Check for updates

Molecular basis for ubiquitin ligase CRL2[FEM1C]-mediated recognition of C-degron

Xiaojie Yan[1,5], Xiaolu Wang[2,5], Yao Li[1], Mengqi Zhou[3], Yanjun Li[3], Lili Song[4], Wenyi Mi[4✉],
Jinrong Min[3✉] and Cheng Dong[1✉]

Proteome integrity depends on the ubiquitin–proteasome system to degrade unwanted or abnormal proteins. In addition to the N-degrons, C-terminal residues of proteins can also serve as degradation signals (C-degrons) that are recognized by specific cullin-RING ubiquitin ligases (CRLs) for proteasomal degradation. FEM1C is a CRL2 substrate receptor that targets the C-terminal arginine degron (Arg/C-degron), but the molecular mechanism of substrate recognition remains largely elusive. Here, we present crystal structures of FEM1C in complex with Arg/C-degron and show that FEM1C utilizes a semi-open binding pocket to capture the C-terminal arginine and that the extreme C-terminal arginine is the major structural determinant in recognition by FEM1C. Together with biochemical and mutagenesis studies, we provide a framework for understanding molecular recognition of the Arg/C-degron by the FEM family of proteins.

The ubiquitin (Ub)–proteasome system (UPS) is a critical protein degradation pathway that maintains protein homeostasis by eliminating damaged, misfolded or unnecessary proteins[1,2]. Aberrations in UPS-associated proteins contribute to various human diseases, including neurodegeneration, abnormal aging and cancers[3,4]. In the canonical UPS, ubiquitination of target substrates is achieved through a cascade of enzymes, including Ub-activating (E1), Ub-conjugating (E2) and Ub ligase (E3) enzymes. The ubiquitinated substrates are subsequently degraded into short peptides by the 26S proteasome[5,6]. In particular, the E3 Ub ligases serve to recognize substrates' degradation signals (degrons), which, in general, are specific sequence motifs located in their cognate substrates[7,8]. One group of degrons are identified at the N termini of proteins, called N-degrons, which are targeted by the N-end rule pathways (recently renamed the 'N-degron pathways')[9–11]. The N-degron pathways govern the in vivo half-lives of proteins based on the destabilizing N-terminal residues of the substrates[10]. All 20 natural amino acids have been characterized as destabilizing N-terminal residues and classified into distinct branches of N-degron pathways, namely, the Arg/N-degron[10], Ac/N-degron[12], Pro/N-degron[13], Gly/N-degron[14] and fMet/N-degron pathways[15]. These N-degron pathways participate in a broad spectrum of biological functions in eukaryotes[10,11,16,17]. The different N-degrons are specifically recognized by their respective E3 Ub ligases or other recognition components (N-recognins), such as UBR proteins, p62 and GID4, and their substrate recognition mechanisms have been elucidated by means of structural biology[18–23].

In addition to the N-degron pathways, recent studies have discovered a new large set of C-degron pathways in which the degrons are located at the extreme C termini of proteins, called C-end degrons, in a pathway termed destruction via C-end degrons (DesCEND)[11,24–26]. The C-degron is generally a motif of fewer than ten residues and can be present in full-length proteins, truncated proteins or proteolytically cleaved forms[24,25]. Multiple distinct C-degrons are identified and targeted by CRLs, including the CRL2 and CRL4 complexes. In these complexes, a variety of CRL substrate receptors are responsible for recognizing their specific C-degrons. For instance, the Kelch domain-containing proteins (KLHDC2, KLHDC3 and KLHDC10) recognize C-terminal glycine residues, the ankyrin (ANK)-repeat proteins (FEM1A, FEM1B and FEM1C) recognize C-terminal arginine residues and the WD40-repeat protein DCAF12 recognizes C-terminal glutamic acid[24–27].

The FEM proteins act as substrate recognition subunits of the CRL2 E3 ligases[28–30], which play important roles in the sex determination pathway and mediate apoptosis in mammalian cells as well as in malignant colon cancer cells[31–34]. However, the molecular mechanisms underlying FEM-mediated recognition in the C-degron pathways remain largely unknown. Here we provide crystal structures of FEM1C in complex with a C-terminal arginine degron (Arg/C-degron). Our structural analysis coupled with biochemical studies sheds light on the substrate recognition specificity for Arg/C-degrons of FEM1C, which would be an attractive target for the development of small-molecule inhibitors.

Results

The N-terminal portion of FEM1C binds Arg/C-degron. CRL2[FEM1A/FEM1B/FEM1C] is responsible for degradation of proteins with an arginine residue at their C termini, and SIL1 has been identified as a substrate of CRL2[FEM1C] (ref. [24]). FEM1C, as a substrate receptor, contains nine ANK repeats (ANK1–ANK9), with two predicted TPR repeats inserted between ANK7 and ANK8 followed by C-terminal BC and Cul2 boxes that are involved in elongin BC and cullin 2 binding, respectively[29] (Fig. 1a). The ANK and TPR repeats are known as scaffolds for mediating protein–protein interactions[35,36].

[1]Department of Biochemistry and Molecular Biology, The Province and Ministry Co-sponsored Collaborative Innovation Center for Medical Epigenetics, School of Basic Medical Sciences, Tianjin Medical University, Tianjin, China. [2]Department of Pharmacology, School of Basic Medical Sciences, Tianjin Medical University, Tianjin, China. [3]Structural Genomics Consortium and Department of Physiology, University of Toronto, Toronto, Ontario, Canada. [4]Department of Immunology, Tianjin Key Laboratory of Cellular and Molecular Immunology, Key Laboratory of Diseases and Microenvironment of Ministry of Education of China, The Province and Ministry Co-sponsored Collaborative Innovation Center for Medical Epigenetics, School of Basic Medical Sciences, Tianjin Medical University, Tianjin, China. [5]These authors contributed equally: Xiaojie Yan, Xiaolu Wang. ✉e-mail: wenyi.mi@tmu.edu.cn; jr.min@utoronto.ca; dongcheng@tmu.edu.cn

To validate which region of FEM1C contributes to binding of Arg/C-degron, we purified several truncated FEM1C proteins and performed glutathione S-transferase (GST) pull-down assays using GST fused with an undecapeptide derived from the C terminus of SIL1 (hereafter referred to as GST–SIL1 degron) as bait (Fig. 1a,b). Among these FEM1C fragments, the fragment of FEM1C consisting of amino acids 1–246 was not stable after removal of the SUMO tag, so we used SUMO-tagged FEM1C$_{1-246}$ in the GST pull-down assay. Our results showed that the following C-terminal truncations were capable of binding the SIL1 degron: FEM1C$_{1-574}$ containing ANK1–ANK9, FEM1C$_{1-371}$ containing ANK1–ANK7 and TPR1-2, and FEM1C$_{1-246}$ containing ANK1–ANK7. In contrast, the N-terminally truncated FEM1C$_{82-556}$, which lacks the first two ANK repeats, failed to bind to the SIL1 degron, indicating that an intact N-terminal fragment covering the first seven ANK repeats of FEM1C is essential for the recognition of Arg/C-degron.

C terminus of FEM1C$_{1-244}$ acts as an Arg/C-degron. Attempts to co-crystallize different FEM1C fragments with the SIL1 degron were unsuccessful, but we could obtain crystals of FEM1C$_{1-244}$ alone, which contains the first seven ANK repeats (ANK1–ANK7) ending with an arginine residue at position 244, and determined its structure at 2.0-Å resolution (Supplementary Table 1). Intriguingly, in this structure, each asymmetric unit contained six FEM1C molecules, which formed two trimers where the three symmetry-related molecules interacted with each other through their C termini (Fig. 1c). Specifically, the 11 C-terminal residues (Thr 234–Arg 244) of one molecule nestled into the groove of another molecule through a series of intermolecular interactions (Fig. 1d,e). Therefore, we hypothesize that this arginine-ended polypeptide could act as an Arg/C-degron of FEM1C. To support this, we fused the 11 C-terminal residues of FEM1C$_{1-244}$ to the C terminus of GST (referred to as the GST–FEM degron) and performed a GST pull-down assay (Fig. 2a,b). Given that FEM1C$_{1-244}$ could recognize its own C terminus from another molecule as observed in the crystal packing, we used the FEM1C$_{1-371}$ construct ending with asparagine (Asn 371) as prey. The pull-down results suggested that the GST–FEM degron could apparently pull down FEM1C$_{1-371}$. To further confirm this interaction, we synthesized an undecapeptide derived from FEM1C (Thr 234–Arg 244, termed the FEM peptide) and performed isothermal titration calorimetry (ITC) assays. The FEM peptide bound to FEM1C$_{1-371}$ with a K_d value of 15 μM (Fig. 2c). As a substrate of CRL2^{FEM1C}, the SIL1 C-terminal undecapeptide (referred to as the SIL1 peptide) displayed a comparable affinity for FEM1C$_{1-371}$ with a K_d value of 41 μM (Fig. 2d), indicating that the FEM peptide serves as a potent artificial Arg/C-degron, which might aid in the identification of more natural substrates.

Characterization of a more potent Arg/C-degron. Previous studies have indicated that substrate recognition by FEM1C is considerably more complex than that of other substrate receptors in the C-degron pathways[24,25]. Therefore, we sought to find a more potent Arg/C-degron sequence based on the electrostatic complementarity of the complex interface from the FEM1C$_{1-244}$ structure. Within the mixed-charge substrate-binding groove, the extreme C-terminal arginine fits well into the negatively charged binding pocket at the end of the groove (P–1). The penultimate position, P–2, is more suitable for accommodating a negatively charged residue, whereas the upstream binding cleft (P–3 to P–10) is a wide acidic path favoring positively charged residues (Fig. 1d,e). Indeed, arginine scanning mutagenesis of the FEM peptide (P–3 to P–10) strengthened the binding to FEM1C$_{1-371}$ (Supplementary Table 2). Thus, we selected an eligible undecapeptide (NRRRRWRERQR) corresponding to the human immunodeficiency virus type 1 (HIV-1) REV protein (amino acids 40–50; ref. [37]) as a candidate (Fig. 2b). As expected, our GST pull-down assay showed that the GST–REV degron bound tightly to FEM1C$_{1-371}$ (Fig. 2a). Additionally, the REV peptide exhibited a significant increase in binding to FEM1C$_{1-371}$ with a K_d value of 0.5 μM (Fig. 2e). In line with these results, the REV peptide was competitive with the SIL1 peptide in binding to FEM1C in a GST pull-down competition assay (Extended Data Fig. 1a). In addition, our global protein stability (GPS) assay (see below) showed that the REV degron was more unstable than the SIL1 degron in vivo (Extended Data Fig. 1b). Therefore, the REV degron is a more potent Arg/C-degron for FEM1C.

