AlphaFold2-based fusion design deciphers crucial role of the E3 UFL1 N-terminal helix in E2 UFC1 binding and ufmylation

Sayanika Banerjee[1*], Julia K Varga[2*], Manoj Kumar[1], Guy Zoltsman[3], Michail N Isupov[4], Rina Rosenzweig[3], Ora Schueler-Furman[2#], Reuven Wiener[1#]

[1] Department of Biochemistry and Molecular Biology, and [2] Department of Microbiology and Molecular Genetics, The Institute for Medical Research Israel-Canada, Hebrew University-Hadassah Medical School, Jerusalem 91120, Israel
[3] Department of Chemical and Structural Biology, Weizmann Institute of Sciences, Rehovot, Israel
[4] The Henry Wellcome Building for Biocatalysis, Biosciences, University of Exeter, Stocker Road, Exeter, EX4 4QD, United Kingdom

[*] These authors contributed equally to this work
[#] To whom correspondence should be addressed: ora.furman-schueler@mail.huji.ac.il, reuvenw@ekmd.huji.ac.il

Abstract

While protein modification by UFM1 (ufmylation) is highly appreciated as an important post-translational modification, little is known about the mechanisms of the enzymes responsible for this modification and in particular on the UFM1 E3 ligase, UFL1, that for its functionality has to form a complex with another protein DDRGK1 (UFBP1). Here we used AlphaFold2 to generate active, easily expressed, fusion proteins encompassing DDRGK1-UFL1. We then solved the crystal structure of this fusion, explaining the dependency of UFL1 on DDRGK1 to form a stable structure. In addition, we deciphered how UFL1, via its N-terminal helix, binds the E2, UFC1, and in turn, allows ufmylation. This mode of binding suggests a competition between E1 and E3 on E2 binding that is required for the proper transfer of UFM1 in the conjugation machinery.

Introduction

Protein modifications by UFM1 (ufmylation) play a role in many cellular processes, including DNA damage repair, the anti-viral and the unfolded protein responses[1]. A three-enzyme cascade involving the E1-UBA5, the E2-UFC1 and the E3-UFL1 is responsible for the attachment of UFM1 to target proteins. Initially, UBA5 activates UFM1 in an ATP-dependent process. Then, UFM1 is transferred from UBA5 to the active site cysteine of UFC1, forming a thioester bond. Finally, with the help of UFL1, UFM1 is transferred from UFC1 to the target protein[2,3].

Surprisingly, UFL1 lacks structural elements that are common to other E3 ligase enzymes, namely a RING domain, a HECT-type catalytic domain or an RBR structure[4–6]. In addition, while some atypical E3 enzymes possess a motif required for interaction with their ubiquitin-like protein[7,8], whether UFL1 has a UFM1-interacting motif is uncertain. Therefore, it remains to be determined whether UFL1 functions in a novel mechanism that does not exist in other E3 ligases. It was shown previously that ufmylation by UFL1 of the nuclear receptor coactivator, ASC1, requires DDRGK1 (also known as UFBP1)[9]. Moreover, a recent model of the interaction between UFL1 and DDRGK1 generated by AlphaFold2 has revealed structural complementation between the two proteins[10]. Besides binding to DDRGK1, UFL1 interacts with LZAP (also known as the adapter protein CDK5RAP3), forming a ternary complex[11]. The latter has been suggested to possess a motif allowing UFM1 binding[12]. Currently, structural data on this UFL1-DDRGK1-LZAP complex are still missing, and how this complex binds UFC1 to facilitate UFM1 conjugation is unclear.

Deep learning of modeling of protein structures is revolutionizing the field of structural biology, spearheaded by AlphaFold2, developed by DeepMind[13–15]. As models are either available, or can be generated within a short time on platforms such as ColabFold[16], they will accelerate studies that previously depended on the expression, purification, and crystallization of one or more proteins, a process that could take years, if successful at all. Structural models provide guidelines for the generation of stable proteins, by identifying disordered regions that hamper protein expression and could be truncated for improved expression. Beyond the study of protein monomers, the structures of many protein complexes can now be modeled[17–19], including interactions mediated by short motifs[20,21]. Besides the assistance of deep learning for the determination of models of the structured

regions of a protein, it is now possible to also study the interaction of regions that adopt a stable structure only upon interaction, as for example the interaction of motifs located within disordered regions with their partners, as well as complementation of a full domain by two proteins.

With these tools in hand, we have set out to study the ufmylation system and reveal yet unsolved challenges in our understanding of this complex regulatory pathway. We report here on two major advances: (1) The establishment of functional fusion proteins, using designed truncations for the UFL1-DDRGK1 complex, as well as an extended UFL1-DDRGK1-LZAP complex. These fusion constructs allow to significantly simplify the study of UFL1 activity, and importantly, have enabled us to solve, for the first time, a crystal structure of UFL1 bound to DDRGK1; (2) The definition of the critical role of the UFL1 N-terminal helix in UFC1 binding and ufmylation. Our model suggests that E2 UFC1 uses the same site to bind both E1 UBA5 as well as E3 UFL1, which we confirm by NMR studies. Of note, this helix-mediated interaction was revealed in a model generated without any prior information, highlighting the contribution of AlphaFold2 to the revelation of new interaction details and regulation.

