bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Atomic-Resolution Prediction of Degrader-mediated Ternary Complex Structures by Combining Molecular Simulations with Hydrogen Deuterium Exchange

Tom Dixon,[†,¶] Derek MacPherson,[‡,¶] Barmak Mostofian,[‡,¶] Taras Dauzhenka,[‡] Samuel Lotz,[‡] Dwight McGee,[‡] Sharon Shechter,[‡] Utsab R. Shrestha,[‡] Rafal Wiewiora,[‡] Zachary A. McDargh,[‡] Fen Pei,[‡] Rajat Pal,[‡] João V. Ribeiro,[‡] Tanner Wilkerson,[‡] Vipin Sachdeva,[‡] Ning Gao,[‡] Shourya Jain,[‡] Samuel Sparks,[‡] Yunxing Li,[‡] Alexander Vinitsky,[‡] Asghar M. Razavi,[‡] István Kolossváry,[‡] Jason Imbriglio,[‡] Artem Evdokimov,[‡] Louise Bergeron,[‡] Alex Dickson,[*,†] Huafeng Xu,[*,‡] Woody Sherman,[*,‡] and Jesus A. Izaguirre[*,‡]

†*Department of Biochemistry and Molecular Biology, Michigan State University, USA*

‡*Roivant Discovery, New York, USA*

¶*These authors contributed equally*

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

Abstract

Targeted protein degradation (TPD) has recently emerged as a powerful approach for removing (rather than inhibiting) proteins implicated in diseases. A key step in

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

TPD is the formation of an induced proximity complex where a degrader molecule recruits an E3 ligase to the protein of interest (POI), facilitating the transfer of ubiquitin to the POI and initiating the proteasomal degradation process. Here, we address three critical aspects of the TPD process using atomistic simulations: 1) formation of the ternary complex induced by a degrader molecule, 2) conformational heterogeneity of the ternary complex, and 3) degradation efficiency via the full Cullin Ring Ligase (CRL) macromolecular assembly. The novel approach described here combines experimental biophysical data with molecular dynamics (MD) simulations to accurately predict ternary complex structures at atomic resolution. We integrate hydrogen-deuterium exchange mass spectrometry (HDX-MS, which measures the solvent exposure of protein residues) directly into the MD simulation algorithm to improve the efficiency and accuracy of the ternary structure predictions of the bromodomain of the cancer target SMARCA2 with the E3 ligase VHL, as mediated by three different degrader molecules. The simulations accurately reproduce X-ray crystal structures–including a new structure that we determined in this work (PDB ID: 7S4E)–with root mean square deviations (RMSD) of 1.1 to 1.6 Å. The simulations also reveal a structural ensemble of low-energy conformations of the ternary complex. Snapshots from these simulations are used as seeds for additional simulations, where we perform 5.7 milliseconds of aggregate simulation time using Folding@home, the world's largest distributed supercomputer. The detailed free energy surface captures the crystal structure conformation within the low-energy basin and is consistent with solution-phase experimental data (HDX-MS and SAXS). Finally, we graft a structural ensemble of the ternary complexes onto the full CRL and perform enhanced sampling simulations. Our results suggest that differences in degradation efficiency may be related to the proximity distribution of lysine residues on the POI relative to the E2-loaded ubiquitin. We make source code and the simulation and experimental datasets from this work publicly available for researchers to further advance the field of induced proximity modulation.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 Introduction

Heterobifunctional degrader molecules are a class of ligands that induce proximity between a target protein of interest (POI) and a E3 ubiquitin ligase, which can ultimately lead to ubiquitination of the POI and its subsequent proteosomal degradation through a complex machinery of proteins.[1] These degrader molecules provide the opportunity of a novel therapeutic modality– single molecules induce catalytic turnover of POI–and potentially offer an avenue for modulation of targets traditionally labeled as undruggable by classical therapeutic strategies.[2–4] The subset of degrader molecules classified as heterobifunctionals consist of two separate moieties joined by a "linker"; the "warhead" binds to the POI and the "ligand" binds to an E3 ligase such as Cereblon (CRBN),[5–7] cIAP,[8] Keap1,[9] and von Hippel-Lindeu protein (VHL).[10–12] In each case it is the ability of the warhead-linker-ligand degrader molecule to induce a ternary complex that is critical for bridging the interaction between the POI and an E3 ligase (which can be the native or non-native degradation partner for the POI).

The formation of the POI-degrader-E3 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 non-native induced proximity complex.[13] X-ray crystallography–the primary experimental technique for determining 3-dimensional structures of the ternary complex[14]–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 is integral to the binding cooperativity (the term used to describe degree to which the binding affinity of ternary complexes are thermodynamically different than the binary counterparts) and degradation efficiency. A study targeting the degradation of Burton Tyrosine Kinase (BTK) by CRBN found that optimal protein removal was achieved through a molecule that induced a non-cooperative ternary complex, demonstrating a disconnect between binding affinity and degradation efficiency.[15] Similarly, BTK was also found to non-cooperatively interact with cIAP but still led to high degradation effi-

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ciency.[16] Interestingly, NMR and crystallography revealed a structural ensemble being sampled by this ternary complex, suggesting specific conformations could be responsible for efficient downstream ubiquitination.[17] In contrast, studies with SMARCA2 and VHL found that more cooperative molecules led to higher degradation efficiency.[18] Furthermore, analysis of the ternary structures revealed a high degree of similarity despite the fact that the heterobifunctional molecules displayed different degrees of degradation efficiency,[18] raising questions about relationship between static structural representations of the the ternary complex and degradation efficiency. These findings and others[19–21] suggest that degradation efficiency is more complex than can be understood through the thermodynamics of binding or static structural analysis. As such, determining the dynamic ensemble of the ternary complex can reveal mechanistic insights to facilitate the design of more effective degrader molecules, especially to understand the relationship between linkers and degradation.[14,22–25]

Previous work to computationally predict ternary structures mostly consists of protein-protein docking protocols, perhaps followed by refinement of the initial structures with molecular dynamics (MD) simulations to assess the stability of the predicted models.[24–30] However, these docking protocols fail to predict high-resolution structures (sub-2.0 Å) with high fidelity. That is, while protein-protein docking protocols have shown some promise in generating structural models of ternary complexes with reasonable resolution (often characterized as sub-10 Å RMSD to an x-ray structure), the best structures typically fall somewhere within a long list of possible poses (often in the hundreds or thousands), demonstrating the challenge associated with the selection of high-accuracy ternary structure models.

Here, we present an integrated workflow that combines solution-state biophysical techniques with advanced molecular dynamics (MD) simulations to produce atomic resolution structural ensembles of the ternary complex. Hydrogen-deuterium exchange (HDX-MS) protection data is used as collective variables in the MD simulation, enhancing both the speed and accuracy of the computational predictions. Furthermore, HDX-MS data are also used as constraints for protein-protein docking when higher

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

throughput and lower resolution models are sought, such as when screening many degrader molecules.

We use the weighted ensemble simulation (WES) approach to perform molecular dynamics simulations at biologically relevant timescales (from microseconds to milliseconds) across multiple graphics processing units (from dozens to thousands of simultaneous GPUs). This approach allows for the speed and throughput needed to sample the conformational free energy landscape at a sufficient level to generate robust, high accuracy predictions of the ternary complex structural ensemble. WES utilizes an adaptive sampling procedure where an ensemble of unbiased trajectories are iteratively simulated and analysed so that computational resources can be optimally reallocated to regions of interest (e.g. unexplored regions of conformational phase space or regions of interest based on data from HDX-MS experiments). Trajectories in sparsely populated regions (i.e. limited data for statistical thermodynamic calculations) are cloned in order to enhance sampling and high-probability regions with sufficient data for computing statistical thermodynamic quantities are merged so computational resources can be reallocated to the sparsely populated regions.[31] The resampling is done such that the probability of the whole simulation ensemble of "walkers" is tracked in a statistically rigorous manner.[32–35] We modified a bin-less algorithm called REVO[36] to more efficiently sample ternary complex formation, which is implemented in the open source software package wepy.[37] The work presented here relies on knowledge of the binding pose of the warhead to the POI and that of the ligand to the E3 ligase, which are typically known from prior experiments or can be generated with computational tools like docking or shape-based alignment. Experimental HDX-MS data is used to determine the level at which residues are shielded from solvent upon ternary complex formation, as compared to the binary complexes (POI plus degrader or E3 ligase plus degrader).

Our ultimate goal is to understand the structural and dynamic basis for differences in degradation among a set of degrader molecules. Here, we focus on three different degrader molecules of the BAF ATPase subunit SMARCA2 isoform 2 that recruit the

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

E3 ligase VHL. The binding affinities and cooperativity of ternary complex formation and the degradation efficiency for these three degrader molecules are summarized in Table 1.

Ternary complex crystal structures of PROTAC 1 (PDB ID: 6HAY) and PROTAC 2 (PDB ID: 6HAX) show slight variations in the interactions and orientation of the proteins in the ternary structure. In addition, we obtain the crystal structure of the highly cooperative and more efficient degrader ACBI1 (PDB ID 7S4E from the work presented here) (Section 2.1). A static analysis of these crystal structures does not explain the difference in cooperativity and degradation of these heterobifunctional degrader molecules. To explain the different degradation profiles of these molecules, we carry out molecular dynamics simulations and solution experiments, which reveal insights beyond what is defined by the crystal structure alone (Section 2.2).

Our results show that by including experimental solution-phase HDX-MS data into the weighted-ensemble simulations (WES+HDX) we obtain improved throughput and accuracy of the ternary structure predictions. Starting from unbound SMARCA2 and VHL structures, WES+HDX is able to produce structural models of ternary complexes with Interface-RMSD below 2 Å from the experimental x-ray crystal structures (Section 2.3). Additionally, WES+HDX generates an ensemble of bound conformations spanning a free energy basin within 3 kcal/mol from the crystal structure (Section 2.6). These dynamic models describe an ensemble of energetically viable structures that could be used to study multiple aspects of the targeted protein degradation process, including binding kinetics, affinity, selectivity, cooperativity, ubiquitination, and degradation. We make prospective ternary structure predictions of the SMARCA2 isoform 1, ACBI1 and VHL:Elongin C:Elongin B, where SMARCA2 isoform 1 has a 17 amino acid extension compared to isoform 2. Our prediction reconciles the HDX-MS data showing interaction of the isoform 1 extension with a beta-strand from VHL (Section 2.6).

We also introduce methodology to determine the conformational free energy landscapes of these ternary complexes, which is the foundation for quantifying the pop-

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ulations of different conformational states. Starting from the crystal structures, we first sample conformations using a Hamiltonian replica exchange molecular dynamics (HREMD) simulation similar to solute tempering and cross-validate these simulations through small-angle X-ray scattering (SAXS) experiments (Section 2.5). From these simulations, we choose structures as seeds to run 10,000 simulations on Folding@Home, totalling approximately 6 ms of accumulated simulation time. We build Markov state models (MSMs)[38] that identify the most probable structures along with their conformational free energies and kinetics of interconversion, all of which can be used to guide the design of novel degrader molecules.