To map the minimal length of the Arg/C-degron for FEM1C recognition, we synthesized a series of peptides based on the REV peptide that retained the extreme C-terminal arginine and ranged from 2 to 11 residues in length (Fig. 2f). Our ITC results showed that the 7-residue degron had almost the same binding affinity as the 9-residue or 11-residue degron, with a low-micromolar affinity. The shorter peptides bearing five to three amino acids showed reduced binding affinities. Nevertheless, the three-residue degron was still able to bind FEM1C, albeit with a 30-fold-reduced affinity, while the two-residue degron displayed much weaker binding to FEM1C than the three-residue degron. Overall, our results suggest that FEM1C recognizes Arg/C-degrons of a certain length, and the seven-residue degron is sufficient for high-affinity binding.

Molecular recognition of Arg/C-degron by FEM1C. To elucidate the molecular details governing Arg/C-degron recognition by FEM1C, we solved the crystal structure of FEM1C$_{1-371}$ in complex with an 11-mer REV peptide (Supplementary Table 1). FEM1C$_{1-371}$ contains seven consensus ANK repeats followed immediately by two TPR repeats (Fig. 3a). In addition to common features, several unique elements were found. First, the loop linking ANK2 and ANK3 (finger 2) adopted a β-hairpin fold unlike other finger-like hairpin loops. In addition, finger 2 bent upward and was in parallel with α-helices of ANK3, forming a slanted, U-shaped structure instead of a typical L-shaped structure as is often seen in ANK repeats[38]. As a result, the finger 2 loop was the key participant in binding the Arg/C-degron (see below). Second, a slight topological difference occurred at the N terminus, where the long loop of ANK1 preceding the hairpin loop formed a helix, not an extended loop. Third, the C terminus includes two canonical TPR motifs, and each was composed of a pair of antiparallel α-helices. In addition, a TPR-like motif consisting of two non-parallel α-helices was inserted between TPR1 and TPR2. A long hairpin loop connecting TPR1 and the TPR-like motif extended to TPR2 along the periphery, forming a compact six-helix bundle structure.

Like the FEM degron, the REV degron interacted tightly with a conserved, concave inner face of the first seven ANK repeats (Fig. 3b). Specifically, finger 2 and the C-terminal inner helix of ANK4 created an acidic, circle-shaped binding pocket shielding the guanidine group of the extreme C-terminal arginine (R–1; Fig. 3c). The carboxyl group of R–1 and the penultimate glutamine (Q–2) were subsequently anchored at a basic ditch. The long side chains of the succeeding R–3 as well as R–7 extended to the next acidic patch (Fig. 3c and Extended Data Fig. 1c). Structural analysis revealed that R–1 was involved in multiple interactions. First, the guanidine group of R–1 was coordinated by hydrogen bonds and a salt bridge from the carboxyl groups of Asp 126 and Asp 77. Simultaneously, the guanidinium moiety of R–1 was sandwiched by Phe 76 and Phe 125 through cation-π interactions (Fig. 3d). Interestingly, when R–1 was inserted between the parallel aromatic moieties of Phe 76 and Phe 125, it constituted a tandem cation-π interaction as Lys 1 59-Phe 125-(R–1)-Phe 76-Arg 50-Tyr 51 (Fig. 3e). Second, the free carboxyl group of R–1 was further stabilized by two hydrogen bonds with the hydroxyl group of Ser 117 and the side-chain NH group of Arg 121 (Fig. 3d). As a result, R–1 fit tightly into this conserved binding pocket created by ANK3 and ANK4 (Fig. 3b–e).
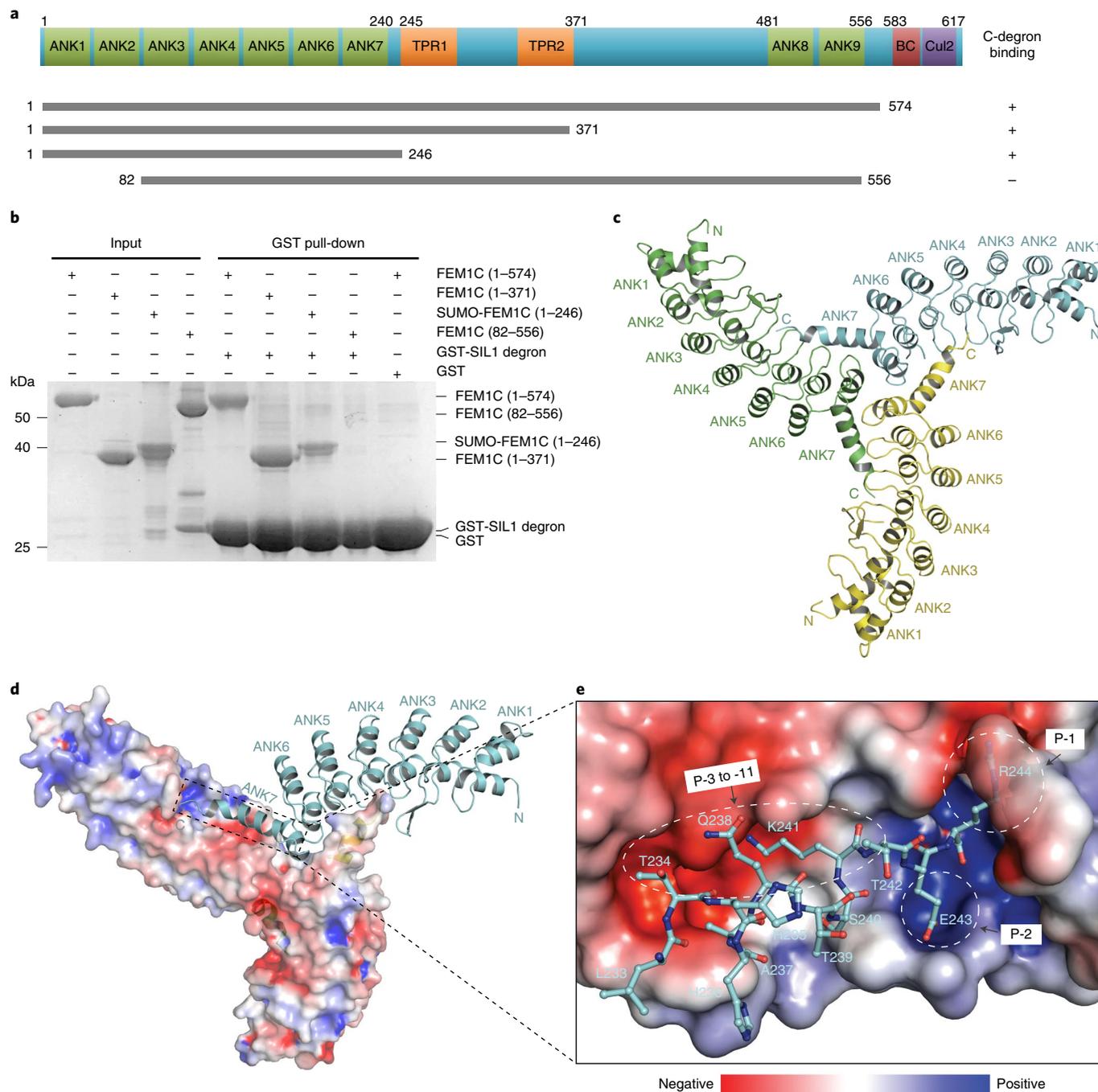
Fig. 1 | Crystal structure of human FEM1C₁₋₂₄₄. a, Domain architecture of human FEM1C. ANK, ankyrin repeat; TPR, tetratricopeptide repeat; BC, elongin BC-binding motif; Cul2, cullin 2-binding motif. **b**, An intact N-terminal region of FEM1C is required for interaction with the SIL1 degron, which was assessed by GST pull-down assay. GST-tagged SIL1₄₅₁₋₄₆₁ peptide was used to pull down purified truncation constructs of FEM1C. **c**, Ribbon diagram of the crystal structure of FEM1C₁₋₂₄₄, which forms a trimer. **d**, The C terminus of ANK7 is inserted into a charged binding pocket formed by a second FEM1C molecule. The electrostatic potential surface is plotted at ±5 kT/e. **e**, Close-up view of the C-terminal residues, which are located in the charged binding pocket.

