

Predicting the structural basis of targeted protein degradation by integrating molecular dynamics simulations with structural mass spectrometry

Received: 1 February 2022

Accepted: 20 September 2022

Published online: 06 October 2022

 Check for updates

A list of authors and their affiliations appears at the end of the paper

Targeted protein degradation (TPD) is a promising approach in drug discovery for degrading proteins implicated in diseases. A key step in this process is the formation of a ternary complex where a heterobifunctional molecule induces proximity of an E3 ligase to a protein of interest (POI), thus facilitating ubiquitin transfer to the POI. In this work, we characterize 3 steps in the TPD process. (1) We simulate the ternary complex formation of SMARCA2 bromodomain and VHL E3 ligase by combining hydrogen-deuterium exchange mass spectrometry with weighted ensemble molecular dynamics (MD). (2) We characterize the conformational heterogeneity of the ternary complex using Hamiltonian replica exchange simulations and small-angle X-ray scattering. (3) We assess the ubiquitination of the POI in the context of the full Cullin-RING Ligase, confirming experimental ubiquitinomics results. Differences in degradation efficiency can be explained by the proximity of lysine residues on the POI relative to ubiquitin.

Heterobifunctional degraders are a class of molecules that induce proximity between a target protein of interest (POI) and a E3 ubiquitin ligase, which can lead to ubiquitination of the POI and its subsequent proteasomal degradation through a complex machinery of proteins¹. Degradation molecules provide the opportunity of a novel therapeutic modality as compared with traditional small molecule inhibitors—single molecules induce catalytic turnover of the POI and potentially offer an avenue for modulation of targets traditionally labeled as undruggable by classical therapeutic strategies^{2–4}. Heterobifunctional degraders consist of two separate protein binding moieties (the warhead and the E3-ligand) joined by a linker. The warhead binds to the POI (and we note that the degrader molecules studied here all have a non-covalently binding warhead) and the E3-ligand binds to an E3 ubiquitin ligase such as Cereblon (CRBN)⁵, cIAP⁶, KEAP1⁷, von Hippel-Lindau protein (VHL)^{8,9}, or, potentially, to any of the more than 600 known E3 ubiquitin ligases³. The ternary complex induced by the E3-ligand-linker-warhead degrader molecule is critical for bridging the interactions between the POI and a ubiquitin ligase (which can be the native or

a non-native degradation partner of the POI). An important consideration when assessing putative degrader molecules is the cooperativity of the ternary complex, i.e., the difference between the binding affinity of the ternary complex and the binary components, which can influence degradation efficiency. The cooperativity is thought to result from interactions across the induced interface of the POI-ligase pair¹⁰.

The formation of the POI-degrader-ligase ternary complex is central to the targeted protein degradation (TPD) process, but how the formation of the ternary structure impacts protein degradation is still poorly understood, especially given the dynamic nature of the complex^{11,12}. X-ray crystallography of the ternary complex¹³ provides a high-resolution structure of a single conformational state, but a growing body of evidence suggests that the dynamic nature of the ternary structure may not be accurately represented by this lowest energy crystallization snapshot. For instance, a study of several heterobifunctional degraders found that different degraders displayed different degrees of efficiency, although the corresponding ternary

✉ e-mail: alexrd@msu.edu; huafeng.xu@roivant.com; woody.sherman@roivant.com; jesus.izaguirre@roivant.com

complex structures are nearly identical, thus raising questions about the static structural representations of the ternary complex and degradation efficiency. Studies targeting the degradation of Burton Tyrosine Kinase (BTK) by CRBN or cIAP found that high degradation efficiencies can also be achieved through degrader molecules that induce a non-cooperative ternary complex, demonstrating a disconnect between binding affinity and degradation efficiency¹⁴. It appears that for degraders that bind with relatively weak affinity ($\sim 1 \mu\text{M}$) to either the target or the ligase, cooperativity is crucial to optimize degradation. On the other hand, for degraders with very high binding affinity (low nM) to the target or the ligase, cooperativity is less crucial.

This and other findings^{15,16} suggest that degradation efficiency is more complex than can be understood through the thermodynamics of binding or the analysis of static structures. As such, determining the dynamic ensemble of the ternary complex may reveal mechanistic insights to facilitate the design of more effective degrader molecules^{13,17,18}. Previous work to computationally predict ternary structures has primarily consisted of protein–protein docking protocols with rigid protein structures, possibly followed by refinement of the initial structures with molecular dynamics (MD) simulations to assess the stability of the predicted models^{18–20}. However, these docking protocols fail to predict experimentally determined structures with high fidelity and they neglect the aforementioned dynamic nature of the ternary structure, highlighting the challenge associated with the generation of ternary structure models.

Recently, Eron et al., using HDX-MS, demonstrated how ternary complex structures of BRD4 do not represent the biologically relevant conformer of the ternary complex induced with CRBN. Molecular modeling revealed the dynamic nature and alternative conformations, which helped explain the dramatically increased cooperativity, ternary complex formation, and degradation of their molecule, CFT-1297, compared to the literature standard, dBET6²¹. The authors use experimental data to improve protein–protein docking predictions, but they admit that the high flexibility of degrader-induced ternary complexes impedes a complete description of the bound conformations using their approach.

The goal of our work here is to understand the structural and dynamic basis of targeted protein degradation and ultimately design molecules for synthesis. We specifically focus on three different VHL-recruiting degraders of SMARCA2, for which crystal structures exist. PROTAC 1 (PDB ID: 6HAY¹⁸) and PROTAC 2 (PDB ID: 6HAX¹⁸) have been solved previously and ACBII (PDB ID: 754E) was solved and deposited as part of this work. The cooperativities and degradation efficiencies for each of these molecules are summarized in Supplementary Table 1. We carry out MD simulations in combination with hydrogen-deuterium exchange mass-spectrometry (HDX-MS), shedding light on the dynamics of the ternary complexes beyond what is provided by static crystal structures. Specifically, we use protection data derived from HDX-MS as collective variables in weighted-ensemble MD simulations that predict ternary complex conformations, enhancing both the speed and accuracy of the computational predictions. We also show the usefulness of HDX-MS data as constraints for protein–protein docking when higher throughput and lower resolution models are sought, such as when screening many degrader molecules. Furthermore, we introduce methods that include long-timescale MD simulations augmented with small-angle X-ray scattering (SAXS) data and Markov state modeling to determine the conformational free energy landscapes of the ternary complexes, which is the foundation for quantifying the populations of different conformational states. Finally, as an example of downstream use of these models, we assemble the entire Cullin-RING ligase (CRL) to explore structural and dynamic factors that may be associated with ubiquitination. Mass spectrometry-based proteomics experiments validate the predicted ubiquitination of several lysines of SMARCA2 induced by

ACBII, supporting the use of the CRL model as a criterion for explaining degradation.

This work offers insights into the dynamic nature of the ternary structure ensemble and that of the full CRL macromolecular assembly that could explain ubiquitination and downstream protein degradation. Our results can be used to guide the design of novel degrader molecules that induce a productive ternary complex ensemble. In particular, having a small set of high-population ternary complex structures can provide an avenue for structure-based degrader discovery, particularly focused on the design of linkers that improve drug-like properties of the degrader molecule while maintaining or improving the aspects of the ternary structure ensemble that lead to ubiquitination.

Results

Different degraders induce similar ternary complex crystal structures

The ternary complexes of the bromodomain of SMARCA2 isoform 2 (iso2-SMARCA2^{BD}) and the VHL/Elongin C/Elongin B (VCB) complex induced by different heterobifunctional degraders have been studied extensively¹⁸. In particular, PROTAC 1, PROTAC 2, and ACBII are three degrader molecules that induce a ternary SMARCA2^{BD}:VCB complex with quite different degradation efficiencies (see Supplementary Table 1). Whereas crystal structures of the ternary complexes induced by PROTAC 1 (PDB ID: 6HAY) and PROTAC 2 (PDB ID: 6HAX) exist, none has been reported to date for ACBII, the most potent degrader among them. Thus, we determined the structure of SMARCA2^{BD}:VCB liganded by ACBII via X-ray crystallography (see electron density map of the ternary complex interface in Supplementary Fig. 1). The structure was obtained by hanging drop vapor diffusion (see X-ray structure determination of iso2-SMARCA2^{BD}:ACBII:VCB complex in the Methods for more details)¹⁸ and solved by molecular replacement to 2.25 Å in the highest resolution shell (Supplementary Table 2), using the PROTAC 2 crystal structure (PDB ID: 6HAX) as the search model (Fig. 1a).

ACBII bridges the induced interface, forming contacts with both proteins. Importantly, the degrader induces favorable contacts across the non-native interface, such as VHL:R69 and SMARCA2^{BD}:F1463 (Fig. 1b, c). SMARCA2^{BD}:N1464 maintains critical bivalent contacts to the aminopyridazine group of ACBII, positioning the terminal phenol group for pi-stacking interactions with residues F1409 and Y1421 (Fig. 1b, c). On the ligase side of the interface, the interactions between Y98 and ACBII are consistent with those between the same residue and PROTAC 1 or PROTAC 2 (Fig. 1b, c)¹⁸.

Despite differences in the linker compositions, the protein–protein interface induced by ACBII is structurally similar to that induced by PROTACs 1 or 2¹⁸ (Fig. 1d). A slight 1.7 Å twist of ACBII compared to the other two degraders, which can be ascribed to their minor differences (e.g. the ACBII linker has one additional ether group compared to the PROTAC 2 linker), results in a subtle swing of the protein in the crystal structure (Fig. 1d). However, the major protein–protein interface contacts are the same (Supplementary Fig. 2) and the structural differences do not align with the markedly different degradation efficiencies obtained¹⁸, suggesting that the corresponding (dynamic) ensembles of induced ternary complexes may be fairly different and therefore responsible for the degradation differential between PROTAC 1, PROTAC 2, and ACBII. Consistent with other studies^{14,22}, this implies that crystallographic snapshots are not suitable to provide a holistic view of the ensemble of all possible ternary complex structures in solution, but merely represent a subset of the relevant conformations favored by crystallization.

Hydrogen-deuterium exchange reveals extended protein–protein interfaces

In order to assess the impact of different degrader molecules on the dynamic nature of interactions at the SMARCA2^{BD}:VHL interface,

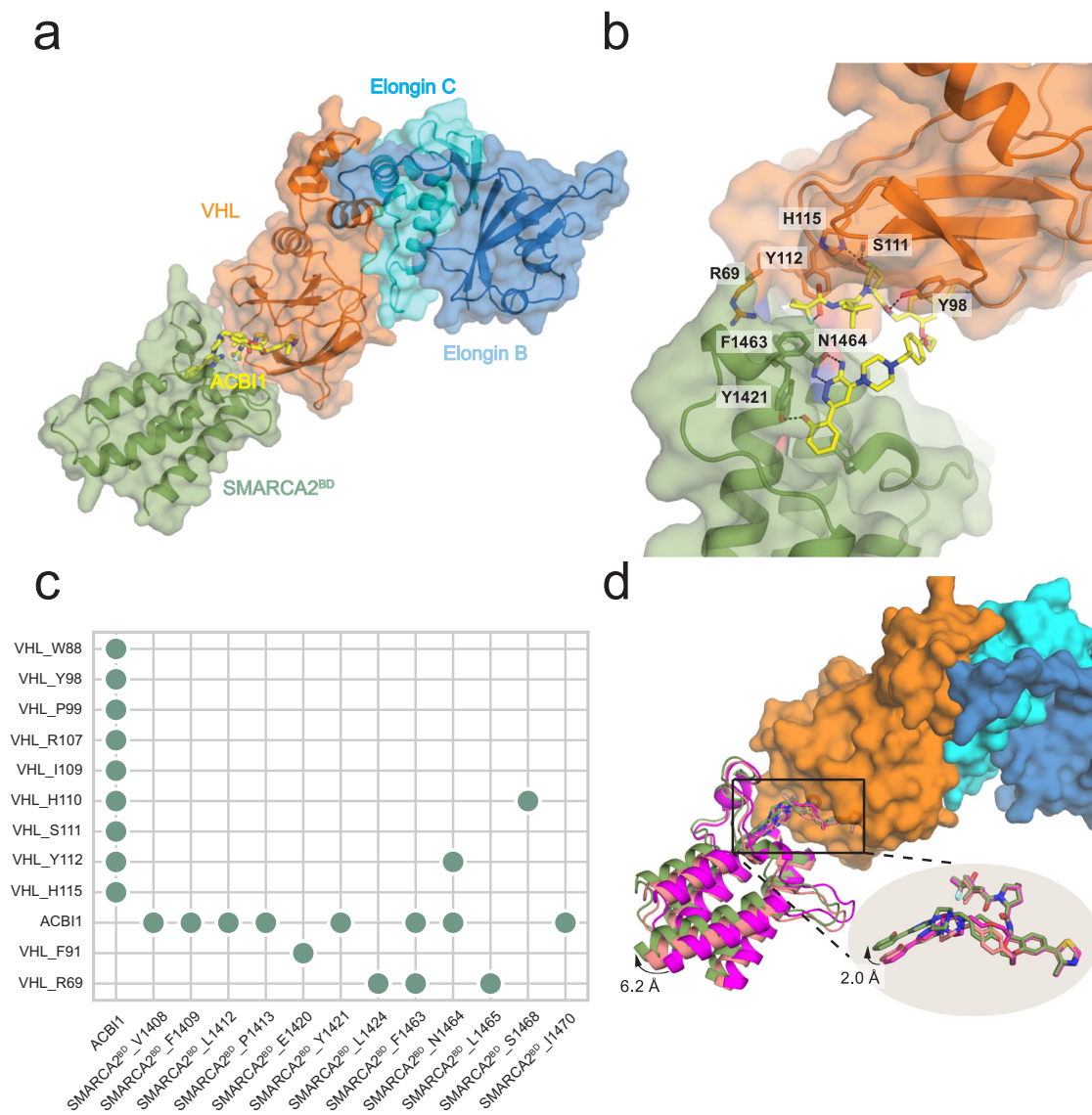


Fig. 1 | Ternary complex of SMARCA2^{BD} and VHL/Elongin C/Elongin B (VCB) induced by ACBII shows structural similarities to those induced by PROTAC 1 and PROTAC 2. **a** Overall perspective of SMARCA2^{BD} and the VCB complex induced by degrader molecule ACBII (shown as yellow stick representation). **b** ACBII-induced interface contacts between SMARCA2^{BD} and VHL. Annotated residues are among those that make the highest number of contacts (see panel c). **c** A contact map for the interface of the crystal structure (obtained by the Arpeggio software⁷⁰).

Contacts are indicated when 10 atomic contacts (i.e., distance ≤ 4.5 Å) are present. **d** Superposition of the crystal structures of PROTAC 1 (PDB ID: 6HAY, purple), PROTAC 2 (6HAX, salmon), and ACBII (7SE4, green) by aligning VHL (orange surface representation) shows varied conformations of the three degraders (up to 2.0 Å at the warhead), resulting in minor alterations of SMARCA2^{BD} within the ternary complex.

we performed hydrogen-deuterium exchange (HDX) mass spectrometry experiments on the respective APO (i.e., either SMARCA2^{BD} or VCB), binary (i.e., either protein bound to a degrader) and ternary (i.e., SMARCA2^{BD}:degrader:VCB complex) species (see Supplementary Tables 3–7) to characterize the induced protein–protein interfaces in solution²¹. This approach is a promising alternative to previous attempts at characterizing degrader ternary complexes that employed multiple crystal structures²², NMR¹⁴, and SAXS coupled with various forms of modeling. In addition, there exists a wealth of knowledge for the integration of HDX-MS coupled with computational modeling²³. Importantly, changes in the rate of deuterium incorporation are dependent on factors like pH, temperature, solvent occlusion and molecular interactions like hydrogen bonding²⁴. Since we control the temperature and pH using robotics systems, that enable precise temporal control over D₂O exposure, we can probe the effects of (binary and ternary) complex formation on hydrogen bonding and solvent

exposure. To ascertain the changes in solvent protection in the binary or ternary complex, the uptake of the APO or binary species is subtracted from that of the corresponding binary or ternary states (referred to as Binary Δ APO and Ternary Δ Binary), respectively. The results are summarized in difference plots that highlight the statistically significant (95% or 98% confidence interval) changes in deuterium uptake (see Supplementary Fig. 37a–d for the SMARCA2^{BD}:VCB complex induced by ACBII).

Figure 2a reveals that large regions of SMARCA2^{BD} become protected upon ternary complex formation induced by ACBII (see Ternary Δ Binary difference plot). These stretches of protected residues, e.g., amino acids 1409–1422 and 1456–1470, overlap with the warhead binding site based on the ternary complex structure published in this work (7S4E) and those published previously (6HAY, 6HAX), which confirms the similarity of the ternary complex interface among the three degrader molecules discussed above. In addition, there are also

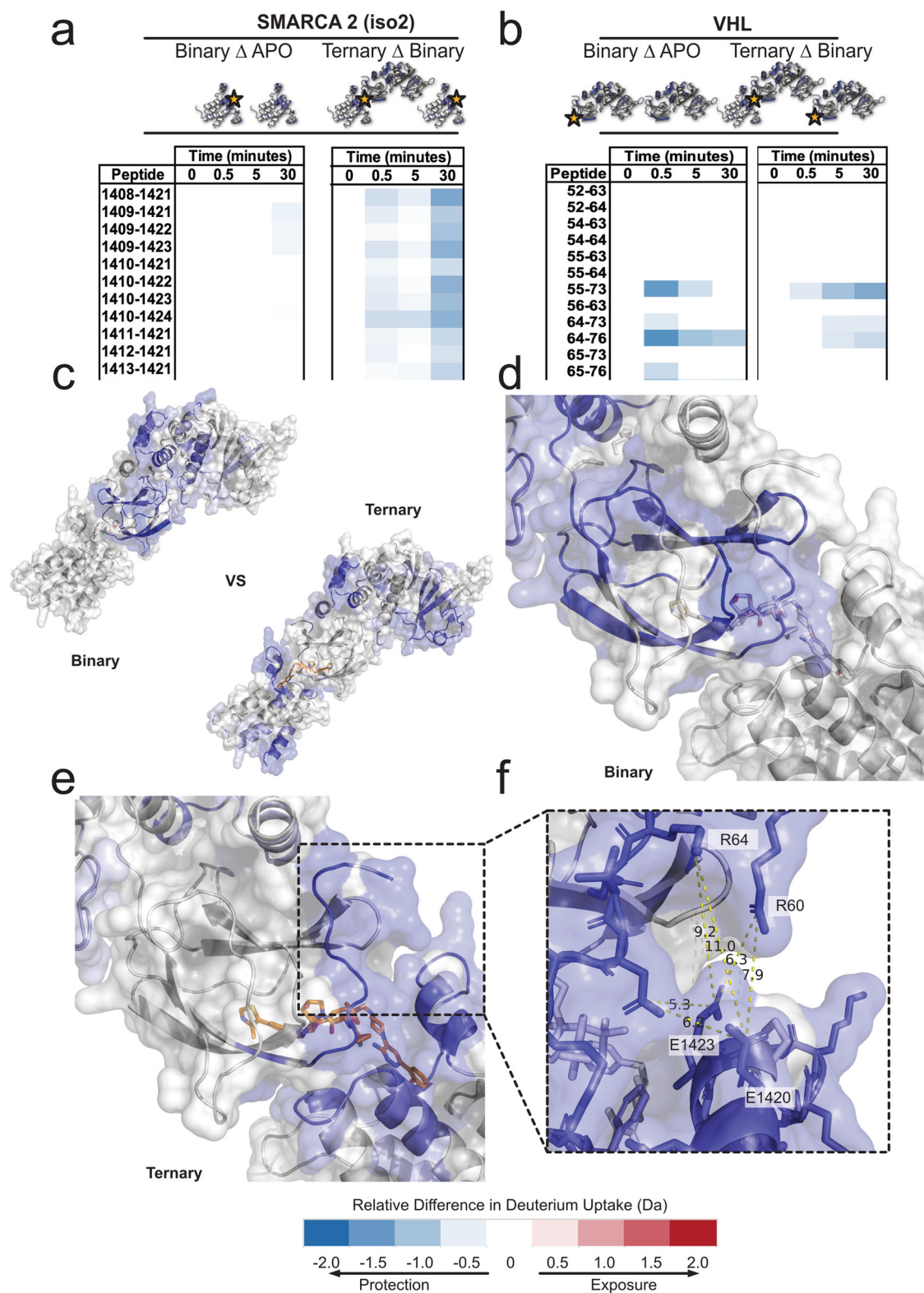


Fig. 2 | HDX-MS reveals an extended interface of the ACBII-induced SMARCA2^{BD}:VHL ternary complex compared to crystallographic data. a SMARCA2^{BD} HDX difference plots covering residues 1408–1424. Binary as compared to the APO, and ternary as compared to the binary states reveal increased protection induced by the presence of ACBII and VCB complex. **b** VHL HDX difference plots covering residues 52–76. Binary compared to APO and ternary compared to binary states of the VHL subunit highlighting extended exchange patterns due to the presence of

the ternary complex. **c** Exchange patterns induced by the binary and ternary forms of the complex superimposed on the crystal structure (PDB ID: 7S4E). **d** Binary-specific induced HD exchange near the E3-ligand and warhead binding sites of VHL and SMARCA2^{BD}. **e** Ternary-specific induced HD exchange near the E3-ligand and warhead binding sites of VHL and SMARCA2^{BD}. **f** Proposed solution-state extended protein interface that may take advantage of salt-bridge interactions to increase cooperativity of the protein–protein complex.

stretches of protected amino acids, 1394–1407, that are too distant from the established binding interface to result from complex formation (Fig. 2a, f).