Finally, we model the entire Cullin-RING E3 ubiquitin ligase (CRL) assembly to explore structural and dynamic factors that may be associated with ubiquitination. We find that the dynamic ternary complex ensemble induced by the most efficient degrader (ACBI1) displays a higher probability density of accessible lysine residues of SMARCA2 within the "ubiquitination zone" of the CRL as compared with the other two degraders and the least efficient degrader (PROTAC 1) showing the lowest probability of accessibly lysine residues within the ubiquitination zone. We postulate that the dynamic structural ensemble induced by ACBI1 is responsible for its greater degradation efficiency (Section 2.7). The mechanistic interpretation is that this ensemble results in a higher probability of forming a productive encounter between the POI and E2-loaded unbiquitin, as defined by density of lysine residues within the ubiquitination zone. Overall, this work offers unique insights into the dynamic nature of the ternary structure ensemble and how dynamics of the full CRL macromolecular assembly could explain downstream protein degradation. We make the simulation and experimental results available to the research community, including source code and the release of a new X-ray crystal structure of ACBI1 with SMARCA2/VHL that has been deposited into the Protein Data Bank (PDB ID: 7S4E).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1: Binding affinity (K_d), efficiencies (IC50, DC50), and cooperativity (α) of PROTAC 1, PROTAC 2, and ACBI1 degraders. Ternary IC50 and binary (SMARCA2) DC50 values are reported; the cooperativity is the ratio of binary over ternary IC50. Table adapted from Farnaby et al.[18]

	$\mathbf{K_d}$, **VHL(nM)**	$\mathbf{K_d}$, **SMARCA2(nM)**	**IC50 (nM)**	**DC50 (nM)**	$\boldsymbol{\alpha}$
PROTAC 1	98 ± 26	4500 ± 480	205 ± 15	300	12
PROTAC 2	100 ± 10	770 ± 51	45 ± 9	N/A	18
ACBI1	250 ± 64	1800 ± 980	26 ± 3	6/3.3	30

2 Results

2.1 Degraders of differential degradation efficiency induce similar ternary complex structures in X-ray crystallography.

The ternary complexes of SMARCA2 isoform 2 (SMC2) and the VHL/ElonginC/ElonginB (VCB) induced by heterobifunctional degraders, including PROTAC 1, PROTAC 2, and ACBI1, have been extensively studied.[24,39] Crytal structures have been published for the ternary complexes induced by PROTAC 1 (PDB ID: 6HAY) and by PROTAC 2 (PDB ID: 6HAX), but no crystal structure has been reported for the ternary complex induced by the most potent degrader ACBI1. To further study the effects of linker composition on the ternary complex ensemble, we determined the structure of SMC2/VHL liganded by ACBI1 via X-ray crystallography. The structure was obtained using conditions from the literature[24] with an alteration to the protocol by including a size exclusion polishing step in between complex formation and crystallization by hanging drop at 4° C. The addition of SEC ensured a homogeneous solution by removing any SMC2 or VCB in the APO or binary states (Supplementary Figure 17). The structure was solved by molecular replacement to 2.25 Å in the highest resolution shell (Supplementary Figure 18), using 6HAX as the search model (Figure 1a). The degrader molecule bridges the induced interface, making several contacts with both proteins.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Importantly, the ligand induces cooperative contacts between several amino acids such as VCB:ARG69 and the backbone of SMC2:PHE1463 (Figure 1 b,c). SMC2:ASN1464 makes critical bivalent contacts to the aminopyridazine group of ACBI1, positioning the terminal phenol group for pi stacking interactions with residues PHE1409 and TYR1421 (Figure 1b,c). On the VHL side of the interface, the interactions made by TYR98 are consistent with the other degraders studied here (Figure 1b,c).[24]

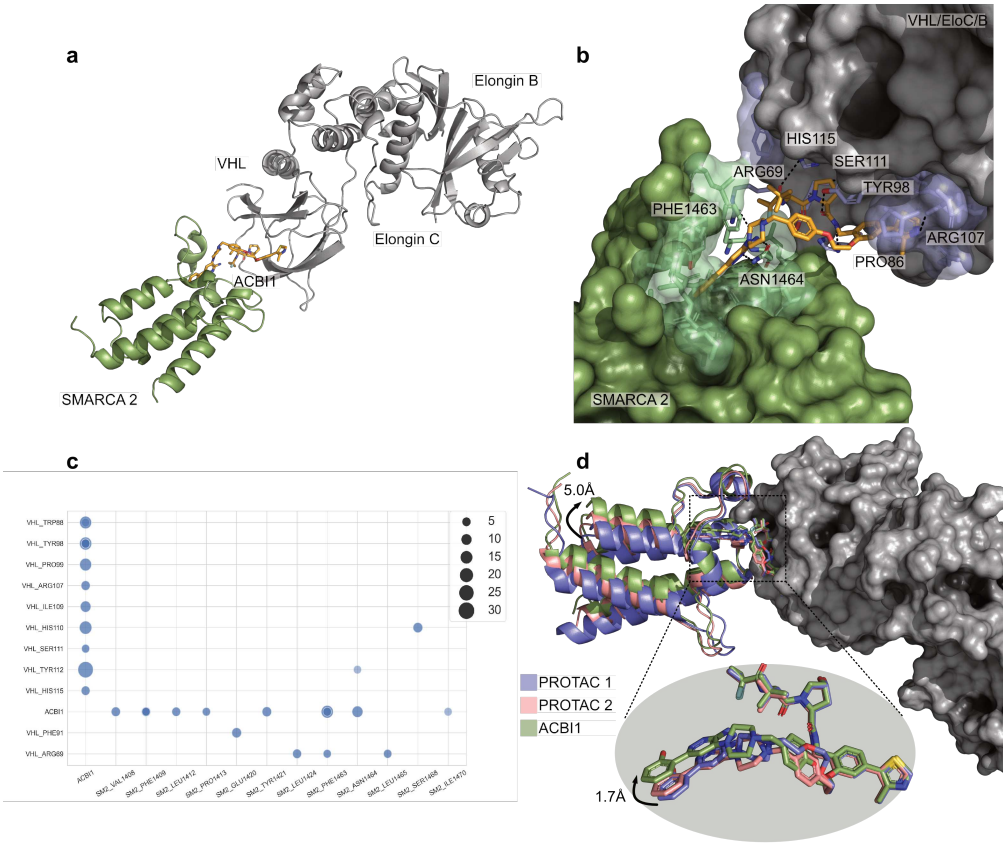

Figure 1: Ternary complex of SMARCA2 and VCB induced by ACBI1 shows structural similarities with PROTAC 1 and PROTAC 2: **a** Overall perspective of SMARCA2 Isoform 2 (smudge) and VHL/ElonginC/ElonginB (grey) induced by degrader molecule ACBI1 (bright orange). **b** Critical interactions induced by ternary complex formation between SMARCA2 and VCB. **c** A contact map for the interface of the crystal structure. The circle size reflects the numbers of atoms (including hydrogen atoms) participating in interactions. **d** Superposition of 6HAY (slate), 6HAX (salmon), 7S4E (smudge) by aligning VHL (grey) shows varied conformations of the ligands PROTAC 1, PROTAC 2, or ACBI1 resulting in alterations of SMARCA2 within the ternary complex.

Comparison of the ligands PROTAC 1, PROTAC 2, and ACBI1, binding to VHL

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

between each of these molecules is largely maintained and superimposable. The differences in the linker composition, however, allow for a slight 1.7 Å twist of the ACBI1 linker by the inclusion of a single ether group. The twist allows for the reorientation of SMC2 terminating in a subtle 5 Å upwards swing of the protein (Figure 1d).

Despite the structural similarity at the protein interface of these degraders,[24] we conclude that the composition of each degrader molecule results in markedly different degradation due to the differences in the dynamic structural ensemble induced by the different linkers. Consistent with other studies,[16,17] this implies that crystallographic snapshots are not able to provide a holistic view of the ensemble being sampled by the ternary complexes in solution, as the structures are collapsing into a single conformation that supports protein crystallization.[40] While the X-ray structures may represent a subset of the relevant conformational states, they do not capture the dynamic nature of the ternary structure ensemble, which ultimately determines to activity/function.[16,17] Thus, the underlying differences in the ternary complex ensembles induced by each of these degrader molecules may be lost. In order to understand the impact of degrader molecules (including the linker) on the dynamic nature of the ternary structure, which would enable more accurate design and prediction of novel degrader molecules, we perform advanced molecular dynamics simulations coupled with solution-state biophysical data to gain a richer understanding of the ternary structure ensemble.

2.2 Hydrogen Deuterium Exchange Reveals Extended Protein-Protein Interfaces

It is clear that targeted protein induced ternary complexes exhibit a vast array of conformations, however sampling and verifying bona fide interfaces remains a challenging hurdle.[16,17] Some attempts at characterizing TPD ternary structures employ multiple crystal forms,[17] NMR[16] and SAXS all coupled with various forms of modeling. In our efforts we sought to employ hydrogen deuterium exchange of the APO, binary and ternary complex species for SMC2 and VCB to map the protein interface in so-

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

lution. Using previously established protocols as the basis for our experiments,[41–43] modifications were then incorporated in order to ensure the majority of the species in solution was the expected APO, binary, or ternary complex, respectively.[44,45] By using the established binding constants, and alpha factors for each degrader (PROTAC 1, PROTAC 2, ACBI1), the ratio of each component of the three−bodied system was established to ensure complex formation. Sequence coverage of both SMC2 and VCB was determined to be 100% for each protein (Supplementary Figure 19). Each protein was also found to have stable deuterium exchange, consistent with proteins that are properly folded throughout the course of the experiment (Supplementary Figures 20-23). To emphasize the protection of the binary and ternary complex states, the APO or binary uptake of like peptides was subtracted from the binary or ternary state, respectively. The resultant difference plots highlight the statistically significant residues that are protected to the 95% or 98% confidence interval. All raw relative uptake plots of the deuterium exchange for each state and experiment can be found in the supplementary material (Supplementary Figures 35−85). Importantly, protection during HDX-MS arises due to changes in the environment around the observed amino acids, which could be a result of direct occlusion of solvent or conformational changes.[46] ~~First~~ The BinaryΔAPO difference plot suggests that the SMC2-ligand binary complex was unstable under our experimental conditions (the ligand concentration is close to the dissociation constant KD =?μM[24]), as there is no substantial difference between the HDX of SMC2 in presence and in absence of the ligand (Figure 2a and e). Importantly, in the presence of VCB (TernaryΔBinary), significant stretches of amino acids become protected, many of which, amino acids 1409-1422 and 1456-1470, overlap with the ligand binding site based on our (7S4E) and previously published crystal structures (6HAY, 6HAX) (Figure 2a and f). Additionally, there are also stretches of amino acids, 1394-1407, beyond the established binding interface that are, based on the structure, too far in distance to result in the observed protection (Figure 2a and f). On the VCB side of the interface, large regions of VHL, which resides at the direct interface of the protein complex, are protected in the presence of the ligand in