Despite the fact that Q−2 was positioned at the basic ditch, only the main-chain O atom of Q−2 formed two hydrogen bonds with the side chain of Arg 121. In contrast, in the basic patch, the guanidinium of R−3 formed two hydrogen bonds and one salt bridge with the side chains of Asn 183, Asp 188 and Glu 191, respectively (Fig. 3d). Interestingly, both the FEM and SIL1 degrons carried a lysine at the −4 position (Fig. 2b), and in the FEM degron structure, K−4 mimicking R−3 of the REV degron occupied the same

positively charged binding pocket of FEM1C (Extended Data Fig. 1d–f). As with Q−2, only the main-chain O atom of E−4 in the REV degron was bound by Arg 121 and Asn 146 through hydrogen bonds. The following R−5 and W−6 were not involved in any apparent interactions with FEM1C. R−7 associated with Glu 191 and Asp 211, which was mediated by electrostatic interactions. The residues upstream of R−7 turned away from the concave binding groove and, therefore, did not contribute directly to Arg/C-degron
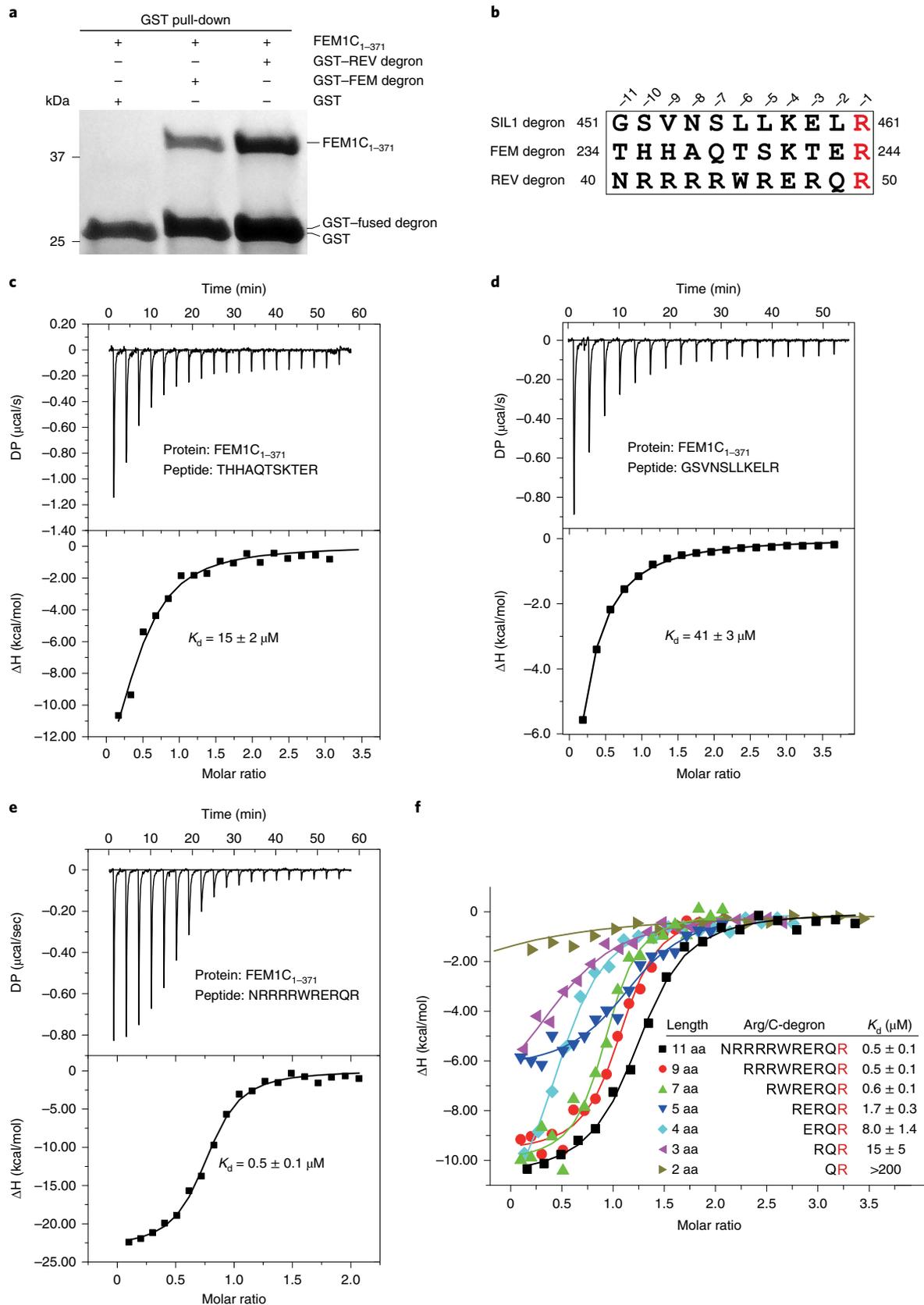
Fig. 2 | Characterization of Arg/C-degron recognition by FEM1C. a, FEM1C$_{234-244}$ or REV$_{40-50}$ peptide, fused to the C terminus of GST, was used to pull down purified FEM1C$_{1-371}$. **b**, Sequence comparison of different Arg/C-degrons. **c**, ITC binding curve of FEM1C$_{1-371}$ with FEM degron. **d**, ITC binding curve of FEM1C$_{1-371}$ with SIL1 degron. **e**, ITC binding curve of FEM1C$_{1-371}$ with REV degron. The K_d error is the fitted error generated from data analysis software (Origin 7.0) using the one-site binding model. **f**, ITC fitting curves of FEM1C$_{1-371}$ titrated with variable lengths of REV peptide; the corresponding peptide lengths, sequences and binding affinities are indicated.
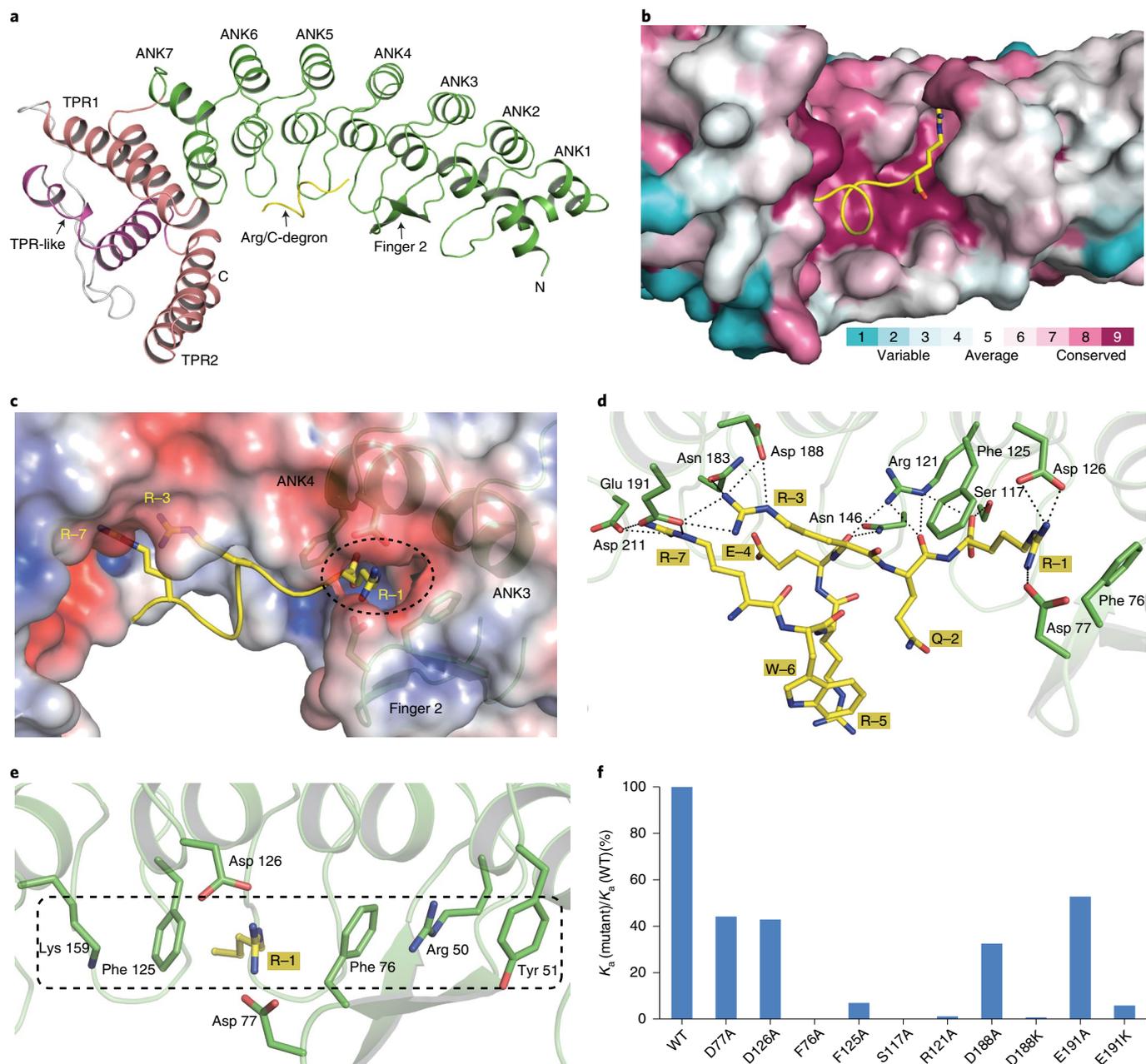
Fig. 3 | Structural basis of Arg/C-degron recognition by FEM1C. a, Ribbon diagram of the overall structure of FEM1C$_{1-371}$ bound to the REV Arg/C-degron peptide. ANK, TPR and TPR-like motifs are colored in green, salmon and magenta, respectively. The Arg/C-degron is shown in yellow. The finger 2 motif between ANK2 and ANK3 forms a β-hairpin fold. **b**, Conservation analysis[48] of the Arg/C-degron binding pocket in FEM1C. The residues of FEM1C are colored based on conservation grades: conserved residues are indicated in maroon and variable residues are indicated in turquoise. **c**, The electrostatic potential surface of the REV-binding pocket in FEM1C$_{1-371}$. Red, negative; blue, positive. The bound REV peptide is shown in yellow, and R−1, R−3 and R−7 are shown as sticks. **d**, Close-up view of the interactions of FEM1C$_{1-371}$ with the REV peptide. Residues of FEM1C are shown as green sticks, the peptide is shown as a yellow stick and the hydrogen bonds are shown as black dashed lines. **e**, The C-terminal arginine of the REV peptide (yellow) participates in a cation-π interaction network with FEM1C (green). **f**, Plot of binding affinities (relative association constant, K_a) of FEM1C$_{1-371}$ mutants for the REV peptide by ITC, with the affinity of wild-type (WT) FEM1C$_{1-371}$ for REV peptide taken as 100%. See also Supplementary Table 3.

recognition, which helps to explain why FEM1C confers selectivity against the C-terminal seven-residue degron for high binding affinity as observed in our ITC assays.