Importantly, the HDX-MS experiments enable us to monitor the impact of cooperativity on ternary complex formation. For instance, the Binary Δ APO plot of SMARCA2^{BD} reveals minimal differences (left column in Fig. 2a), suggesting a comparable amount of protection upon ACBII-binding to SMARCA2^{BD}. However, it should be kept in mind that, in our experiments, the concentrations of SMARCA2^{BD} and ACBII (both at $-15\ \mu\text{M}$ after dilution, see “Methods”) are close to their dissociation constant, $K_D = 10\ \mu\text{M}^{18}$, which, based on standard enzyme complex calculations, means that $\sim 45\%$ of the solutes form a binary complex. This is far less than the 80% threshold required to measure the effect of binding events on deuterium exchange (see “Methods”), providing a rational explanation for the aforementioned highly similar protection patterns of the APO and the binary states that, as a matter of fact, result from the frequent binding and unbinding events in the binary system. Central to the theme of cooperativity, adding only VCB ($-15\ \mu\text{M}$, see “Methods”) to SMARCA2^{BD} and ACBII, under the very same experimental conditions as in the SMARCA2^{BD}:ACBII binary system, yields strong protection of distinct regions on SMARCA2^{BD} (Fig. 2c). Obviously, the presence of VCB alone limits the observed deuterium exchange (indicated by the Ternary Δ Binary plot in the right column of Fig. 2a), suggesting that the ligase forms (non-native) contacts with SMARCA2^{BD}, which stabilize the interactions with ACBII in a cooperative fashion, and, consequently, greatly reduce the amount of ACBII required to elicit a binding response on SMARCA2^{BD}. It is important to note that, while cooperativity induced by VHL is enhancing the ability of SMARCA2^{BD} to bind ACBII, the resultant protection on SMARCA2^{BD} within the ternary complex is a mixture of both binary and ternary protection effects. Nevertheless, these considerations illuminate the crucial role of cooperativity in degrader-induced targeted protein degradation.

Large regions of VHL are protected in the presence of the E3-ligand too, as indicated by the corresponding Binary Δ APO difference plot (see left column in Fig. 2b). The most protected residues in the binary state are centered around amino acids 87–116, which include all nine residues in the E3-ligand binding site of VHL. In the presence of SMARCA2^{BD} (see Ternary Δ Binary difference plot in the right column in Fig. 2b), much of the allosteric network due to E3-ligand binding can be subtracted away leaving only the most significantly protected residues induced by ternary complex formation (Fig. 2b, d). In particular, residues 60–72, which house the critical interaction of R69, show significant protection due to ternary complex formation (Fig. 2b, d). Moreover, we observe protection of residues 166–176 and residues 187–201 on VHL (see Supplementary Fig. 37 b, f) as well as some regions on Elongin B and C upon ternary complex formation (see Supplementary Fig. 37c, d). Although these sites are distal from the binding interface, they can be proximal to each other as their mapping onto the ternary complex crystal structure reveals (Fig. 2c). This suggests that conformational changes of the VCB ligase in ternary complex may exist, that differ from those in its APO or binary state, potentially giving rise to allosteric effects²⁵ that play a role in the ACBII-induced arrangement of SMARCA2 in the full Cullin-RING ligase complex.

The difference between HDX-MS binary and ternary SMARCA2^{BD} experiments reveals that the interactions at the protein–protein interface help stabilize the ternary complex. Many of the charged interface residues, that are solvent-exposed and outside the range of traditional hydrogen bonding or salt-bridge interactions ($>6.3\ \text{\AA}$) in the corresponding X-ray crystal structure (e.g. K1416, E1420, E1423 on SMARCA2^{BD} and R60, R64 on VHL) are determined to be protected based on the HDX-MS results (Fig. 2e). In fact, the protected, charged interface residues of SMARCA2^{BD} lie outside the direct warhead binding pocket in the crystal structure of the ternary complex. Interestingly, R60 through R64 on VHL are protected in the ternary complex

for a longer duration than in the binary complex alone. This enhanced protection across the interface suggests that conformational rearrangements are responsible for protein–protein interactions. Our simulations presented below (see “Results” section “The structural determinants of degrader ternary complex formation”) support this hypothesis, finding contacts between several of these charged interface residues. These results further underscore the importance of cooperativity driving the formation of the ternary complex for ligases with poor binding affinity to the POI.

Interestingly, we find that iso1-SMARCA2^{BD}:ACBII:VCB shows a slightly different protection pattern from iso2-SMARCA2^{BD}:ACBII:VCB, mainly in that residues G104 through L116 of VHL show significant protection in the former compared to the latter ternary complex. In our crystal structure of the iso2-SMARCA2^{BD}:ACBII:VCB system, these protected residues are close to the site where the additional 17 residues of iso1-SMARCA2^{BD} appear, suggesting that the protected residues in VHL may be interacting with these residues that are not present in iso2-SMARCA2^{BD}. Consistent with this hypothesis, residues I1414–N1417 of the iso1-SMARCA2^{BD} extension show some protection in the ternary complex.

Studying the solution-state dynamics of degrader ternary complexes uncovers key details that are missed by crystallographic snapshots alone. As many of the crystallographic contacts are nearly identical between the different degrader molecules, important interactions may be underrepresented in the crystal structure, such as those between charged interface residues or residues distal from the binding site identified to be significantly protected upon ternary complex formation. Therefore, utilizing HDX-MS information, or other data derived from solution-state experiments, as restraints in modeling and simulation opens a pathway from a single accepted protein structure to a vast ensemble of conformations. Production of accurate ternary complex ensembles enables alternative routes for the design, optimization, and mechanism-of-action studies of heterobifunctional degraders.

HDX data enhance weighted ensemble simulations of ternary complex formation

We simulate the formation of iso2-SMARCA2^{BD}:VHL degrader ternary complexes using weighted ensemble (WE) simulations, where a set of weighted trajectories are evolved in parallel along pre-defined collective variables, providing a means to compute non-equilibrium properties and predict likely binding pathways²⁶. This path-sampling strategy can sample rare events by orders of magnitude more efficiently than conventional MD simulations and it has been employed before for tasks such as protein–protein²⁷ and protein–ligand binding²⁸. It is noteworthy, however, that our simulations are not informed by any structural data about the ternary complex interface from X-ray crystallography experiments.

Starting from a dissociated configuration, in which the degrader molecule is bound to VHL, yet both are clearly apart from SMARCA2^{BD} (initial separation distance $\sim 20\ \text{\AA}$), the formation of ternary aggregates is simulated yielding complexes with interface structures well comparable to those obtained experimentally, which are generally located in the low free energy basins. For these simulations, we use a collective variable defined by the number of atomic contacts and the warhead-RMSD (w-RMSD) with respect to the crystal structure of the target-warhead complex (see the “Methods” section “WE-HDX simulations”).

As HDX experiments show, and our simulations of ternary complexes below confirm, the ternary complex exists as a dynamic ensemble of multiple conformations, of which the X-ray structure is a snapshot. Thus, we assess the quality of bound complexes by the minimum interface-RMSD (I-RMSD)²⁹ of each simulated aggregate with respect to a set of structurally diverse reference ternary structures (see Supplementary Fig. 5). This set of distinct structures is obtained from

long-timescale (>1 μ s) brute-force MD simulations, thus allowing a comparison to a variety of possible ternary complexes and not merely to a single experimental reference structure. We provide detailed descriptions of the methodology and the evaluation of all simulations performed in the Supplementary Information.

Protein–protein encounter complexes, i.e., the formation of protein contacts, are usually observed within 500 ns of aggregate simulation time. An ensemble of about 500 bound ternary complexes with a minimum I-RMSD < 2 Å can usually be obtained after ~2 μ s, which takes ~12 days using a single A40 GPU per simulation but is highly parallelizable to more GPUs.

Remarkably, when introducing as a collective variable the number of contacts formed by the protected residues (see Supplementary Table 8), as determined by the HDX-MS experiments described above, the prediction accuracy of ternary complex formation is significantly improved compared to simulations in which any protein–protein contacts were considered (see Supplementary Figs. 3 and 4). Supplementary Movie 1 (see Supplementary Table 15) shows the continuous trajectory of one such ternary complex binding event, where the addition of protected-residue contacts enhances the ternary complex binding.

We note that the use of HDX-MS data in our approach is rather qualitative, as the simulations are solely informed by the existence of specific interaction sites and not by the degree of those interactions. HDX rate constants are not estimated during the simulations as often performed in quantitative approaches that combine HDX-MS experiments with simulation³⁰. Rather, our method falls in the category of simulations guided by HDX-MS data, in which qualitative correlations between simulation and experiment are attempted to be established²³ (see Supplementary Note 1). We present a particularly interesting example of synergy between molecular simulations and HDX-MS experiments, in which the path-sampling algorithm is furnished with a fairly simple parameter derived from the experimental measurements, i.e., the contact numbers between distinct sites. We call this integrated approach WE-HDX. Despite its simplicity, WE-HDX seems particularly appropriate for the formation of ternary complexes that have distinct contacts across their binding interface.

To systematically study the formation of SMARCA2^{BD}:VHL ternary complexes with all three degraders, we run seven independent WE-HDX simulations with PROTAC 2 for an aggregate simulation time of 12.5 μ s and three such simulations totaling ~6 μ s for both PROTAC 1 and ACBII. The difference in the number of simulations is due to the greater flexibility of the PROTAC 2 ternary complexes, compared to the other two degraders. Ensembles of bound ternary complexes were formed with minimum I-RMSDs of 0.5 Å for ACBII, 0.7 Å for PROTAC 1, and 1.1 Å for PROTAC 2, respectively.

To highlight the sampling ability of WE-HDX simulations, Figure 3a compares the minimum I-RMSD of the SMARCA2^{BD}:PROTAC 2:VHL simulation with that from vanilla MD simulations of the same system as a function of aggregate simulation time. While the minimum I-RMSD converges to 2.5 Å in the WE-HDX simulations within 0.5 μ s of aggregate simulation time, that for the vanilla MD remains as high as 10 Å after 1.4 μ s of simulation.

The very high prediction accuracy of the WE-HDX simulations is illustrated for the SMARCA2^{BD}:PROTAC 2:VHL system in Fig. 4. One example of a predicted structure is visualized in Fig. 4a, b. The contact maps presented in Fig. 4c compare the ternary interface of the experimental crystal structure to that of the minimum I-RMSD structure produced by the WE-HDX simulations. Each point reflects the degree of interaction, revealing an interaction pattern from the WE-HDX simulations that is comparable to that from experiment. The near-perfect alignment (minimum I-RMSD = 1.1 Å) of one sampled conformation with the crystal structure shown in Fig. 4d further emphasizes that the interactions of degrader ternary complexes observed experimentally can be recaptured by WE-HDX.

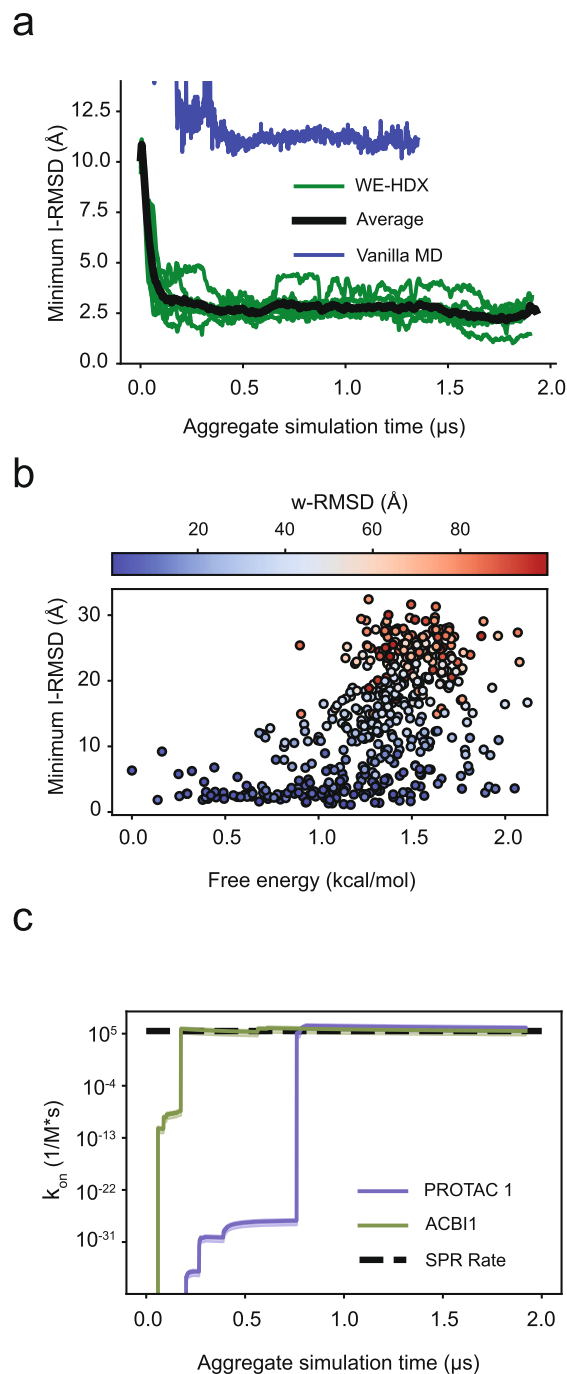


Fig. 3 | Assessing ternary complex formation. **a** The minimum I-RMSD over time during the WE-HDX simulations of the PROTAC 2 system. Each green line indicates one replica ($n = 7$) and the black line is the average between all runs. The blue line indicates the minimum I-RMSD for a vanilla molecular dynamics simulation. **b** A scatter plot of the free energy vs the minimum I-RMSD of each of the 500 clusters from the PROTAC 2 simulations. The circles are colored by w-RMSD. **c** The predicted binding rates for the PROTAC 1 system (purple) and the ACBII system (green). The black line is the experimental binding rate determined via SPR. Source data are provided as a Source Data file.

Six out of seven of the SMARCA2^{BD}:PROTAC 2:VHL simulations produced binding events for a total of 3278 uniquely bound ternary complexes. In order to assess the degree of heterogeneity within this ensemble, we clustered the WE-HDX results into 500 macrostates with a k -means algorithm using the C_{α} - C_{α} distances between the ligase and target protected residues. As expected, all states with a low minimum

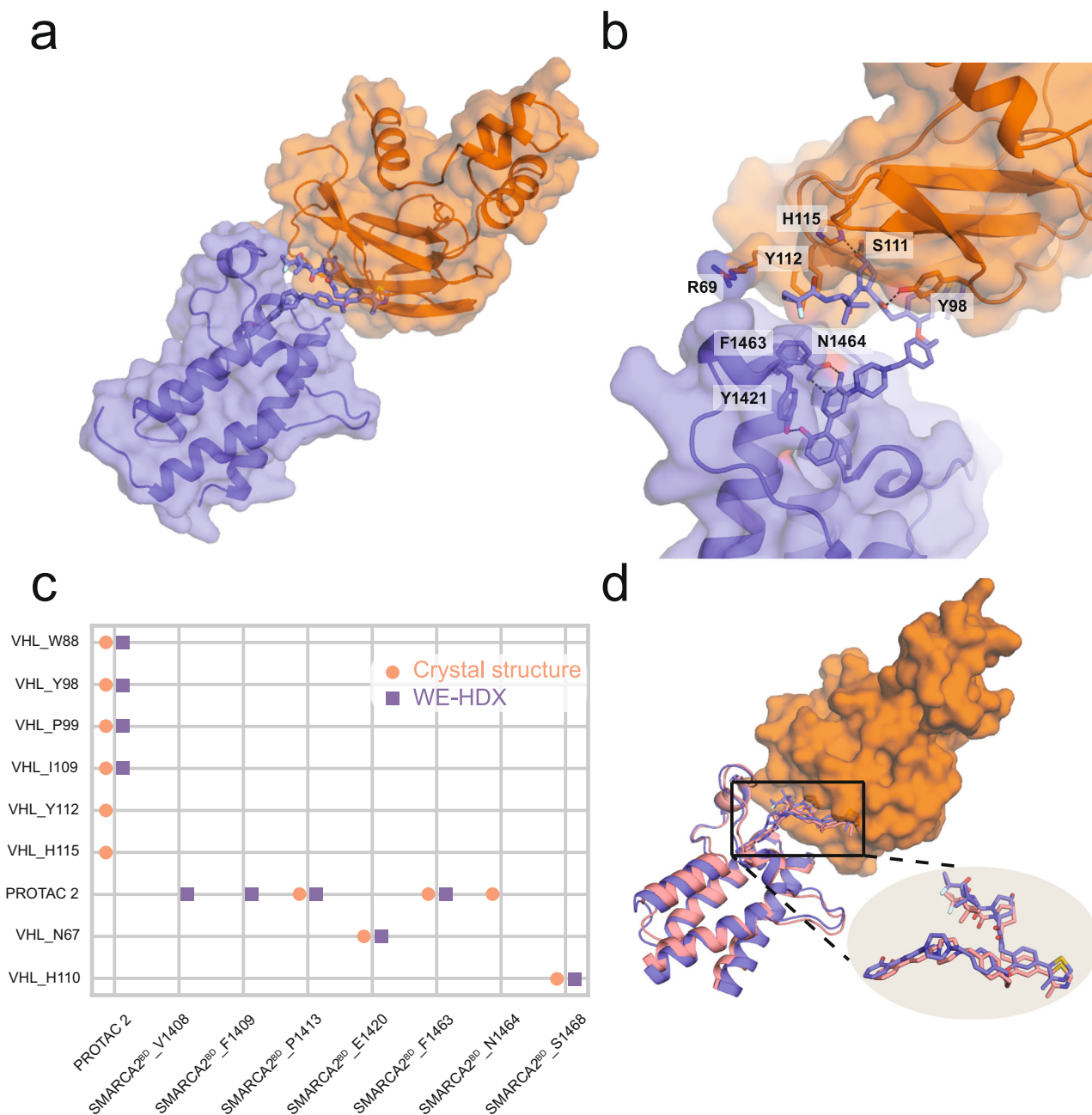


Fig. 4 | Illustration of one representative prediction of SMARCA2^{BD}:PROTAC 2:VHL produced by WE-HDX simulations and its comparison to the crystal structure (PDB ID: 6HAX). a A simulated ternary structure with minimum I-RMSD = 1.1 Å. SMARCA2^{BD} (purple) and VHL (orange) are shown in cartoon and transparent surface representations and PROTAC 2 is shown in stick

representation. **b** Structural details of the binding interface. Annotated residues are among those that make the highest number of contacts (see panel c). **c** A contact map of the interfaces from the crystal (salmon) and the simulated structure (purple). **d** Structural alignment of the simulated (purple) with the crystal structure (salmon) with a detailed PROTAC 2 comparison.