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

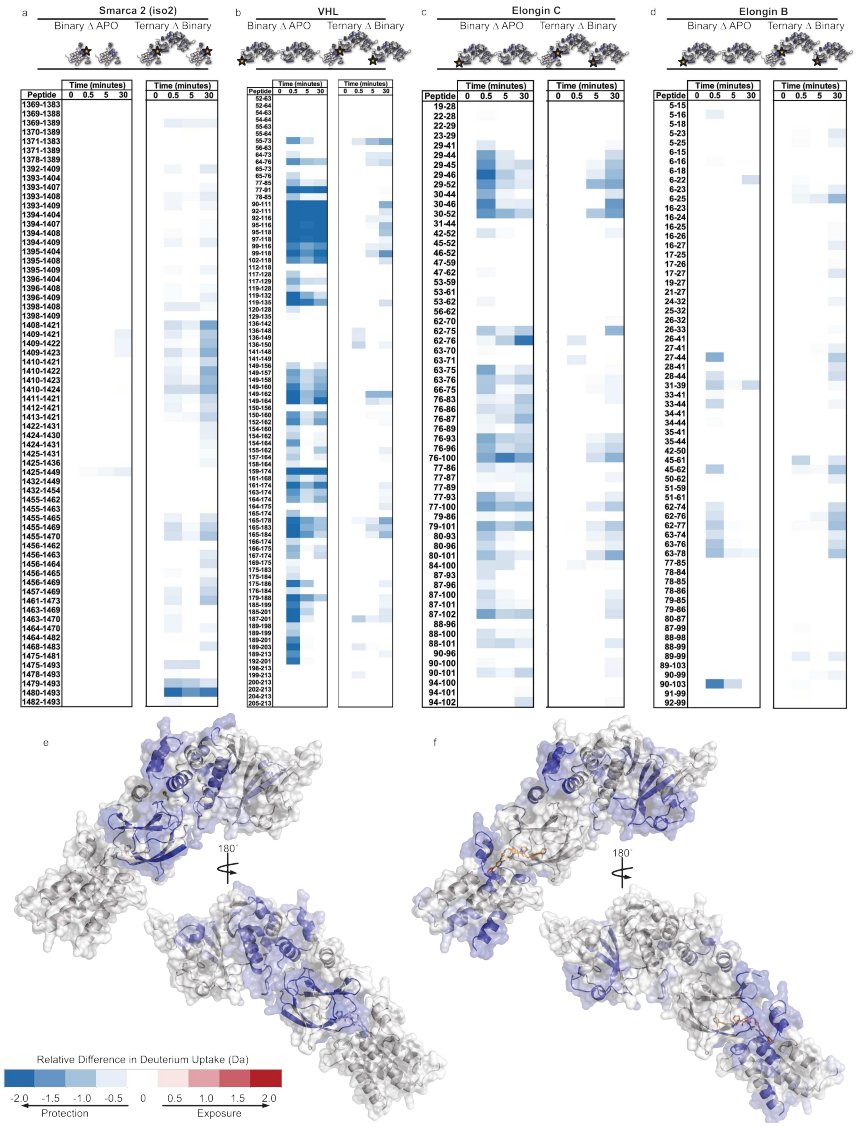

Figure 2: Induction of ternary complex of SMARCA2 and VHL by ACBI1 induces ensemble specific protection: a-d, SMARCA2 isoform 2(a), VHL(b), Elongin C(c), Elongin B(d) monitored for exchange over time. The difference plots of each protein in the binary and ternary states generated by subtracting the deuterium exchange of like peptides of the APO or binary from the binary or ternary states, respectively. Regions of significantly different exchange, depicted in blue (negative), are exchanging less than the comparative state, while regions that appear in red (positive) are exchanging more based on the error of their respective datasets to the 95% confidence interval. The resultant difference plots of the binary modes (e), or ternary complex (f) were plot on to the structure 7S4E. The datasets were repeated on 2 separate days, Δ represents the difference, where the APO is subtracted from the binary state, BinaryΔAPO, or the binary is subtracted from the ternary state, TernaryΔBinary.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the BinaryΔAPO state (Figure 2b and e). The most protected residues in the binary state are centered around amino acids 87-116, which include all 9 residues in the ligand binding site of VHL. However, there is a large amount of protection across the entire protein indicating the presence of a large allosteric network across the protein.[47] In the presence of SMC2 (TernaryΔBinary), much of the allosteric network due to ligand binding can be subtracted away leaving only the most significantly protected residues due to ternary complex formation (Figure 2b and f). Residues 60-72, which house the critical interaction of ARG69 show significant protection due to ternary complex formation (Figure 2b and f). Additionally, traveling upward in sequence away from the binding site we see continued protection of residues 166-176, again stretching from 187-201 (Figure 2b and f). These stretches of amino acids are distal from the direct binding interface of the SMC2:VCB complex. Moreover, Elongin B and C, which are both distal from the interface, show some additional minor changes due to the presence of the ternary complex (Figure 2c and d). When grafted on to the structure, these distal protection patterns observed in VHL, ElonC, and ElonB all spatially align with one another potentially uncovering a critical network of allosteric changes being induced by ACBI1. These changes that are not observed in the snapshot provided by crystallography help explain how this molecule vastly improves biological function, i.e., perhaps by changing the orientation of the ligase to target lysine residues on SMC2, as suggested by our simulations of the SMC2:VCB complex (see below).

2.3 Efficient simulation of ternary complex formation using REVO Weighted Ensemble simulations

We simulate the formation of a degrader ternary complex with the weighted ensemble path sampling approach. This method has been employed before for tasks such as protein-protein binding.[48] It is noteworthy, however, that the pre-defined CVs in the current simulations are not informed by structural data about the ternary complex interface from X-ray crystallography experiments. In particular, we employ the WES

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

variant 'Resampling of Ensembles by Variation Optimization' (or REVO),[36] which optimizes an objective function called the trajectory variation (see Methods). Here we compare the performance of the REVO algorithm with three different distance metrics: 1) differences in the warhead RMSD, (RMSD between the target binding domain of the PROTAC between the reference structure and a simulation frame, hereby called w-RMSD); 2) differences in target-ligase contacts; 3) a weighted combination of metrics 1, 2, and the differences in contacts between the target and the PROTAC for the PROTAC 2 system, hereby called the triple distance metric. The residues selected to compute the target-ligase contacts used in distance metrics 2 and 3 were selected based on residues that showed increased protection from hydrogen-deuterium exchange based on the HDX-MS experiments. Between three and seven REVO simulations are run for each distance metric on the SMARCA2-VHL-PROTAC 2 system, and a summary of their performance is given in Table 2.

Table 2: A summary of the performance of REVO simulations run with different distance metrics. Each REVO simulation ran with 48 walkers. The number of binding events (N_{binding}) counts the barrier crossings into the bound state, defined using an I-RMSD < 2.0 Å. The number of simulations with binding events (Sims. w/ binding) shows the probability of binding success. The total simulation time (Sim. time) aggregates the length of all trajectories in each REVO simulation.

Distance Metric	N_{binding}	Sims. w/ binding	Sim. time (μs)
W-RMSD	3378	2/3	5.8
Target-Ligase Contacts	4283	1/3	5.8
Triple	3278	6/7	12.5

Encouragingly, we observe a large number of binding events with all three distance metrics examined here. The w-RMSD and triple distance metrics found approximately the same number of binding events (about 3,300), whereas the target-ligase contacts distance metric sees over 4,000 binding events. We find that the low I-RMSDs are achieved early in the REVO simulations, with the minimum I-RMSD reaching ∼2.5 Å and stabilizing after about 500 ns of aggregate simulation time, whereas the vanilla MD simulation we ran plateaued at 10 Å after 1.35μs of simulation (Figure 3a,b). When comparing between the distance metrics both w-RMSD and the triple distance

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

metrics were able to find sub-2 Å structures within an aggregate simulation time of 200 ns, whereas it took the target-ligase contacts distance metric 800 ns to find structures of the same quality (Figure 3b). This indicates that using the degrader orientation to the binding site may be required to quickly and consistently generate low I-RMSD structures.

We performed clustering on all structures produced by the 6HAX simulations by a k-means clustering algorithm into 500 macrostates using the $C\alpha - C\alpha$ distances on residues determined from the HDX-MS experiments. Low I-RMSD states all have low values of w-RMSD (as expected) (Figure 3c). High free energy states have a large range of both I-RMSD values. However, the low free energy states are coincidentally below 10 Å. Five of the lowest free energy states have I-RMSD values below 3 Å.

Determining the I-RMSD of the ensemble is only possible when we have a crystal structure for the ternary complex. When trying to filter many possible degraders, it is not always feasible to solve these structures for every compound. Therefore, we need to rely on other physical quantities for predicting acceptable structures in such cases. We developed two definitions of the bound ensemble for the REVO simulations: 1) Using w-RMSD below 2 Å and 2) Using w-RMSD below 2 Å and more than 30 residue-residue contacts between the target and ligase residues that showed increased protection from hydrogen-deuterium exchange obtained from the HDX-MS experiments. The first definition was used on the simulations where we used the warhead distance metric and the second definition was used for the target-ligase contacts and triple distance metrics. Using these definitions of the bound ensemble to filter our simulations, the I-RMSD probability distributions for the REVO simulations with and without HDX-MS data to guide them were similar (Figure 5a), with an average value at 2 Å for both metrics. However, using this definition for the bound ensemble, 43% of structures that met this criteria had I-RMSD values below 2 Å, whereas the bound ensemble generated from REVO simulations not using HDX only selected conformations with an I-RMSD below 2 Å at 33% (Figure 5b). Both metrics had a high percentage of structures with I-RMSDs below 2.5 Å, and 97% of structures from the REVO simulations using

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

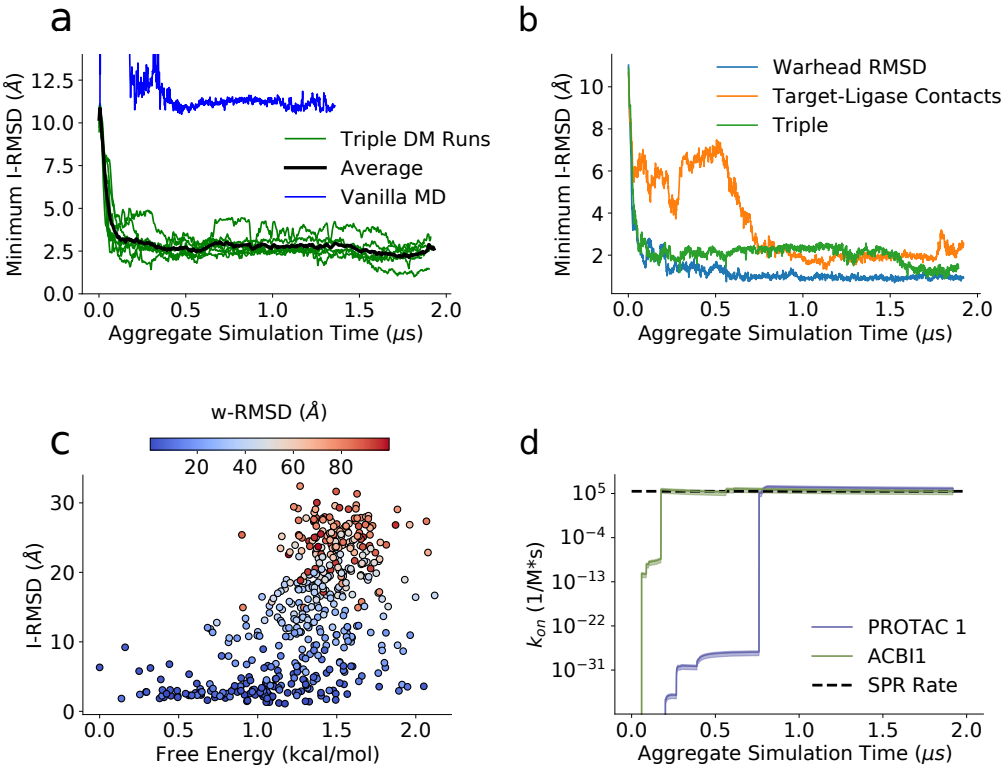

Figure 3: Comparing the w-RMSD, number of target-ligase contacts, and triple distance metrics (Linear combination of w-RMSD, target-ligase contacts and number of target-PROTAC contacts). (**a**) The minimum I-RMSD over time during the simulation for the triple distance metric. Each green line indicates one replica and the black line is the average between all runs. (**b**) The minimum I-RMSD for each distance metric. (**c**) A scatter plot of the free energy vs the I-RMSD after clustering the 6HAX simulations. The circles are colored by w-RMSD. (**d**) The predicted binding rates for PROTAC 1 system (purple) and the ACBI1 system (green). The black line is the experimental on rate determined via SPR.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

HDX-MS data had I-RMSD values below 3 Å. It is worth noting that both the target-ligase contacts and triple distance metric both use HDX-MS data to help guide the simulations. However, the target-ligase contacts metric did not produce low w-RMSD structures, the lowest being just below 6 Å and thus did not contribute the the bound ensemble via this definition.