Key residues of FEM1C in Arg/C-degron binding. To further validate the key residues of FEM1C in mediating Arg/C-degron binding, we generated a series of point mutants of FEM1C and performed ITC assays to examine their binding affinities toward the

REV peptide. As expected, the alterations caused variable reductions in binding affinity, ranging from mild effects to apparent loss of binding ability (Fig. 3f and Supplementary Table 3). Notably, six residues of FEM1C (Phe 76, Asp 77, Ser 117, Arg 121, Phe 125 and Asp 126) were involved in direct interactions with R−1 (Fig. 3d) and are absolutely conserved in the FEM family of proteins (Supplementary Fig. 1), implicating the strict conservation in recognition of Arg/C-degrons by the FEM family. Substitution of the R−1
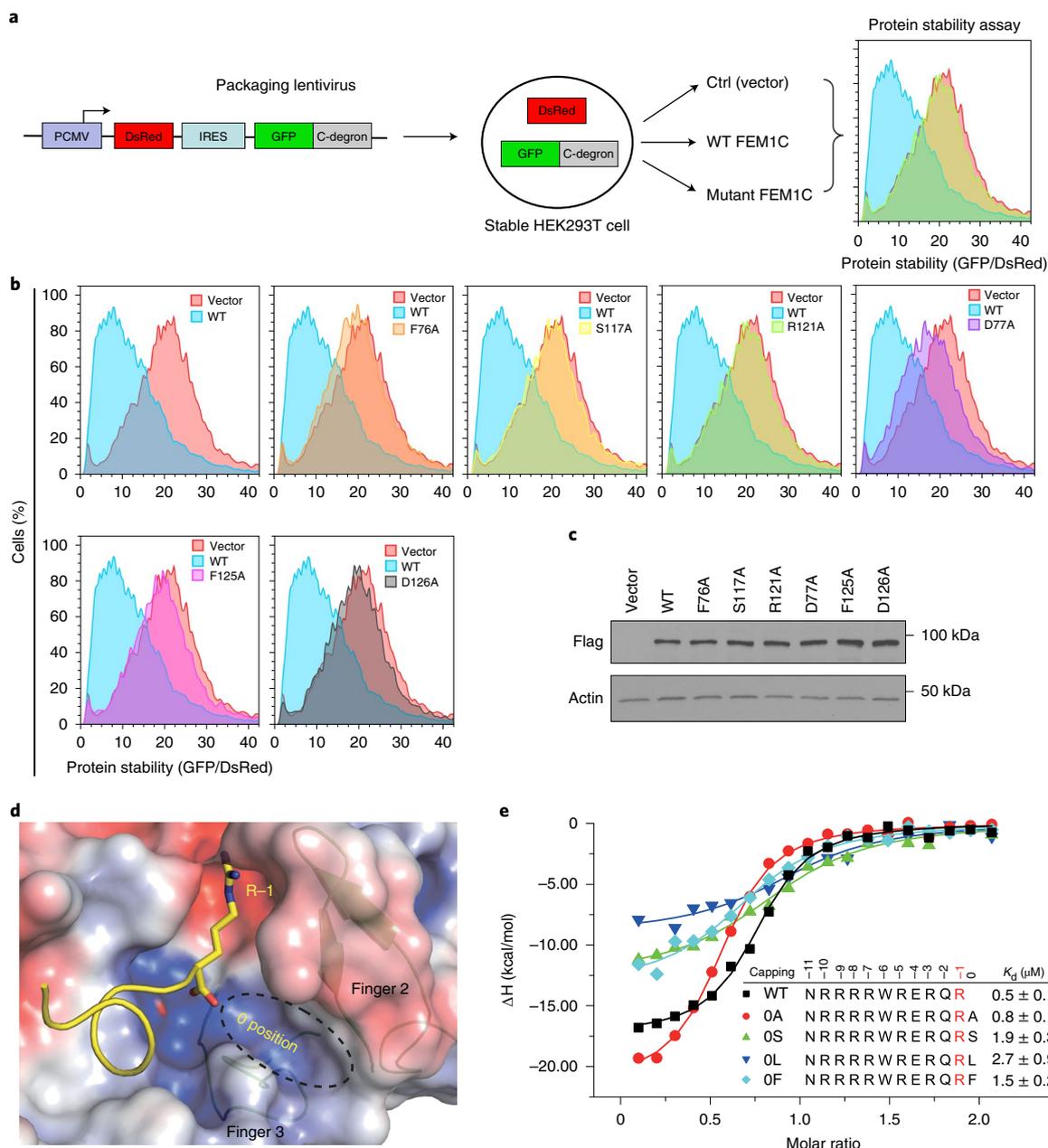
Fig. 4 | Mutagenesis studies of the key interactions between FEM1C and Arg/C-degron. a, Schematic representation of the GPS assay. P_CMV, cytomegalovirus promoter; IRES, internal ribosome entry site; Ctrl, control; WT, wild type. **b**, Stability analysis of the GFP-fused SIL1 degron with exogenous expression of wild-type and mutant FEM1C proteins. The GFP/DsRed ratio was analyzed by flow cytometry. **c**, Western blot analysis of wild-type and mutant FEM1C expression in HEK293T GPS reporter cell lines. **d**, Surface representation of the free C-terminal arginine-binding pocket. A hypothetical 0 position following the R–1 position is encircled by the dotted line. **e**, ITC fitting curves of FEM1C_{1–371} titrated with REV peptide capped with different residues at the 0 position. The corresponding peptide sequences and binding affinities are indicated.

guanidinium-interacting Asp 77 or Asp 126 with alanine reduced the binding affinity by ~2-fold. The F125A and F76A mutants, in which the cation-π interaction was impaired, showed dramatically decreased binding affinity by ~15-fold or abolished degron binding, suggesting a critical role of cation-π interactions in R–1 recognition. Not surprisingly, the FEM1C_{82–556} construct, lacking the two crucial residues Phe 76 and Asp 77, could not bind to the SIL1 peptide in the GST pull-down assay (Fig. 1b). Additionally, the S117A substitution, which would disrupt the hydrogen bond with the carboxyl group of R–1, abrogated degron binding, highlighting the importance of this hydrogen bond in mediating C-end degron recognition.

Likewise, the R121A mutant only weakly bound to the REV peptide, which is in agreement with our structural observations that Arg 121 is involved in multiple interactions with R–1, Q–2 and E–4. Alanine mutation of Glu 191 or Asp 188 diminished the binding affinity by 2- to 3-fold, whereas mutating the negatively charged residues to lysine resulted in 20-fold-decreased binding (E191K) or an almost complete loss of degron binding (D188K), suggesting that this acidic patch is favored by the positively charged residues embedded in the upstream sequence of the Arg/C-degron.

To further corroborate whether recognition of R–1 in the SIL1 degron by FEM1C is required for SIL1 degradation, the in vivo

stability of the SIL1 peptide was examined by GPS technology[39–41]. We introduced the SIL1 peptide at the C terminus of green fluorescent protein (GFP) in the lentiviral GPS vector and assessed degradation of the GFP fusion protein by monitoring the GFP/DsRed ratio using DsRed as an internal reference (Fig. 4a). Overexpression of Flag-tagged wild-type FEM1C dramatically reduced the GFP/DsRed ratio and promoted the degradation of the GFP-fused SIL1 degron. In contrast, the degron-binding-deficient mutants, including D77A, D126A, F76A, F125A, S117A and R121A, did not promote the degradation of the GFP-fused SIL1 degron (Fig. 4b,c), further confirming the essential role of R–1 recognition in FEM1C-mediated SIL1 degradation.

R–1 serves as the principal recognition determinant. To explore the key elements required for FEM1C recognition within Arg/C-degrons, we synthesized a series of mutant peptides derived from the REV peptide and examined their binding affinities to FEM1C by ITC (Supplementary Table 2). Not surprisingly, amidation of the C-terminal carboxyl group resulted in 14-fold-decreased binding affinity, further confirming that FEM1C favors a free C terminus. Replacement of R–1 with alanine substantially weakened the binding by 34-fold, and even mutating R–1 to positively charged lysine also reduced the binding affinity by 20-fold, suggesting that FEM1C specifically recognizes arginine at the –1 position. Accordingly, in the complex structure, the –1 position binding cage consisted of a pair of aspartate residues (Asp 126 and Asp 77) and a pair of phenylalanine residues (Phe 76 and Phe 125; Fig. 3d), which preferentially recognize the guanidinium group by a combination of hydrogen bonding and salt bridge and cation-π interactions. Consistent with previous studies, lysine is disfavored at the –1 position[25], probably due to the fact that arginine forms stronger cation-π interactions with phenylalanine as well as more extensive electrostatic interactions with aspartate owing to its geometric structure compared to lysine[42,43]. This is reminiscent of N-terminal arginine recognition by the UBR proteins in which lysine substitution failed to form hydrogen bonds with arginine-interacting residues due to its shorter side chain[18], although lysine in the N-degron could occupy the arginine degron-binding site.

The important role of the –3 position of the Arg/C-degron. Unlike substitution of R–1, peptide with substitution of Q–2 with alanine had a comparable binding affinity to the wild-type REV peptide (Supplementary Table 2), in agreement with our findings that only the main chain of Q–2 participates in the interactions (Fig. 3d). However, replacement of R–3 with alanine in the REV peptide, which disrupts electrostatic interactions, greatly decreased the binding affinity by tenfold. Conversely, arginine substitution of T–3 in the FEM peptide or E–3 in the SIL1 peptide increased the binding affinity by around fourfold. The E–3R mutant peptide of SIL1 was less stable in the GPS assay (Extended Data Fig. 2a), supporting the importance of R–3 in mediating interactions with FEM1C. In contrast to R–1 and R–3, other alanine mutations of the REV degron spanning from positions –4 to –10 had minor or no effects on degron binding. Intriguingly, in vivo, FEM1B is responsible for the selective degradation of CDK5R1 (ref. [24]), which carries a leucine at the –3 position (Supplementary Table 2). We hypothesize that L–3 is important for recognition of the CDK5R1 C-degron by FEM1B. Based on the FEM1C–REV complex structure, R–3 is located in a negatively charged pocket created by Asn 183, Asp 188 and Glu 191 (Extended Data Fig. 2b), which prefers binding to the positively charged R–3. In contrast to FEM1C and FEM1A, FEM1B has hydrophobic alanine and phenylalanine residues instead of the hydrophilic asparagine and aspartic acid residues of FEM1C and FEM1A at positions 188 and 193, respectively (Extended Data Fig. 2c). The resultant hydrophobic pocket in FEM1B is more suitable for binding to a hydrophobic residue in the –3 position, such as L–3

of CDK5R1. Indeed, our ITC binding results showed that FEM1B had more than tenfold greater selectivity toward wild-type CDK5R1 than the L–3R or L–3A mutant (Supplementary Table 2). Therefore, FEM1B might primarily eliminate proteins with a hydrophobic residue located at the –3 position of the Arg/C-degron. Nevertheless, FEM1C was still able to bind the CDK5R1 peptide in vitro, albeit with threefold-reduced affinity compared to FEM1B. But the L–3R alteration strengthened the binding affinity of FEM1C by tenfold, converting it to a FEM1C-selective Arg/C-degron (Supplementary Table 2). Thus, Arg/C-degrons can be targeted for degradation by FEM1A, FEM1B and FEM1C, and the upstream sequence context of R–1 further fine-tunes substrate selectivity to achieve exquisite regulation of protein turnover in vivo.