I-RMSD have low values of w-RMSD too (Fig. 3b). States with high free energies, i.e., above 1.5 kcal mol⁻¹, have large minimum I-RMSDs, ranging from 1.5 to 30 Å. However, the minimum I-RMSD distribution among the 20 low free energy states below 0.5 kcal mol⁻¹ is significantly tighter, ranging from 1.1 to 9.2 Å with an average value of 3.7 Å and 12 out of the 20 states even having a minimum I-RMSD below 3 Å.

We predict ternary complex binding rate constants for the three different degraders directly from WE-HDX simulations using the probability flux into a bound state (minimum I-RMSD < 2 Å). While the predicted rates for PROTAC 1 and ACBII are on the same order of magnitude as in experiments (Fig. 3c), we predict a significantly slower

binding rate for PROTAC 2, which is not yet determined experimentally (see Supplementary Table 9). However, for all three rates there are large uncertainties, as has previously been observed in WE rate calculations³¹. Better statistics can be achieved by longer simulation times or the use of recently proposed algorithms that converge these rates more efficiently^{32,33}, which is beyond the scope of this work.

In most of the analysis above, we have used the minimum I-RMSD with respect to a set of reference structures, as described, to assess the quality of structures obtained from WE-HDX simulations. Alternatively, the C_α-RMSD of the entire ternary complex has been used before as a parameter to gauge their prediction accuracy²⁷. Supplementary Figure 4b shows that the interface-RMSD, and, in particular, the threshold

at 2 Å is indeed an appropriate metric for the identification of ternary complexes, as all such complexes formed in our WE-HDX simulations of the system with PROTAC 2 have a minimum I-RMSD < 2 Å for a C_{α} -RMSD \leq 5 Å, which is clearly below the threshold used in other studies (e.g. C_{α} -RMSD \leq 10 Å used by Drummond et al.²⁰).

As in most design projects X-ray structures may not be readily available, it is important to determine the usefulness of predictive features that do not depend on ternary complex X-ray structures. To this end, we filtered the ensemble of simulated SMARCA2^{BD}:PROTAC 2:VHL structures for bound complexes with warhead-RMSD < 2 Å and >30 contacts between protected residues (Fig. 5a). Among these, the bulk of the density was limited to minimum I-RMSD values between 1 and 4 Å, with 90% below 3 Å and 43% even below 2 Å (Fig. 5b), indicating that observables such as the warhead-RMSD and the number of contacts between protected residues can be used to characterize bound ternary complexes.

Knowledge of a large number of degrader-induced ternary complexes is essential to understanding the structural and dynamic features that lead to targeted protein degradation. As the WE-HDX results reveal, the level of detail associated with such simulations allows an entire ensemble of ternary complexes, including many conformations with a pronounced protein interface, to be generated ab initio, i.e., even from a fairly dissociated state and with no additional information on the protein–protein binding pose. This is a significant achievement with regard to the design of effective degrader molecules, for which ternary complex structures are not obtained experimentally.

HDX-MS improves prediction of ternary complexes using docking

Several docking procedures to predict ternary complexes of degrader molecules have been described. Most of them have stages for generating protein–protein complexes in the absence of the degrader, linker, alignment of linker or whole degrader to the protein–protein complexes, and some sort of scoring^{19,20}. We used an approach comparable to that published by Bai et al.¹⁹

In contrast to recent work²¹, our docking method uses HDX-MS data to impose additional distance restraints at the sampling stage (instead of post-sampling scoring). Also, differently from the distance restraints derived from chemical cross-linking experiments³⁴, our approach is based on the statistics of the length of the linker in a degrader molecule. Application of the HDX-MS data for re-ranking of the docking predictions, as described by Eron et al.²¹, may lead to a more quantitative assessment of structures. Discussion of the interplay of HDX-MS-derived restraints and HDX-MS-based re-rankings in docking is beyond the scope of the present work.

We show that incorporating experimentally retrieved distance restraints into the docking protocol significantly improves its ability to predict ternary complexes of high quality (see detailed comparisons in Supplementary Figs. 6 and 7). In particular, it is striking how strongly the incorporation of HDX-MS data can boost the accuracy of the docking protocol among the highest-ranked docking poses.

Although WE-HDX simulations consistently outperform the HDX-enhanced docking routine (Fig. 5), docking, in combination with HDX-MS (Docking-HDX), is a useful tool for the quick filtering of a large number of degrader designs considering the significantly less computational cost of this approach (25 CPU hours for the generation of one ensemble compared to ~12 A40 GPU days for the WE-HDX method).

HREMD and SAXS reveal highly flexible ternary complexes

The HDX-MS measurements revealed substantial flexibility, which is consistent with the structural diversity obtained from WE-HDX simulations and from the docking protocol of the SMARCA2^{BD}:VHL ternary degrader-protein complexes studied here. To further enhance the exploration of their conformational heterogeneity, we perform

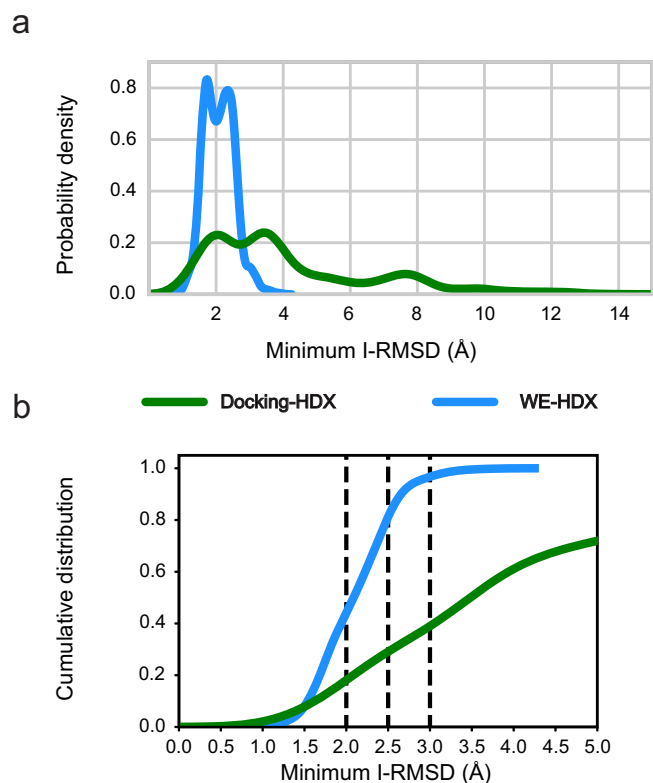


Fig. 5 | Comparing the bound ensembles determined by docking and WE simulations with information from HDX-MS for the PROTAC 2-induced ternary complex. Simulated structures with a warhead-RMSD \leq 2 Å and >30 contacts between the SMARCA2^{BD} and VHL interface are considered bound, whereas the docked bound structures are determined as the top-100 from Rosetta-scoring. **a** Probability densities of minimum I-RMSD values for the bound ensembles from WE-HDX (blue) and Docking-HDX (green). **b** Cumulative distributions of minimum I-RMSD values for the probability densities shown in panel a, illustrating that larger ensembles of bound ternary complexes can be obtained from WE-HDX (blue) compared to Docking-HDX (green). The dashed vertical lines indicate three specific thresholds of minimum I-RMSD (2, 2.5, and 3 Å), below which a complex can be considered bound. Source data are provided as a Source Data file.

atomistic Hamiltonian replica-exchange MD (HREMD) simulations based on the X-ray structures. HREMD is a parallel tempering simulation method that efficiently samples large conformational changes of proteins in aqueous solution and, therefore, is a promising strategy to study the protein–protein interactions and the flexibility of degraders in ternary complexes (see “Methods” section “HREMD simulation” for more details). In particular, we simulate ternary complexes of both isoforms of SMARCA2^{BD} connected only to the VHL subunit or, in order to be consistent with our experiments, to the larger VCB complex by PROTAC 1, PROTAC 2, or ACB11 (see Supplementary Table 13 for a list of all HREMD simulations performed). The structure of iso1-SMARCA2^{BD}, which is not experimentally resolved, is obtained by homology modeling with the iso2-SMARCA2^{BD} structure used as template (see the “Methods” section “Isoform 1 homology model” for more details). HREMD simulations with iso1-SMARCA2^{BD} were performed to test whether they could explain the ternary complex protection differential observed between that isoform and Isoform 2. To ensure the HREMD-generated ensembles are accurate and reliable, we validate the simulations by directly comparing results against those from size-exclusion chromatography coupled to small-angle X-ray scattering (SEC-SAXS) data (see Fig. 6a and Supplementary Tables 11, 12).

The agreement ($\chi^2=1.55$ and $\chi^2=1.23$ for iso1- and iso2-SMARCA2^{BD}:ACB11:VCB respectively, where χ^2 is defined in Eq. (14))

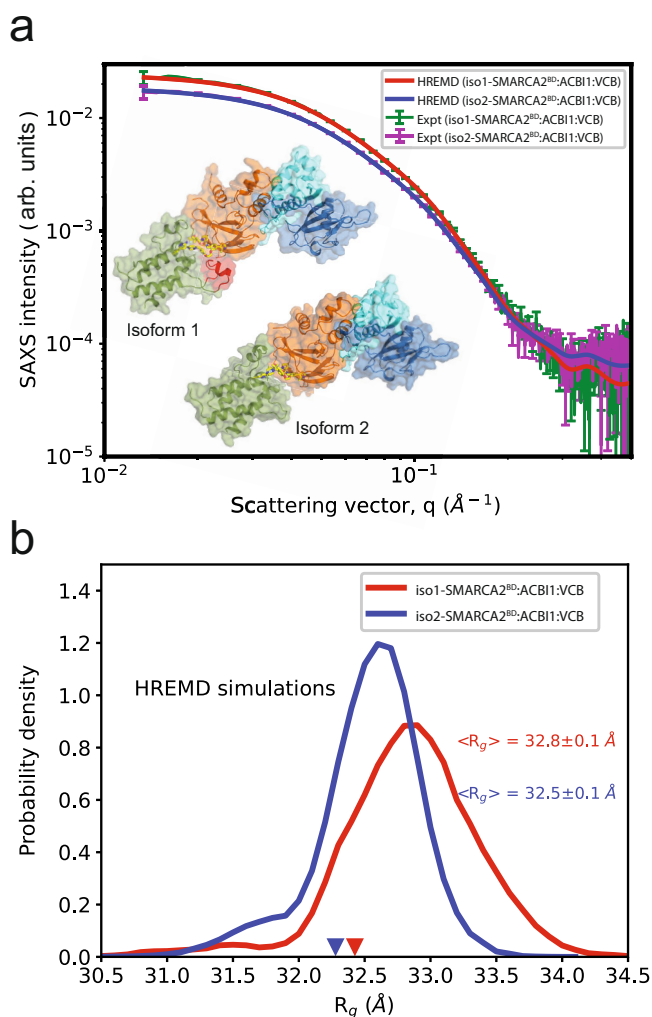


Fig. 6 | SAXS profiles and structural ensembles of iso1-/iso2-SMARCA2^{BD}:ACBII:VCB complexes. **a** Comparison of theoretical (red and blue) and experimental (green and purple) SAXS profiles, SAXS intensity vs. q . The experimental profiles are the averaged (buffer-subtracted) intensities from elution peak 1 of the SEC trace (see Supplementary Fig. 15). Error bars describe the standard deviation in the experimental scattering intensity at a given q (see “Methods”). **b** The histograms of R_g of iso1-SMARCA2^{BD}:ACBII:VCB (red) and iso2-SMARCA2^{BD}:ACBII:VCB (blue) complexes calculated from HREMD simulations. The inverted red and blue triangles are the R_g values of starting structures of iso1-/iso2-SMARCA2^{BD}:ACBII:VCB from homology model and crystallography respectively. Source data are provided as a Source Data file.

between SAXS profiles obtained from experiment and such calculated from simulations shows that the HREMD simulations capture the long timescale conformational ensembles to experimental accuracy. Furthermore, the ensemble-averaged R_g of the two complexes from simulation are in agreement to R_g values obtained by Guinier approximation (Eq. (2)) to experimental SAXS data (Supplementary Fig. 16), $R_g = 33.4 \pm 0.4 \text{ \AA}$ and $32.3 \pm 0.3 \text{ \AA}$ for iso1- and iso2-SMARCA2^{BD}:ACBII:VCB, respectively. The histograms of R_g (calculated from atomic coordinates using Equation (3)) suggest that ternary complexes are flexible in solution leading to a change in overall conformation compared to their corresponding simulation starting structures, i.e., a homology model of iso1-SMARCA2^{BD}:ACBII:VCB and the crystal structure of iso2-SMARCA2^{BD}:ACBII:VCB (Fig. 6b). These results illustrate the need for enhanced sampling methods, such as HREMD, to rigorously probe the conformational changes of the inherently flexible ternary degrader complexes.

To demonstrate the value of the HREMD simulations in aiding in the prediction of degrader efficacy, we analyze the thermodynamics of ternary complex formation by estimating a conformational free energy penalty for the binding of a fully dissolved PROTAC 1, PROTAC 2, or ACBII to SMARCA2^{BD} and VHL in a ternary complex. To this end, we simulate the individual degraders in solution (see the “Methods” section “MD simulation of degraders” for more details), in addition to the ternary complex simulations presented above, and compare, as an observable proxy, the average linker end-to-end distance (normalized by the number of backbone atoms in the linker) of each degrader when fully dissolved to the corresponding value obtained when bound in a ternary complex. We observe that, in both environments, PROTAC 2 and ACBII adopt a significantly more expanded linker conformation compared to PROTAC 1 (Supplementary Fig. 21), which has a lower SMARCA2-degradation efficiency than the other two degraders (Supplementary Table 1). This suggests, in accord with previous empirical findings²⁰, that degraders with extended linkers in solution more easily induce SMARCA2^{BD}:VHL ternary complexes (Supplementary Fig. 21). In the next section, we further explore the ternary complex ensembles sampled with HREMD.

The structural determinants of degrader ternary complex formation

We quantify the free energy landscapes of several of the ternary complexes sampled in the HREMD simulations, namely iso2-SMARCA2^{BD}:PROTAC 1:VHL, iso2-SMARCA2^{BD}:PROTAC 2:VHL, iso2-SMARCA2^{BD}:ACBII:VHL, and iso1-SMARCA2^{BD}:ACBII:VCB. We begin this analysis by performing principal component analysis (PCA) decomposition of the distances between interface residues to identify high-variance collective variables (see “Methods” section “Conformational free energy landscape determination” for more details). The probability distribution of these high-variance features allows us to determine a more easily interpretable free energy landscape from our simulation data. We find that the landscape of each ternary complex contains several local minima differing by only a few kcal mol⁻¹ (Fig. 7a and Supplementary Fig. 22).

Using k -means clustering in the PCA feature space, we then identify distinct clusters of conformations. Cluster centers roughly correspond to local minima in the free energy landscape (Fig. 7a and Supplementary Fig. 22). These clusters of simulated conformations are consistent with our HDX-MS protection data: Figure 8 shows that interface residues that were found to be protected in HDX-MS experiments are observed to interact in either the most populated or second most populated cluster. Notably, this analysis shows that in representative structures (namely the second most populated cluster centers) of iso1-SMARCA2^{BD}:ACBII:VCB, the helix formed by the 17-residue extension of iso1-SMARCA2^{BD} interacts with a beta sheet of the VHL subunit (Fig. 8b) in accordance with our HDX-MS experiments that found this beta sheet to be protected in presence of iso1-SMARCA2^{BD}, but not in the presence of iso2-SMARCA2^{BD} (Fig. 8a). Similarly, representative structures from highly populated clusters of iso2-SMARCA2^{BD}:ACBII:VHL and iso2-SMARCA2^{BD}:PROTAC 2:VHL show contacts between residues that were observed to be protected in HDX-MS experiments (see blue-colored regions in Fig. 8a, c), whereas such from the most populated cluster of iso2-SMARCA2^{BD}:PROTAC 1:VHL (Fig. 8e) do not show these contacts. Representative structures from the most populated cluster of iso2-SMARCA2^{BD}:degrader:VHL with all three degraders are displayed in Supplementary Fig. 23.