All the above analysis was done using the PROTAC 2 system. We also performed three 1.96 μs simulations for the PROTAC 1 and ACBI1 systems using the triple distance metric, totaling 5.88 μs for each system. All the simulations for these two systems were able to produce structures of quality I-RMSD, the lowest being 0.69 Å for PROTAC 1 and 0.47 Å for ACBI1.

We next predict the on rates on the three different PROTACs (Table 3) using the flux into the bound state as defined when the state reaches an I-RMSD below 2 Å. For PROTAC 1 and ACBI1, our predicted rates are on the same order of magnitude as experiment (Figure 3d). For PROTAC 2, we were unable to experimentally determine the on rate so we simply report the rate predicted by simulation. However, for all three rates there were large errors. This is due to the weighted ensemble algorithm being slow to converge. To obtain better statistics, more simulation time is needed.

Table 3: Comparison of k_{on} rates between simulation and experiment for the ACBI1 PROTAC 1, and PROTAC 2 systems. The experimental rate for PROTAC 2 has not been determined yet.

PROTAC	Predicted Rate ($M^{-1}s^{-1}$)	Experimental Rate ($M^{-1}s^{-1}$)
ACBI1	$3 * 10^5 \pm 2 * 10^5$	$2.4 * 10^5$
PROTAC 1	$10 * 10^5 \pm 8 * 10^5$	$2.9 * 10^5$
PROTAC 2	$2.2 * 10^2 \pm 1.7 * 10^2$	N/A

Figure 4 shows an example of a structural prediction obtained for SMC2-PROTAC 2-VHL (PDB ID 6HAX[24]). The contact maps presented in Figure 4c have been obtained by the Arpeggio software[49] applied to the ternary interfaces. Each point on the contact maps reflects the degree of interaction. As can be seen from both the aligned prediction and co-crystallized structure (Figure 4d) and from the contact map (panel (Figure 4c), the accuracy of prediction is very high.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

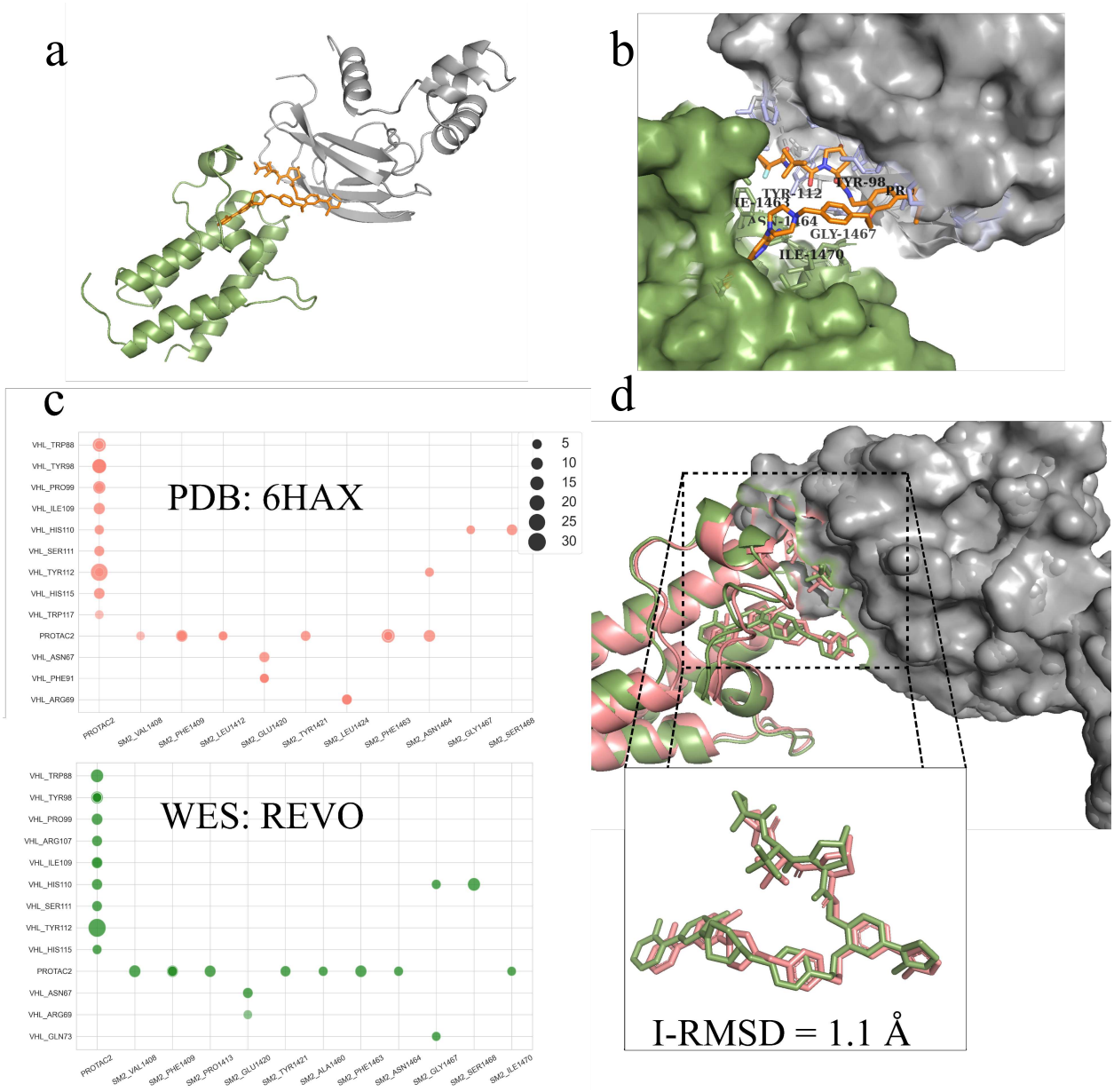

Figure 4: Illustration of the representative prediction produced by REVO simulation and its comparison to the co-crystallized structure (PDB ID: 6HAX) (**a**) predicted ternary structure with I-RMSD=1.1 Å; (**b**) detail of the binding interface; (**c**) contact maps for the interfaces of co-crystallized and predicted structures. The circle size reflects the number of atoms (including hydrogens) participating in interactions; (**d**) structurally aligned prediction (green) and co-crystallized structure (pink) with a detailed PROTAC 2 comparison shown.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The simulation results presented above come from the physics-based formulation of molecular dynamics. This level of detail and the associated accuracy of predictions comes with an elevated computational cost: three simulation replicas required around 300 A40 GPU hours. Thus, when the goal is to filter a large set of degrader molecule designs, less expensive methods with acceptable accuracy are needed.

2.4 HDX-MS improves prediction of ternary complex using docking

Molecular docking methods are based on three cornerstones: sampling of conformations, searching for optimal solutions and scoring of predicted binding modes. Advantages from the assumptions about sampling, searching and scoring come together with limitations on the accuracy of predictions: the quality of produced predictions tend to be variable and even when good predictions are present in a generated set, selecting them is a challenging problem. Here we demonstrate that including HDX-MS data in the form of restraints greatly improves the accuracy of docking predictions.

Following the idea of using HDX-MS to interrogate degrader-induced ternary complexes in solution,[50] we identify the interface residues as those having the largest change in deuterium uptake. In contrast to recent work,[50] 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 experiments,[51] our restrained docking approach is based on the statistics of the length of the linker in a degrader molecule.

We have performed docking experiments with 4 systems: 1) PROTAC 2 in complex with the bromodomain of human SMARCA2 and pVHL (PDB ID: 6HAX[24]); 2) PROTAC 1 in complex with the bromodomain of human SMARCA2 and pVHL (PDB ID: 6HAY[24]); 3) ACBI1 complex with the bromodomain of human SMARCA2 isoform-1 and pVHL; 4) ACBI1 complex with the bromodomain of human SMARCA2 isoform-2 and pVHL (PDB ID: 74SE, this publication). The assessment metrics used

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

are the standard CAPRI parameters: fnat – fraction of native contacts, L-RMSD – ligand RMSD, and I-RMSD – interface RMSD.[52] I-RMSD has been chosen to compare the accuracy of predictions produced by docking and REVO for the first system (results for the other systems and the corresponding CAPRI metrics are reported in the Supplementary Materials).

Figure 5 presents the values of I-RMSD calculated for structures predicted by docking and REVO in two separate scenarios: with and without the information from HDX-MS experiments. Panel **b** shows the fraction of predicted structures that have values of I-RMSD within the specified limits. REVO simulations consistently outperform docking in the accuracy of predictions. Also, both computational methods demonstrate improved accuracy when the information from HDX-MS experiments is used. Interestingly, the relative performance of the docking protocol is improved more by the availability of the restraints derived from HDX-MS as compared with the REVO method, as seen by the gaps between orange and green bars are larger than the corresponding gaps between the black and blue bars. However, even without HDX data the REVO simulations still outperform the docking calculations *with* HDX data. Similar observations are made when the whole sets of Top-100 docking predictions and the bound ensemble from REVO simulations are considered – see Figure 5a. The spread in distributions for the docking predictions illustrate the limitations of semi-flexible docking and the used scoring method. At the same time, the docking protocol used here required around 75 CPU hours for 3 independent replicas, which is significantly less computationally expensive than the 300 A40 GPU hours required by the REVO approach. The achieved accuracy of the docking protocol in combination with information from HDX-MS makes it a useful tool for quick filtering of large numbers of degrader designs.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