Effects of Arg/C-degron capping. FEM1C functions as the receptor for the extreme C-terminal arginine of substrates[24,25]. Furthermore, FEM1C was shown to recognize substrates carrying an arginine at the –2 or –3 position[25]. In our crystal structure, we found that the side chain of R–1 inserted into the semi-open binding pocket, but the terminal carboxyl group of R–1 was not buried in this pocket. One O atom of the carboxyl group was pointing out toward a shallow groove formed by finger 2 and finger 3 (Fig. 4d). In other words, the otherwise free carboxylate of R–1 might be capped by an additional residue in a hypothetical 0 position. Indeed, our ITC analysis showed that FEM1C could bind REV peptide with an additional capping residue at the 0 position. Specifically, capping with the small alanine residue caused a mild reduction in FEM1C binding, whereas leucine capping reduced the binding affinity by fivefold (Fig. 4e). Leucine capping also resulted in a more stable peptide in our degradation assay (Extended Data Fig. 2d). In addition to the REV peptide, we also examined the effects of capping the FEM and SIL1 peptides with a single residue at the 0 position. ITC binding results revealed that capping the FEM or SIL1 peptide with alanine could be tolerated, but serine, leucine or phenylalanine capping caused an almost complete loss of binding to FEM1C (Supplementary Table 2). In contrast, capping of the REV peptide with two or three serine residues could be tolerated, although there was a 50- to 100-fold reduction in FEM1C binding (Extended Data Fig. 2e,f). This is possibly due to the fact that the REV peptide has a low-micromolar affinity, which is 100-fold stronger than that for the SIL1 or FEM peptide; therefore, the REV peptide capped with 1–3 residues could still be detected with modest binding. Collectively, our data suggest that the free R–1 acts as the principal recognition determinant for FEM1C. Adding any additional residues following R–1 would weaken the binding to FEM1C. The residues upstream of R–1 are not strictly conserved and substitutions can be tolerated, although the presence of arginine at the –3 position is preferred and the intramolecular interaction networks are required for robust binding of an Arg/C-degron to FEM1C.

Discussion

In the UPS, the E3 Ub ligases specifically recognize their degrons through a sophisticated interaction network[44]. In this study, we characterized the interactions of FEM1C with its Arg/C-degron. This degron-binding model is considerably different from that of the Gly/C-degron. Specifically, KLHDC2 recognizes Gly/C-degron through a deep binding pocket[45], similar to Pro/N-degron recognition by GID4 (refs. [22,23]). However, FEM1C utilizes a concave groove to capture its Arg/C-degron.

Virtually all proteins have unique N- and C-terminal sequence contexts, which can become degradation signals and be recognized by respective E3 ligases for degradation through distinct N- and C-degron pathways[11]. In this sense, arginine can act as a destabilizing residue no matter whether it is located at the N terminus or C terminus of proteins. When it is present at the N terminus, the resulting Arg/N-degron can be recognized by the UBR

proteins[18,19], while at the C terminus, the resulting Arg/C-degron can be recognized by the FEM proteins[24,25]. In both cases, the substrate-binding pockets use an acidic binding surface to engage the arginine degron, and either the free α-amino group or the free carboxyl group is oriented by hydrogen-bonding interactions (Extended Data Fig. 3a–d). Remarkably, FEM1C further stabilizes the C-terminal arginine via specific cation-π interactions. Moreover, the side chain of the C-terminal arginine is inserted into a semi-open binding pocket of FEM1C, while the whole N-terminal arginine lies on a shallow binding groove of UBR proteins (Extended Data Fig. 3a,b). Besides the arginine residue, the UBR proteins favor a hydrophobic residue at the second position of the Arg/N-degron, whereas FEM1C does not exhibit an analogous preference at the penultimate position but favors a positively charged residue at the antepenultimate position in the Arg/C-degron. Additionally, the UBR proteins mainly interact with the first two residues of their degrons, whereas FEM1C recognizes a relatively long degron sequence context. Thus, the unique degron recognition mechanisms of these proteins may boost their substrate selectivity and fine-tune protein turnover by the UPS.

The human proteome includes more than 2,000 different proteins with arginine as their last or penultimate residue at the C terminus. In addition to full-length proteins, Arg/C-degrons could be exposed by proteolysis, suggesting that FEM1C could have many potential substrates and participate in a very broad spectrum of biological processes. Multiple somatic mutations of *FEM1C* occur in various cancer samples[46,47]. Importantly, the cancer-associated alterations D77Y, D77G, D126N, N146S and E191Q of FEM1C are located in the degron-binding groove, and our binding results reveal that these cancer-associated substitutions impair FEM1C binding to the REV peptide (Supplementary Table 3).

In summary, the UPS is the primary mechanism for protein turnover, and the proper proteasome-mediated removal of abnormal cellular proteins is an evolutionarily conserved mechanism for maintaining human health and well-being. Unsurprisingly, defects associated with the UPS and its upstream substrate recognition pathways are frequently associated with disease predisposition. Therefore, our study not only elucidates the substrate recognition mechanism of Arg/C-degron by FEM1C, but may also facilitate the identification of physiological substrates and development of therapeutic inhibitors.

Online content

References

1. Hershko, A. & Ciechanover, A. The ubiquitin system. *Annu. Rev. Biochem.* **67**, 425–479 (1998).
2. Balchin, D., Hayer-Hartl, M. & Hartl, F. U. In vivo aspects of protein folding and quality control. *Science* **353**, aac4354 (2016).
3. Johnson, B. M. & DeBose-Boyd, R. A. Underlying mechanisms for sterol-induced ubiquitination and ER-associated degradation of HMG CoA reductase. *Semin. Cell Dev. Biol.* **81**, 121–128 (2018).
4. Sontag, E. M., Samant, R. S. & Frydman, J. Mechanisms and functions of spatial protein quality control. *Annu. Rev. Biochem.* **86**, 97–122 (2017).
5. Pickart, C. M. Mechanisms underlying ubiquitination. *Annu. Rev. Biochem.* **70**, 503–533 (2001).
6. Finley, D. Recognition and processing of ubiquitin–protein conjugates by the proteasome. *Annu. Rev. Biochem.* **78**, 477–513 (2009).
7. Varshavsky, A. Naming a targeting signal. *Cell* **64**, 13–15 (1991).
8. Lucas, X. & Ciulli, A. Recognition of substrate degrons by E3 ubiquitin ligases and modulation by small-molecule mimicry strategies. *Curr. Opin. Struct. Biol.* **44**, 101–110 (2017).
9. 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).
10. Varshavsky, A. The N-end rule pathway and regulation by proteolysis. *Protein Sci.* **20**, 1298–1345 (2011).
11. Varshavsky, A. N-degron and C-degron pathways of protein degradation. *Proc. Natl Acad. Sci. USA* **116**, 358–366 (2019).
12. Hwang, C. S., Shemorry, A. & Varshavsky, A. N-terminal acetylation of cellular proteins creates specific degradation signals. *Science* **327**, 973–977 (2010).
13. 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).
14. Timms, R. T. et al. A glycine-specific N-degron pathway mediates the quality control of protein N-myristoylation. *Science* https://doi.org/10.1126/science.aaw4912 (2019).
15. Kim, J. M. et al. Formyl-methionine as an N-degron of a eukaryotic N-end rule pathway. *Science* https://doi.org/10.1126/science.aat0174 (2018).
16. Dissmeyer, N. Conditional protein function via N-degron pathway-mediated proteostasis in stress physiology. *Annu. Rev. Plant Biol.* https://doi.org/10.1146/annurev-arplant-050718-095937 (2019).
17. Liu, Y., Liu, C., Dong, W. & Li, W. Physiological functions and clinical implications of the N-end rule pathway. *Front. Med.* **10**, 258–270 (2016).
18. Choi, W. S. et al. Structural basis for the recognition of N-end rule substrates by the UBR box of ubiquitin ligases. *Nat. Struct. Mol. Biol.* **17**, 1175–1181 (2010).
19. Matta-Camacho, E., Kozlov, G., Li, F. F. & Gehring, K. Structural basis of substrate recognition and specificity in the N-end rule pathway. *Nat. Struct. Mol. Biol.* **17**, 1182–1187 (2010).
20. Kwon, D. H. et al. Insights into degradation mechanism of N-end rule substrates by p62/SQSTM1 autophagy adapter. *Nat. Commun.* **9**, 3291 (2018).
21. Zhang, Y. et al. ZZ-dependent regulation of p62/SQSTM1 in autophagy. *Nat. Commun.* **9**, 4373 (2018).
22. Dong, C. et al. Molecular basis of GID4-mediated recognition of degrons for the Pro/N-end rule pathway. *Nat. Chem. Biol.* **14**, 466–473 (2018).
23. Dong, C. et al. Recognition of nonproline N-terminal residues by the Pro/N-degron pathway. *Proc. Natl Acad. Sci. USA* https://doi.org/10.1073/pnas.2007085117 (2020).
24. Koren, I. et al. The eukaryotic proteome is shaped by E3 ubiquitin ligases targeting C-terminal degrons. *Cell* **173**, 1622–1635 (2018).
25. Lin, H. C. et al. C-terminal end-directed protein elimination by CRL2 ubiquitin ligases. *Mol. Cell* **70**, 602–613 (2018).
26. 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).
27. Timms, R. T. & Koren, I. Tying up loose ends: the N-degron and C-degron pathways of protein degradation. *Biochem. Soc. Trans.* https://doi.org/10.1042/BST20191094 (2020).
28. Kamura, T. et al. VHL-box and SOCS-box domains determine binding specificity for Cul2–Rbx1 and Cul5–Rbx2 modules of ubiquitin ligases. *Genes Dev.* **18**, 3055–3065 (2004).
29. 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).
30. 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).
31. Starostina, N. G. et al. A CUL-2 ubiquitin ligase containing three FEM proteins degrades TRA-1 to regulate *C. elegans* sex determination. *Dev. Cell* **13**, 127–139 (2007).
32. Chan, S. L., Yee, K. S., Tan, K. M. & Yu, V. C. The *Caenorhabditis elegans* sex determination protein FEM-1 is a CED-3 substrate that associates with CED-4 and mediates apoptosis in mammalian cells. *J. Biol. Chem.* **275**, 17925–17928 (2000).
33. Subauste, M. C. et al. Fem1b, a proapoptotic protein, mediates proteasome inhibitor-induced apoptosis of human colon cancer cells. *Mol. Carcinog.* **49**, 105–113 (2010).
34. Wang, S. et al. Atlas on substrate recognition subunits of CRL2 E3 ligases. *Oncotarget* **7**, 46707–46716 (2016).
35. 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).
36. Blatch, G. L. & Lassle, M. The tetratricopeptide repeat: a structural motif mediating protein–protein interactions. *Bioessays* **21**, 932–939 (1999).
37. Malim, M. H., Bohnlein, S., Hauber, J. & Cullen, B. R. Functional dissection of the HIV-1 Rev trans-activator—derivation of a trans-dominant repressor of Rev function. *Cell* **58**, 205–214 (1989).