Results

AlphaFold2-assisted engineering of an active ufm1 E3-ligase

UFL1 has been suggested to be only active in the presence of DDRGK1[22]. In line with a parallel recent study[10], our starting point was a model of the UFL1-DDRGK1 interaction (see **Figure 1a** and **Supplementary Figure 1)** that we generated using AlphaFold2 (see Methods). The model shows nicely the crucial contribution of DDRGK1 to complement the first winged helix domain repeat of UFL1 (residues 27-58) and explains why neither DDRGK1 nor UFL1 are folded when expressed alone. This model, as well as additional information about the importance of different regions[3], assisted us in the design of a fusion construct encompassing DDRGK1-UFL1 (**Figure 1b**), in which we removed the N-terminal region of DDRGK1 and the C-terminal region of UFL1 (*i.e.,* DDRGK1:207-314 - UFL1:1-200). Furthermore, we noted the predicted low confidence of the N-terminal helix of UFL1 (average AlphaFold2 pLDDT <70) (**Figure 1c**), and therefore generated a second construct in which we also truncated this N-terminal UFL1 helix, DDRGK1-UFL1ΔN (DDRGK1:207-306 - UFL1:27-200) .

To test the AlphaFold2-based design of the above fusion proteins comprising UFL1-DDRGK1, we first verified that they do not form soluble aggregates. We purified the fusion proteins (**Supplementary Figure 2**) and tested their elution profile using gel filtration (**Figure 1d**). Indeed, the fusion proteins did not elute as soluble aggregates. Overall, our results imply that, similar to the co-expression of UFL1 with DDRGK1 that allows purification of a soluble UFL1-DDRGK1 complex[10], the fusion protein is also soluble.

To date, structural data on DDRGK1-UFL1 complexes are based on AlphaFold models[10]. This motivated us to exploit our fusion proteins for determination of their crystal structure. Indeed, we successfully solved the crystal structure of DDRGK1-UFL1ΔN to 3.1 Å resolution (**Supplementary Table 1**). The structure reveals four repeats of winged helix (WH) domains, as expected: the first is contributed by DDRGK1, the second is formed partially by DDRGK1 and partially by UFL1, while the last two are from UFL1 (**Figure 1e**). This structure is very similar to our corresponding AlphaFold2 model (backbone RMSD = 1.4 Å). Most of the structural differences are concentrated in the first WH domain (aa 1-65 in the fusion; backbone RMSD = 2.4 Å). Interestingly, less structural differences are observed in the combined WH domain (aa 66-130 in the fusion; backbone RMSD = 2.1 Å), although both of its parts are connected in the fusion protein (**Figure 1e**). In the crystal structure the last 18 amino acids are flexible and are not detected in the electron density. In the AlphaFold2 model these residues form an alpha helix that belongs to the next WH domain that is missing in our structure, suggesting that this helix is not stable on its own. Overall, our crystal structure suggests that the fusion protein maintains the overall architecture of the UFL1-DDRGK1 complex as observed in the AlphaFold2 model.

