic of the detailed interactions between FEM1 proteins and degrons. a, FEM1C and NS11; **b**, FEM1C and OR51B2; **c**, FEM1C and Clone13; **d**, FEM1C and SIL1_R/RG; **e**, FEM1B and CDK5R1; **f**, FEM1C and CDK5R1. Intermolecular hydrogen bonds and hydrophobic interactions are indicated by black dashes and black arrows, respectively.

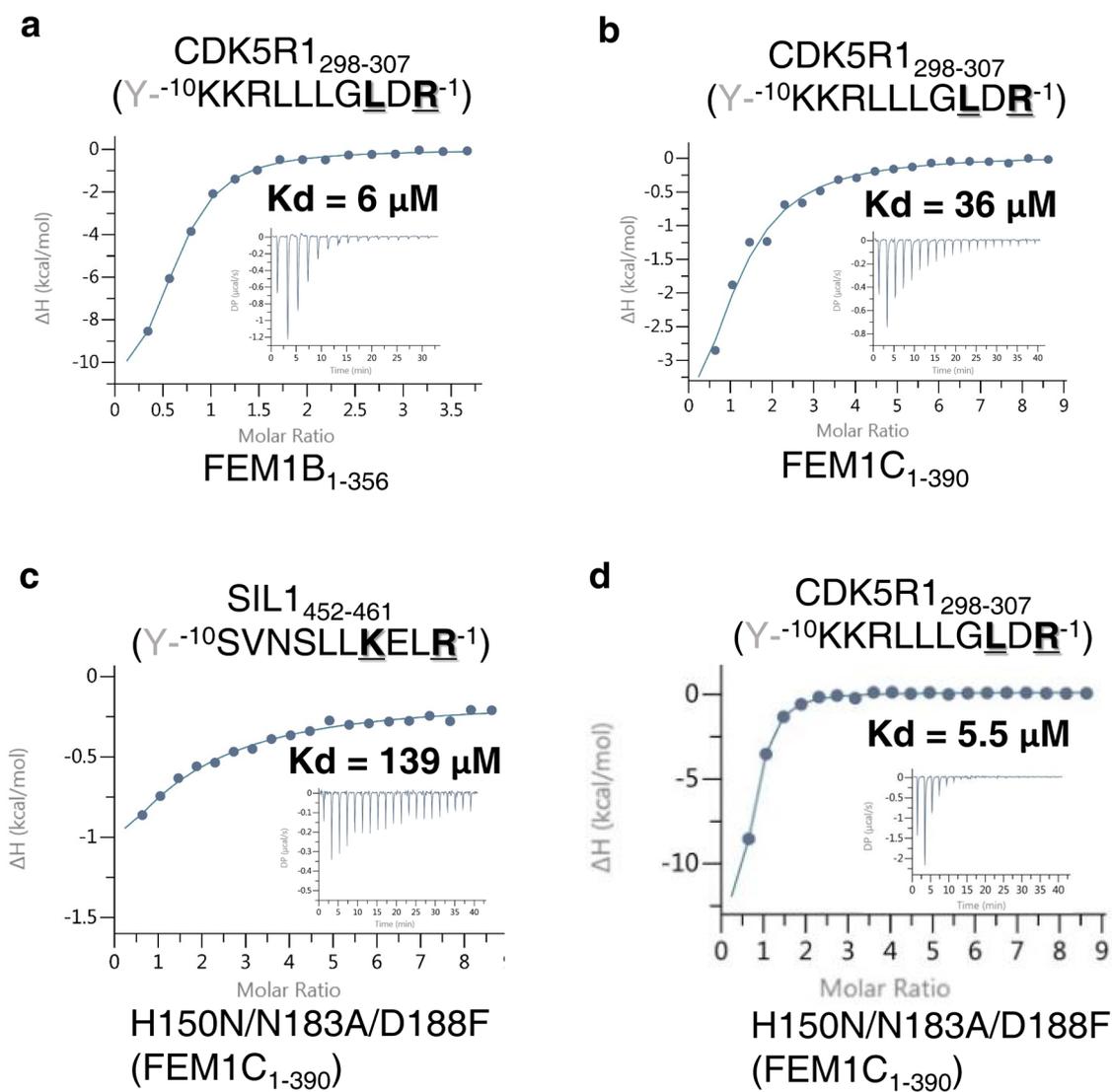


Extended Data Fig. 5 | See next page for caption.