38. Sedgwick, S. G. & Smerdon, S. J. The ankyrin repeat: a diversity of interactions on a common structural framework. *Trends Biochem. Sci.* **24**, 311–316 (1999).

39. 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).

40. Yen, H. C. & Elledge, S. J. Identification of SCF ubiquitin ligase substrates by global protein stability profiling. *Science* **322**, 923–929 (2008).

41. Emanuele, M. J. et al. Global identification of modular cullin-RING ligase substrates. *Cell* **147**, 459–474 (2011).

42. Sokalingam, S., Raghunathan, G., Soundrarajan, N. & Lee, S. G. A study on the effect of surface lysine to arginine mutagenesis on protein stability and structure using green fluorescent protein. *PLoS ONE* **7**, e40410 (2012).

43. Gallivan, J. P. & Dougherty, D. A. Cation-π interactions in structural biology. *Proc. Natl Acad. Sci. USA* **96**, 9459–9464 (1999).

44. Zheng, N. & Shabek, N. Ubiquitin ligases: structure, function, and regulation. *Annu. Rev. Biochem.* **86**, 129–157 (2017).

45. Rusnac, D. V. et al. Recognition of the diglycine C-end degron by CRL2^{KLHDC2} ubiquitin ligase. *Mol. Cell* **72**, 813–822 (2018).

46. Forbes, S. A. et al. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.* **45**, D777–D783 (2017).

47. Simonyan, V. & Mazumder, R. High-performance integrated virtual environment (HIVE) tools and applications for big data analysis. *Genes* **5**, 957–981 (2014).

48. Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* **38**, W529–W533 (2010).

Methods

Cloning, expression and purification of human FEM1C and FEM1B. The sequences encoding human FEM1C$_{1-574}$, FEM1C$_{82-556}$, FEM1C$_{1-246}$ and FEM1B$_{1-357}$ were subcloned into the pET28-MKH8SUMO expression vector with an 8×His–SUMO tag followed by a TEV cleavage site. The sequence encoding FEM1C$_{1-244}$ was cloned into a pET28GST-LIC vector containing a GST tag and a thrombin cleavage site. The sequence encoding FEM1C$_{1-371}$ was cloned into a pET28-MHL vector with a 6×His tag followed by a TEV cleavage site. These plasmids were transformed into *Escherichia coli* BL21 (DE3) cells under kanamycin selection. The cells were grown in Terrific Broth medium at 37 °C until an OD$_{600}$ of ~1.0 was reached, and then the temperature was reduced to 18 °C. The proteins were expressed by induction with 0.2 mM IPTG overnight. The cells were harvested by centrifugation and were lysed by sonication in lysis buffer (20 mM Tris-HCl pH 7.5, 500 mM NaCl, 5% (vol/vol) glycerol and 2 mM β-mercaptoethanol). The lysate was cleared by centrifugation, and the supernatant containing FEM1C$_{1-244}$ was loaded onto a GST column and washed with lysis buffer. The GST tag was removed by the addition of 60 U thrombin protease (Sigma) to the column at 4 °C overnight. The other fusion proteins were applied to Ni-NTA columns, washed with 25 mM imidazole in lysis buffer and then eluted with 250 mM imidazole in lysis buffer. The samples were digested by homemade TEV protease with a molar ratio of 1:20 in dialysis buffer (20 mM Tris-HCl pH 7.5 and 500 mM NaCl) at 4 °C overnight to remove the SUMO tag or His tag and were then reloaded onto the Ni-NTA column. All proteins were further purified using a Superdex 200 size exclusion column (GE Healthcare) equilibrated with gel-filtration buffer (20 mM Tris-HCl pH 7.5, 250 mM NaCl and 1 mM DTT). The peak fractions were collected and concentrated to 8 mg ml^{-1} using an Amicon Ultra centrifugal filter. Purity was assessed by SDS–PAGE. The purified proteins were flash frozen in liquid nitrogen and stored at −80 °C. The mutants were generated by QuikChange using the sequence encoding FEM1C$_{1-371}$ as a template and verified by DNA sequencing. The mutant proteins were expressed and purified similarly to the wild-type proteins.

Cell culture and viral transduction. Human HEK293T cells (ATCC) were maintained in DMEM (Cellgro) supplemented with 10% (vol/vol) FBS (Sigma). pMD2.G (Addgene) and psPAX2 (Addgene) were cotransfected into HEK293T cells with either pCDH-puro-cMyc (Addgene, 46970) or pHAGE-GPS3.0-DEST lentiviral transfer vector (gifted by S.J. Elledge, Howard Hughes Medical Institute). Viral supernatants were harvested after 48 h and used to infect HEK293T cells in the presence of 8 μg ml^{-1} polybrene.

Global protein stability assay. The oligonucleotide encoding the 11-mer SIL1 or REV peptide was cloned into pENTR3C and subsequently cloned into the lentiviral vector pHAGE-GPS3.0-DEST using Gateway techniques (Invitrogen). The DsRed- and GFP-fused C-degron proteins can be coexpressed by the pHAGE-GPS3.0-DEST vector carrying a single promoter and an internal ribosome entry site. The stability of the GFP-fused C-degron was determined by measuring the cellular GFP/DsRed ratio, which can be analyzed by flow cytometry with DsRed as an internal control. To test the activity of FEM1C mutants, human *FEM1C* cDNA was cloned into pENTR3C and subsequently cloned into pCDH-Flag destination vectors using Gateway techniques (Invitrogen). pCDH-Flag Gateway Destination vector was constructed based on pCDH-puro-cMyc by adding a Flag tag and Gateway cassette. The wild-type and mutant FEM1C were overexpressed in GPS reporter cells by lentivirus infection after selection with blasticidin (10 μg ml^{-1}) for 4 d. Cells were analyzed by flow cytometry using a Beckman Coulter CytoFLEX S (CytExpert 2.3 software). Expression of wild-type and mutant FEM1C in HEK293T GPS reporter cell lines was examined by western blot with anti-Flag M2 (1:5,000; Sigma-Aldrich, F3165); anti-actin (1:5,000; Sigma-Aldrich, A1978) was used as an internal control.

Protein crystallization. The crystallization trials were performed at 18 °C using the sitting drop vapor diffusion method by mixing 1 μl of protein and 1 μl of reservoir solution. The FEM1C$_{1-244}$ crystals were crystallized in solutions consisting of 15% (wt/vol) PEG3350 and 17% (vol/vol) Tacsimate pH 8.0. For crystallization of the FEM1C$_{1-371}$–REV peptide complex, FEM1C$_{1-371}$ was mixed with the REV peptide at a molar ratio of 1:1.5 for 1 h on ice. The crystals were obtained in precipitant conditions consisting of 16% (vol/vol) Jeffamine M-600 pH 7.0 and 0.1 M HEPES 7.0. The crystal was mounted in the respective reservoir solution with the addition of 20% (vol/vol) glycerol as a cryoprotectant and flash frozen in liquid nitrogen.

Data collection and structure determination. Diffraction images of FEM1C$_{1-244}$ were collected at beamline 24-ID-C of the Advanced Photon Source (APS) and processed with HKL3000 (ref. [49]). The initial structure was solved by molecular replacement with Phaser[50] and the search model from Protein Data Bank entry 5MA4 (amino acids 13–165; ref. [51]). Iterations of phase improvement were performed with Parrot[52], automated building was performed with Buccaneer[53], additional phase modification was performed with RESOLVE inside phenix.autobuild, manual building was performed in Coot[54], further iterations of rebuilding were performed in Coot and refinement was performed in autoBUSTER. The diffraction data of the FEM1C–REV peptide complex were collected at APS beamline 24-ID-E and processed with HKL3000 (ref. [49]). The structure was solved by molecular replacement with the program Phaser and coordinates from the FEM1C$_{1-244}$ structure. The model was refined with REFMAC[55] and rebuilt with Coot.

Isothermal titration calorimetry. ITC measurements were performed at 25 °C using a MicroCal Auto-iTC200 instrument (GE Healthcare). All proteins and peptides were prepared in an ITC buffer consisting of 20 mM Tris-HCl pH 7.5, 250 mM NaCl and 1 mM DTT with concentration ranges from 20 to 60 μM and 0.2 to 1 mM, respectively. For each experiment, the peptide was titrated into the protein with 20 injections of 2 μl each spaced by 180 s with a reference power of 10 μcal s^{-1}. The titration data were processed using Origin 7.0 software, and the K_d error was the fitted error in a one-site binding model.