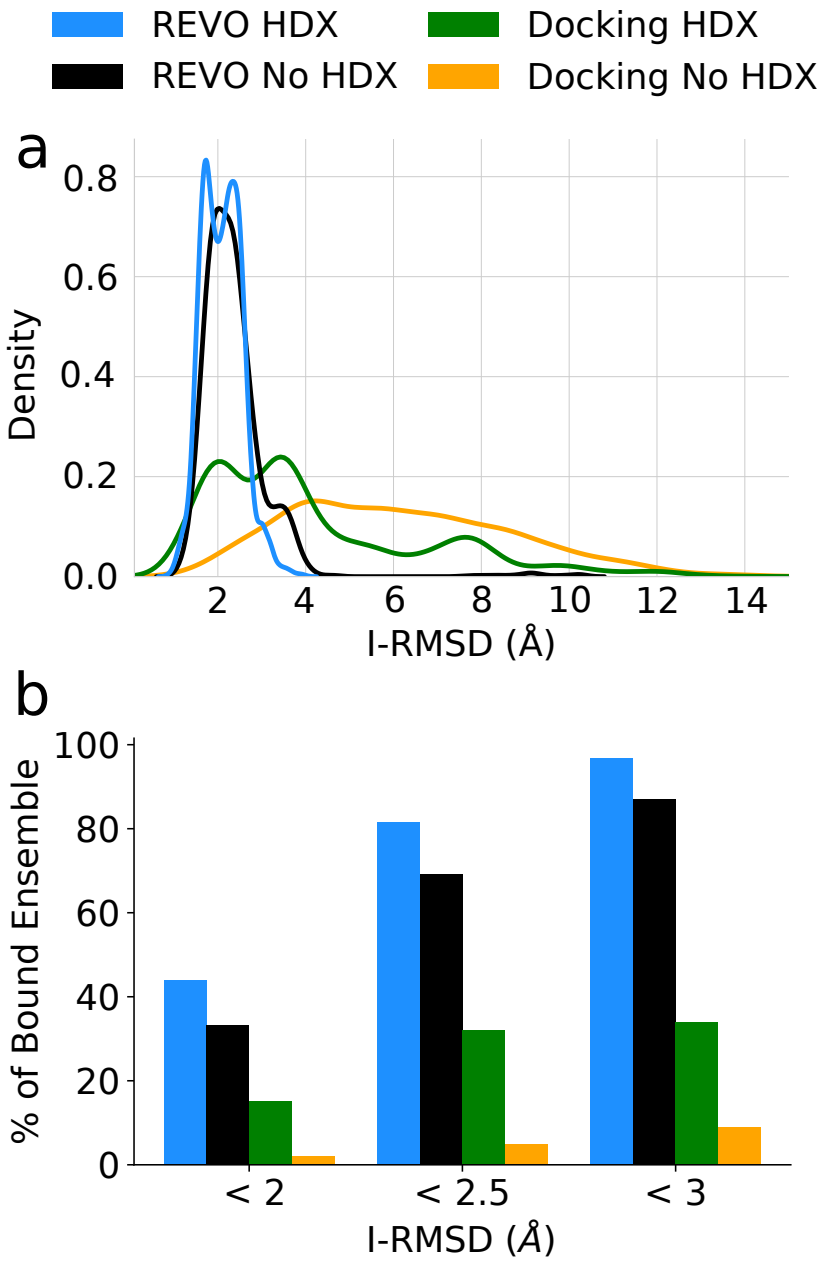

Figure 5: Comparing the bound ensembles determined by docking and REVO simulations with and without information from HDX-MS for the PDB ID 6HAX ternary complex. The REVO bound ensemble is defined as structures below a warhead RMSD of 2 Å and more than 30 contacts between the target and ligase interface. The docking bound basin is defined as the 100 top structures as determined by Rosetta-scoring. (**a**) Probability density function distributions of I-RMSD values for the bound ensembles. (**b**) The percent of structures in the predicted bound ensembles below specific I-RMSD thresholds (2 Å, 2.5 Å, and 3 Å).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2.5 Flexibility of Ternary Complex Measured by SAXS and HREMD Simulations

HDX-MS measurements revealed substantial flexibility of protein complexes studied here. To further study the conformational heterogeneity of the ternary complexes, we collected small-angle x-ray scattering (SAXS) data of SMARCA2-iso1/iso2:ACBI1:VCB, abbreviated to Iso1 and Iso2 complexes hereafter respectively, in solution near physiological condition. We ensured the resulting scattering profile was solely due to a complex with all four protein chains and a degrader, and devoid of contributions from binary or uncomplexed protein using size exclusion chromatography coupled to SAXS (SEC-SAXS). The elution peaks 1 and 2 of the SEC profile were assigned to the ternary complex and uncomplexed or binary complex respectively (Supplementary Figure 10). The final SAXS profile of each complex (Figure 6**a**) was determined from the average scattering signal from the sample in the elution peak 1, where we see a variability in calculated radius of gyration, R_g (red solid/open circles in Supplementary Figure 10). Therefore, SAXS data also confirms the dynamics of Iso1 and Iso2 complexes. Using Guinier approximation (Eq. 1), we determined R_g as 33.4±0.4 Å and 32.3±0.3 Å for Iso1 and Iso2 respectively from experimental SAXS data (Supplementary Figure 11). SAXS data only provides ensemble-averaged structural information (i.e., R_g, overall size), but it lacked any atomistic detail of protein-protein interactions or contacts.

We thus performed Hamiltonian replica-exchange MD (HREMD) simulations to explore the conformational free energy landscape of Iso1, Iso2 and other protein complexes (see Supplementary Table 2), and to further understand the protein-protein interactions. First, we validated HREMD-generated ensembles of two large protein complexes (Iso1 and Iso2) by directly comparing to the measured SAXS data (Figure 6**a**). An excellent agreement between the experimental and calculated SAXS profiles ($\chi^2 = 1.55$ and 1.23 for Iso1 and Iso2, respectively) ensured that the unbiased HREMD simulations of protein complexes are accurately capturing the long timescale conformational ensembles to experimental accuracy. χ^2 is defined in Eq. 11. The HREMD-generated

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ensembles suggest each of the complexes is flexible in solution leading to a change in overall conformation compared to starting structure (homology model of Iso1 and crystal structure of Iso2, respectively) as confirmed by the histograms of R_g (Figure 6**b**). R_g from simulation is calculated from atomic coordinates using Eq. 2.

2.6 Conformational sampling of ternary complexes

Our HDX-MS and SAXS data suggest that the protein complexes studied here are dynamic and sample several distinct bound conformations. We use HREMD simulations to identify these structures and quantify the free energy landscape of these complexes. First, we perform principle component analysis (PCA) of the interface distances observed in our HREMD simulations in order to reduce the dimensionality of the simulation data. The probability distribution of the highest-variance features allow us to measure a more easily interpretable free energy landscape from our simulation data than would be possible otherwise. We find that the landscape of each protein complex contained several local minima differing by only a few kcal/mol.

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, see Supplementary Figure 14. The clusters identified by k-means are consistent with our HDX-MS protection data. Figure 7 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 identified by k-means. Notably, this analysis shows that in the second most populated structure of Iso1-ACBI1-VCB, the helix formed by the 17 residue extension of isoform 1-SMARCA2 interacts with a beta sheet of VHL, Figure 7d, in accordance with HDX-MS experiments that found this beta sheet to be protected in presence of Iso1, but not in the presence of Iso2. Similarly, highly populated structures of Iso2-ACBI1-VHL and Iso2-PROTAC2-VHL show contact between residues that were observed to be protected in HDX-MS experiments with these PROTACs, but not with PROTAC 1, while the most populated structure of PROTAC 1 does not show these contacts.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

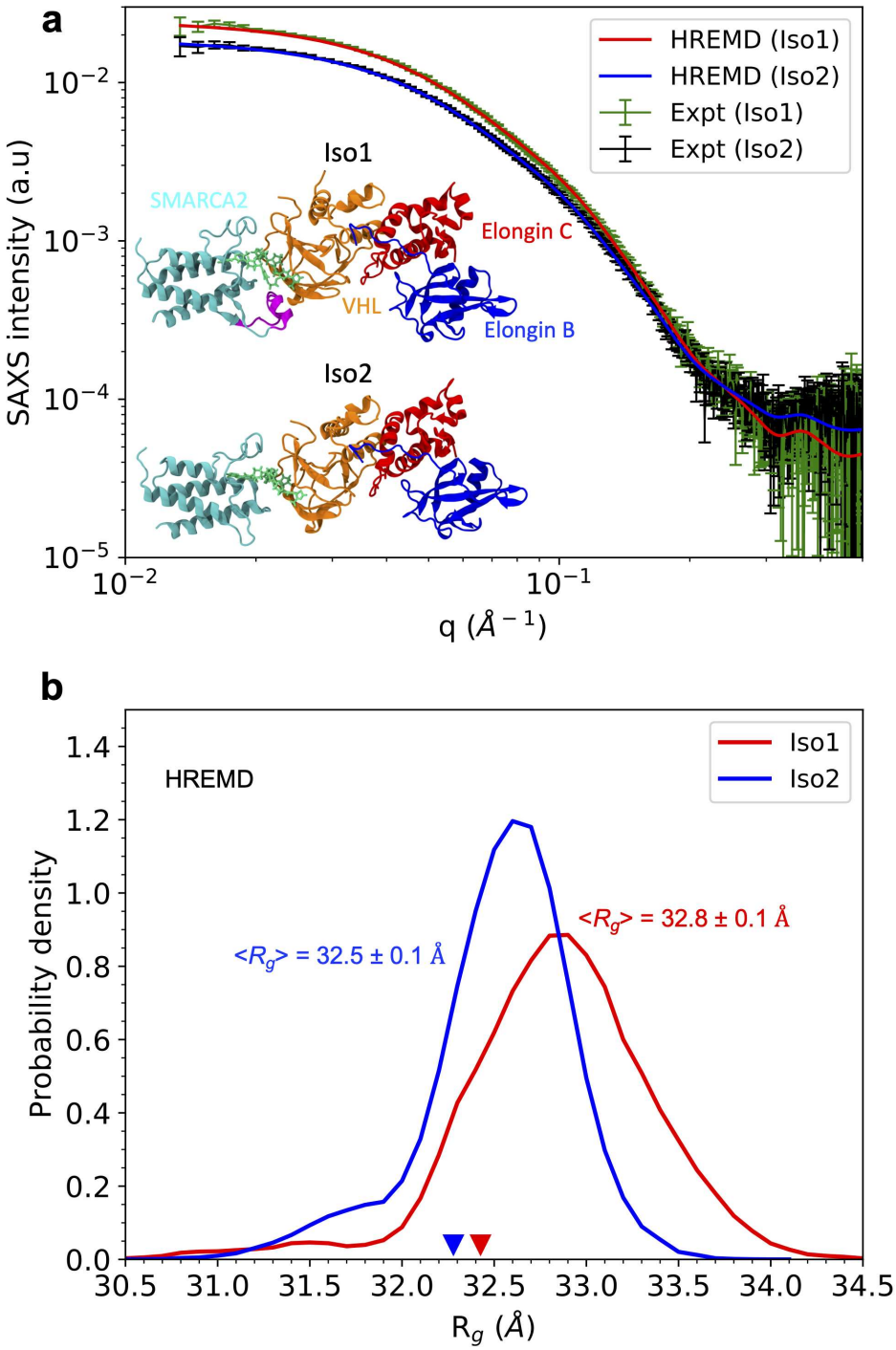

Figure 6: SAXS profiles and structural ensembles of Iso1 and Iso2 complexes. **a** Comparison of theoretical and experimental SAXS profiles, $< I(q) >$ vs. q. **b** The histograms of R_g of Iso1 (red) and Iso2 (blue) complexes calculated from HREMD simulations. The inverted triangles are the R_g values of starting structures of Iso1 (red) complexes and Iso2 (blue) from homology model and crystallography respectively.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