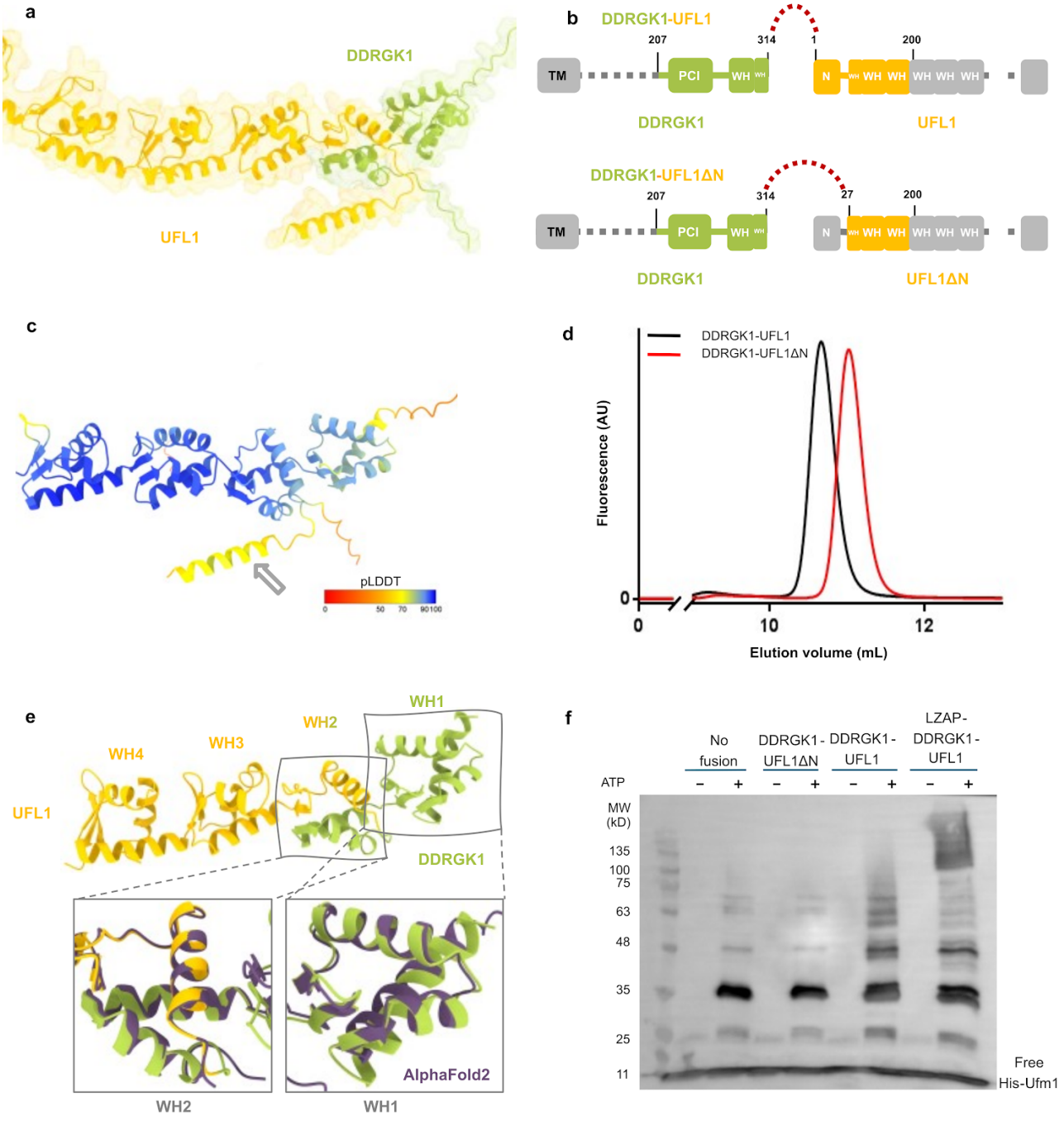

Figure 1: AlphaFold2-assisted generation of an active fusion protein for ufmylation. **a** AlphaFold2 structural model of the DDRGK1-UFL1 complex (similar to Peters *et al.*[10]), DDRGK1 (colored in green); UFL1 (colored in yellow). **b** Details of the proteins and the designed fusion constructs. DDRGK1 was connected to UFL1, removing the N-terminal region of DDRGK1 and the C-terminal region of UFL1. In a second construct we also removed the N-terminal helix of UFL1, due to its flexibility suggested by AlphaFold2 (see Text). Regions removed from the parent proteins are shown in gray (WH: winged helix domains, PCI: proteasome-COP9-initiation factor 3 domain). **c** Model of the DDRGK1-UFL1 complex, colored according to pLDDT, highlighting the low confidence in the structure of the N-terminal helix (gray arrow). **d** Gel filtration elution profiles of fusion proteins. **e**. Crystal structure of the DDRGK1-UFL1ΔN fusion. The four winged helix (WH) domains are indicated and numbered. The blow ups are superimposed onto the AlphaFold2 model of the indicated WH domains (purple). **f** The fusion protein is active. Western blot shows changes in ufmylation only for fusion constructs that include the N-terminal helix of UFL1 (see **Supplementary Figure 3** for protein controls).

With the above fusion proteins in hand, we tested their functionality as E3 ligases. To that end we incubated pure UBA5, UFC1 with and without fusion proteins and analyzed the ufmylation pattern (**Figure 1f, Supplementary Figure 3**). Indeed, addition of our full fusion construct (DDRGK1-UFL1) resulted in changes in the ufmylation pattern. To our surprise, however, the fusion protein lacking the UFL1 N-terminus (DDRGK1-UFL1ΔN) did not show any such changes, suggesting that the UFL1 N-terminus is essential for E3 ligase activity.

A structural model of the UFC1-UFL1 interaction reveals a critical role of the helix in the N-terminal tail of UFL1

Motivated by the contribution of the AlphaFold2 model to the successful design of a fusion protein with E3 ligase activity, we decided to use AlphaFold2 to also help uncover the underlying details of the critical role of the UFL1 N-terminal helix to its ligase activity. To this end, we modeled the binding of UFC1 to the UFL1-DDRGK1 complex (**Figure 2a,b**). This model explains the dominant contribution of the N-terminal helix in this interaction, and identifies the residues in UFC1 that are crucial for its binding (Y36, I40, R55, **Supplementary Table 2**). The importance of the interface hotspot residues in the UFL1 helix is emphasized by their high degree of evolutionary conservation (W5 , I8, L11 and F15; **Figure 2c)**. Reassuringly, as mentioned above, this N-terminal region was modeled as a helix by AlphaFold2, albeit with low confidence in the apo structure (**Figure 2d**, see also **Figure 1c**) which indicated that it might be disordered and only fold into a helix upon binding. Indeed, in our binding experiments using strep-tag UFC1, only fusion proteins possessing the UFL1 N-terminal helix showed binding to UFC1 (**Figure 2e**), and accordingly, we detected E3 activity only in fusion proteins that contain the N-terminus (as shown above in **Figure 1f**).

To support our model of the interaction of UFL1 N-terminus with UFC1, we used NMR spectroscopy to define the changes in UFC1 chemical shifts upon binding to DDRGK1-UFL1 or DDRGK1-UFL1ΔN. To that end, we exploited the reported assigned (^{1}H, ^{15}N)-HSQC NMR spectra for UFC1[23]. As expected from our activity and pull-down assays **(Figures 1f and 2e**), the addition of DDRGK1-UFL1, but not DDRGK1-UFL1ΔN, to ^{15}N-labeled UFC1 caused strong attenuations (**Figure 3a,b**). In line with the AlphaFold2 model, these attenuations include residues from UFC1 **α**-helix I (amino acids 26-48) and from **β**-strand I (amino acids 54-48), that directly interact with the UFL1 N-terminus (**Figure 3c**). Interestingly, besides the above residues, the N-terminal half of UFC1 **α**-helix

II (amino acids 135-145), which is located on the other side and not directly involved in UFL1 N-terminus binding based on our AlphaFold2 model, showed chemical perturbations. This raises the possibility that these residues are allosterically regulated by binding of the UFL1 N-terminus to UFC1.