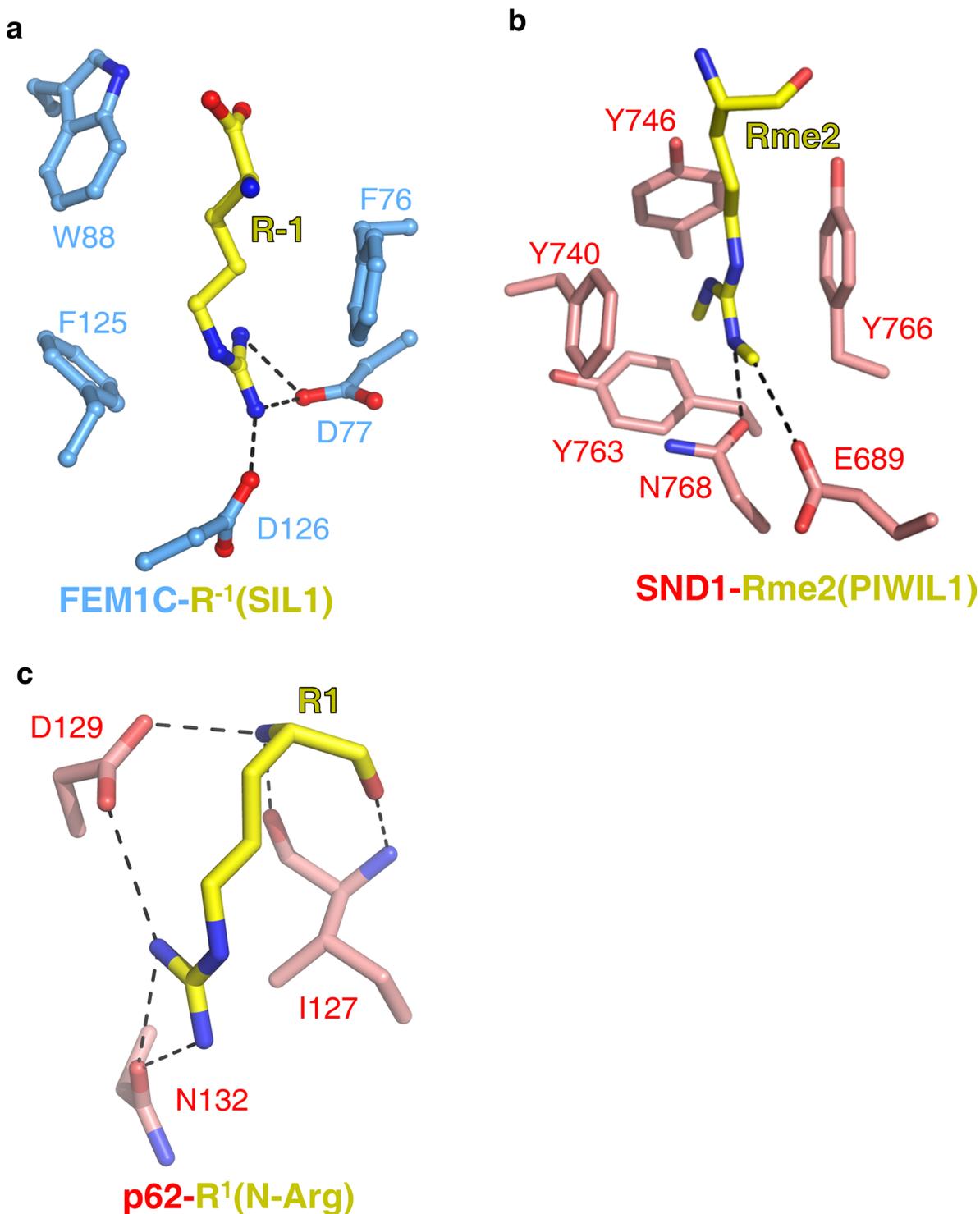
Extended Data Fig. 5 | Detailed interactions between FEM1C and C-degrons. **a**, The electrostatic surface of the FEM1C ankyrin repeats (1-240) bound with OR51B2, with the peptides shown in yellow sticks and labelled. **b**, Recognition of $^{-4}\text{RFSR}^{-1}$ (R-X-X-R) by FEM1C. The peptide residues and FEM1C residues involved in the intermolecular interactions are labelled, and shown in yellow and blue sticks, respectively. **c**, The electrostatic surface of the FEM1C ankyrin repeats (1-240) bound with the Clone13 $^{-6}\text{TQGRAR}^{-1}$. **d**, Recognition of $^{-4}\text{GRAR}^{-1}$ (R-X-R) by FEM1C. **e**, The electrostatic surface of the FEM1C ankyrin repeats (1-240) bound with the peptide $^{-7}\text{LLKELRG}^{-1}$. **f**, Recognition of $^{-5}\text{KELRG}^{-1}$ (K-X-X-RG) by FEM1C. **g**, The electrostatic surface of the FEM1C ankyrin repeats (1-240) bound with CDK5R1. **h**, Recognition of $^{-9}\text{KRLLGLDR}^{-1}$ by FEM1C. **(c)**, **(e)**, and **(g)** are shown in a manner similar to that shown in Extended Data Fig. 5a, while **(d)**, **(f)**, and **(h)** are shown in a way similar to that shown in Extended Data Fig. 5b.