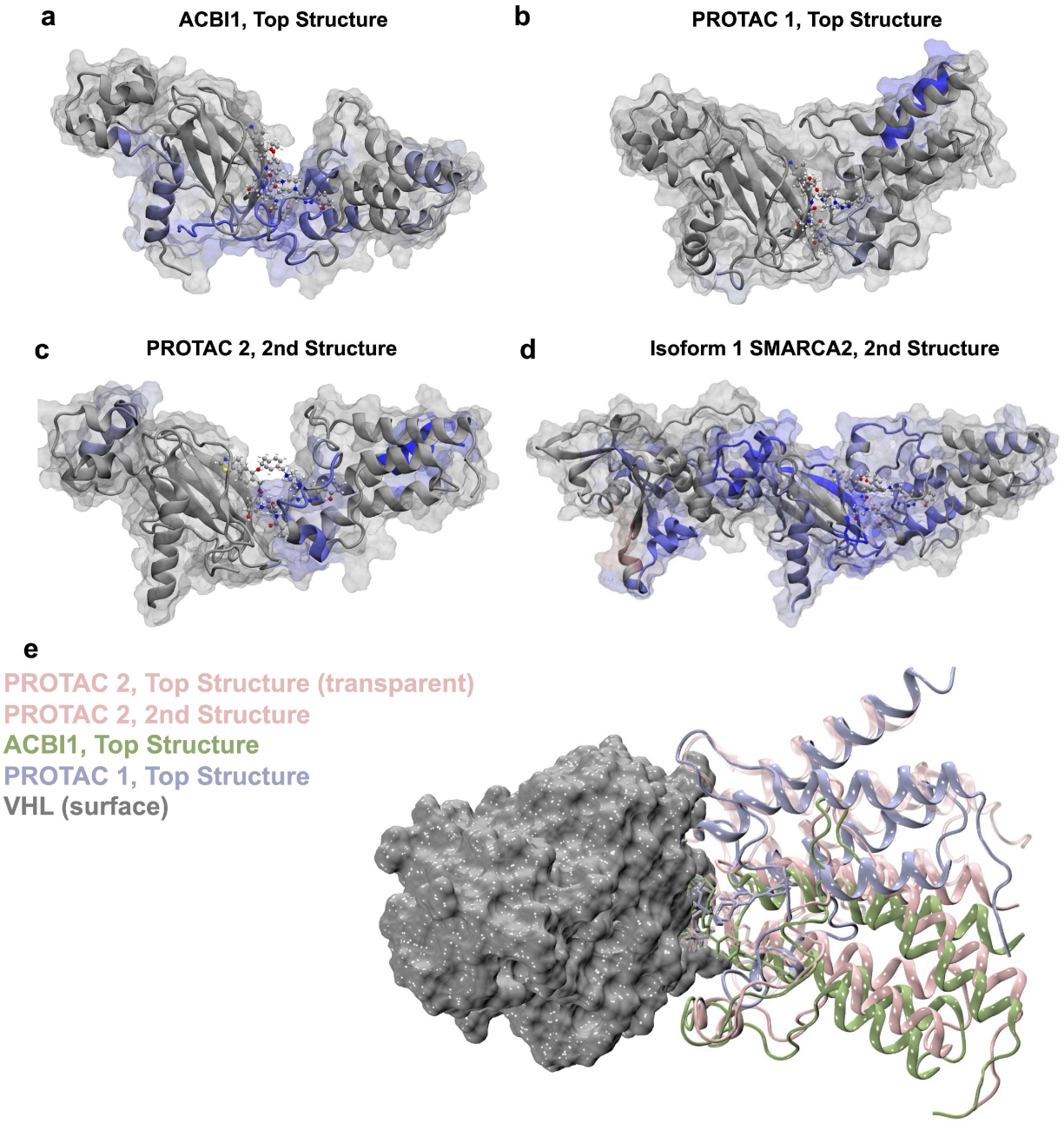

Figure 7: Most populated structures of SMARCA2 bound to VHL with different degrader molecules, identified by dimension reduction and clustering of HREMD simulation data. **(a-d)** Colors of VHL and SMARCA2 represent HDX-MS protection in the presence of the degrader molecules relative to the situation in the absence of the degrader. The second ranked structures of **c** PROTAC 2 and **d** isoform 1 SMARCA2 are displayed that support HDX-MS data, whereas the top three structures are included in Supplementary Figure 15. Elongin B and Elongin C are also included in panel **d**. **e** The top structures of ternary complexes are compared after aligning VHL to illustrate conformational differences among top structures of ternary complexes.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

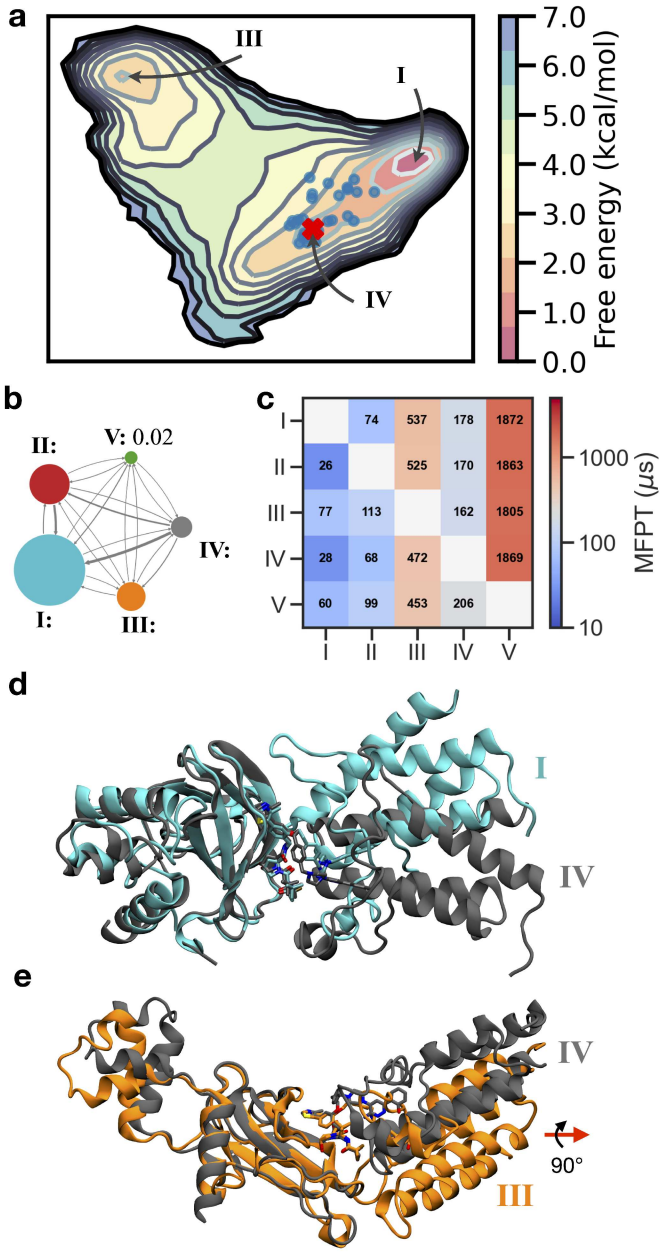

Figure 8: **a** Conformational free energy landscape as a function of the first two tICA features of SMARCA2-PROTAC2-VHL ternary complex inferred from a Markov state model (MSM). The ensemble of bound states from REVO simulations is shown as blue points; the crystal structure (PDB ID 6HAX) is shown as a red X. In this projection, states II and V are close to state I. **b** Network diagram of the coarse-grained MSM calculated using a lag time of 50 ns, with the stationary probabilities associated with each state indicated. **c** Mean first-passage times to transition (MFPT) from one state in the MSM to another. Numbers indicate predicted MFPTs in μs. **d-e** Comparison of the crystal structure (gray) with the lowest free energy state (cyan) and a metastable state (orange) predicted by the MSM. Arrows indicate a change of orientation relative to **d**.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We selected 98 representative structures from HREMD data to use as initial configurations for Folding@home (F@H) simulations of Iso2-PROTAC2-VHL. Each initial condition was cloned 100 times and run for $\sim$ 650 ns, for a total of $\sim$ 6 ms of simulation data. These independent MD trajectories provide the basis for fitting a Markov state model (MSM),[53] which provides a full thermodynamic and kinetic description of the system and allow for the prediction of experimental observables of interest.[38] First, we used time-lagged independent component analysis (tICA)[54] to determine a projection of the underlying MD data. The distance between points in the tICA feature space corresponds roughly to kinetic distance.

The MSM predicts a stationary probability distribution on tICA space that is in general different from the empirical distribution of our simulation data. Interestingly, the MSM predicts that the the crystal structure is 1.5 kcal/mol higher in free energy than the global free energy minimum, while the bound structures obtained from our REVO simulations are $\sim 1.5 - 3.5$ kcal/mol above the global minimum, Figure 8**a-c**. The model also predicts a metastable state with free energy 2.2 kcal/mol (Figure 8**e**).

This model is coarse-grained to obtain a five-state MSM, of which the following three states are of particular interest: the global minimum state (or state I) with a stationary probability of 0.63, the 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. The global minimum state differs from the crystal structure 6HAX by an I-RMSD of 3.6 Å, while the metastable state has an I-RMSD of 4.8 Å relative to the crystal structure. The global minimum state is stabilized by a large number of protein-protein contacts (Supplementary Figure 16). 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 PROTAC. On the other hand, the metastable state lacks contacts between PROTAC 2 and ARG29, ASN90, and ILE96 of SMARCA2. The area of the binding interface was substantially increased in both the metastable and global minimum states relative to the crystal structure: the global minimum state had a buried surface area of 2962 $Å^2$, compared to 2800 $Å^2$ for

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the metastable state and 2369 $Å^2$ for the crystal structure.