Our analysis shows that both iso2-SMARCA2^{BD}:ACBII:VHL and iso2-SMARCA2^{BD}:PROTAC 1:VHL assume quite stable conformations: in both cases, the majority of snapshots fall into the largest cluster of conformations (Supplementary Fig. 23). The ground state (lowest free energy) structures are also quite similar to the corresponding crystal structures (C_α -RMSD of $1.7 \pm 0.3 \text{ \AA}$ for iso2-SMARCA2^{BD}:PROTAC 1:VHL and $0.8 \pm 0.1 \text{ \AA}$ for iso2-SMARCA2^{BD}:ACBII:VHL). However, iso2-

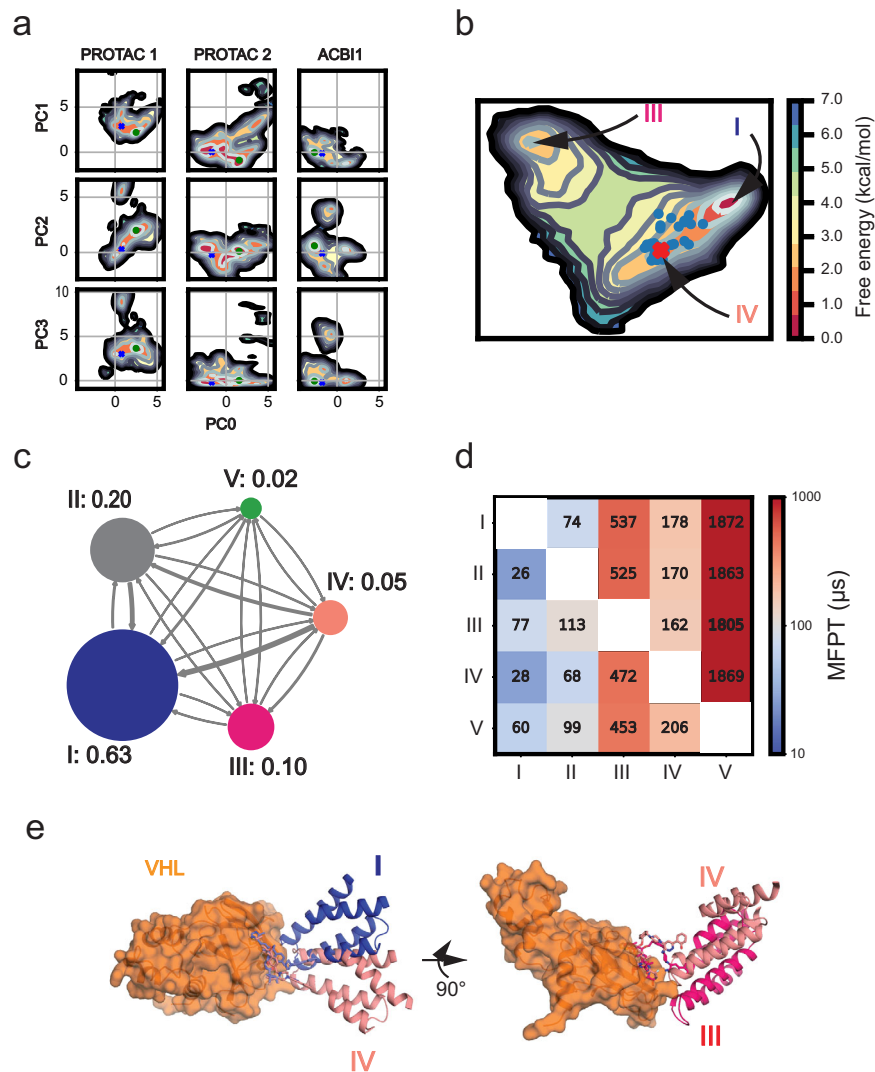


Fig. 7 | Thermodynamic and kinetic analysis of ternary complex simulations. **a** Conformational free energy landscapes of the iso2-SMARCA2^{BD}:PROTAC 1:VHL, iso2-SMARCA2^{BD}:PROTAC 2:VHL, and iso2-SMARCA2^{BD}:ACBII:VHL systems in the PCA space defined by our analysis of HREMD simulations. The crystal structure of each system is shown as a dark blue X, while the center of the largest *k*-means cluster is shown as a green point. Energy scale bar shown in panel **b**.

b Conformational free energy landscape as a function of the first two tICA features of iso2-SMARCA2^{BD}:PROTAC 2:VHL ternary complex inferred from a Markov state model (MSM) determined using long-timescale Folding@home simulations. The

ensemble of bound states from WE-HDX simulations is shown as blue points; the crystal structure (PDB ID: 6HAX) is shown as a red X. In this projection, states II and V are not indicated but are close to state I. **c** Network diagram of the coarse-grained MSM calculated using a lag time of 50 ns, with the stationary probabilities associated with each state indicated. **d** Mean first-passage times (MFPTs) to transition between MSM states. Numbers indicate predicted MFPTs in μ s. **e** Comparison of the crystal structure (salmon) with the lowest free energy state (blue) and a metastable state (red) predicted by the MSM. Arrows indicate the change of orientation by 90°.

SMARCA2^{BD}:PROTAC 2:VHL shows a much more dynamic landscape, and samples conformations similar to both the ground state of iso2-SMARCA2^{BD}:ACBII:VHL and iso2-SMARCA2^{BD}:PROTAC 1:VHL. This result, based on the enhanced sampling of ternary complexes, allows us to rationalize the differential in degradation efficiencies observed among the three degraders (Supplementary Table 1). We suggest that PROTAC 1 may fail to mediate the degradation of SMARCA2 because the (stable) conformation adopted by the ternary complex cannot be productively ubiquitinated. ACBII, on the other hand, induces a productive conformation of the ternary complex, facilitating ubiquitination. Hence, PROTAC 2 would then fall between the two, as the corresponding ternary complexes sample both the productive conformation induced by ACBII and the non-productive PROTAC 1-like conformation (see free energy landscapes in Fig. 7a).

To characterize the free energy landscape of iso2-SMARCA2^{BD}:PROTAC 2:VHL more comprehensively, we select 98 representative structures from the corresponding HREMD simulation as initial configurations for simulations on Folding@home (F@H), one of the largest distributed computing networks. Each initial condition was cloned 100 times and run for ~650 ns, for a total of ~6 ms of simulation time. These independent MD trajectories provide the basis for fitting a Markov state model (MSM)³⁵, which provides a full thermodynamic and kinetic description of the system and allows for the prediction of experimental observables of interest. We use time-lagged independent component analysis (tICA) to determine the collective variables with the slowest dynamics. The distance between points in the tICA feature space corresponds roughly to a kinetic distance³⁶.

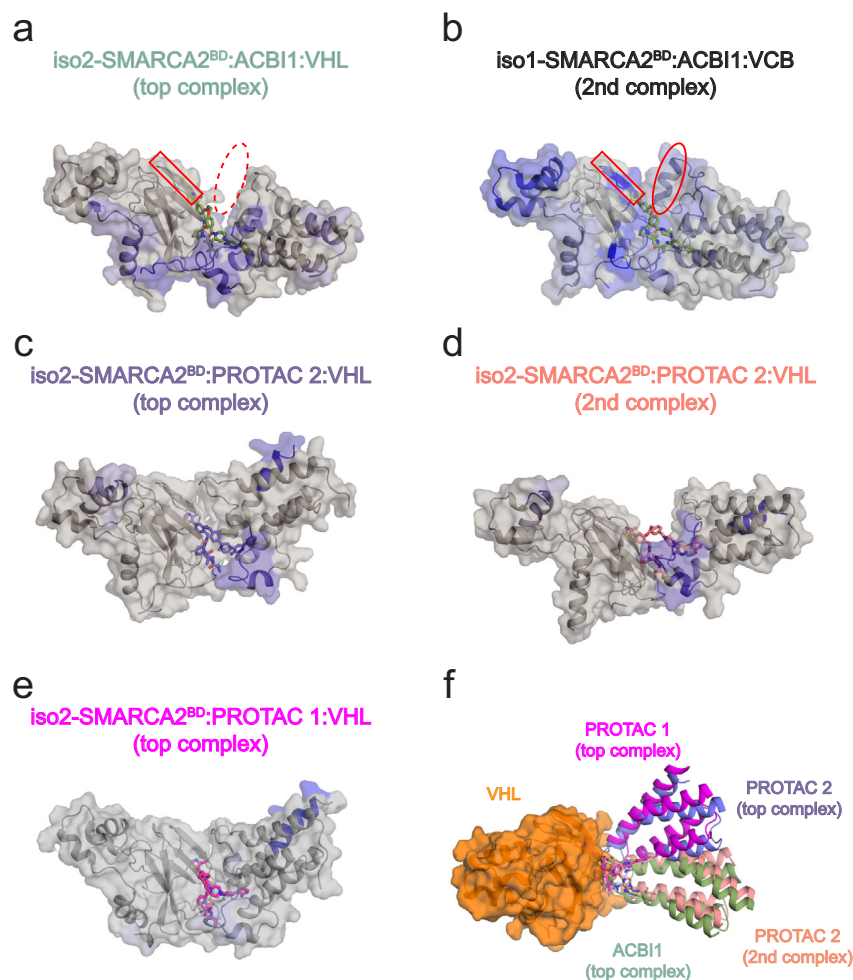


Fig. 8 | Most populated structures of SMARCA2^{BD} bound to VHL with different degrader molecules, identified by dimension reduction and clustering of HREMD simulation data. **a–e** The blue-colored regions of SMARCA2^{BD} and VHL represent HDX-MS protection in the presence of the corresponding degrader molecule relative to SMARCA2^{BD}:VHL or SMARCA2^{BD}:VCB complexes in the absence of the degrader. Representative structures from the second most populated clusters (“2nd complex”) of iso1-SMARCA2^{BD}:ACBI1:VCB (panel **b**) and iso2-SMARCA2^{BD}:PROTAC 2:VHL (panel **d**) support our HDX-MS results. In panels **a** and **b**, note that the beta sheet highlighted by a red rectangle does not show HDX-MS

protection in iso2-SMARCA2^{BD}:ACBI1:VHL, and does not contact VHL in simulations of that system. This region does show HDX-MS protection in iso1-SMARCA2^{BD}:ACBI1:VCB, and we find in simulations that it forms contacts with an alpha helix that is only present in iso1-SMARCA2^{BD} (indicated by a red oval). Note that Elongin B and Elongin C are included in the simulations associated with the structure presented in panel **b**, but omitted here for clarity. **f** The ternary complexes from panels **a**, **c**, **d**, and **e** are compared after aligning VHL (orange surface representation) to illustrate the conformational heterogeneity among highly populated structures of ternary complexes with different degraders.

The MSM uses the observed dynamics of the simulations to predict a stationary probability distribution on tICA space that is, in general, different from the empirical distribution of our simulation data. The result is shown in Figure 7b. This model is coarse-grained to obtain a five-state MSM, of which the following three states are of particular interest: the ground state I with a stationary probability of 0.63, a metastable state III with 0.10 probability, and state IV, to which the experimental crystal structure can be assigned and which has a stationary probability of 0.05 (Fig. 7c, d).

Importantly, the MSM predicts that the ternary complex crystal structure with PROTAC 2 is 1.5 kcal mol⁻¹ higher in free energy than the global free energy minimum and that they differ by an I-RMSD of 3.6 Å (Fig. 7b, e), thus lending credence to our approach of extensive conformational sampling to identify previously undetermined structures. The model further predicts a relative free energy of 2.2 kcal mol⁻¹ for the metastable state with an I-RMSD of 4.4 Å relative to the crystal structure (Fig. 7b, e). Interestingly, the SMARCA2^{BD}:PROTAC 2:VHL ternary complex structures simulated by the WE-HDX strategy described above can be well identified on this free energy

landscape too (blue points on the projection in Fig. 7b), demonstrating how the simulation of ternary complex formation yields valid conformations.

The classification into five macro-states can be attributed to structural differences at the ternary complex interface. For instance, the global minimum state is stabilized by a number of protein–protein contacts and, furthermore, contacts between PROTAC 2 and R1403, N1464, and I1470 of SMARCA2^{BD}, that are missing in the metastable state (Supplementary Fig. 24). On the other hand, contacts between VHL and PROTAC 2 are largely unchanged between the metastable and global minimum states, likely due to the tight interaction between VHL and the degrader. The area of the binding interface is substantially increased in both the metastable and global minimum states relative to the crystal structure: the global minimum state has a buried surface area of 2962 Å², compared to 2800 Å² for the metastable state and only 2369 Å² for the crystal structure. We note that these differences observed at the interfaces of distinct ternary complexes further support the adequacy of the minimum I-RMSD metric we used above to measure the prediction accuracy.

We also performed F@H simulations of iso2-SMARCA2^{BD}:PROTAC 1:VHL (900 μ s of aggregate simulation time across 1000 trajectories from 99 initial structures coming from HREMD) and iso2-SMARCA2^{BD}:ACBII:VHL (500 μ s of aggregate simulation time across 2000 trajectories from 100 initial structures coming from HREMD). These simulations were used to fit MSMs for these systems using the same procedure described above (Supplementary Figs. 25 and 26). The resultant MSMs predict that the crystal structure of the iso2-SMARCA2^{BD}:PROTAC 1:VHL system is 2.2 kcal mol⁻¹ higher than its global free energy minimum, while the crystal structure of the iso2-SMARCA2^{BD}:ACBII:VHL system is only 0.7 kcal mol⁻¹ higher in energy than its ground state. Coarse-graining the PROTAC 1 model yields a two-state MSM, while a three-state MSM is obtained for the ACBII system. In both cases, the crystal structure falls into the most probable macro-state. Interestingly, in the predicted ground state of the ternary complex with PROTAC 1, SMARCA2^{BD} is oriented relative to VHL (Supplementary Fig. 25d, e) in a similar fashion as in the predicted ground state with PROTAC 2 (Fig. 7e), while in the ground state of the ACBII system, the position of SMARCA2^{BD} relative to VHL (Supplementary Fig. 26d, e) is more similar to that in the crystal structure, which, as described above, is comparable among all three ternary complexes. This illustrates that notable conformational changes can be induced by different degrader molecules.

Interestingly, the simulations of ternary complexes of iso2-SMARCA2^{BD} and VHL mediated by the three degraders confirm important interactions between charged residues at the SMARCA2^{BD}:VHL interface that were suggested by the HDX-MS experiments presented above. In particular, R60 on VHL, which is experimentally found to be protected for a longer duration in the ternary complex, preferentially forms contacts with E1420 on the SMARCA2^{BD} interface (Supplementary Fig. 34) for ACBII and PROTAC 2 but not for PROTAC

1. ACBII also induces contacts between K1416 of SMARCA2^{BD} and N90/D92 of VHL, which are notably reduced in the presence of PROTAC 1 and PROTAC 2. This consistent observation in both experiment and simulation may contribute to the stronger cooperativity observed for ACBII compared to PROTAC 1 and PROTAC 2.

The millisecond-long simulations presented here are, to the best of our knowledge, the most extensive sampling of ternary degrader complexes to date, permitting examination of their free energy landscapes in detail. Remarkably, these simulations capture key structural determinants observed experimentally, such as HDX-MS residue protection and ternary complex stability and, furthermore, reveal structural differences between the energetically most favorable states of SMARCA2^{BD}:VHL induced by different degraders that may contribute to cooperativity.

Large-scale simulations yield accurate ubiquitination predictions

In addition to simulating the ternary complex formation and associated dynamics, a more complete understanding of the ubiquitination process should involve the full Cullin-RING E3 ubiquitin ligase (CRL). To this end, we probe the different ternary degrader complexes in the context of the full CRL macromolecular assembly by examining the separation of different solvent-exposed POI lysine residues from the ubiquitination zone of the CRL³⁷ (Fig. 9a), specifically focusing on the probability of POI lysine residue density within this zone. The hypothesis is that the ubiquitination rate depends on the probability of finding a lysine residue in the ubiquitination zone. As such, this analysis can provide insight into the degradation potency of degrader molecules. First, we build an entire E2-E3 complex for CRL-VHL in its activated form using a recently obtained structure of the active form of the closely related CRL- β TrCP as reference³⁸ (see the “Methods” section “Cullin-RING E3 ubiquitin

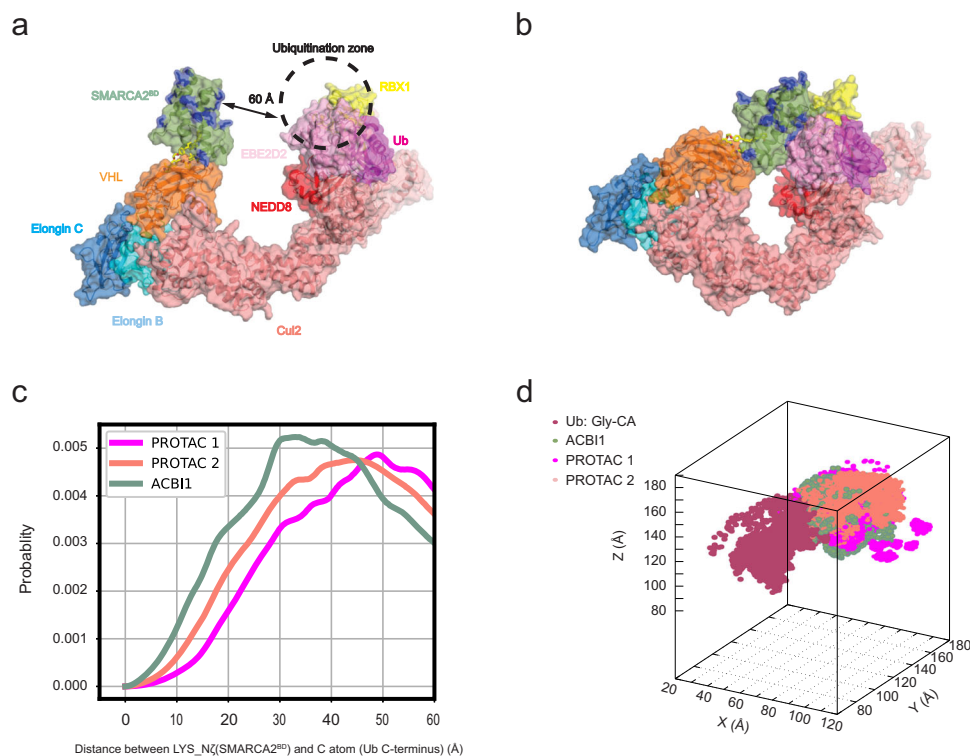


Fig. 9 | Degradable-dependent SMARCA2^{BD} lysine densities in the CRL-VHL ubiquitination zone. **a** Active form of CRL-VHL with bound SMARCA2^{BD} and E2-ubiquitin in the open CRL conformation. Lysine residues on SMARCA2^{BD} are highlighted in blue. **b** Same as panel **a** with a closed conformation of CRL generated by meta-eABF simulations. **c** Probability of distances of SMARCA2^{BD} lysine residues

(side-chain nitrogen atom) to the C-terminal glycine C atom of ubiquitin derived from ternary complex HREMD simulations with the three different degraders PROTAC 1, PROTAC 2, and ACBII. **d** Density of lysine residues in 3D space near the ubiquitination zone of CRL-VHL. Source data are provided as a Source Data file.

ligase (CRL) simulations to explore activation” for more details). Second, we use the meta-eABF simulation approach, (see the “Methods” section “Meta-eABF simulations on full Cullin-RING E3 ubiquitin ligases (CRL) complex” for more details), to sample CRL open-closed conformations in the presence of SMARCA2^{BD} (Fig. 9b). These conformations are then used as reference states to superimpose structures from HREMD simulations of ternary SMARCA2^{BD}:VHL complexes on the active state of the CRL-VHL, allowing us to obtain lysine densities from SMARCA2^{BD} in the ubiquitination zone of the CRL-VHL.

Comparing the lysine densities of the three degraders (Fig. 9c), we observe that ACBII places the most lysine density in the ubiquitination zone of CRL-VHL, followed by PROTAC 2 and PROTAC 1. This order of lysine density in the ubiquitination zone agrees with the experimentally observed degradation data between ACBII and PROTAC 1¹⁸, and also places PROTAC 2 between these two, thus establishing a procedure to qualitatively predict the ubiquitination likelihood of the target protein.

Interestingly, applying the same analysis on the three SMARCA2^{BD}:degrader:VHL crystal structures (Supplementary Fig. 35), the distribution of lysine distances from the ubiquitination zone is indistinguishable among PROTAC 1, PROTAC 2, and ACBII up to relatively large separations (30 Å). This strong contrast between the set of distributions in Fig. 9c and in Supplementary Fig. 35 further emphasizes the need for long-timescale simulations, as performed in the current study, to reach reliable conclusions on degrader-induced ternary complex structure and function.