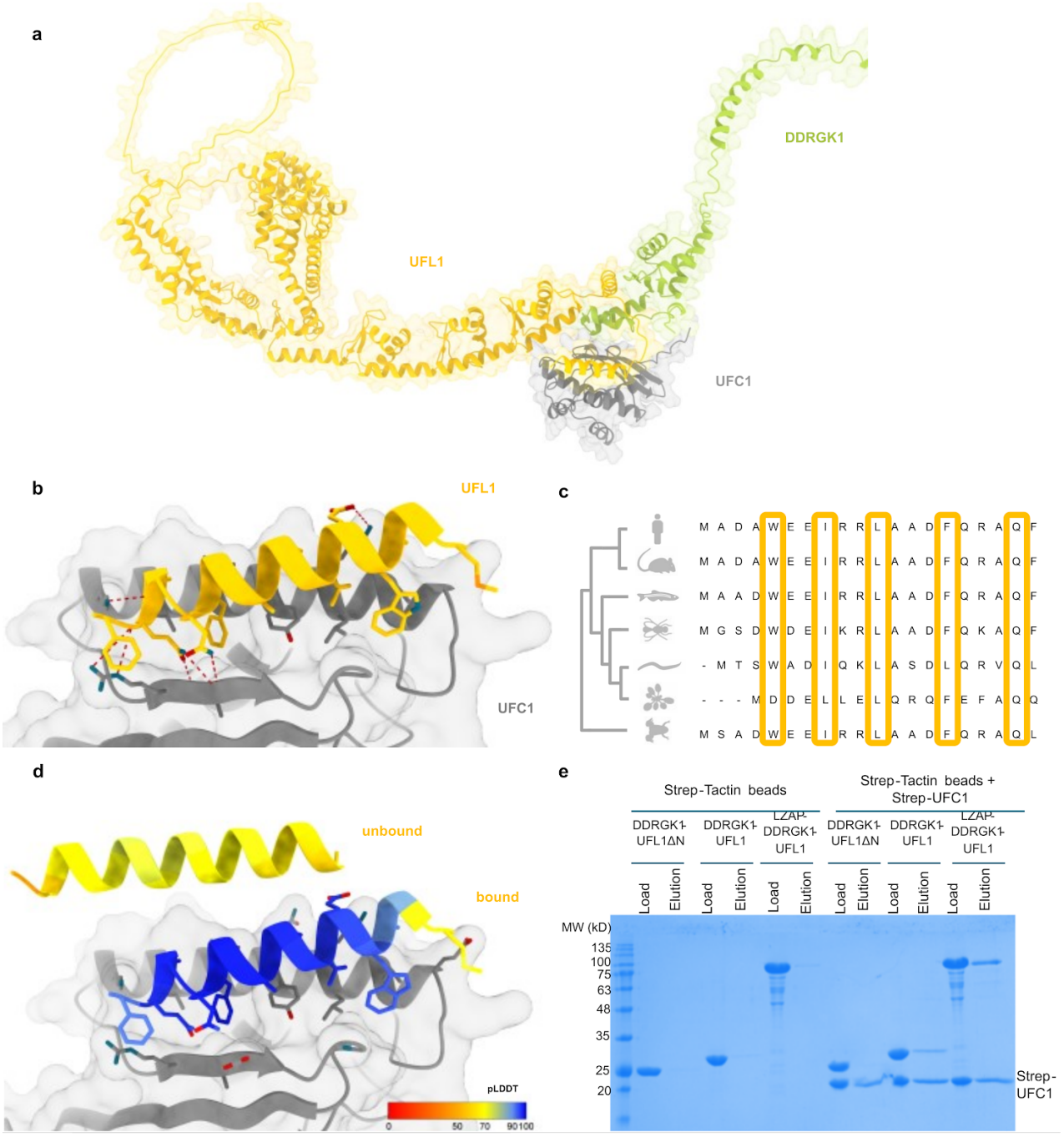

Figure 2: The N-terminal helix of UFL1 is crucial for binding of UFC1 and activity. a Overall view of the UCF1-DDRGK1-UFL1 ternary complex. **b** Details of the interaction: UFL1 N-terminal helix bound to UFC1. **c** Analysis of conservation of UFL1 N-terminal helix shows evolutionary conservation of the residues predicted to be involved in binding. **d** The N-terminal helix (colored according to pLDDT) is modeled with high confidence when bound to UFC1 (in gray), in contrast to the low confidence for this helix in the unbound structure (see also **Figure 1c**). **e** Experimental validation: Coomassie stain gel shows that UFC1 binds to DDRGK1-UFL1 but not DDRGK1-UFL1ΔN, demonstrating that this interaction depends on the presence of the N-terminal

helix of UFL1. For panels (**b and d**) AMBER relaxation was performed after AlphaFold2 structure prediction, to optimize side-chain orientations.