2.7 Identifying the ubiquitination zone for Cullin Ring Ligase with VHL and SMARCA2

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 do this, we study different degrader molecules in the context of the full Cullin-RING E3 ubiquitin ligase (CRL) along with the POI. Here, we study the position different solvent-exposed Lys residues from the POI relative to the ubiquitination zone of the CRL macromolecular assembly, specifically focusing on the probability of POI lysine residue density within this zone. The hypothesis is that a higher density of POI Lys residues in the ubiquitination zone would lead to a higher chance of ubiquitination. As such, this analysis can provide insights on the degradation potency of degrader molecules. First we built 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-βTrCP as reference[55] (see Methods). Second, we used the meta-eABF simulation approach (see Methods) to sample CRL open-closed conformations in the presence of SMARCA2. These conformations were then used as reference states to superimpose structures from HREMD simulations of VHL:degrader:SMARCA2 on the active state of the CRL-VHL, allowing us to obtain Lys densities from SMARCA2 in the ubiquitination zone of the CRL-VHL. Comparing the Lys densities of the three degraders (9), we observe that ACBI1 places the most Lys density in the ubiquitination zone of CRL-VHL, followed by PROTAC 2 and PROTAC 1. This order of Lys density in the ubiquitination zone agrees with the experimentally observed degradation data.[18]

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

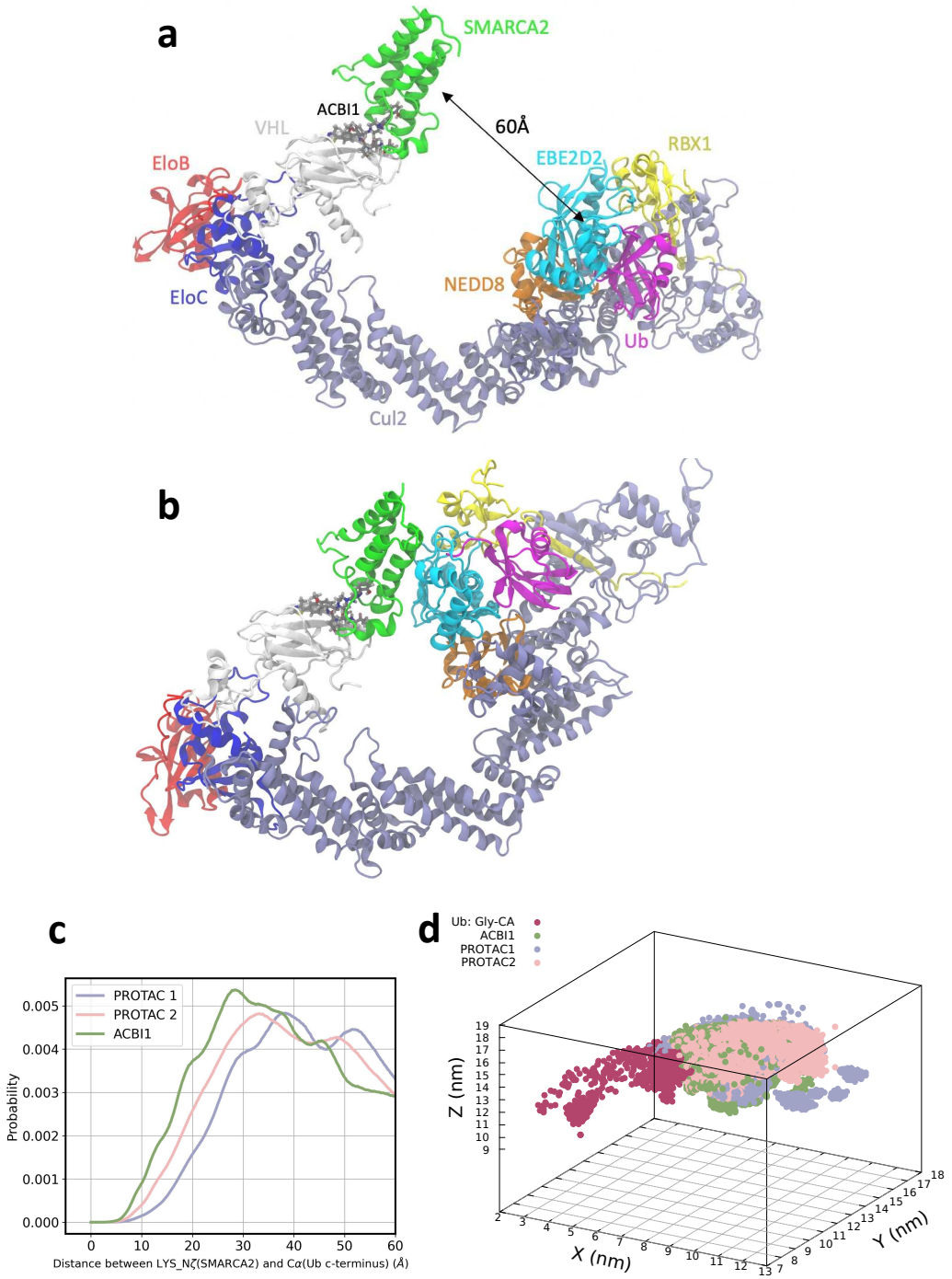

Figure 9: Degrader dependent SMARCA2 Lys densities in the CRL-VHL ubiquitination zone. **a**. Active form of CRL-VHL with bound SMARCA2 and E2-ubiquitin with open CRL conformation. **b**. same as **a** with a closed conformation of CRL generated by meta-eABF simulations. **c**. Distance of Lys residues (side-chain nitrogen atom) from SMARCA2 to the C-terminus glycine Cα of ubiquitin for three different degraders. **d**. Density of Lys residues in 3D space near ubiquitination zone of CRL-VHL.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

3 Discussion

Ternary complex formation is a critical step in the process of targeted protein degradation. However, studying the dynamic nature of ternary degrader complexes has posed challenges to the field due to the size of the system, degrees of flexibility, timescales for biological motions, and limited solution-phase experimental data. Here, we studied the structure and dynamics of three different degrader molecules of SMARCA2 that have similar thermodynamic binding profiles and crystal structure conformations but different degradation efficiencies. We solved the crystal structure of the ternary complex bridged by ACBI1 (PDB ID: 7S4E), which revealed a potential structural and dynamic aspect of the degradation profiles, which led to a series of solution-phase experimental studies and molecular dynamics (MD) simulations in an effort to explain and predict the degradation profiles. Hydrogen deuterium mass spectrometry (HDX-MS) and small-angle X-ray scattering (SAXS) coupled with extensive MD simulations enabled the generation of atomic resolution representations of the dynamic ensembles. We highlighted the conformational landscape of SMARCA2:PROTAC 2:VHL, for which we carried out 6 ms of accumulated simulation time using the Folding@home distributed supercomputer cluster. We propose an enhanced-seeding method that includes HREMD simulations of the ternary complex, extraction of a modest number of seed structures (100 in the work presented here), and independent simulations starting from each seed using Folding@home. We applied time-lagged independent component analysis (tICA) for dimensionality reduction and built a Markov state model (MSM) to assess the conformational free energy landscape and transitions between low-energy states. We found that the experimentally determined crystal structure is close to, but not coinciding with, the global minimum in the free energy landscape, although it is within 2 kcal/mol from the global minimum. By coarse-graining the MSM, we were able to identify five low-energy structures that contain different protein-protein interfaces between SMARCA2 and VHL. A less computationally costly solution was also explored, where we found that only 0.5 μs of lowest rank/unbiased replica of

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

HREMD (aggregate 12 μs with 24 replicas) gave us a qualitatively similar conformational landscape with the same global minimum, albeit a lower resolution free energy surface. Thus, we proceeded to run HREMD on other systems of interest (PROTAC 1, ACBI1, and Isoform 1 of SMARCA2; see Supplementary Table 2 for number of replicas used and aggregate length for each system). Simulation analyses of the most probable structures for each of these ternary complexes show that ACBI1 and PROTAC 1 have different orientations of SMARCA2 relative to VHL, with PROTAC 2 sampling both orientations. We also cross-validated these HREMD simulations by comparing with SAXS data, where we observed excellent agreement. We propose that this sampling procedure can be replicated for other target and E3 ligase complexes, since no target-specific information was used for the simulations in this work other than the starting x-ray structure coordinates.

We then explored the prediction of the ternary complex itself, using the binary structures of the known binding mode of the ligase ligand binding to VHL and warhead binding to SMARCA2, which is typically known in advance of designing heterobifunctional degrader molecules. We developed a protocol using Weighted Ensemble Simulations (WES) that is able to produce a 3-dimensional structural ensemble of high accuracy structures using the RMSD to the bound pose of the warhead as collective variable: we find structures less than 2 Å I-RMSD within 2-3 kcal/mol from the most probable conformation of the ternary complex. The addition of experimental solution-phase HDX-MS protection factor data, which identifies residues that are most likely to be precluded from solvent interactions in the context of the ternary structure, further increased the quality of the predictions. Similarly, we discovered that HDX-MS data can improve docking results, although, overall docking has lower quality than the WES simulations, likely due to the minimal sampling of internal protein degrees of freedom and lack of explicit solvent. We made publicly available all relevant simulation data and the source code for running WES+HDX.

To further validate the WES+HDX procedure, we performed a prospective prediction of the ternary complex of isoform 1 SMARCA2, with a 17 amino acid extension,

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the structure of which was previously unknown. Our HREMD simulations agree with experimental SAXS data and our ternary complex predictions suggest ways that the SMARCA2 extension is interacting with a beta-strand from VHL, explaining the observed HDX-MS protection pattern. The WES+HDX method provides an opportunity to visualize, at an atomistic level, the molecular interactions that guide ternary structure formation for complexes with previously known crystal structures. The ultimate goal is to uncover the solution-phase structural ensembles adopted by the target ligase pair, which appear to extend beyond what is observed using conventional structure determination methods. This information may provide a better representation of the factors that influence ternary complex formation, ultimately leading to downstream ubiquitination. Moreover, knowledge of critical atomic interactions provides a basis for alternative strategies to improve degrader designs, such as modifying linkers to induce specific conformational ensembles of the ternary complex that are associated with higher degradation. WES+HDX is also able to recapitulate experimental K_{on} for ternary and binary complex formation.

Finally, incorporating all of our integrative efforts, we propose a model of the VCB Cullin-RING E3 ubiquitin ligase (CRL) complex that we use to predict the impact of different degraders on the proximity of solvent-exposed lysine residues of SMARCA2 to approach the E2-loaded ubiquitin on the CRL. We built an active model of CRL with VHL and SMARCA2, and used the MD simulation method meta-eABF to sample the distance distribution between ubiquitin in the E2 enzyme and SMARCA2. We find that the conformations sampled in the ternary complex with the ACBI1 degrader place more of the exposed lysine residues on SMARCA2 within the ubiquitination zone of the CRL, as compared with those sampled by complexes formed with PROTAC 1 or PROTAC 2. It is thought-provoking to speculate that the differences in conformational ensembles induced by each of these degraders may produce differential rates of ubiquitination, leading to the observed differences in degradation. Indeed, these three heterobifunctional degrader molecules are highly similar, with identical POI warheads and E3 ligands – the only difference being a few atoms in the linker – yet there are dif-

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ferences in degradation despite the similarities in binding thermodynamics and X-ray crystal structures. To further validate these results, it will be important to understand how each of these degrader complexes influence the location and magnitude of the ubiquitination of the POI. Future efforts will also include studies of more degrader molecules for the SMARCA2/VHL system and the addition of other POI/E3 pairs.

4 Methods

4.1 Cloning, expression and purification of SMARCA2 and VHL/EloB/C

The SMARCA2 gene from Homo sapiens was custom-synthesized at Genscript with N-terminal GST tag (Ciulli 2019 Nature ChemBio) and thrombin protease cleavage site. The synthetic gene comprising the SMARCA2 (UniProt accession number P51531-1; residues 1373-1511) 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, 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 tag[18] 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. ElonginB and ElonginC gene from Homo sapiens was custom-synthesized with AVI-tag at C-terminus of EloB.[23] 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 protein expression, the plasmid was transformed into BL21(DE3) 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 x 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 minutes 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 5,000 xg 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 ^{o}C, and the crude extract was centrifuged at 39,000 xg (JA-17 rotor, Beckman-Coulter) for 30 min at 4 ^{o}C. Two ml Glutathione Sepharose 4 B (Cytiva) was added into the supernatant and mixed at 4 ^{o}C for 1 hour, 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 ^{o}C overnight during dialysis against 1 L of Buffer B (20 mM HEPES, pH 7.5, 150 mM NaCl, 1mM 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 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 ^{o}C.