To experimentally validate degrader-induced changes in global protein and ubiquitination levels, we treated HeLa cells with 300 nM of ACBII for 1 h, followed by global mass spectrometry-based proteome and ubiquitinomics analysis. In total, we quantified >12,500 ubiquitination sites on 5300 proteins (see Supplementary Data 2). As expected, our results confirm ACBII-induced degradation of the SMARCA2 protein (Fig. 10a). The loss of SMARCA2 protein abundance was rescued by

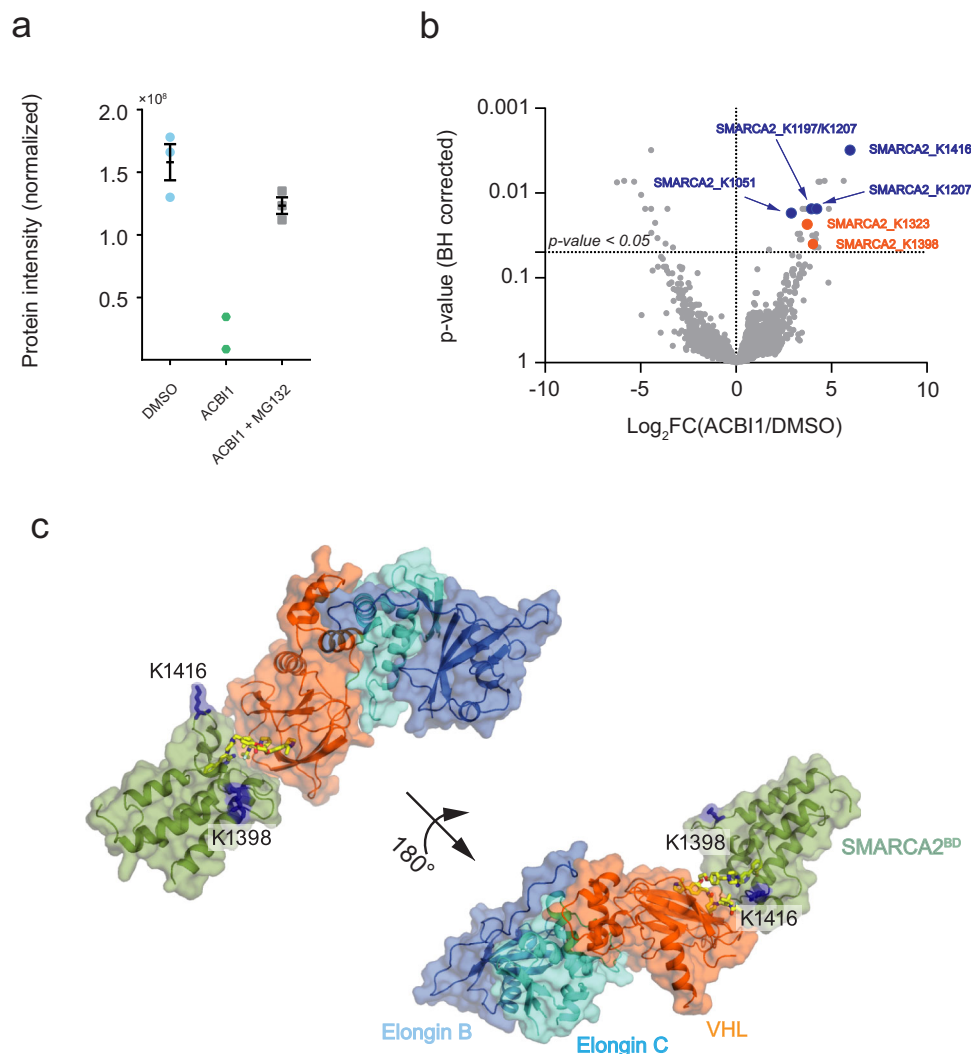


Fig. 10 | Changes in ubiquitination levels on the proteome of HeLa cells upon treatment with ACBII at 300 nM for 1 h. a Change in SMARCA2 protein abundance upon treatment with DMSO (cyan circles), ACBII (green hexagons), and ACBII + MG132 (gray squares). The ACBII treatment significantly decreases the SMARCA2 protein abundance compared to the DMSO alone and, upon co-treatment with the proteasomal inhibitor (MG132), the abundance is rescued to levels almost similar to the DMSO alone. The data ($n = 3$; 2 for ACBII-only) are visualized as the arithmetic mean \pm one standard error (for $n > 2$). Source data are provided as a Source Data file. **b** Distribution of changes in ubiquitination levels plotted as Log_2 fold change in ACBII versus DMSO control against the Benjamini-

Hochberg corrected p-value for all quantified ubiquitinated sites ($n = 12,569$) from triplicate measurements. The SMARCA2 sites with significant changes in ubiquitination levels ($p\text{-value} < 0.05$ and $\text{Log}_2\text{FC}(\text{ACBII}/\text{DMSO}) \geq 1$) are marked. The sites unique to SMARCA2 are marked as solid orange circles and SMARCA2/4 shared sites are shown as solid blue circles. A pairwise t-test, with Benjamini-Hochberg multiple comparison, was used for comparing two conditions in triplicate measurements. Source data are provided as Supplementary Data 2. **c** Location of the two SMARCA2^{BD} lysine residues K1398 and K1416 (shown in blue stick representation) on the SMARCA2^{BD}:ACBII:VHL crystal structure.

co-treatment with 1 μM proteasomal inhibitor MG132 that impedes the targeted degradation. In addition, global ubiquitination profiling identified several SMARCA2 lysine sites, some of which show a statistically significant increase in ubiquitination levels after the ACBII treatment, compared with the vehicle control DMSO (Fig. 10b, Supplementary Table 14, Supplementary Data 2). Ubiquitinated lysine residues were detected both on (e.g., K1398 and K1416) and outside the bromodomain (e.g., K1101, K1197/K1207, K1323), with the most significantly ubiquitinated residue (K1416) located on the SMARCA2^{BD} (Fig. 10c). Not all of the lysine residues from bromodomain can be detected in this experiment due to the repeated occurrence of lysine and arginine residues within short intervals, hence the cleaved peptide is too small to be detected by the mass spectrometer. However, among those detected, the general trend is in agreement with the above-mentioned prediction that ACBII tends to position lysine residues closer to ubiquitin (Supplementary Fig. 36). Our results are in agreement with recent data from Arvinas and Genentech showing that a potent and selective SMARCA2 degrader is most significantly inducing ubiquitination of a lysine residue on the bromodomain of SMARCA2, although other lysine residues are ubiquitinated to lesser degrees³⁹. These results further validate the hypothesis that degraders like ACBII directly influence ubiquitination of lysine residues in the ubiquitination zone of CRL-VHL by modulating their global proximity to ubiquitin.

Discussion

The formation of a ternary complex is a critical step in targeted protein degradation. However, accurately predicting the structural ensemble of the ternary complex is challenging due to the size of the multi-protein system, the inherent conformational flexibility associated with forming non-native protein-protein interactions, the relevant timescales for biological motions, and the limited data associated with the solution-phase dynamics of ternary structures. The ability to accurately predict the formation of degrader-induced ternary complexes and the corresponding structural ensembles would provide a better understanding of TPD and enable more precise optimization of degrader molecules (e.g. linker length, composition, and attachment points).

Here, we studied three different degrader molecules in complex with SMARCA2^{BD} and VHL that have similar thermodynamic binding profiles and protein-protein interactions observed in the crystal structures but different degradation efficiencies. The crystal structure determined in this work of ACBII complexed with SMARCA2^{BD} and VHL (PDB ID: 7S4E) reveals a similar conformation to previously published and close degrader analogs: PROTAC 1 (PDB ID: 6HAY) and PROTAC 2 (PDB ID: 6HAX). The similar binding thermodynamics and crystal structure complexes, yet different degradation efficiencies, motivated our work to explore the dynamic nature of the ternary structure, which might be the source of the differing degradation efficiencies (although other factors such as permeability may also play a role). The approach we describe here combines MD simulations with solution-phase biophysical experiments to produce dynamic ternary structure predictions that could be helpful in elucidating the characteristics that impact binding cooperativity and degradation efficiency.

We apply enhanced Hamiltonian replica exchange molecular dynamics (HREMD) simulations, validated by experimental SAXS data, to derive heterogeneous ensembles of ternary complex conformations that constitute the basis for millisecond-long MD simulations on Folding@home. Detailed free energy landscapes predict that the experimental crystal structures are -1 – 2 kcal mol⁻¹ higher in free energy than the lowest energy (most favorable) conformations, confirming that they are snapshots in low free energy basins, but not the global minima of those basins. Simulation global minima reveal notable differences in the orientation between SMARCA2^{BD} and VHL induced by the three degraders.

To put the simulation results in a larger context, we examine the likelihood of ubiquitination for specific SMARCA2^{BD}:VHL degrader

ternary complexes by deploying the entire Cullin-RING E3 ubiquitin ligase (CRL). The orientation of SMARCA2^{BD} with respect to the CRL changes dramatically in these global minima from simulation compared to the crystal structures: in particular, we find that ACBII positions lysines of SMARCA2^{BD} closer on average to ubiquitin in the E2 ligase than does PROTAC 1, with PROTAC 2 shifting between an ACBII-like position and a PROTAC 1-like position, suggesting that ACBII has the highest propensity to facilitate ubiquitination of SMARCA2^{BD}. We employed proteomics and ubiquitinomics experiments to determine ubiquitinated lysine residues for SMARCA2 in HeLa cells. The results confirm the hypothesis that ACBII positions several lysine residues in closer proximity of E2-ubiquitin, enhancing ubiquitination probability, and hence, degradation efficiency. For example, we predict that K1416 in SMARCA2^{BD} is most likely to be ubiquitinated, which is also the case in the ubiquitinomics experiment.

HDX-MS experiments revealed charged interface residues that are protected and yet are not in contact in the crystal structures. Our long-timescale ternary complex simulations revealed that several of these residues form contacts in the ternary complex ensembles. Some of these are common to all three degraders, whereas some are absent from PROTAC 1 (e.g., VHL:R60 and SMARCA2^{BD}:E1420). These contacts may underlie the differences in cooperativity, placement in the ubiquitination zone, and ultimately degradation.

We developed a protocol that incorporates information about protected residues as contact collective variables in weighted ensemble simulations that seek to form interactions among the protected residues and to bind the warhead portion of the degrader to SMARCA2^{BD} (WE-HDX). This method reliably produced ternary complex structures that were in low free energy basins of the ternary complex landscape, and that were similar to the conformations accessible when starting simulations from the crystal structure. This method also provides estimates of the k_{on} for ternary complex formation. We compared WE-HDX to docking using HDX-MS protected residues as constraints. We find that HDX constraints improve the quality of ternary complexes for docking; yet WE-HDX is more accurate than docking using HDX constraints. Further usage of the HDX-MS data could be done by computing HDX-MS observables from simulation, and then reweighting the ternary complex landscape accordingly. Even though many models are proposed in the literature, we did not estimate that those models would give us accurate reweighting at this point in time, although clearly this would be a fruitful avenue for future research.

Our integrative approach provides a richer understanding of the dynamics of ternary complex ensembles, which could improve the design of degrader molecules for new systems of interest. For the three degraders studied here, the global minima from HREMD and F@H simulations showed that the orientation of SMARCA2 lysines with respect to the E2-loaded ubiquitin is a discriminating feature, particularly of ACBII/PROTAC 2 with respect to PROTAC 1, suggesting this to be critical for a productive ternary complex. From the conformational landscape we also find that the stability of the ternary complex differs among these 3: PROTAC 1 and ACBII are more stable than PROTAC 2; however PROTAC 1 is in a non-productive configuration. Thus, the stability of the ternary complex induced by ACBII might distinguish it from PROTAC 2. Furthermore, we consider the conformational free energy penalty for the degrader to go from its conformation in solution to ternary complex, and again we find that this penalty is larger for PROTAC 1 than it is for PROTAC 2 and ACBII. We also found protected charged residues from HDX-MS that, while not in contact in the crystal structures, appear in simulations such as VHL:R69 and SMARCA2^{BD}:E1420, giving clues to potential structural determinants of cooperativity. And we found that ACBII had the highest ubiquitination probability based on our CRL modeling, followed by PROTAC 2 and PROTAC 1, which was confirmed by ubiquitinomics experiments presented here.

The methodologies described here rely on advanced physics-based simulations and solution-phase biophysical experiments. Since this approach is based on physical principles without the need for training data, we expect it to be transferable to other POI-ligase ternary complexes with induced proximity degrader molecules, and possibly to other induced proximity systems (e.g. phosphorylation, methylation, and acetylation). Efforts in our group are underway to expand the application to more potential degraders in the SMARCA2^{BD}:VHL system and to other POI-ligase combinations. We have used the simulation methods outlined here in a prospective manner as follows: we have predicted ternary complex ensembles of potential heterobifunctional degraders using WE-HDX; used HREMD and F@H simulations out of HREMD to select the lowest free energy structures; then calculated the ubiquitination probability of these structures by modeling them in the full CRL. We have then optimized for short and rigid linkers against the ternary complex structures selected for preferential ubiquitination, also using the conformational free energy penalty for the degrader to go from solution to the ternary complex as a design objective. Based on that we have selected the heterobifunctional molecules that optimize these properties. We expect to report on this larger data set in a future publication.

Methods

Cloning, expression, and purification of SMARCA2^{BD} and VHL/Elongin B/C

The bromodomain of the *SMARCA2* gene from *Homo sapiens* (UniProt accession number P51531-1; residues 1373-1511) was custom-synthesized at Genscript with N-terminal GST tag¹⁸ and thrombin protease cleavage site. The synthetic gene comprising *SMARCA2* was cloned into pET28 vector to create plasmid pL-477. The second construct of *SMARCA2* with deletion 1400-1417 (UniProt accession number P51531-2) was created as pL-478. For biotinylated SMARCA2^{BD}, AVI-tag was gene-synthesized at C-terminus of pL-478 to create pL-479. The *VHL* gene from *Homo sapiens* was custom-synthesized with N-terminal His6 tag¹⁸ and thrombin protease cleavage site. The synthetic gene comprising the *VHL* (UniProt accession number P40337; residues 54-213) was cloned into pET28 vector to create plasmid pL-476. *EloB* and *EloC* genes from *Homo sapiens* were custom-synthesized with AVI-tag at the C-terminus of Elongin B¹⁷. The synthetic genes comprising the *EloB* (UniProt accession number Q15370; residues 1-104) and *EloC* (UniProt accession number Q15369; residues 17-112) were cloned into pCDFDuet vector to create plasmid pL-474. For protein structural study, AVI-tag was deleted in pL-474 to create pL-524.

For SMARCA2^{BD} protein expression, the plasmid was transformed into BL21(DE3; One Shot™, catalog number: C600003) chemically competent *E. coli* cells and plated on Luria-Bertani (LB) medium containing 50 µg/ml kanamycin at 37 °C overnight. A single colony of BL21(DE3)/pL-477 or BL21(DE3)/pL-478 was inoculated into a 100-ml culture of LB containing 50 µg/ml kanamycin and grown overnight at 37 °C. The overnight culture was diluted to OD₆₀₀ = 0.1 in 2 × 1-liter of Terrific Broth medium containing 50 µg/ml kanamycin and grown at 37 °C with aeration to mid-logarithmic phase (OD₆₀₀ = 1). The culture was incubated on ice for 30 min and transferred to 16 °C. IPTG was then added to a final concentration in each culture of 0.3 mM. After overnight induction at 16 °C, the cells were harvested by centrifugation at 5000 × g for 15 min at 4 °C. The frozen cell paste from 2 L of cell culture was suspended in 50 ml of Buffer A consisting of 50 mM HEPES (pH 7.5), 0.5 M NaCl, 5 mM DTT, 5% (v/v) glycerol, supplemented with 1 protease inhibitor cocktail tablet (Roche Molecular Biochemical) per 50 ml buffer. Cells were disrupted by Avestin C3 at 20,000 psi twice at 4 °C, and the crude extract was centrifuged at 39,000 × g (JA-17 rotor, Beckman-Coulter) for 30 min at 4 °C. Two ml Glutathione Sepharose 4 B (Cytiva) was added into the supernatant and mixed at 4 °C for 1 h, washed with Buffer A and eluted with 20 mM reduced glutathione (Sigma). The protein concentration was measured by Bradford assay, and GST-tag

was cleaved by thrombin (1:100) at 4 °C overnight during dialysis against 1 L of Buffer B (20 mM HEPES, pH 7.5, 150 mM NaCl, 1 mM DTT). The sample was concentrated to 3 ml and applied at a flow rate of 1.0 ml/min to a 120-ml Superdex 75 (HR 16/60) (Cytiva) pre-equilibrated with Buffer B. The fractions containing SMARCA2^{BD} were pooled and concentrated by Amicon® Ultracel-3K (Millipore). The protein concentration was determined by OD₂₈₀ and characterized by SDS-PAGE analysis and analytical LC-MS. The protein was stored at -80 °C.

For VHL/Elongin B/C protein expression, the plasmids were co-transformed into BL21(DE3) and plated on Luria-Bertani (LB) medium containing 50 µg/ml kanamycin and 50 µg/ml streptomycin at 37 °C overnight. A single colony of BL21(DE3)/pL-476/474 or BL21(DE3)/pL-476/524 was inoculated into a 100-ml culture of LB containing 50 µg/ml kanamycin and 50 µg/ml streptomycin and grown overnight at 37 °C. The overnight culture was diluted to OD₆₀₀ = 0.1 in 6 × 1-liter of Terrific Broth medium containing 50 µg/ml kanamycin and 50 µg/ml streptomycin and grown at 37 °C with aeration to mid-logarithmic phase (OD₆₀₀ = 1). The culture was incubated on ice for 30 min and transferred to 18 °C. IPTG was then added to a final concentration of 0.3 mM in each culture. After overnight induction at 18 °C, the cells were harvested by centrifugation at 5000 × g for 15 min at 4 °C. The frozen cell paste from 6 L of cell culture was suspended in 150 ml of Buffer C consisting of 50 mM HEPES (pH 7.5), 0.5 M NaCl, 10 mM imidazole, 1 mM TCEP, 5% (v/v) glycerol, supplemented with 1 protease inhibitor cocktail tablet (Roche Molecular Biochemical) per 50 ml buffer. Cells were disrupted by Avestin C3 at 20,000 psi twice at 4 °C, and the crude extract was centrifuged at 17,000 × g (JA-17 rotor, Beckman-Coulter) for 30 min at 4 °C. Ten ml Ni Sepharose 6 FastFlow (Cytiva) was added into the supernatant and mixed at 4 °C for 1 h, washed with Buffer C containing 25 mM imidazole and eluted with 300 mM imidazole. The protein concentration was measured by Bradford assay. For protein crystallization, His-tag was cleaved by thrombin (1:100) at 4 °C overnight during dialysis against 1 L of Buffer D (20 mM HEPES, pH 7.5, 150 mM NaCl, 1 mM DTT). The sample was concentrated to 3 ml and applied at a flow rate of 1.0 ml/min to a 120-ml Superdex 75 (HR 16/60) (Cytiva) pre-equilibrated with Buffer D. The fractions containing VHL/Elongin B/C were pooled and concentrated by Amicon® Ultracel-10K (Millipore). The protein concentration was determined by OD₂₈₀ and characterized by SDS-PAGE analysis and analytical LC-MS. The protein was stored at -80 °C. For SPR assay, 10 mg VHL/Elongin B/C protein complex was incubated with BirA (1:20), 1 mM ATP and 0.5 mM Biotin and 10 mM MgCl₂ at 4 °C overnight, removed free ATP and Biotin by 120-ml Superdex 75 (HR 16/60) with the same procedure as above, and confirmed the biotinylation by LC/MS.