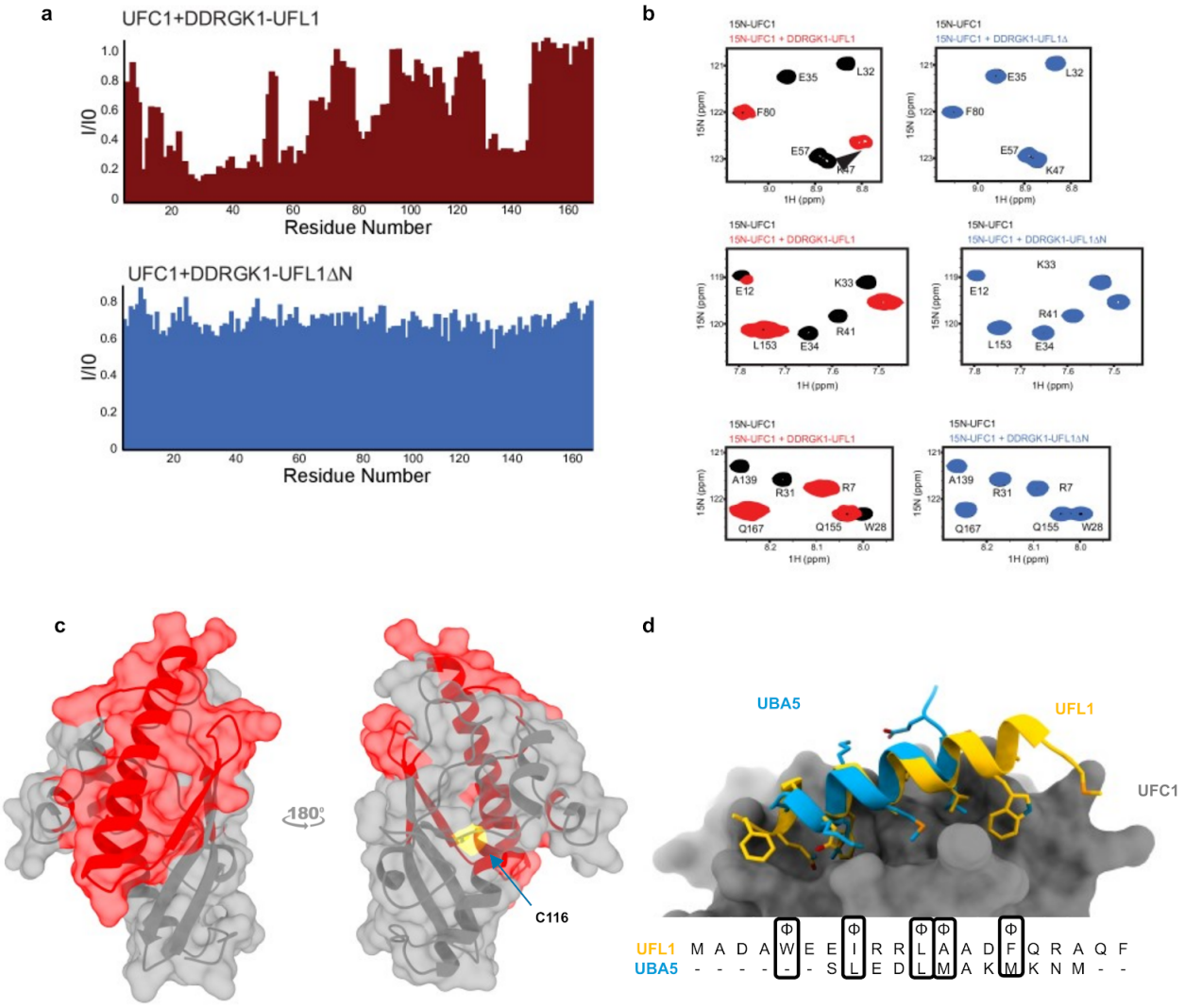

Figure 3: E3 UFL1 and E1 UBA5 bind to the same site on E2 UFC1, as demonstrated by NMR experiments and structural models. a Intensity changes to UFC1 residues upon addition of 1.5-fold excess (300 µM) of DDRGK1-UFL1 (red) or 2-fold excess (400 µM) DDRGK1-UFL1ΔN (blue). Removal of the N-terminal UFL1 regions significantly impairs UFC1- DDRGK1-UFL1 complex formation. **b** Selected regions of the $^{1}H–^{15}N$ HSQC spectrum: 0.2 mM UFC1 alone (black) and in the presence of two-fold excess of DDRGK1-UFL1 (red) or DDRGK1-UFL1ΔN (blue). **c** structure of UFC1 (PDB ID: 7NW1[24]) with residues displaying significant intensity changes (I/IO<0.4) upon addition of DDRGK1-UFL1 are colored in red. **d** Superposition of the model of the UFL1 N-terminal helix-UFC1 complex onto the solved structure of the UBA5 N-terminal helix-UFC1 complex (PDB ID: 7NW1) suggests that both bind to the same binding site (as suggested also for other types of ubiquitination[25]).

According to our model and NMR data, UFC1 binds the N-terminal helix of UFL1 with high confidence, using the very same binding pocket that UFC1 uses also to bind the

C-terminal tail of E1 activating enzyme UBA5 (as shown by a crystal structure solved previously by us[24], **Figure 3d**). To estimate the relative binding strength between the helices of UFL1 and UBA5, we applied an approach suggested by *Chang et al.* in which AlphaFold2 is run using a sequence that contains both peptides, assuming that the stronger binding peptide will bind to the receptor binding site[26]. As UFL1 was predicted to bind as a much longer helix (20 residues *vs.* 12 residues of UBA5), we anticipated that it would be also predicted to bind stronger, due to the larger hydrophobic interaction surface. Indeed, the UFL1 helix invariably outcompeted the UBA5 helix in the binding site and was modeled with significantly higher confidence (**Supplementary Figure 4**).

While binding of UFL1 to DDRGK1 is essential for the correct folding of these proteins[10], this complex is also known to bind LZAP[27]. To investigate the contribution of the latter to ufmylation, we generated a model of the UFL1-DDFRK1-LZAP complex bound to UFC1 (**Supplementary Figure 5**). This model suggests that binding of LZAP would not affect binding of UFL1 and ufmylation. Based on this model, we generated a fusion protein, along similar lines as the fusion proteins described above (**Figure 1b**). We connected DDRGK1 to UFL1 (1-200) the same way. In order to connect the C-terminus of LZAP to DDRGK1, we extended DDRGK1 to residue 87 (**Supplementary Figure 5**). We also removed LZAP residues 251-341 that according to the AlphaFold2 model form an unstructured loop. This fusion protein covering LZAP, DDRGK1-UFL1, showed binding to UFC1 **(Figure 2e**) as well as E3 activity (**Figure 1f**).

Discussion and Conclusion

Beyond providing models for many of today's known proteins, AlphaFold2 models also include information about the reliability of these models, allowing us to identify stable parts, as well as regions that will need to interact with partners to be stabilized. In particular, despite the availability of monomer structures, biological activity is often achieved by multiprotein assemblies, which still await to be modeled. In these cases, the ternary model generated by AlphaFold2 can be invaluable for the design of a fusion protein that encompasses all the individual components of such a complex.

Mechanistic studies on UFL1 as the E3 ligase of UFM1 have remained elusive until not long ago. While working with the intact UFL1 complex would be ideal, this is not always feasible for detailed structural research. The design of E3 constructs that are suitable for such study is therefore of high interest. However, making these constructs turns to be

highly challenging once the E3 of interest needs other partners for its folding and function as in the case of UFL1. Using AlphaFold2 we have generated a structural model for the E3 UFL1 together with its partners DDRGK1 and LZAP, and used it as the basis for our fusion protein designs. These fusion proteins are highly expressed in *E. coli* and can easily be purified, thereby bypassing the intrinsic challenges embedded in working with the UFL1 complex. But more importantly, they retain the E3 activity, and are thus suitable for a mechanistic study focusing on E2 - E3 interaction and function.