Glutathione S-transferase pull-down and competition binding assays. GST-fusion C-degrons were produced in *E. coli* BL21 cells and purified using GST and Superdex 200 size exclusion columns (GE Healthcare). Approximately 90 μg of purified GST or GST-fusion degrons was applied to 20 μl of GST bead slurry, washed with binding buffer (20 mM Tris-HCl pH 7.5, 300 mM NaCl and 3 mM DTT) and resuspended in 180 μl of binding buffer. Purified FEM1C fragments were added to the reaction solutions and incubated on a rotating wheel for 30 min at 4 °C. After washing three times with binding buffer, the pulled-down samples were eluted with 10 mM glutathione and subjected to SDS–PAGE followed by Coomassie blue staining. For the competition binding assay, after immobilization of GST-tagged SIL1 peptide on 20 μl of GST beads, the beads were incubated with 80 μg of purified FEM1C$_{1-371}$ and different amounts of REV peptide with final concentrations of 0, 0.15, 0.3, 0.4 and 0.5 mM. After incubation at 4 °C for 30 min, the beads were washed three times with binding buffer. The captured proteins were eluted with 10 mM glutathione and analyzed by SDS–PAGE followed by Coomassie blue staining.

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

Data availability

The atomic coordinates and structure factors of FEM1C and the FEM1C–peptide complex have been deposited in the Protein Data Bank (https://www.rcsb.org/) with the accession codes 6XKC and 7JYA, respectively. Source data are provided with this paper.

References

49. 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 Biol. Crystallogr.* **62**, 859–866 (2006).
50. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
51. Hansen, S. et al. Design and applications of a clamp for green fluorescent protein with picomolar affinity. *Sci. Rep.* **7**, 16292 (2017).
52. Cowtan, K. Recent developments in classical density modification. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 470–478 (2010).
53. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1002–1011 (2006).
54. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
55. Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367 (2011).

Author contributions

C.D., J.M. and W.M. conceptualized the project and analyzed the data. X.Y. performed GST pull-down assays, protein purification and crystallization with help from Yao Li. C.D. determined the crystal structures. X.Y. and M.Z. conducted the ITC assays. X.W. performed the GPS assays under the supervision of W.M. X.Y., X.W., L.S. and Yanjun Li cloned the constructs. C.D. wrote the manuscript with critical input from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41589-020-00703-4.

Supplementary information is available for this paper at https://doi.org/10.1038/s41589-020-00703-4.

Correspondence and requests for materials should be addressed to W.M., J.M. or C.D.

Peer review information *Nature Chemical Biology* thanks David Dougan 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.
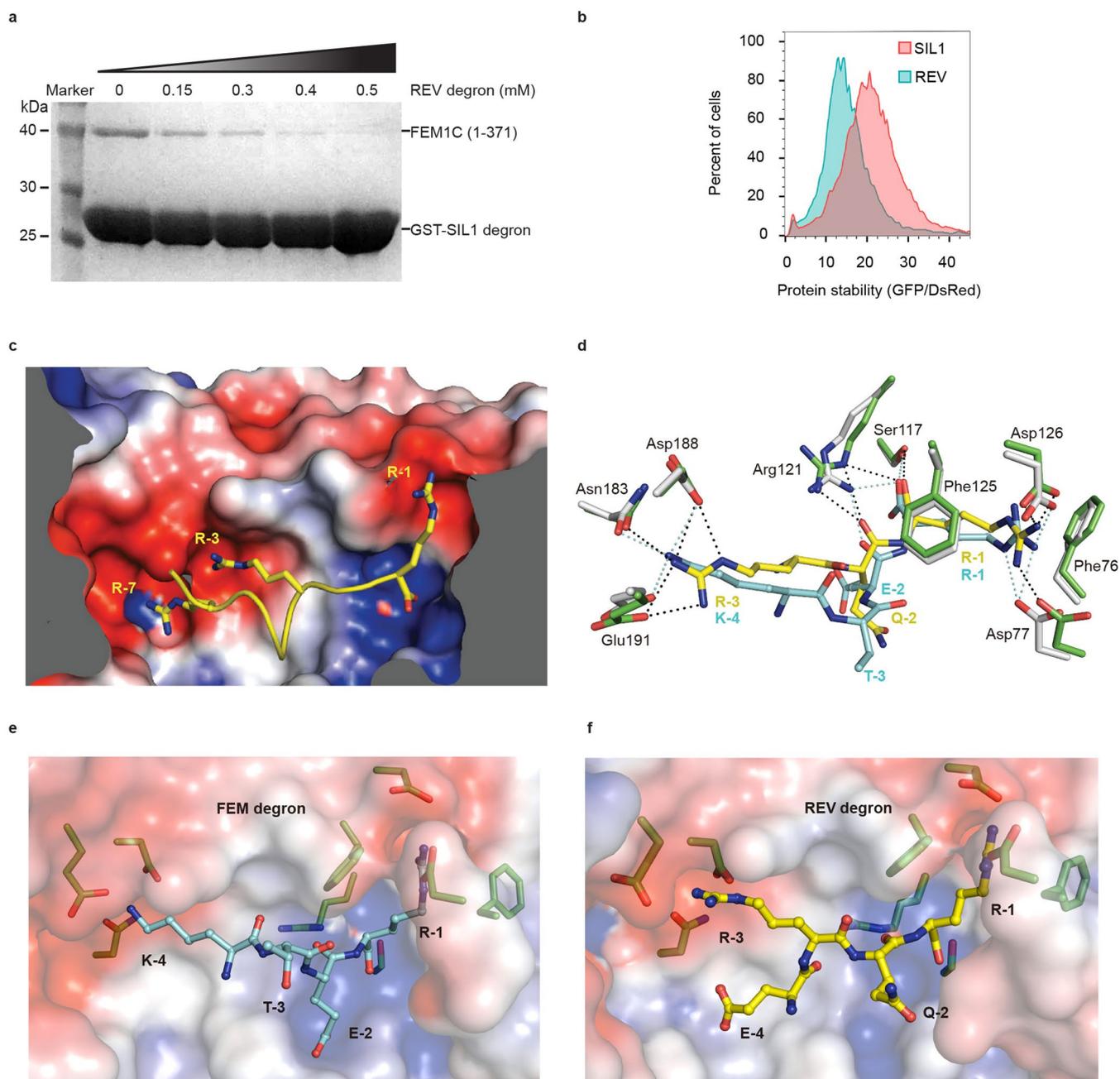
Extended Data Fig. 1 | Comparison of different C-degrons recognition by FEM1C. a, GST fusion SIL1 peptide immobilized on GST beads, and the pull-down was performed by incubating purified FEM1C (aa 1-371) in the presence of increasing concentrations of REV peptide. **b**, Stability comparison of the GFP-fused SIL1 and REV degrons by global protein stability assay (GPS experimental design in Fig. 4a). **c**, Cross-section view of the Arg/C-degron binding pocket. **d**, Overlay of REV degron ($^{-3}$RQR^{-1}) and FEM degron ($^{-4}$KTER^{-1}) binding pockets. The FEM peptide and its interacting residues of FEM1C are shown as cyan and gray sticks, respectively. Hydrogen bonds between FEM peptide and FEM1C are indicated as blue dash lines. The representation of REV-binding mode is same as Fig. 3d. **e,f**, The electrostatic potential surfaces of the FEM degron (cyan) and the REV degron (yellow) binding pockets. The K-4 of FEM degron and R-3 of REV degron share the same negatively charged binding pocket.
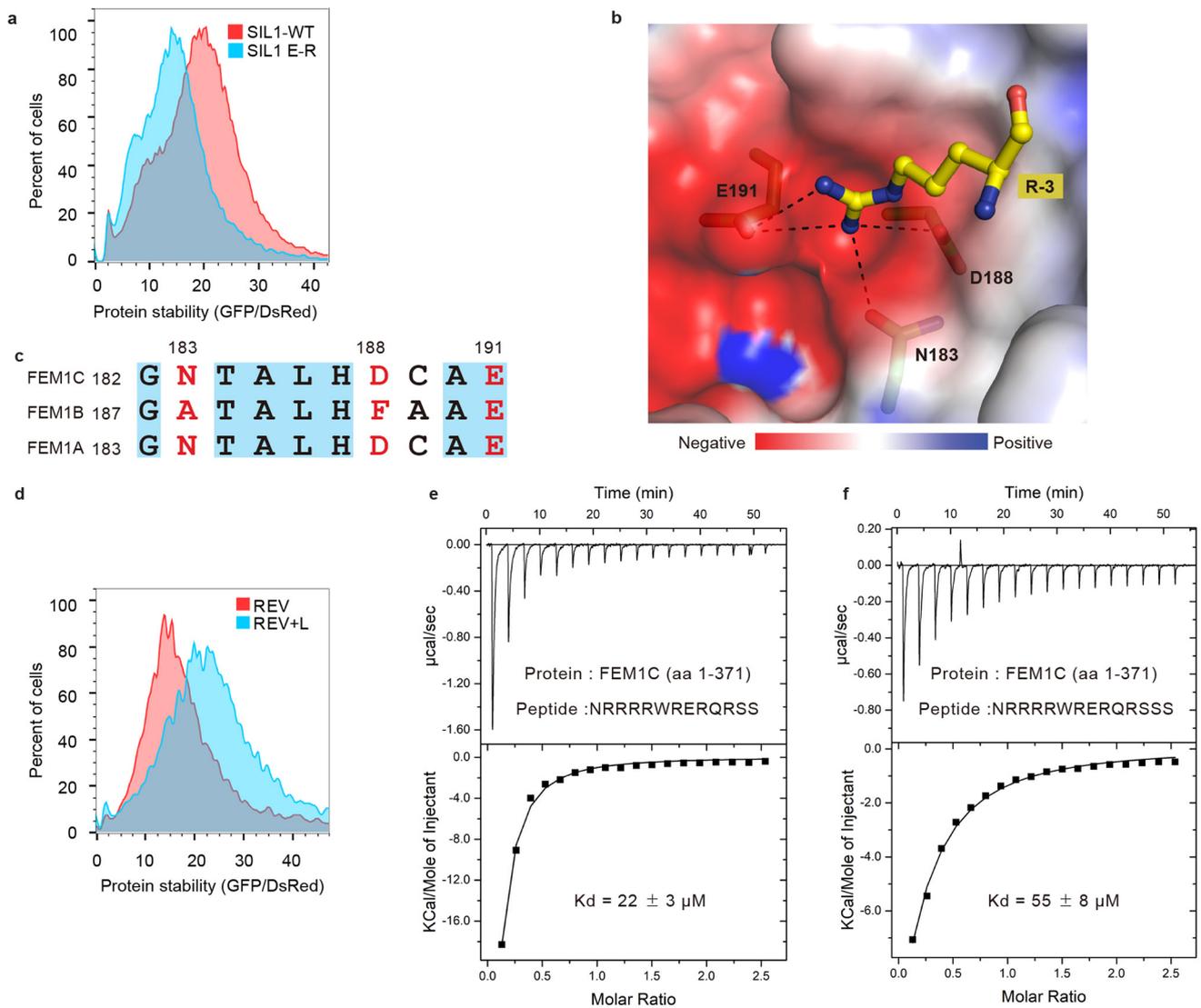
Extended Data Fig. 2 | The effects of C-degron sequence contexts on the binding of FEM proteins. a, Stability comparison of the GFP-fused SIL1 and its E-3R mutant by global protein stability assay. **b**, The electrostatic properties of R-3 binding pocket in FEM1C (red, negative; blue, positive). R-3 of REV degron is shown as yellow stick, and its hydrogen-bonding residues in FEM1C are indicated. **c**, Sequence alignment of FEM1C (aa 182-191), FEM1B (aa 187-196) and FEM1A (aa 183-192). The R-3 interacting residues are colored in red. **d**, Stability comparison of the GFP-fused REV degron and REV degron capped with a leucine by global protein stability assay. **e**, **f**, ITC curve of FEM1C (aa 1-371) binding to the REV degron capped with two serine (e) or three serine residues (f).
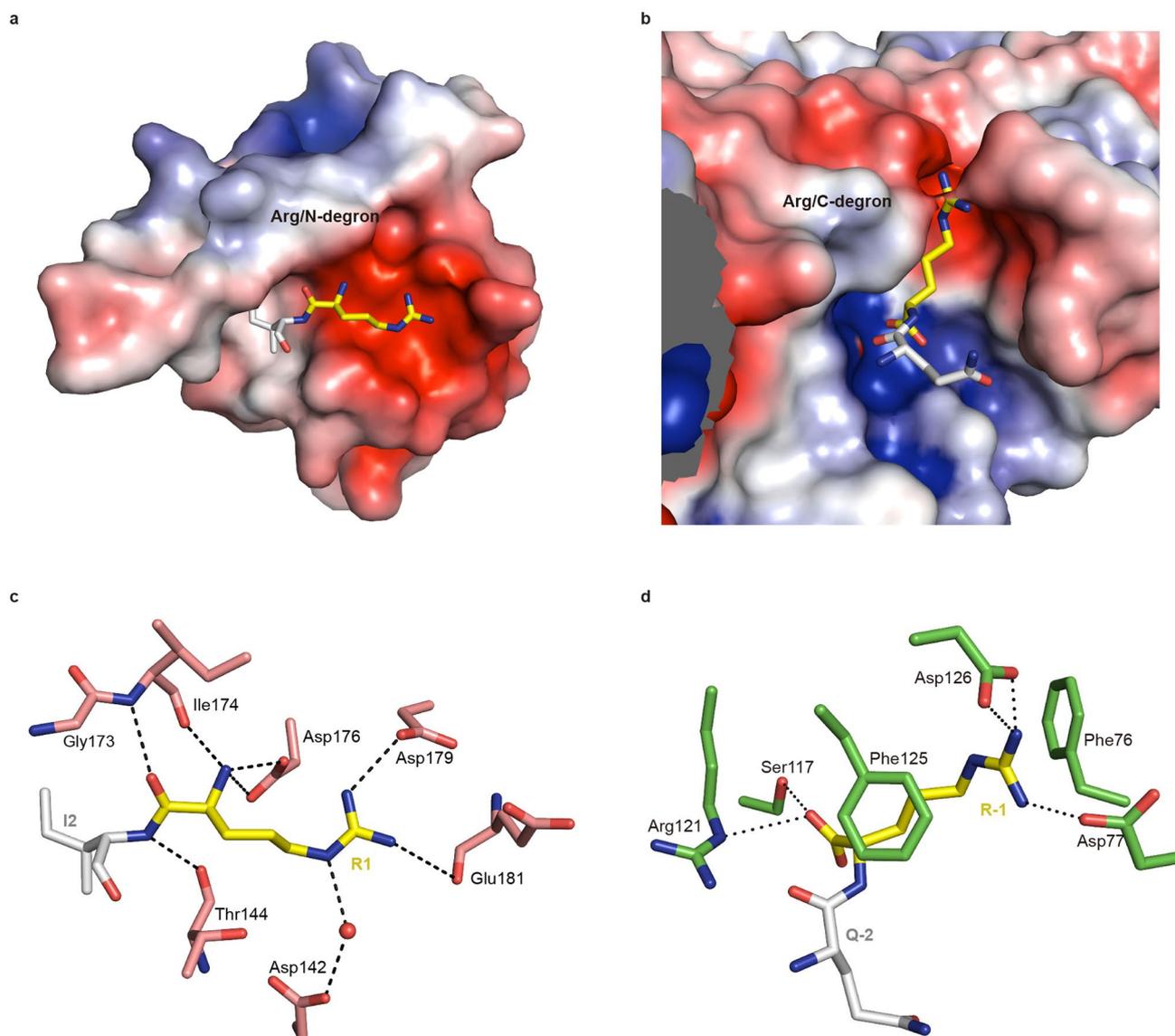
Extended Data Fig. 3 | Structural comparison of Arg/C-degron and Arg/N-degron recognitions. a, The electrostatic potential surface of the UBR domain (PDB: 3NIH) bound to an Arg/N-degron plotted at ± 5 kT/e (red, negative; blue, positive). **b**, The electrostatic potential surface of FEM1C bound to an Arg/C-degron plotted at ± 5 kT/e. **c**, Interaction of the UBR domain with a Arg/N-degron. The N-terminal arginine is shown as yellow stick and its interacting residues in UBR are shown as salmon sticks. **d**, Interactions of FEM1C with an Arg/C-degron. The C-terminal arginine is shown as yellow stick and its interacting residues in FEM1C are shown as green sticks.