X-ray structure determination of iso2-SMARCA2^{BD}:ACBII:VCB complex

Purified SMARCA2 and VCB in 50 mM HEPES, pH 7.5, 150 mM NaCl, 1 mM DTT were incubated in a 1:1:1 molar ratio with ACBII for 1 h at room temperature. Incubated complex was subsequently injected on to a Superdex 10/300 GL increase (Cytiva) pre-incubated with 50 mM HEPES, pH 7.5, 150 mM NaCl, 1 mM DTT, 2% DMSO at a rate of 0.5 ml/min to separate any noncomplexed partners from the properly formed ternary complex. Eluted fractions corresponding to the full ternary complex were gathered and spun concentrated to 14.5 mg/mL using an Amicon Ultrafree 10K NMWL Membrane Concentrator (Millipore). Crystals were grown 1-3 µL hanging drops by varying the ratio of protein to mother liquor from 0.5-2:0.5-2 respectively. Crystals were obtained in buffer consisting of 0.1 M HEPES, pH 7.85, 13% PEG 3350, 0.2 M sodium formate incubated at 4 °C. Crystals grew within the first 24 h but remained at 4 °C for 5 days until they were harvested, cryo-protected in an equivalent buffer containing 20% glycerol and snap frozen in LN₂. Diffraction data were collected at NSLS2 beamline FMX (λ = 0.97932 Å) using an Eiger X 9M detector. Crystals were found to be in the P 21 21 21 space group with unit cell dimensions of a = 80.14 Å,

$b = 116.57 \text{ \AA}$, $c = 122.23 \text{ \AA}$, where $\alpha = \beta = \gamma = 90^\circ$. Crystal contained two copies of the SMARCA2:ACBII:VCB (VHL, EloC, EloB) complex within the asymmetric unit cell. The structure was solved by performing molecular replacement with CCP4i243 PHASER using PDB ID 6HAX as the replacement model. MR was followed by iterative rounds of modeling (COOT44) and refinement (REFMAC545-53) by standard methods also within the CCP4i2 suite. Structures were refined to $R_{\text{work}}/R_{\text{free}}$ of 23.7%/27.5%.

Hydrogen-deuterium exchange mass spectrometry

HDX experiments were performed using a protein stock at the initial concentration of 200 μM of SMARCA2^{BD}, VCB in the APO, binary (200 μM ACBII), and ternary (200 μM PROTAC ACBII) states in 50 mM HEPES, pH 7.4, 150 mM NaCl, 1 mM TCEP, 2% DMSO in H₂O. With the knowledge of binding constants for each of the three degraders, the assays were designed to optimize the complex formation of 80% or greater in the D₂O labeling solution after the 1:13 dilution (94% ACBII, 93% PROTAC 2, 89% PROTAC 1) to obtain maximal exchange of the ternary complexes. Maximizing complex formation in solution ensures that the ratio of liganded to free protein in solution does not complicate the downstream analysis⁴⁰. The protein samples were injected into the nanoACQUITY system equipped with HDX technology for UPLC separation (Waters Corp.⁴¹) to generate mapping experiments used to assess sequence coverage. Generated maps were used for all subsequent exchange experiments. HDX was performed by diluting the initial 200 μM protein stock 13-fold with D₂O (Cambridge Isotopes) containing buffer (10 mM phosphate, pD 7.4, 150 mM NaCl) and incubated at 10 °C for various time points (0.5, 5, 30 min). At the designated time point, an aliquot from the exchanging experiment was sampled and diluted 1:1 into D₂O quenching buffer containing (100 mM phosphate, pH 2.1, 50 mM NaCl, 3M GuHCl) at 1 °C. The process was repeated at all time points, including for non-deuterated samples in H₂O-containing buffers. Quenched samples were injected into a 5- μm BEH 2.1 \times 30-mm Enzymate-immobilized pepsin column (Waters Corp.) at 100 $\mu\text{l}/\text{min}$ in 0.1% formic acid at 10 °C and then incubated for 4.5 min for on-column digestion. Peptides were collected at 0 °C on a C18 VanGuard trap column (1.7 μm \times 30 mm) (Waters Corp.) for desalting with 0.1% formic acid in H₂O and then subsequently separated with an in-line 1.8 μm Hss T3 C18 2.1 \times 30-mm nanoACQUITY UPLC column (Waters Corp.) for a 10-min gradient ranging from 0.1% formic acid to acetonitrile (7 min, 5–35%; 1 min, 35–85%; 2 min hold 85% acetonitrile) at 40 $\mu\text{l}/\text{min}$ at 0 °C. Fragments were mass-analyzed using the Synapt G2Si ESL-Q-ToF mass spectrometer (Waters Corp.). Between injections, a pepsin-wash step was performed to minimize peptide carryover. Mass and collision-induced dissociation in data-independent acquisition mode (MSE) and ProteinLynx Global Server (PLGS) version 3.0 software (Waters Corp.) were used to identify the peptides in the non-deuterated mapping experiments and analyzed in the same fashion as HDX experiments. Mapping experiments generated from PLGS were imported into the DynamX version 3.0 (Waters Corp.) with quality thresholds of MSI signal intensity of 5000, maximum sequence length of 25 amino acids, minimum products 2.0, minimum products per amino acid of 0.3, minimum PLGS score of 6.0. Automated results were inspected manually to ensure the corresponding m/z and isotopic distributions at various charge states were assigned to the corresponding peptides in all proteins (SMARCA2^{BD}, VHL, Elongin C, Elongin B). All raw relative uptake plots of the deuterium exchange for each state and experiment can be found in Supplementary Data 1. DynamX was utilized to generate the relative deuterium incorporation plots and HDX heat map for each peptide (Supplementary Fig. 39) of each protein within the complex and stable deuterium exchange (Supplementary Figs. 40–43). The relative deuterium uptake of common peptides was determined by subtracting the weighted-average mass of the centroid of the non-deuterated control samples from the deuterated samples at each time

point. Time points in the middle of the initial incorporation rate were chosen to avoid fully exchanged states and thus increase the significance of the measurements and protected residue determination. All experiments were made under the same experimental conditions negating the need for back-exchange calculations but therefore are reported as relative²⁴. All HDX experiments were performed twice, on 2 separate days, and a 98 and 95% confidence limit of uncertainty was applied to calculate the mean relative deuterium uptake of each data set. Differences in deuterium uptake that exceeded the error of the datasets were considered significant⁴².

SEC-SAXS experiments

SAXS data were collected with an AKTAmicro (GE Healthcare) FPLC coupled to a BioXolver L SAXS system (Xenocs) that utilized an Excillum MetalJet D2 + X-ray source operating at a wavelength of 1.34 \AA . We measured two protein complex samples,

- (i) iso1-SMARCA2^{BD}:ACBII:VCB, and
- (ii) iso2-SMARCA2^{BD}:ACBII:VCB.

The scattering data was detected on PILATUS3 300 K (Dectris) detector with a resulting q range of 0.0134–0.5793 \AA^{-1} . To ensure the resulting scattering profile is solely due to complexes with all four protein chains and a degrader, and devoid of contributions from binary or uncomplexed proteins, size exclusion chromatography is coupled to SAXS (SEC-SAXS). The elution peak 1 of the SEC profile is assigned to the ternary complexes, whereas peak 2 is attributed to binary or uncomplexed proteins, respectively (Supplementary Fig. 15). The SEC-SAXS data for each sample was collected by loading 500 μL of the ternary complex formed by the addition of equimolar concentrations (275 μM) of SMARCA2^{BD}, VCB and ACBII, onto a Superdex 200 Increase 10/30 equilibrated with 20 mM HEPES pH 7.5, 150 mM NaCl and 1 mM DTT at 20 °C. The solution scattering data was collected as a continuous 60 s data-frame measurements with a flow rate of 0.05 mL/min. The average scattering profile of all frames within the elution peak 1 was calculated and subtracted from the average buffer scattering to yield the scattering data of the protein complex. The final SAXS profile of each ternary complex (Fig. 6a) was determined from the average scattering signal from the sample in the elution peak 1, where the relatively large variability in the calculated radius of gyration, R_g (red solid/open circles in Supplementary Fig. 15) indicates that complexes are dynamic or flexible. Data reduction was performed using the BioXTAS RAW 2.0.3 software⁴³. The uncertainty in the experimental scattering intensity is described by the standard deviation of the detection signal at a given q

$$\sigma(q) = \frac{1}{N(q)} \sqrt{\sum_{i=1}^{N(q)} [n_i - I(q)]^2}, \quad (1)$$

where $I(q)$ is the measured SAXS intensity for a given q , the magnitude of scattering vector given by $q = 4\pi \sin \theta / \lambda$ with 2θ being the scattering angle and λ the wavelength of the incident beam. n_i is the count of detected signals per pixel i in the raw SAXS image and $N(q)$ are all pixels belonging to the same q bin.

R_g was estimated from an experimental SAXS curve using the Guinier approximation,

$$I(q) \approx I(0)e^{-\frac{q^2 R_g^2}{3}}, \text{ for } q \rightarrow 0 \quad (2)$$

where $I(0)$ is the forward scattering intensity at $q = 0$. The linear region in $\ln(I(q))$ vs. q^2 was fitted at low- q values such that $q_{\text{max}} * R_g \leq 1.3$ to estimate R_g , where q_{max} is the maximum q -value in the Guinier fit (Supplementary Fig. 16). On the other hand, R_g of the protein complex

in simulation was directly calculated from atomic coordinates using following relation,

$$R_g = \sqrt{\frac{\sum_i m_i \|\mathbf{r}_i\|^2}{\sum_i m_i}} \quad (3)$$

where m_i is the mass of i^{th} atom and \mathbf{r}_i is the position of i^{th} atom with respect to the center of mass of the molecule.

Molecular dynamics simulations

The initial coordinates of the systems were obtained from X-ray crystal structures with PDB IDs 6HAX, 6HAY, and 7S4E, respectively. The missing atoms were added using the LEaP module in AMBER20. The AMBER ff14SB force field⁴⁴ was employed for the protein and the degrader force field parameters were generated using in-house programs for all MD simulations in this study. The explicit solvent was modeled using TIP3P water encapsulating the solute in a rectangular box. Counter ions were added to the system to enforce neutrality. Langevin dynamics were used to maintain the temperature at 300 K and the collision frequency was set to 2.0 ps⁻¹. The SHAKE algorithm was utilized so that a 2 fs time step could be achieved.

A step-wise equilibration protocol was used prior to running the production phase of the MD simulations. First, a minimization was performed with a positional restraint of 5 kcal mol⁻¹ Å⁻² applied to all solute heavy atoms followed by a fully unrestrained minimization. Each minimization was composed of 500 steps of the steepest descent followed by 2000 steps of conjugate gradient. Using 5 kcal mol⁻¹ Å⁻² positional restraint on the heavy atoms of the solute, the system was linearly heated from 50 to 300 K for a duration of 500 ps (NVT ensemble) followed by a density equilibration of 750 ps (NPT ensemble). Over the course of five 250 ps simulations, the restraints on the heavy atoms of the systems were reduced from 5 to 0.1 kcal mol⁻¹ Å⁻². Then, a 500 ps simulation was run with a positional restraint of 0.1 kcal mol⁻¹ Å⁻² on the backbone atoms followed by a fully unrestrained 5 ns simulation.

Three independent regular MD simulations were performed for each of the three bound degrader complexes for up 1 μs. Structures obtained from these simulations were clustered into 25 groups based on interface residue distances. One representative structure from each cluster (along with the experimentally obtained crystal structure) were used as the set of reference ternary complexes for the evaluation of bound complex predictions by WE simulations or docking.

All simulation data were analyzed by in-house analysis scripts and visualized/analyzed by the VMD (<https://www.ks.uiuc.edu/Research/vmd/vmd-1.9.3/>) and PyMol (<https://pymol.org/2/>) software packages.

Isoform 1 homology model

Since no suitable X-ray structure for iso1-SMARCA2^{BD} is available in the PDB, we have used the YASARA (“Yet Another Scientific Artificial Reality Application”) homology modeling module (YASARA Biosciences GmbH) to build a high-resolution model of iso1-SMARCA2^{BD} based on its amino acid sequence. The sequence that was used is Uniprot P51531-1 (residues 1373–1493), which has an additional 17 aa loop compared to P51531-2 (missing loop at 1400–1417). As a template for homology modeling, we used the structure from the PDB ID 6HAY. Once the model was completed, an AMBER minimization, which restrained all heavy atoms except the loop residues, was run. This ensured that the residues in the loop do not overlap and assume a stable secondary structure conformation. Minimization did not show major side-chain movements in the final minimized output which further suggested that the structure was stable.

WE-HDX simulations

WE-HDX simulations of the formation of ternary complexes were run with both bin-less and binned WE variants (see Supplementary Note 1).

These binding simulations were run with iso2-SMARCA2^{BD} and the degrader-bound VHL subunit. The Elongin C and Elongin B subunits were omitted in these path-sampling simulations as the process of ternary complex formation is mainly determined by interactions at the SMARCA2^{BD}:degrader:VHL interface.

Initially, the ternary complexes were unbound manually by separating the corresponding VHL-degrader complex from SMARCA2^{BD} by 20–40 Å (depending on the system). The (rectangular) simulation boxes of these unbound systems were then solvated with explicit water molecules and counter ions were added to neutralize their net charge. The PROTAC 1 system had 21,191 water molecules and 10 chlorine ions. The PROTAC 2 simulations had 31,567 water atoms and 9 chlorine ions. The ACBII system had 24,093 water molecules, 9 chlorine ions. The dimensions of the simulation systems were 131 Å × 84 Å × 84 Å for the PROTAC 1 system, 144 Å × 89 Å × 91 Å for the PROTAC 2 system, and 123 Å × 76 Å × 98 Å for the ACBII system.

REVO-epsilon weighted ensemble method. As a bin-less WE variant, we applied the REVO algorithm. We will describe the application of the REVO algorithm as it pertains to this study, but a more detailed explanation can be found in previous works. The goal of the REVO resampling algorithm is to maximize the variation function defined as:

$$V = \sum_i V_i = \sum_i \sum_j \left(\frac{d_{ij}}{d_0} \right)^\alpha \phi_i \phi_j \quad (4)$$

where V_i is the trajectory (or walker) variation, d_{ij} is the distance between walkers i and j determined using a specific distance metric, d_0 is the characteristic distance used to make the distance term dimensionless, set to 0.148 for all simulations, the α is used to determine how influential the distances are to the walker variation and was set to 6 for all the simulations. The novelty terms ϕ_i and ϕ_j are defined as: $\phi_i = \log(w_i) - \log\left(\frac{p_{\min}}{100}\right)$. The minimum weight, p_{\min} , allowed during the simulation was 10⁻³⁰. Cloning was attempted for the walker with the highest variance, V_i when the weights of the resultant clones would be larger than p_{\min} , provided it is within distance ϵ of the walker with the maximal progress towards binding of the ternary complex. The two walkers selected for merging were within a distance of 2 Å and have a combined weight larger than the maximal weight allowed, p_{\max} , which was set to 0.1 for all REVO simulations. The merge pair also needed to minimize:

$$\frac{V_j w_i - V_i w_j}{w_i + w_j} \quad (5)$$

If the proposed merging and cloning operations increase the total variance of the simulation, the operations are performed and we repeat this process until the variation can no longer be increased.

Three different distance metrics were used while simulating the PROTAC 2 system: Using the warhead RMSD to the crystal structure, maximizing the contact strength (defined below) between protected residues identified by HDX data, and a linear combination of the warhead RMSD, contact strength between HDX-protected residues, and the contact strength between SMARCA2^{BD} and the degrader. The simulations for the other systems used the last distance metric exclusively. To compute the warhead RMSD distance metric, we aligned to the binding site atoms on SMARCA2^{BD}, defined as atoms that were within 8 Å of the warhead in the crystal structure. Then the RMSD was calculated between the warhead in each frame and the crystal structure. The distance between a set of walkers i and j is defined as:

$$d_{\text{warhead_rmsd}} = \left| \frac{1}{\text{RMSD}_i} - \frac{1}{\text{RMSD}_j} \right| \quad (6)$$

The contact strength is defined by determining the distances between residues. We calculate the minimum distance between the residues and use the following to determine the contact strength:

$$\text{contact strength} = \frac{1}{1 + e^{-k(r-r_0)}} \quad (7)$$

where k is the steepness of the curve, r is the minimum distance between any two residues and r_0 is the distance we want a contact strength of 0.5. We used 10 for k and 5 Å for r_0 . The total contact strength was the sum of all residue-residue contact strengths. The distance between walkers i and j was calculated by:

$$d_{\text{contacts}} = |\text{contact strength}_i - \text{contact strength}_j| \quad (8)$$

where contact strength $_i$ is the contact strength of walker i .

All REVO simulations were run using OpenMM v.7.5.0. Simulation details are as described above. The degrader-VHL interface was restrained to maintain the complex during the simulation by using a OpenMM custom centroid force defined as:

$$\text{Centroid Force} = k * (\text{dist} - \text{edist})^2 \quad (9)$$

where the dist is the distance between the center of mass of the degrader and the center of mass of VHL and the edist is the distance between the center of mass of the degrader and center of mass of VHL of the crystal structure, and k is a constant set to 2 kcal mol $^{-1}$ * Å 2 .

Binned weighted ensemble method. We also applied a variant of the WE simulation, in which the pre-defined collective variable is divided into bins, using the WESTPA software⁴⁵. Each bin may contain a number (M) of walkers, i , that carry a certain weight (w_i). The simulations were run for a relatively short time ($\tau = 50$ ps), after which walkers are either replicated, if their number per bin is $< M$, or they are merged, if there are $> M$ walkers per bin. Similar to REVO, the sum of all w_i equals 1 in any iteration, i.e., the trajectory replication and merging operations correspond to an unbiased statistical resampling of the underlying distribution⁴⁶.

The unbound systems described above were taken as the starting configuration for each binding simulation with the GPU-accelerated version of the AMBER molecular dynamics package⁴⁷. To ensure the degrader remains bound to the VHL protein during these simulations, a modest (1 kcal mol $^{-1}$ Å $^{-2}$) flat-bottom position restraint was enforced between the center of masses of the E3-ligand and protein binding site heavy atoms. All other MD simulation parameters were as described above.

M was set to 5 and two collective variables (CV1 and CV2) were defined to assess progress during the simulations or ternary complexes with each of the three degraders. CV1 was defined as the warhead-RMSD, or w-RMSD, of the degrader warhead with respect to the corresponding crystal structure of the bound complex. CV2 was a combination of two observables; it was either defined to be the number of native atom contacts between the warhead and the SMARCA2^{BD} binding interface or, if the binding sites were so distant that no contacts were formed, it was defined as the distance of the binding partners, i.e., SMARCA2^{BD} and the VHL-degrader binary complex. Contacts were counted between non-hydrogen atoms within a radius of 4.5 Å and, to ensure that CV2 is defined along one linear dimension, the contact counts were scaled by -1 . This selection of CV1 and CV2 with an appropriate binning allowed the separated binding partners to assemble, during the WE simulations, into ternary complexes that are similar to the corresponding crystal structures, which were used for w-RMSD and native contact calculations.

When augmenting the WE simulations with HDX-MS data, i.e., in the WE-HDX simulations, only the protected residues of the two

proteins, as informed by the corresponding experiments, were taken into consideration for the contact counts of CV2.

The ensemble of predicted bound structures was evaluated by comparing the distributions of minimum interface-RMSDs (I-RMSDs) with respect to the aforementioned set of reference ternary complexes, where the interface is defined by SMARCA2^{BD} and VHL residues within 10 Å. Furthermore, to obtain a subset of reliable predictions, these I-RMSD distributions only contain structures with w-RMSD < 2 Å and > 30 contacts between any residues of the two proteins or, in the case of WE-HDX simulations, between the protected residues of the two proteins.