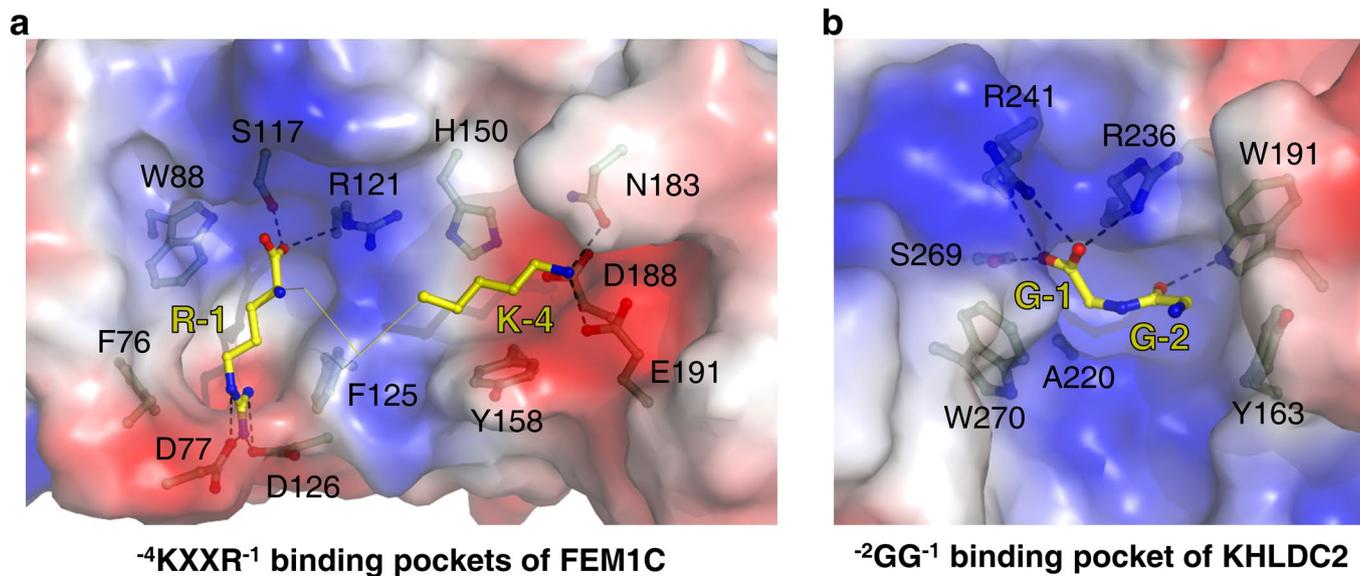
Extended Data Fig. 6 | Comparison of the binding of K-X-X-R binding mode (SIL1) with that of **a**, R-F-S-R binding mode (OR51B2), **b**, G-R-A-R (Clone13), and **c**, K-X-X-RG (SIL1_R/RG). In **(a)-(c)**, the SIL1 peptide is shown in yellow ribbon, with key residues shown in sticks. The SIL1 interaction residues of FEM1C are shown in blue sticks. The peptides of OR51B2, Clone13, and SIL1_R/RG are shown in gray ribbon, with key residues shown in sticks. The FEM1C residues that interact with OR51B2, Clone13, or SIL1_R/RG are shown in grey sticks. Different conformations of key residues are highlighted by red circles.



Extended Data Fig. 7 | Distinct binding properties of FEM1B and FEM1C. a, b, ITC binding curves of CDK5R1 binding to FEM1B₁₋₃₅₆ (**a**) and FEM1C₁₋₃₉₀ (**b**). **c, d,** ITC binding curves of FEM1C₁₋₃₉₀ triple mutant H150N/N183A/D188F titrated with SIL1₄₅₂₋₄₆₁ (**c**) and CDK5R1₂₉₈₋₃₀₇ (**d**).