We should, however, not ignore that such constructs might miss some of the details of these interactions. In addition, while removal of unstructured regions, or regions previously reported not to be important for an interaction, simplifies experiments, subtle regulation in these regions will be missed. For example, we removed the flexible loop in LZAP (**Supplementary Figure 5**), which may affect its binding to the DDRGK1-UFL1 E3 ligase complex, and its possible regulation. This could explain why in our hands, LZAP addition did not significantly affect ufmylation (**Figure 1f**), whereas another study showed a negative influence of LZAP on ufmylation[10]. Targeted inclusion of these regions in follow up studies will reveal the importance of skipped parts.

As a matter of fact, it is the removal of such elements of poor structural confidence that led us to the most important finding of the present study - the identification of the crucial role of the N-terminal helix of UFL1 for proper ufmylation. Once located, we made use of another important feature of AlphaFold2, namely its ability to model short motif-mediated interactions at high confidence, as we have previously demonstrated[20]. The resulting model provides the atomic details of the UFL1 helix - UFC1 interaction (**Figure 2b**) and allows to identify the interface hotspot residues that are critical for this interaction (**Figure 3** and **Supplementary Table 2**). As we proceed to the study of additional components involved in the regulation of ufmylation, models generated by AlphaFold2 will continue to guide us and reveal additional details of regulation.

Methods

AlphaFold2 predictions

In general, structural models of individual proteins and complexes were generated using colabfold[16]. Due to the size of the UFL1-UFC1-LZAP-DDRGK1 complex, all AlphaFold2 predictions on this complex were performed locally, using the LocalColabFold installation (downloaded on 17/07/2022 from https://github.com/YoshitakaMo/localcolabfold). Unless indicated otherwise, the predictions were run with all 5 models and default seed, default multiple sequence alignment generation using the MMSeqs2 server and with 3 recycles, without linkers between the

monomers. The "computational competition assay" was run as suggested by Chang *et al.*[26], providing both competing peptides in a single prediction run, provided before and after the receptor UFC1 sequence. All structure visualizations were created with ChimeraX v1.3[28].

Calculation of sequence conservation

Conservation of UFL1 was calculated with the ConSurf server[29], using default parameters. Alignment of the human and model animal sequences of UFL1 was performed using ClustalOmega[30] on the UniProt server[31].

Computational alanine scanning

To estimate the contribution of different residues to the binding of UFL1 N-terminal helix to UFC1, we applied alanine scanning using the Robetta server[32]. Residues with predicted effect of $\Delta\Delta G_{binding}$>1.0kcal/mol were retained as hotspot residues.

Cloning

The fusion constructs DDRGK1$_{207-314}$-UFL1$_{1-200}$ (corresponding to DDRGK1-UFL1 in the main text) and DDRGK1$_{207-305}$-UFL1$_{27-200}$ (corresponding to DDRGK1-UFL1ΔN in the main text) (**Figure 1b**) were generated using Gibson assembly (Gibson assembly master mix, New England Biolabs) according to the manufacturer's protocol. To generate the fragments by PCR, we used the Primers detailed in **Table Ia**, to generate fragments detailed in **Table Ib.** DDRGK1$_{207-314}$-UFL1$_{1-200}$ was cloned in pET15b by Gibson assembly of fragment I, II and linear pET15b. DDRGK1$_{207-305}$-UFL1$_{27-200}$ was cloned in pET15b by Gibson assembly of fragment III, IV and linear pET15b. The fusion construct LZAP(Δ251-341)_DDRGK1$_{87-314}$-UFL1$_{1-200}$ (corresponding to LZAP-DDRGK1-UFL1 in the main text) was synthesized and cloned in pET15b by Gibson assembly. All of the constructs were verified by DNA sequencing.

Table I: Primers and Fragments used/generated in this study

a. List of Primers

Primer #	**Sequence (5'-3')**
475	GGCTTTGTTAGCAGCCGGATCCTCGAGTCAAGGCCGGGTAATAGCACTG
1017	CTAATCTCTTCCCAGGCGTCCGCCATGGCTGGGGCTTGGGCAGG
1018	GAGTCCCCTGCCCAAGCCCCAGCCATGGCGGACGCCTGGGAAG
1044	TATGGAGAATCTTTACTTTCAGGGGATGACTGAGGAACAGTCCCAG
1045	GCCCCAGGCGATGAGGGAGTTGC
1046	AACTCCCTCATCGCCTGGGGCTTGTCCGAGCGGAACTGCATTGAG

b. Details of Fragments

Fragment	**Name**	**Forward Primer**	**Reverse Primer**
Fragment I	DDRGK1 207-314	1044	1017

Fragment II	UFL1 1-200	1018	475
Fragment III	DDRGK1 207-305	1044	1045
Fragment IV	UFL1 27-200	1046	475