Ternary complex docking protocol

Following the previously reported applications of molecular docking to predictions of ternary complexes (i.e., Methods 4 and 4b from Drummond et al.²⁰ as well as the approach from Bai et al.¹⁹), we assume that high fidelity structures of SMARCA2^{BD}:warhead and VHL:E3-ligand are known and available to be used in protein-protein docking. This docking of two proteins with bound degrader moieties is performed in the absence of the linker. The conformations of the linker are sampled independently with an in-house developed protocol that uses implementation of fast quantum mechanical methods, CREST⁴⁸. Differently from the docking protocols described in refs. [19,20], we make use of distance restraints derived either from the end-to-end distances of the sampled conformations of linker, or from the HDX-MS data. Thus, before running the protein-protein docking, we generate an ensemble of conformers for linkers and calculate the mean (x_0) and standard deviation (sd) for the end-to-end distance. This information is then used to set the distance restraints in the RosettaDock software⁴⁹:

$$f_1(x) = \left(\frac{x - x_0}{\text{sd}} \right)^2, \quad (10)$$

where x is the distance between a pair of atoms in a candidate docking pose (the pair of atoms is specified as the attachment points of the linker to warhead and E3-ligand).

When information about the protected residues is available from HDX-MS experiments, we used them to set up a set of additional distance restraints:

$$f_{2,i}(x) = \frac{1}{1 + \exp(-m \cdot (x - x_0))} - 0.5, \quad (11)$$

where i is the index of a protected residue, x_0 is the center of the sigmoid function and m is its slope. As above, x_0 value was set to be the mean end-to-end distance calculated over the ensemble of linker conformers. The value of m was set to be 2.0 in all the performed docking experiments. The type of RosettaDock-restraint is *SiteConstraint*, with specification of C_α atom for each protected residue and the chain-ID of partnering protein (i.e., x in Eq. (11)) is the distance of C_α atom from the partnering protein). Thus, the total restraint-term used in docking takes the form:

$$f_{\text{restr.}}(x) = w \cdot \left(f_1(x) + \sum_i f_{2,i}(x) \right), \quad (12)$$

where $w = 10$ is the weight of this additional score function term.

RosettaDock implements a Monte Carlo-based multi-scale docking algorithm that samples both rigid-body orientation and side-chain conformations. The distance-based scoring terms, Eq. (12), bias sampling towards those docking poses that are compatible with specified restraints. This limits the number of output docking structures, as only those ones that pass the Metropolis criterion with the additional term of Eq. (12) will be considered.

Once the docking poses are generated with RosettaDock, all the pre-generated conformations of the linker are structurally aligned onto each of the docking predictions¹⁹. Only those structures that satisfy the RMS-threshold value of ≤ 0.3 Å are saved as PDB files. All the docking predictions are re-ranked by the values of Rosetta Interface score (I_{sc}). The produced ternary structures are examined for clashes, minimized and submitted for further investigations with Molecular Dynamics methods. Details about running the described docking protocol can be found in the Supplementary Note 2.

HREMD simulation

The simulation box of a ternary complex was solvated with explicit water and counter ions were added to neutralize the net charge of the system. We chose the Amber ff14SB force field⁴⁴ for protein and TIP3P water model⁵⁰. For the degrader molecules, force field parameters were generated using in-house force field generator. The LINC algorithm⁵¹ was used to constrain all bonds including hydrogen atoms. The equation of motions was numerically integrated with a time step of 2 fs using the Verlet leapfrog algorithm⁵². The particle-mesh Ewald summation⁵³ with a fourth-order interpolation and a grid spacing of 1.6 Å was employed to calculate the long-range electrostatic interactions. A cutoff of 12 Å was imposed for the short-range electrostatic and Lennard-Jones interactions. The solute and solvent were coupled separately to a temperature bath of 300 K using the Velocity-rescale thermostat⁵⁴ with a relaxation time of 0.1 ps. The Parrinello-Rahman algorithm⁵⁵ with a relaxation time of 2 ps and isothermal compressibility of 4.5×10^{-5} bar⁻¹ was utilized for a pressure coupling fixed at 1 bar. We started with minimizing the energy of a system using the steepest descent algorithm. Then, the system was equilibrated at the NVT and NPT ensembles for 1 ns each. Finally, we ran the production runs in the NPT ensemble.

The details of Hamiltonian replica-exchange MD (HREMD) can be found in Supplementary Note 3, Supplementary Figs. 17 and 18 and Supplementary Table 13). For all HREMD simulations, we chose the effective temperatures, $T_0 = 300$ K and $T_{max} = 425$ K such that the Hamiltonian scaling parameter, $\lambda_0 = 1.00$ and $\lambda_{min} = 0.71$ for the lowest and the highest rank replicas, respectively. We estimated the number of replicas (n) in such a way that the average exchange probabilities (p) between neighboring replicas were in the range of 0.3 to 0.4. We used $n = 20$ and $n = 24$ for SMARCA2^{BD}:degrader:VHL and SMARCA2^{BD}:degrader:VCB respectively. Each simulation was run for 0.5 μ s/replica, and a snapshot of a complex was saved every 5 ps (total 100,001 frames per replica). Finally, we performed all the analyses on only the lowest rank replica that ran with original/unscaled Hamiltonian.

We assessed the efficiency of sampling by observing (i) the values of p , (ii) a good overlap of histograms of potential energy between adjacent replicas (Supplementary Fig. 17), and (iii) a mixing of exchange of coordinates across all the replicas (Supplementary Fig. 18). Furthermore, to show the convergence of HREMD simulation, we computed three metrics, radius of gyration of a ternary complex, center of mass (COM) distance between SMARCA2^{BD} and VHL, and heavy atom contacts within 5 Å between SMARCA2^{BD} and VHL. The distributions of these metrics are plotted with cumulative length of HREMD simulation (Supplementary Figs. 19 and 20). We noted that the distributions are similar for the last 0.3 μ s (0–0.3 μ s, 0–0.4 μ s and 0–0.5 μ s) of the lowest rank replica implying the convergence of the simulation.

MD simulation of degraders

PROTAC 1, PROTAC 2, and ACBII were solvated in a simulation box with 1002, 1207, and 3169 TIP3P water⁵⁰ molecules respectively, along with counter ions to neutralize the system. All other simulation parameters were same as described in the “Methods” section “HREMD simulation”. The production MD simulation of each degrader was run in the NPT ensemble for 1 μ s.

Conformational free energy landscape determination

In order to quantify the conformational free energy landscape, we performed dimension reduction of our simulation trajectories using principle component analysis (PCA). First, the simulation trajectories were featurized by calculating interfacial residue contact distances. Pairs of residues were identified as part of the interface if they passed within 7 Å of each other during the simulation trajectory, where the distance between two residues was defined as the distance between their C_α atoms. PCA was then used to identify the features that contributed most to the variance by diagonalizing the covariance matrix of the iso2-SMARCA2:PROTAC 2:VHL system; four PCA features were used in our analysis, chosen because these many features were needed to explain >95% of the variance.

After projecting the simulation data onto the resultant feature space, snapshots were clustered using the k -means algorithm. The number of clusters k was chosen using the elbow-method, i.e., by visually identifying the point at which the marginal effect of an additional cluster was significantly reduced. In cases where no elbow could be unambiguously identified, k was chosen to be the number of local maxima of the probability distribution in the PCA feature space. The centroids determined by k -means approximately coincided with such local maxima, consistent with the interpretation of the centroids as local minima in the free energy landscape, see Supplementary Fig. 23.

To prepare the Folding@home simulations, HREMD data were featurized with interface distances and its dimensionality reduced with PCA as described above. The trajectory was then clustered into 98 k -means states for PROTAC 2, and 100 states for both PROTAC 1 and ACBII, whose cluster centers were selected as seeds for Folding@home massively parallel simulations. The simulation systems and parameters were kept the same as for HREMD and loaded into OpenMM where they were energy-minimized and equilibrated for 5 ns in the NPT ensemble ($T = 310$ K, $p = 1$ atm) using the openmmtools Langevin BAOAB integrator with 2 fs timestep. 100 trajectories with random starting velocities were then initialized on Folding@home for each of the seeds. The final dataset consists of 9800 trajectories, 5.7 ms of aggregate simulation time, and 650 ns median trajectory length. This dataset is made publicly available at: <https://console.cloud.google.com/storage/browser/paperdata>.

For computational efficiency, the data was strided to 5 ns/frame, featurized with closest heavy atom interface distances, and projected into tICA space at lag time 5 ns using commute mapping. The dimensionality of the dataset was chosen to keep the number of tICs necessary to explain 95% of kinetic variance: 219 for PROTAC 1, 339 for PROTAC 2, and 197 for ACBII. The resulting tICA space was discretized into microstates using k -means: we used 30 microstates for PROTAC 1, 1000 microstates for PROTAC 2, and 40 microstates for ACBII. The Markov state models (MSM) were then estimated from the resulting discretized trajectories at lag time 50 ns. For the PROTAC 2 MSM, we used a minimum number of counts for ergodic trimming (i.e., the ‘min-count_connectivity’ argument in PyEMMA) of 4, as the default setting resulted in a trapped state whose connectivity between simulation sub-ensembles starting from two different seeds was observed only due to clustering noise. The validity of the MSM was confirmed by plotting the populations from raw MD counts vs. equilibrium populations from the MSM, which is a useful test, especially when multiple seeds are used and the issue of connectivity is paramount. A hidden Markov model (HMM) was then computed to coarse-grain the transition matrix using 2 macrostates for PROTAC 1, 5 macrostates for PROTAC 2, and 3 macrostates for ACBII. Chapman-Kolmogorov tests using the transition matrices from these HMMs are shown in Supplementary Figs. 30–32. A better alternative to build macrostate models might be to construct memory kernels⁵⁶ rather than fuzzy assignments of states as in HMMs. This may also reduce the computational resources needed to estimate free energies and transition kinetics of macrostates.

During analysis of our PROTAC1 simulations, we found that one of our initial 100 seeded structures was kinetically separated from the others, such that reversible transitions to this state were not observed in our F@H trajectories. Transitions between this state and the ground state were therefore identified as the slowest mode by tICA. Since transitions to this state were never observed in our F@H simulations, we simply removed all trajectories seeded from this initial structure and omitted the first tIC from our analysis.

Comparison of HREMD to SAXS experiment

We validated the HREMD-generated ensembles of iso1/iso2-SMARCA2:ACBII:VCB complexes by directly comparing to the experimental SAXS data. The theoretical SAXS profile was computed from each snapshot from the HREMD simulation trajectory using CRY SOL⁵⁷ available in a software package ATSAS⁵⁸. The following CRY SOL command was used: *crysol <filename.pdb> -lm 20 -sm 0.5 -ns 201 -un 1 -eh -dro 0.03*. To expedite the writing of PDBs from HREMD trajectory and calculation of SAXS profiles, we used the multiprocessing functionality implemented in a Python package *idpflex*⁵⁹. The ensemble-averaged theoretical SAXS profile was determined as below,

$$\langle I(q) \rangle = \frac{1}{n} \sum_{i=1}^n I_i(q) \quad (13)$$

where $n = 100,001$ is the total number of frames in HREMD trajectory of each complex. The ensemble-averaged theoretical SAXS profile was compared to the experiment (Fig. 6c) by minimizing chi-square (χ^2) given by,

$$\chi^2 = \frac{1}{(m-1)} \sum_{i=1}^m \left\{ \frac{[\langle I_{\text{expt}}(q_i) \rangle - (c \langle I_{\text{calc}}(q_i) \rangle + b)]^2}{\sigma_{\text{expt}}(q_i)} \right\} \quad (14)$$

where $\langle I_{\text{expt}}(q) \rangle$ and $\langle I_{\text{calc}}(q) \rangle$ are the ensemble-averaged experimental and theoretical SAXS intensities respectively, m is the number of experimental q points, c is a scaling factor, b is a constant background, and σ_{expt} is the error in $I_{\text{expt}}(q)$.

Cullin-RING E3 ubiquitin ligase (CRL) simulations to explore activation

To study the impact of different degraders on ubiquitination, first we constructed an active form of the Cullin-RING E3 ubiquitin ligase (CRL) with VHL and grafted it onto the ternary structures from the SMARCA2^{BD}:degraded:VHL simulations described above. We used targeted MD simulations (TMD)⁶⁰ to drive the activation of the CLR based on the active structure of a homologous E3 ligase, CRL- β TrCP (PDB ID: 6TTU)³⁸. The full CRL-VHL system was built using PDB IDs 1LQB⁹ and 5N4W⁶¹ including VHL, ElonginB, ElonginC, Cullin2, and RBX1. NEDD8 was placed near residue Lys689 of the CRL where neddylation occurs.

As the collective variable for TMD, we used the residue-based RMSD of the last -70 C α atoms of the Cullin C-terminus (where neddylation and subsequent activation occur) of Cullin1 from the 6TTU structure³⁸ as the reference state and modeled Cullin2 from its inactive form in the 5N4W structure to this reference state. In addition, the C α atoms of the entire NEDD8 protein from the 6TTU structure was also used as a reference structure during TMD. Residues 135–425 from Cullin2 and corresponding residues from Cullin1 were used for alignment during TMD. The force constant for TMD was set to 30 kJ mol⁻¹ nm⁻². The system in a rectangular simulation box with a total number of ~500,000 atoms and an ionic concentration of 0.120 M using KCl. Hydrogen mass repartitioning (HMR) was used to enable 4 fs timestep simulations using the the AMBER ff14SB force field parameters. The TMD structure was then used to build the entire complex for CRL:VHL:degrader:SMARCA2^{BD}. The system also included E2 and ubiquitin from the 6TTU structure.

This system was solvated in a truncated octahedral box to avoid protein rotation during simulation and it was equilibrated for about 30 ns before subsequent meta-eABF simulations for identifying the ubiquitination zone.

Meta-eABF simulations on full Cullin-RING E3 ubiquitin ligases (CRL) complex

We employ an advanced path-based simulation method that combines metadynamics with extended adaptive biasing force (meta-eABF) to study the dynamic nature of the full CRL:VHL:degrader:SMARCA2^{BD} complex and generate a diverse set of putative closed conformations that place the E2-loaded ubiquitin close to lysine residues on SMARCA2^{BD}. The results from the meta-eABF simulation are used to seed additional simulations for unbiased ensemble-scale sampling.

Similar to adaptive biasing force (ABF) methods, meta-eABF simulations also utilize adaptive free energy biasing forces to enhance sampling along one or more collective variables (CVs), but the practical implementation is different^{62,63}. Meta-eABF evokes the extended Lagrangian formalism of ABF whereby an auxiliary simulation is introduced with a small number of degrees of freedom equal to the number of CVs, and each real CV is associated with its so-called fictitious counterpart in the low-dimensional auxiliary simulation. The real CV is tethered to its fictitious CV via a stiff spring with a large force constant and the adaptive biasing force is equal to the running average of the negative of the spring force. The biasing force is only applied to the fictitious CV, which in turn "drags" the real simulation along the real CV via the spring by periodically injecting the instantaneous spring force back into the real simulation. Moreover, the main tenet of the meta-eABF method is employing metadynamics (MtD) or well-tempered metadynamics (WTM) to enhance sampling of the fictitious CV itself. The combined approach provides advantages from both MtD/WTM and eABF.

For CRL-VHL closure we chose a single CV, the center-of-mass (COM) distance between SMARCA2^{BD} and E2 ligase-ubiquitin (E2-Ub) complex. The initial COM distance after relaxation was ~65 Å, and we ran 40 ns of meta-eABF simulation biasing the COM distance between 25 and 75 Å. During this simulation we saw multiple ring closing-opening events with the last frame representing a slightly open conformation with COM distance ~36 Å. We then continued the meta-eABF simulation for another 80 ns but narrowing the bias range on the COM distance to 25–40 Å in order to focus the sampling on closed or nearly closed conformations. The simulations were run using OpenMM 7.5⁶⁴ interfaced with PLUMED 2.7⁶⁵.

Mass spectrometry-based proteomics and ubiquitin analysis

Hela cells (ATCC) were cultured at a seeding density of 6E6 cells per 150 cm dish the day before in IMDM + 10% FCS. Next day, the cells were treated for 1 h with either (i) 300 nM of ACBII, (ii) 300 nM of ACBII + 10 μ M MG132 or, (iii) vehicle (DMSO) alone. Three plates of cells were treated for triplicate measurement in each condition. The cell pellets were collected after 1 h and lysed in 50 mM TEAB (pH 7.5) buffer containing 5% (w/v) SDS. Protein amounts were quantified using a BCA assay (Thermo Fisher Scientific) according to manufacturers' instructions. A total of 5 mg of each sample was processed and digested overnight using the S-trap-based approach⁶⁶ with S-trap midi-columns (ProtiFi) according to a manufacturers' protocol. Enrichment of ubiquitinated peptides (GG-remnants) was performed⁶⁷ using an anti-Gly remnant antibody (CST, PTMScan®Ubiquitin Remnant Motif (K-e-GG) Kit) following manufacturers' instructions. We used 10 μ L of slurry beads (corresponding to 62.5 μ g antibody) for each ubiquitin pull-down. Each sample was desalted and dried down before separating into four fractions using basic reversed-phase tip columns as previously reported⁶⁸. For separation, dried peptides were reconstituted in 200 μ L of buffer A (25 mM NH₄F₄, pH 10 in ddH₂O) and loaded on self-packed 200 μ L C-18 Stage Tips (Empore Octadecyl C18 47 mm

Solid Phase Extraction Disks #2215, 3 M Purification, Egan, MN, USA). After loading, the peptides were washed with 200 μ L of buffer A and eluted sequentially with buffer A containing 2.5%, 7.5%, 12.5% ACN, and 50% ACN. The sample flow-through fraction was combined with the 50% ACN fraction for a total of four fractions. Fractions were dried down and stored at -20 °C until further analysis.

LC-MS/MS analysis. Peptides were dissolved in 0.1% formic acid (FA) and analyzed on a Q-Exactive Plus mass spectrometer (Thermo Scientific) coupled to an Ultimate 3000 RSLCnano ultra HPLC system (Thermo Scientific). The samples were separated in a 120 min gradient (from 4% solvent B to 32% solvent B over 100 min; Solvent A 0.1% FA, 5% DMSO in water; solvent B 0.1% FA, 5% DMSO in acetonitrile). The loading buffer was 0.1% formic acid in water. The mass spectrometer was operated in a data-dependent acquisition (DDA) mode with an MS1 scan from 360 to 1300 m/z , acquired at 70,000 resolution. The MS1 scan was followed by 20 m/z dependent MS2 scans. The precursor ions were fragmented by higher energy collision dissociation (HCD) and acquired at a resolution of 17,500. The automatic gain control (AGC) targets for MS1 and MS2 were set at 3×10^6 ions and 1×10^5 ions, respectively. The maximum ion injection time for MS1 was set to 25 ms for MS1 and 50 ms for MS2 acquisition, with a dynamic exclusion of 35 s. The normalized collision energy was set at 28%. Peptide and protein identification. MaxQuant software v.2.0.1.0 was used for protein identification and label-free quantification (LFQ). The raw mass spectrometry data files were searched against the Human UniProt database using trypsin as the digestion enzyme with up to two missed cleavages allowed. Carbamidomethylated cysteine was set as a static modification. Oxidation of methionine, protein N-terminal acetylation, and GlyGly on lysine were set as variable modifications. The match-between-run option in MaxQuant was switched on. To control for false positives, a 1% false discovery rate was used on the PSM and the protein level.