Extended Data Fig. 8 | Comparison of the Arg⁻¹ binding pocket of FEM1C with other arginine binding modules. a, Recognition of Arg⁻¹ (yellow) by FEM1C (blue). **b**, Recognition of Rme2 (yellow) of PIWIL1 by SND1 (red) (PDB: 3OMC). **c**, Recognition of Arg/N-degron (yellow) by p62 ZZ domain (red) (PDB: 6MIU).



Extended Data Fig. 9 | Comparison of the Arg/C-degron recognition by FEM1C with that of Gly/C-degron by KHLDC2. The electrostatic surface of FEM1C and KHLDC2 are shown. a, SIL1 binding surface of FEM1C. The SIL1 peptide is shown in yellow ribbon, with Arg⁻¹ and Lys⁻⁴ shown in sticks. FEM1C residues involved in the interaction with Arg⁻¹ and Lys⁻⁴ are also shown in sticks. **b,** di-glycine binding surface of KHLDC2 (PDB: 6DO3). The di-glycine are shown in yellow sticks, and the KHLDC2 residues involved in binding to di-glycine are also shown in sticks.

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 (n) 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
<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. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<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 d , Pearson's r), 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

Western blot data were imaged by Amersham Imager 680 (Cytiva) using ImageQuant TL software v8.2.
Flow data were collected using LSRFortessa cell analyzer (BD Biosciences) using BD FACSDiva software v8.0.2.
ITC binding experiments were carried out on a MicroCal iTC200 calorimeter (GE Healthcare)

Data analysis

Crystallographic data processing and refinement was done using HKL2000/3000 or XDS56. The low resolution model of SeMet FEM1B (1-356) was solved by CRANK2, built manually by COOT (Wincoot 0.8.9.2), and refined by Phenix 1.14-3260. Other peptide bound structures were solved by molecular replacement with FEM1C(1-403) as the search model and were further refined by Phenix 1.14-3260. Structure figures were made with PyMOL 1.7.
ITC binding data were analyzed by MicroCal PEAQ-ITC Analysis software (Malvern Panalytical, UK).
Flow data were analyzed with FlowJo v.10 (Becton Dickinson).

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

The structure data were deposited into PDB (<http://www.rcsb.org/>) under accession codes 6LBN, 6LBG, 6LE6, 6LDP, 6LEN, 6LEY, 6LFO, 6LBF and 7CNG. Source data are provided with this paper.

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://www.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	No statistical methods were used to predetermine sample sizes. The sample size was chosen so that desired structure features were obtained by X-ray crystallography. For GPS assays minimum of 20,000 DsRed positive cells were collected to enable accurate analysis. Number of replicates for each experiment is provided in the figure captions.
Data exclusions	No data were excluded
Replication	ITC and GPS experiments shown in source data were reproduced at least twice. Flow cytometry data were repeated independently in triplicates with similar results.
Randomization	Sample randomization is not relevant to our study, because for X-ray crystallography, the best crystal has to be used for data collection and structure determination.
Blinding	Blinding was not employed during data collection or analysis. However, controls and samples were analyzed in the exact same manner using the same software settings. For GPS data, automatic data collection and analysis was conducted without human intervention where possible.

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

Methods

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

Antibodies

Antibodies used	Flag (F3165, Sigma-Aldrich), Tubulin (MS581P1, Fisher Scientific), Peroxidase AffiniPure Goat Anti-Rabbit (111-035-144) and Goat Anti-Mouse (#115-035-003) (Jackson ImmunoResearch)
Validation	All antibodies are commercial and were validated by the manufacturer for the purpose used in this study, i.e. immunoblotting using human cells. Also see references (PMIDs: 26260981, 31365523, 30194296)

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	HEK293T (ATCC CRL-3216)
Authentication	HEK293T cell line were purchased from ATCC that routinely performs STR analysis on its human cell lines.
Mycoplasma contamination	Mycoplasma negative.
Commonly misidentified lines (See ICLAC register)	None of the commonly misidentified lines was used in this study.

Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Cells were detached with trypsin, washed once with PBS and then directly analysed on the Flow cytometer.
Instrument	LSRFortessa cell analyzer (BD Biosciences).
Software	FlowJo V.10
Cell population abundance	Live cells were ~85%, singlets were ~96%, DsRed positives were ~65%-70%
Gating strategy	Cells were first gated for live cells using forward and side scatter (FSC-A/SSC-A). Single cells were discriminated based on height vs. width of the forward scatter (FSC-H/FSC-W). Finally, GPS reporter positive cells were gated based on the DsRed expression. A figure for gating strategy was provided as Supplementary Fig. 3.

- Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.