Protein expression and purification

UBA5, UFC1, UFM1 were expressed and purified as previously described[33]. All the fusion constructs (DDRGK1-UFL1, DDRGK1-UFL1ΔN, LZAP-DDRGK1-UFL1) were expressed in *E. coli* T7 express (New England Biolabs). The transformed cells were grown in 2xYT and induced at 16 °C overnight with 0.3 mM isopropyl-β-thio-galactoside (IPTG). The induced cells were harvested by centrifugation at 8000×*g* for 15 min. Pellets were resuspended in lysis buffer (50 mM NaH2PO4 pH 8.0, 500 mM NaCl, 10 mM imidazole, and 5mM β-mercaptoethanol), supplemented with 1 mM phenyl-methyl sulphonyl fluoride (PMSF) and DNase. Cells were disrupted using a microfluidizer (Microfluidics). Lysate was cleared by centrifugation at 20000 x g for 45 min and was subjected to 5 ml His-Trap columns (GE Healthcare). The protein was eluted with a linear imidazole gradient of 15–300 mM in 30 column volumes. Fractions containing the purified protein were pooled and dialyzed overnight at 4°C against dialysis buffer (25 mM NaH2PO4 pH 8.0, 300 mM NaCl, and 5mM β-mercaptoethanol) in the presence of TEV protease. Cleaved protein was then subjected to a second round of His-Trap column and flow-through containing the cleaved protein was collected. Further purification was done using 16/60 Superdex 75 pg for DDRGK1-UFL1 and DDRGK1-UFL1ΔN or 16/60 Superdex 200 pg size exclusion chromatography for LZAP-DDRGK1-UFL1, equilibrated in buffer containing Tris-Cl pH 7.5 (20 mM), NaCl (200 mM), and DTT (2 mM). The purified proteins were concentrated and flash-frozen in liquid N2 and stored at −80 °C.

In vitro ufmylation assay

UBA5 (0.5 µM), His-UFM1 (10 µM), UFC1 (5 µM) and fusion fragments (5 µM each) were mixed together in a buffer containing Hepes (50 mM pH 8.0), NaCl (100 mM) and $MgCl_2$ (10 mM). Reactions were initiated by the addition of ATP (5 mM) and were incubated at 30°C for 1 hour. The negative control sample was incubated without ATP. After incubation the reactions were stopped by adding 5X SDS-sample buffer containing β-mercaptoethanol. The samples along with the control were then loaded on 10% SDS-PAGE followed by immunoblot with anti-6x His antibody (Abcam).

In vitro pull down assay

Recombinant purified strep-UFC1 (5 µM) and fusion fragments (5 µM each) were mixed in PBS in total volume of 50 µL for 1 h at RT and subsequently precipitated with Strep-Tactin beads (Iba Lifesciences). The mixtures were washed twice with PBS. The bound proteins were eluted using 7.5 mM desthiobiotin in 50 mM Hepes, pH 8.0 and 300 mM NaCl buffer. Then the samples were analyzed by 15% SDS-PAGE followed by Coomassie Brilliant Blue staining.

Fluorescence-detection size-exclusion chromatography (FSEC)

For the FSEC assay 40 µl of the fragments (10 µM) were injected at flow rate of 0.4 ml min^{-1} to a Superdex 75 Increase 10/300 GL column (GE Healthcare) equilibrated with buffer containing 20

mM Tris-Cl pH 7.5, 200 mM NaCl with 1 mM DTT. Fluorescence was detected using the RF-20A fluorescence detector for HPLC (Shimadzu, Japan) (for Trp, excitation: 285 nm, emission: 330 nm).

Crystallography

Crystals of DDRGK1-UFL1ΔN were grown at 20ºC using the hanging drop vapor diffusion method. Protein was concentrated to 100 mg/ml and crystalized in a solution containing 0.7 M Ammonium tartrate dibasic and 0.1 M Sodium acetate trihydrate, at pH 4.6. The crystals were cryoprotected using a reservoir solution containing 25% glycerol and flash frozen in liquid nitrogen.

Diffraction data for the DDRGK1-UFL1ΔN crystals were collected on beamline ESRF ID30A-3 at 100ºK. Data were processed using Dials[34] and scaled using Aimless[35]. The putative space group was determined to be hexagonal $P6_2$ or its enantiomorph. The structure was solved by molecular replacement (MR) with MOLREP[36] using the AlphaFold2 model of the UFL1-DDRGK1 heterodimer (**Figure 1a**). The translation function confirmed the space group to be $P6_4$. The asymmetric unit contains two chains of the chimeric protein. The MR model was refined in REFMAC5[37] and BUSTER. The electron density was subject to density modification with NCS averaging using Parrot[38,39]. The model was further refined using REFMAC5 with input density modification phases[40]. The model was rebuilt using COOT[39] and ISOLDE[41] implemented in ChimeraX[28]. Details of the quality of the refined model are presented in **Supplementary Table 1**.

NMR spectroscopy

All NMR experiments were carried out at 25°C on a 23.5T (1000 MHz) Bruker spectrometer equipped with triple resonance (x,y,z) gradient cryoprobe. The experiments were processed with NMRPipe[58] and analyzed with NMRFAM-SPARKY[59]. The interaction of UFC1 with DDRGK1-UFL1 fragments was monitored by 2D $^{1}H–^{15}N$ HSQC experiments with the assignments for UFC1 transferred from the BMRB (entry 6546[23]). DDRGK1-UFL1 (100-400 μM) or DDRGK1-UFL1ΔN (400 μM) were titrated into 200 μM of ^{15}N-labeled UFC1 in 20 mM TRIS pH 7.6.

Acknowledgment

This work was supported, in whole or in part, by the Israel Science Foundation, founded by the Israel Academy of Science and Humanities (grant number 491/2021 to R.W. and grant number 301/2021 to O.S.-F.), and by the Israel Cancer Research Fund (award ID 21-113-PG to R.W.). J.K.V. is supported by a Marie Sklodowska-Curie European Training Network Grant #860517 (Ubimotif).