![nature research logo]

Corresponding author(s): Cheng Dong

Last updated by author(s): Oct 5, 2020

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 our Editorial Policies 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	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	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)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	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	Diffraction data were collected at APS-24-IDC/E
Data analysis	HKL3000 v720, PHASER 2.7.17, CCP4 Software Suite 7.0.078 (PARROT, BUCCANEER and RESOLVE programs), PHENIX 1.11.1, COOT 0.8.7.1, Origin software 7.0, PyMOL 1.7.0.1.

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 and 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 atomic coordinates and structure factors of FEM1C and FEM1C-peptide complex have been deposited in the Protein Data Bank (https://www.rcsb.org/) with the accession code 6XKC and 7JYA, respectively.

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

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.
Data exclusions	No data were excluded.
Replication	Each assay was performed at least twice. All attempts at replication were successful.
Randomization	N/A
Blinding	N/A

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
[ ]	[×] Antibodies
[ ]	[×] Eukaryotic cell lines
[×]	[ ] Palaeontology and archaeology
[×]	[ ] Animals and other organisms
[×]	[ ] Human research participants
[×]	[ ] Clinical data
[×]	[ ] Dual use research of concern

Methods

n/a	Involved in the study
[×]	[ ] ChIP-seq
[ ]	[×] Flow cytometry
[×]	[ ] MRI-based neuroimaging

Antibodies

Antibodies used	Monoclonal ANTI-FLAG® M2 antibody produced in mouse, clone M2, Sigma-Aldrich, catalog number F3165; Monoclonal Anti-β-Actin antibody, catalog number A1978, lot number 043M4840V, Sigma-Aldrich.
Validation	Anti Flag M2 antibody is used for the detection of Flag fusion proteins. This monoclonal antibody is produced in mouse and recognizes the FLAG sequence at the N-terminus, Met N-terminus, and C-terminus, and also recognizes FLAG at an internal site. In our experiment, Flag tag is fused to FEM1C at N-terminus. The western blot data shows this antibody recognized flag-fused WT and mutants FEM1C but not control sample. Anti-β-Actin antibody recognizes an epitope located on the N-terminal end of the β-isoform of actin. Monoclonal anti-beta-actin antibody can be used for microarray, indirect immunofluorescence, and immunohistochemical analyses. Furthermore, the product has been used for immunocytochemistry at 10-40 μg/mL using human foreskin fibroblasts. The antibody has also been used for western blot at 0.5-1 μg/mL using cell extract of human foreskin fibroblasts or chicken fibroblasts. Monoclonal mouse anti-actin antibody was used as a loading control for western blot analysis of FEM1C expression in our manuscript.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HEK293T cells from ATCC
Authentication	HEK293T was authenticated by STR Fingerprinting Service at MD Anderson Characterized Cell Line Core Facility.
Mycoplasma contamination	HEK293T was tested negative for mycoplasma contamination.

Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used in this study.

Flow Cytometry

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	Single cell suspension was generated from HEK-293T trypsinization.
Instrument	Beckman Coulter CytoFLEX S
Software	CytExpert 2.3 software (Beckman Coulter, Brea, CA, USA)
Cell population abundance	Purity of sorted populations was more than 98% in all samples and was assessed by flow cytometry.
Gating strategy	Gating for HEK-293T cells: 1. Gate on fsc-h vs. ssc-h was set to include all cell populations, but excluding debris. 2. gate on ssc-a vs. ssc-h was set to exclude doublets. 3. gate on fsc-h vs. DAPI was set to exclude dead cells (DAPI+). 4. The selected cells was calculated for GFP/DsRed ratio.

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