Data analysis. For the ubiquitination profiling data, the distribution of ubiquitination sites/protein intensity ratios between each sample and the vehicle samples were computed. A constant scaling factor per sample was determined so that the median of this distribution becomes 1, assuming the intensities of most ubi-site proteins do not change. All intensities in the samples were multiplied with this scaling factor for normalization. Statistical testing was performed using the limma R package⁶⁹ on normalized log₂-transformed intensities. Missing values were imputed as long as no replicate of the same treatment had an intensity larger than the median intensity of the treatment. Proteome-corrected ubiquitination values were calculated using imputed intensities. Statistical significance was determined using pairwise *t*-test with Benjamini-Hochberg correction for multiple testing.

Reporting summary

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

Data availability

All data presented in this study as main text or Supplementary Figures and Tables are provided in the Source Data file.

The crystal structure of the ternary complex SMARCA2^{BD}:ACBII:VHL resolved in this study has been deposited in the Protein Data Bank under accession code [754E](#).

The proteomics data obtained in this study as part of the HDX-MS experiments have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository under accession code [PXD033849](#).

The proteomics data obtained in this study as part of the ubiquitinomics experiments have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository under accession code [PXD033763](#).

The small-angle X-ray scattering data obtained in this study for the ternary complex iso1-SMARCA2^{BD}: ACBII:VCB has been deposited in the Small Angle Scattering Biological Data Bank under accession code [SASDPE8](#).

The small-angle X-ray scattering data obtained in this study for the ternary complex iso2-SMARCA2^{BD}: ACBII:VCB has been deposited in the Small Angle Scattering Biological Data Bank under accession code [SASDPF8](#).

The crystal structure of the ternary complex SMARCA2^{BD}:PROTAC 1:VHL, used in this study for comparison and for simulation, is available in the Protein Data Bank under accession code [6HAY](#).

The crystal structure of the ternary complex SMARCA2^{BD}:PROTAC 2:VHL, used in this study for comparison and for simulation, is available in the Protein Data Bank under accession code [6HAX](#).

The crystal structure of the NEDD8 protein, used in this study for simulation, is available in the Protein Data Bank under accession code [6TTU](#).

The crystal structures of the Cullin2 and RBX1 proteins, used in this study for simulation, are available in the Protein Data Bank under accession code [5N4W](#).

The crystal structures of the VHL, Elongin C, and Elongin B proteins, used in this study for simulation, are available in the Protein Data Bank under accession code [1LQB](#).

The gene sequence of the bromodomain of SMARCA2 (isoform 1), used in this study for cloning and protein expression as well as for homology modeling, is available in the Uniprot database under accession code [P51531-1](#).

The gene sequence of the bromodomain of SMARCA2 (isoform 2), used in this study for cloning and protein expression, is available in the Uniprot database under accession code [P51531-2](#).

The gene sequence of VHL, used in this study for cloning and protein expression, is available in the Uniprot database under accession code [P40337](#).

The gene sequence of Eloc, used in this study for cloning and protein expression, is available in the Uniprot database under accession code [Q15369](#).

The gene sequence of Elob, used in this study for cloning and protein expression, is available in the Uniprot database under accession code [Q15370](#).

The initial and final configurations of all Molecular Dynamics simulation trajectories performed in this study can be accessed at <https://doi.org/10.5281/zenodo.7017262>.

Full-length simulation trajectories can be accessed at <http://console.cloud.google.com/storage/browser/paperdataand> require a Google account. Source data are provided with this paper.

Code availability

All simulation codes applied and movies prepared in this study can be accessed at <https://doi.org/10.5281/zenodo.7017262>.

References

1. Wu, T. et al. Targeted protein degradation as a powerful research tool in basic biology and drug target discovery. *Nat. Struct. Mol. Biol.* **27**, 605–614 (2020).
2. Schneider, M. et al. The PROTACtable genome. *Nat. Rev. Drug Discov.* **20**, 789–797 (2021).
3. Schapira, M., Calabrese, M. F., Bullock, A. N. & Crews, C. M. Targeted protein degradation: expanding the toolbox. *Nat. Rev. Drug Discov.* **18**, 949–963 (2019).
4. Coleman, K. G. & Crews, C. M. Proteolysis-targeting chimeras: harnessing the ubiquitin-proteasome system to induce degradation of specific target proteins. *Annu. Rev. Cancer Biol.* **2**, 1–18 (2017).
5. Matyskiela, M. E. et al. A cereblon modulator (CC-220) with improved degradation of ikaros and aiolos. *J. Med. Chem.* **61**, 535–542 (2018).

6. Ohoka, N. et al. In vivo knockdown of pathogenic proteins via specific and nongenetic inhibitor of apoptosis protein (IAP)-dependent protein erasers (SNIPERs). *J. Biol. Chem.* **292**, 4556–4570 (2017).
7. Wei, J. et al. Harnessing the E3 ligase KEAP1 for targeted protein degradation. *J. Am. Chem. Soc.* **143**, 15073–15083 (2021).
8. Rodriguez-Gonzalez, A. et al. Targeting steroid hormone receptors for ubiquitination and degradation in breast and prostate cancer. *Oncogene* **27**, 7201–7211 (2008).
9. Hon, W.-C. et al. Structural basis for the recognition of hydroxyproline in hif-1 α by pvh1. *Nature* **417**, 975–978 (2002).
10. Cowan, A. D. & Ciulli, A. Driving e3 ligase substrate specificity for targeted protein degradation: lessons from nature and the laboratory. *Annu. Rev. Biochem.* **91**, 295–319 (2022).
11. Casement, R. et al. *Mechanistic and Structural Features of PROTAC Ternary Complexes*, 79–113 (Springer US, New York, NY, 2021).
12. Li, W., Zhang, J., Guo, L. & Wang, Q. Importance of three-body problems and protein–protein interactions in proteolysis-targeting chimera modeling: insights from molecular dynamics simulations. *J. Chem. Inf. Model.* **62**, 523–532 (2022).
13. Hughes, S. & Ciulli, A. Molecular recognition of ternary complexes: a new dimension in the structure-guided design of chemical degraders. *Essays Biochem.* **61**, 505–516 (2017).
14. Schiemer, J. et al. Snapshots and ensembles of BTK and cIAP1 protein degrader ternary complexes. *Nat. Chem. Biol.* **17**, 152–160 (2021).
15. Bondeson, D. P. et al. Lessons in protac design from selective degradation with a promiscuous warhead. *Cell Chem. Biol.* **25**, 78–87 (2018).
16. Ward, C. C. et al. Covalent ligand screening uncovers a rnf4 e3 ligase recruiter for targeted protein degradation applications. *ACS Chem. Biol.* **14**, 2430–2440 (2019).
17. Gadd, M. S. et al. Structural basis of PROTAC cooperative recognition for selective protein degradation. *Nat. Chem. Biol.* **13**, 514–521 (2017).
18. Farnaby, W. et al. BAF complex vulnerabilities in cancer demonstrated via structure-based PROTAC design. *Nat. Chem. Biol.* **15**, 672–680 (2019).
19. Bai, N., Kirubakaran, P. & Karanicolas, J. Rationalizing PROTAC-mediated ternary complex formation using Rosetta. *J. Chem. Inf. Model.* **61**, 1368–1382 (2021).
20. Drummond, M. L., Henry, A., Li, H. & Williams, C. I. Improved accuracy for modeling PROTAC-mediated ternary complex formation and targeted protein degradation via new in silico methodologies. *J. Chem. Inf. Model.* **60**, 5234–5254 (2020).
21. Eron, S. J. et al. Structural characterization of degrader-induced ternary complexes using hydrogen-deuterium exchange mass spectrometry and computational modeling: implications for structure-based design. *ACS Chem. Biol.* **16**, 2228–2243 (2021).
22. Nowak, R. P. et al. Plasticity in binding confers selectivity in ligand-induced protein degradation. *Nat. Chem. Biol.* **14**, 706–714 (2018).
23. Devaurs, D., Antunes, D. A. & Borysik, A. J. Computational modeling of molecular structures guided by hydrogen-exchange data. *J. Am. Soc. Mass Spectrom.* **33**, 215–237 (2022).
24. Wales, T. E. & Engen, J. R. Hydrogen exchange mass spectrometry for the analysis of protein dynamics. *Mass Spectrom. Rev.* **25**, 158–170 (2006).
25. Gallagher, E. S. & Hudgens, J. W. Mapping protein-ligand interactions with proteolytic fragmentation, hydrogen/deuterium exchange-mass spectrometry. *Methods Enzymol.* **566**, 357–404 (2016).
26. Zuckerman, D. M. & Chong, L. T. Weighted ensemble simulation: Review of methodology, applications, and software. *Annu. Rev. Biophys.* **46**, 43–57 (2017).
27. Saglam, A. S. & Chong, L. T. Protein-protein binding pathways and calculations of rate constants using fully-continuous, explicit-solvent simulations. *Chem. Sci.* **10**, 2360–2372 (2018).
28. Dickson, A. Mapping the ligand binding landscape. *Biophys. J.* **115**, 1707–1719 (2018).
29. Méndez, R., Leplae, R., De Maria, L. & Wodak, S. J. Assessment of blind predictions of protein-protein interactions: current status of docking methods. *Proteins* **52**, 51–67 (2003).
30. Huang, L., So, P.-K. & Yao, Z.-P. Protein dynamics revealed by hydrogen deuterium exchange mass spectrometry: correlation between experiments and simulation. *Rapid Commun. Mass Spectrom.* **33**, 83–89 (2018).
31. Dixon, T., Uyar, A., Ferguson-Miller, S. & Dickson, A. Membrane-mediated ligand unbinding of the pk-11195 ligand from tsps. *Biophys. J.* **120**, 158–167 (2021).
32. Copperman, J. & Zuckerman, D. M. Accelerated estimation of long-timescale kinetics from weighted ensemble simulation via non-markovian "microbin" analysis. *J. Chem. Theory Computation* **16**, 6763–6775 (2020).
33. DeGrave, A. J., Bogetti, A. T. & Chong, L. T. The red scheme: Rate-constant estimation from pre-steady state weighted ensemble simulations. *J. Chem. Phys.* **154**, 114111 (2021).
34. Zhang, M. M. et al. An integrated approach for determining a protein-protein binding interface in solution and an evaluation of hydrogen-deuterium exchange kinetics for adjudicating candidate docking models. *Anal. Chem.* **91**, 15709–15717 (2019).
35. Scherer, M. K. et al. Pyemma 2: A software package for estimation, validation, and analysis of markov models. *J. Chem. Theory Comput.* **11**, 5525–5542 (2015).
36. Husic, B. E., McGibbon, R. T., Sultan, M. M. & Pande, V. S. Optimized parameter selection reveals trends in markov state models for protein folding. *J. Chem. Phys.* **145**, 194103 (2016).
37. Buhimschi, A. D. & Crews, C. M. Evolving rules for protein degradation? insights from the zinc finger degrome. *Biochemistry* **58**, 861–864 (2019).
38. Baek, K. et al. Nedd8 nucleates a multivalent cullin–ring–ube2d ubiquitin ligation assembly. *Nature* **578**, 461–466 (2020).
39. Yauch, R. et al. Selective protac-mediated degradation of smarca2 is efficacious in smarca4 mutant cancers, Preprint at <https://doi.org/10.21203/rs.3.rs-1133243/v1> (2022).
40. Kochert, B. A., Iacob, R. E., Wales, T. E., Makriyannis, A. & Engen, J. R. Hydrogen-deuterium exchange mass spectrometry to study protein complexes. *Methods Mol. Biol.* **1764**, 153–171 (2018).
41. Wales, T. E., Fadgen, K. E., Gerhardt, G. C. & Engen, J. R. High-speed and high-resolution UPLC separation at zero degrees celsius. *Anal. Chem.* **80**, 6815–6820 (2008).
42. Dagbay, K. B. & Hardy, J. A. Multiple proteolytic events in caspase-6 self-activation impact conformations of discrete structural regions. *Proc. Natl Acad. Sci. USA* **114**, E7977–E7986 (2017).
43. Hopkins, J. B., Gillilan, R. E. & Skou, S. BioXTAS RAW: improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. *J. Appl. Crystallogr.* **50**, 1545–1553 (2017).
44. Maier, J. A. et al. ff14sb: Improving the accuracy of protein side chain and backbone parameters from ff99sb. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).
45. Zwier, M. C. et al. Westpa: An interoperable, highly scalable software package for weighted ensemble simulation and analysis. *J. Chem. Theory Comput.* **11**, 800–809 (2015).
46. Zhang, B. W., Jasnow, D. & Zuckerman, D. M. The "weighted ensemble" path sampling method is statistically exact for a broad class of stochastic processes and binning procedures. *J. Chem. Phys.* **132**, 054107 (2010).
47. Pearlman, D. A. et al. Amber, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* **91**, 1–41 (1995).

48. Pracht, P., Bohle, F. & Grimme, S. Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Phys. Chem. Chem. Phys.* **22**, 7169–7192 (2020).
49. Marze, N. A., Roy Burman, S. S., Sheffler, W. & Gray, J. J. Efficient flexible backbone protein-protein docking for challenging targets. *Bioinformatics* **34**, 3461–3469 (2018).
50. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
51. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. Lincs: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).
52. Van Gunsteren, W. F. & Berendsen, H. J. C. A leap-frog algorithm for stochastic dynamics. *Mol. Simul.* **1**, 173–185 (1988).
53. Darden, T., York, D. & Pedersen, L. Particle mesh ewald: an $n \log(n)$ method for ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
54. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).
55. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
56. Cao, S., Montoya-Castillo, A., Wang, W., Markland, T. E. & Huang, X. On the advantages of exploiting memory in markov state models for biomolecular dynamics. *J. Chem. Phys.* **153**, 014105 (2020).
57. Svergun, D., Barberato, C. & Koch, M. H. J. CRYSOLE – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768–773 (1995).
58. Manalastas-Cantos, K. et al. ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis. *J. Appl. Crystallogr.* **54**, 343–355 (2021).
59. Borreguero, J. M., Islam, F. F., Shrestha, U. R. & Petridis, L. idpflex: Analysis of intrinsically disordered proteins by comparing simulations to small angle scattering experiments. *J. Open Source Softw.* **3**, 1007 (2018).
60. Cheng, X., Wang, H., Grant, B., Sine, S. M. & McCammon, J. A. Targeted molecular dynamics study of c-loop closure and channel gating in nicotinic receptors. *PLoS Comput. Biol.* **2**, e134 (2006).
61. Cardote, T. A., Gadd, M. S. & Ciulli, A. Crystal structure of the cul2-rbx1-elobc-vhl ubiquitin ligase complex. *Structure* **25**, 901–911 (2017).
62. Comer, J. et al. The adaptive biasing force method: everything you always wanted to know but were afraid to ask. *J. Phys. Chem. B* **119**, 1129–1151 (2015).
63. Fu, H., Shao, X., Cai, W. & Chipot, C. Taming rugged free energy landscapes using an average force. *Acc. Chem. Res.* **52**, 3254–3264 (2019).
64. Eastman, P. et al. Openmm 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **13**, e1005659 (2017).
65. Bonomi, M. et al. Promoting transparency and reproducibility in enhanced molecular simulations. *Nat. Methods* **16**, 670–673 (2019).
66. Zougman, A., Selby, P. J. & Banks, R. E. Suspension trapping (STrap) sample preparation method for bottom-up proteomics analysis. *Proteomics* **14**, 1006–1000 (2014).
67. Udeshi, N. D. et al. Refined preparation and use of anti-diglycine remnant (K- ϵ -GG) antibody enables routine quantification of 10,000s of ubiquitination sites in single proteomics experiments. *Mol. Cell. Proteom.* **12**, 825–831 (2013).
68. Ruprecht, B., Zecha, J., Zolg, D. P. & Kuster, B. Proteomics. *Methods Mol. Biol.* **1550**, 83–98 (2017).
69. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
70. Jubb, H. C. et al. Arpeggio: A web server for calculating and visualising interatomic interactions in protein structures. *J. Mol. Biol.* **429**, 365–371 (2017).

Acknowledgements

This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725. A.R.D. acknowledges support from the National Institutes of Health (R01GM130794) and the National Sciences Foundation (DMS 1761320). We thank the University of Massachusetts Institute of Applied Life Sciences Mass Spectrometry Core (RRID:SCR_019063) and Stephen J. Eyles for their support and mentorship during the collection, and processing of all Hydrogen Deuterium Exchange Data. We thank Helix Biostructures LLC for their assistance with X-Ray data collection and raw data reduction. SAXS measurements were based upon research conducted at the Structural Biology Platform of the Université de Montréal, which is supported by the Canadian Foundation for Innovation award #30574. We would also like to thank OmicScouts GmbH for their support on Proteomics data generation. We are grateful to all the citizen scientists who contributed their compute power to make parts of this work possible, and members of the Folding@home community who volunteered to help with technical support to run these simulations. contribution numbers may be acknowledged.

Author contributions

T.Di and B.M. ran and analyzed WE-HDX simulations. D.Ma performed crystallography and HDX-MS experiments. T.Da ran docking simulations. S.L. wrote software to support simulations and analysis. D.Mc ran M.D. simulations and analyses. S.Sh and R.P. performed homology modeling and analyzed data. U.R.S. ran HREMD simulations and compared to SAXS. R.W. and Z.A.M. ran FAH simulations and performed conformational landscape analyses. F.P. analyzed HDX data. J.V.R. assisted with visualization and analyses. T.W. and V.S. helped scale WESTPA on Summit and run all simulations efficiently in our HPC cluster. N.G. and S.J. performed protein production. S.Sp performed SAXS analyses. Y.L. and A.V. performed SPR experiments. X.Z. oversaw synthesis of degrader molecules. A.M.R. and I.K. performed CRL simulations and ubiquitination analyses. J.A. and B.R. designed and analyzed proteomics experiments. J.Im, A.E., and L.B. helped edit the paper. A.D., H.X., W.S., and J.A.I. directed the research presented in this paper. All authors wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-33575-4>.

Correspondence and requests for materials should be addressed to Alex Dickson, Huafeng Xu, Woody Sherman or Jesus A. Izaguirre.


















Peer review information *Nature Communications* thanks Daniel Deredge, Daniel Zaidman and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

Tom Dixon ^{1,2,3,4}, **Derek MacPherson** ^{1,4}, **Barmak Mostofian** ^{1,4}, **Taras Dauszhenka** ¹, **Samuel Lotz** ¹, **Dwight McGee**¹, **Sharon Shechter**¹, **Utsab R. Shrestha** ¹, **Rafal Wiewiora**¹, **Zachary A. McDargh**¹, **Fen Pei** ¹, **Rajat Pal**¹, **João V. Ribeiro** ¹, **Tanner Wilkerson**¹, **Vipin Sachdeva**¹, **Ning Gao**¹, **Shourya Jain**¹, **Samuel Sparks**¹, **Yunxing Li**¹, **Alexander Vinitsky**¹, **Xin Zhang** ¹, **Asghar M. Razavi**¹, **István Kolossváry**¹, **Jason Imbriglio**¹, **Artem Evdokimov**¹, **Louise Bergeron**¹, **Wenchang Zhou**¹, **Jagat Adhikari** ¹, **Benjamin Ruprecht**¹, **Alex Dickson** ^{2,3} , **Huafeng Xu** ¹ , **Woody Sherman**¹  & **Jesus A. Izaguirre** ¹ 

¹Roivant Discovery, New York City, NY 10036, USA. ²Department of Computational Mathematics, Science, and Engineering, Michigan State University, East Lansing, MI 48824, USA. ³Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824, USA. ⁴These authors contributed equally: Tom Dixon, Derek MacPherson, Barmak Mostofian.  e-mail: alexrd@msu.edu; huafeng.xu@roivant.com; woody.sherman@roivant.com; jesus.izaguirre@roivant.com