References

1. Banerjee, S., Kumar, M. & Wiener, R. Decrypting UFMylation: How Proteins Are Modified with UFM1. *Biomolecules* **10**, (2020).
2. Komatsu, M. *et al.* A novel protein-conjugating system for Ufm1, a ubiquitin-fold modifier. *EMBO J.* **23**, 1977–1986 (2004).
3. Tatsumi, K. *et al.* A novel type of E3 ligase for the Ufm1 conjugation system. *J. Biol. Chem.* **285**, 5417–5427 (2010).
4. Berndsen, C. E. & Wolberger, C. New insights into ubiquitin E3 ligase mechanism. *Nat. Struct. Mol. Biol.* **21**, 301–307 (2014).
5. Wenzel, D. M., Lissounov, A., Brzovic, P. S. & Klevit, R. E. UBCH7 reactivity profile reveals parkin and HHARI to be RING/HECT hybrids. *Nature* **474**, 105–108 (2011).
6. Zheng, N. & Shabek, N. Ubiquitin ligases: structure, function, and regulation. *Annu. Rev. Biochem.* **86**, 129–157 (2017).
7. Cappadocia, L., Pichler, A. & Lima, C. D. Structural basis for catalytic activation by the human ZNF451 SUMO E3 ligase. *Nat. Struct. Mol. Biol.* **22**, 968–975 (2015).
8. Reverter, D. & Lima, C. D. Insights into E3 ligase activity revealed by a SUMO-RanGAP1-Ubc9-Nup358 complex. *Nature* **435**, 687–692 (2005).
9. Yoo, H. M. *et al.* Modification of ASC1 by UFM1 is crucial for ERα transactivation and breast cancer development. *Mol. Cell* **56**, 261–274 (2014).
10. Peter, J. J. *et al.* A non-canonical scaffold-type E3 ligase complex mediates protein UFMylation. *EMBO J.* e111015 (2022) doi:10.15252/embj.2022111015.
11. Wu, J., Lei, G., Mei, M., Tang, Y. & Li, H. A novel C53/LZAP-interacting protein regulates stability of C53/LZAP and DDRGK domain-containing Protein 1 (DDRGK1) and modulates NF-kappaB signaling. *J. Biol. Chem.* **285**, 15126–15136 (2010).
12. Picchianti, L. *et al.* Shuffled ATG8 interacting motifs form an ancestral bridge between UFMylation and C53-mediated autophagy. *BioRxiv* (2022) doi:10.1101/2022.04.26.489478.
13. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
14. Tunyasuvunakool, K. *et al.* Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590–596 (2021).
15. Varadi, M. *et al.* AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444 (2022).
16. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
17. Evans, R. *et al.* Protein complex prediction with AlphaFold-Multimer. *BioRxiv* (2021) doi:10.1101/2021.10.04.463034.
18. Humphreys, I. R. *et al.* Computed structures of core eukaryotic protein complexes. *Science* **374**, eabm4805 (2021).
19. Bryant, P., Pozzati, G. & Elofsson, A. Improved prediction of protein-protein interactions using AlphaFold2. *Nat. Commun.* **13**, 1265 (2022).
20. Tsaban, T. *et al.* Harnessing protein folding neural networks for peptide-protein docking. *Nat. Commun.* **13**, 176 (2022).
21. Johansson-Åkhe, I. & Wallner, B. Benchmarking Peptide-Protein Docking and Interaction Prediction with AlphaFold-Multimer. *BioRxiv* (2021) doi:10.1101/2021.11.16.468810.
22. Witting, K. F. & Mulder, M. P. C. Highly Specialized Ubiquitin-Like Modifications: Shedding Light into the UFM1 Enigma. *Biomolecules* **11**, (2021).
23. Liu, G. *et al.* GFT NMR based resonance assignment for the 21 kDa human protein UFC1. *J.*

Biomol. NMR **32**, 261–261 (2005).
24. Kumar, M. *et al.* Structural basis for UFM1 transfer from UBA5 to UFC1. *Nat. Commun.* **12**, 5708 (2021).
25. Eletr, Z. M., Huang, D. T., Duda, D. M., Schulman, B. A. & Kuhlman, B. E2 conjugating enzymes must disengage from their E1 enzymes before E3-dependent ubiquitin and ubiquitin-like transfer. *Nat. Struct. Mol. Biol.* **12**, 933–934 (2005).
26. Chang, L. & Perez, A. AlphaFold encodes the principles to identify high affinity peptide binders. *BioRxiv* (2022) doi:10.1101/2022.03.18.484931.
27. Stephani, M. *et al.* A cross-kingdom conserved ER-phagy receptor maintains endoplasmic reticulum homeostasis during stress. *eLife* **9**, (2020).
28. Pettersen, E. F. *et al.* UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
29. 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. *GENOMICS AND COMPUTATIONAL BIOLOGY* **38**, W529-33 (2010).
30. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
31. UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).
32. Kortemme, T., Kim, D. E. & Baker, D. Computational alanine scanning of protein-protein interfaces. *Sci. STKE* **2004**, pl2 (2004).
33. Oweis, W. *et al.* Trans-Binding Mechanism of Ubiquitin-like Protein Activation Revealed by a UBA5-UFM1 Complex. *Cell Rep.* **16**, 3113–3120 (2016).
34. Waterman, D. G. *et al.* Diffraction-geometry refinement in the DIALS framework. *Acta Crystallogr. D Struct. Biol.* **72**, 558–575 (2016).
35. Evans, P. R. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 282–292 (2011).
36. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 22–25 (2010).
37. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367 (2011).
38. Cowtan, K. Recent developments in classical density modification. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 470–478 (2010).
39. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
40. Pannu, N. S., Murshudov, G. N., Dodson, E. J. & Read, R. J. Incorporation of prior phase information strengthens maximum-likelihood structure refinement. *Acta Crystallogr. D Biol. Crystallogr.* **54**, 1285–1294 (1998).
41. Croll, T. I. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr. D Struct. Biol.* **74**, 519–530 (2018).
42. Kortemme, T. & Baker, D. A simple physical model for binding energy hot spots in protein-protein complexes. *Proc Natl Acad Sci USA* **99**, 14116–14121 (2002).