For VHL/EloB/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 x 1-liter of Terrific Broth medium containing 50 μg/ml kanamycin and 50 μg/ml streptomycin and grown at 37 °C with

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

aeration to mid-logarithmic phase (OD600 = 1). The culture was incubated on ice for 30 minutes 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 5,000 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 ºC, and the crude extract was centrifuged at 17000 g (JA-17 rotor, Beckman-Coulter) for 30 min at 4 ºC. Ten ml Ni Sepharose 6 FastFlow (Cytiva) was added into the supernatant and mixed at 4 ºC for 1 hour, 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 ºC 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 3ml 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/EloB/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 ºC. For SPR assay, 10 mg VHL/EloB/C protein complex was incubated with BirA (1:20), 1 mM ATP and 0.5 mM Biotin and 10mM MgCl2 at 4 ºC 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.

4.2 Hydrogen Deuterium Exchange Mass Spectrometry

Our HDX analyses were performed as reported previously with minor modifications.[41,43,56] HDX experiments were performed using a protein stock at the initial concentration of 200 μM of SMARCA2, VCB in the APO, binary (200 μM PROTAC ACBI1) and ternary (200 μM PROTAC ACBI1) states in 50 mM HEPES, pH 7.4, 150 mM NaCl,

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 mM TCEP, 2% DMSO in H2O. The protein samples were injected into the nanoACQUITY system equipped with HDX technology for UPLC separation (Waters Corp.[57]) 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 1:13 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 X 30-mm Enzymate-immobilized pepsin column (Waters Corp.) at 100 μl/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 X 30 mm) (Waters Corp.) for desalting with 0.1% formic acid in H2O and then subsequently separated with an in-line 1.8µMHss T3 C18 2.1 X 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 μl/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 MS1 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

proteins (SMARCA2, VHL, ElonC, ElonB). DynamX was utilized to generate the relative deuterium incorporation plots and HDX heat map for each peptide. 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. All experiments were made under the same experimental conditions negating the need for back-exchange calculations but therefore are reported as relative.[46] 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. Mean relative deuterium uptake thresholds were calculated as described previously.[41,43,56] Differences in deuterium uptake that exceeded the error of the datasets were considered significant.

4.3 Structural Determination of SMARCA2:ACBI1:VHL 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 ACBI1 for 1 hour 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 hours 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 was collected at NSLS2 beamline FMX (λ=0.97932 Å) using an Eiger X 9M detector.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Crystals were found to be in the P 21 21 21 space group with unit cell dimensions of a= 80.14, b= 116.57, c= 122.23 Å, where α= β= γ=90°. Crystal contained two copies of the SMARCA2:ACBI1:VCB (VHL, ElonC, ElonB) complex within the asymmetric unit cell. The structure was solved by performing molecular replacement with CCP4i2[58] PHASER using PDB ID 6HAX as the replacement model. MR was followed by iterative rounds of modeling (COOT[59]) and refinement (REFMAC5[60–68]) by standard methods also within the CCP4i2 suite. Structures were refined to R_{work}/R_{Free} of 23.7%/27.5%.

4.4 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 Å. We measured two protein complex samples,

(i) SMARCA2-iso1:ACBI1:VCB (Iso1), and

(ii) SMARCA2-iso2:ACBI1:VCB (Iso2).

The scattering data was detected on PILATUS3 300 K (Dectris) detector with a resulting q range of 0.0134 – 0.5793 Å^{-1}. The SEC-SAXS data for each sample was collected by loading 500 μL of the ternary complex formed by addition of equimolar concentrations (275 μM) of SMARCA2, VCB and ACBI1, 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 second 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. Data reduction was performed using the BioXTAS RAW 2.0.3 software.[69] R_g was estimated from experimental an SAXS curve using the Guinier approximation,

$$I(q) \approx I(0)e^{\frac{-q^2 {R_g}^2}{3}}, for\ q \to 0 \tag{1}$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

where $I(q)$ and $I(0)$ are the measured SAXS intensity and forward scattering intensity at q=0, respectively. q is the magnitude of scattering vector given by, $q = 4\pi sin2\theta/\lambda$, where 2θ is the scattering angle and λ is the wavelength of incident beam. The linear region in $ln(I(q))$ vs. q^2 was fitted at low-q values such that $q_{max}R_g \leq 1.3$ to estimate R_g, where q_{max} is the maximum q-value in the Guinier fit (Fig. S3). On the other hand, R_g of the protein complex in simulation was directly calculated from atomic coordinates using following relation,

$$R_g = \sqrt{\frac{\sum_i m_i \left\|\mathbf{r_i}\right\|^2}{\sum_i m_i}} \tag{2}$$

where m_i is the mass of ith atom and $\mathbf{r_i}$ is the position of ith atom with respect to the center of mass of the molecule.

4.5 Molecular dynamics simulations

The initial coordinates of the system were obtained from X-ray crystal structures PDB ID 6HAX, 6HAY, or 7S4E, respectively. The missing atoms were added using the LEaP module in AMBER20. The AMBER ff14SB force field[70] was employed for the protein and the ligand 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 was 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 2 fs time step could be achieved.

A step-wise equilibration protocol was used prior to running the production phase of the Molecular Dynamics 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 decent 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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.

4.6 Isoform 1 homology model

Since no suitable X-ray structure for SMARCA2 isoform 1 BRD domain 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 SMARCA2 isoform 1 from its amino acid sequence.This was used as a model for the binding of the warhead to SMARCA2 isoform1, although the final prediction reported here came from our simulations. The sequence that was used is Uniprot P51531-1 (residues 1373- 1493) 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

4.7 REVO-epsilon Weighted Ensemble method

To observe binding of the VHL-PROTAC complex and SMARCA2 we apply a variant of the weighted ensemble algorithm REVO. Each cycle of the REVO algorithm is comprised of two parts: semi-independent MD trajectories performed in parallel and resampling. Each of the MD trajectories (called "walkers") has a statistical weight (w) that contributes to statistical observables. After a trajectory time of τ we perform resampling. In resampling similar walkers are merged together and unique walkers are

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

cloned, as defined by a distance metric. During cloning, the weight is evenly divided between the resultant clones and, when walkers are merged, the weights are combined to ensure the conservation of probability.

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:

$$V = \sum_i V_i = \sum_i \sum_j \left(\frac{d_{ij}}{d_0}\right)^{\alpha} \phi_i \phi_j \tag{3}$$

where V_i is the 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^{-50}. The walker with the highest variance, V_i and when the weights of the resultant clones would be larger than p_{min}, and is within distance ϵ of the walker with the maximal progress towards binding of the ternary complex was proposed to be cloned. The two walkers selected for merging were within a distance of 2 $\mathring{(}$A) 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:

$$\frac{V_j w_i - V_i w_j}{w_i + w_j} \tag{4}$$

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. After resampling is complete, we begin a new cycle.

Three different distance metrics were used while simulating the PROTAC 2 system: Using the warhead RMSD to the crystal structure, maximizing the contact strength between protected residues identified by HDX data, and a linear combination of the

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

warhead RMSD, contact strength between HDX-protected residues, and the contact strength between SMARCA2 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, 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: $d = |\frac{1}{RMSD_i} - \frac{1}{RMSD_j}|$. 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:

$$strength = \frac{1}{1 + e^{-k(r-r_0)}} \tag{5}$$

where k is the steepness of the curve, r is the minimum distance between any 2 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 the walkers was calculated by: $d = |cs_i - cs_j|$ where cs is the contact strength of a given walker.

All REVO simulations were run using OpenMM v.7.5.0. The timestep for each simulation was 2 fs. To ensure our simulations maintained 300 K, we used the Langevin integrator found in OpenMM with a friction coefficient of 1 ps^{-1} which was coupled to a Monte Carlo barostat set to 1 atm and volumes moves were attempted every 100 fs to maintain a constant pressure. The nonbonded forces were computed using the CutoffPeriodic method in OpenMM using a cutoff of 10 Å. The atomic positions and velocities were saved every 10,000 timesteps or 20 ps of simulation time, which is the resampling period (τ) used here. The degrader-VHL complex was constrained to maintain the complex during the simulation by using a OpenMM custom centroid force defined as:

$$CentroidForce = k * (dist - edist)^2 \tag{6}$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

4.8 Ternary complex docking protocol

For the purpose of quick filtering through a large number of degrader designs, we take advantage of the conventional restriction of molecular flexibility used in molecular docking methods. Following[28,30] (Methods 4 and 4b) and,[27] we assume that high fidelity structures of POI:warhead (i.e., SMC2:PROTAC binding moiety) and E3:ligand (i.e., VHL:PROTAC binding moiety) are known and available to be used in protein–protein docking. This docking of two proteins with bound PROTAC moieties is performed in the absence of the linker. The conformations of linker are sampled independently with an in-house developed protocol that uses implementation of fast quantum mechanical methods, CREST.[71–73] Differently from the docking protocols described in,[27,28,30] we make use of the 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 values of 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 software:[74,75]

$$f_1(x) = (\frac{x - x_0}{sd})^2, \tag{7}$$

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 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:

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$f_{2,i}(x) = \frac{1}{1 + \exp(-m \cdot (x - x0))} - 0.5, \tag{8}$$

where i is the index of a protected residue, $x0$ is the center of the sigmoid function and m is its slope. As above, $x0$ 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 Cα atom for each protected residue and the chain-ID of partnering protein (i.e., x in Eq.(8) is the distance of Cα atom from the partnering protein). Thus, the total restraint-term used in docking takes the form:

$$f_{restr.}(x) = w \cdot (f_1(x) + \sum_i f_{2,i}(x)), \tag{9}$$

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. (9), bias sampling towards those docking poses that are compatible with specified restraints. This allows to limit the number of output docking structures, as only those ones that pass the Metropolis criterion with the additional term of Eq. (9) 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 predictions.[27] 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 Supplementary Material.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

4.9 HREMD simulation

The details of Hamiltonian replica-exchange MD (HREMD)[76,77] can be found in the Supplementary Material (Supplementary Figures 12 and 13, and Tables 2 and 3). For all HREMD simulations, we chose the effective temperatures, T_0 = 300 K and T_{max} = 425 K such that the Hamiltonian scaling parameter, λ_0 = 1.00 and λ_{min} = 0.71 for the lowest and the highest rank replicas respectively. The effective temperatures of intermediate replicas are listed in Supplementary Table 3. 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:degrader:VHL and SMARCA2:degrader:VCB respectively. Each simulation was run for 0.5 μs/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 (Supplementary Table 3), (ii) a good overlap of histograms of potential energy between adjacent replicas (Supplementary Figure 12), and (iii) a mixing of exchange of coordinates across all the replicas (Supplementary Figure 13).

4.10 Conformational free energy landscape determination

In order to quantify to 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 5 Å of each other during the simulation trajectory, where the distance between two residues was defined as the distance between their closest heavy atoms. PCA was then used to identify the features that contributed most to the variance by diagonalizing the covariance matrix; for each simulated system, the number of features used in our analysis was chosen as that which explained at least 95% of the variance.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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. Interestingly, 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.

To prepare the Folding@home simulations, HREMD data was featurized with interface distances and its dimensionality reduced with PCA as described above. The trajectory was then clustered into 98 k-means states, 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 milliseconds 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 (as described above), and projected into tICA space at lag time 5 ns using commute mapping. The dimensionality of the dataset was reduced to 339 dimensions, keeping the number of tICs necessary to explain 95% of kinetic variance. The resulting tICA space was discretized into 1000 microstates using k-means. The Markov state model (MSM) was then estimated from the resulting discretized trajectories at lag time 50 ns using a minimum number of counts for ergodic trimming (i.e. the 'mincount_connectivity' argument in PyEMMA) of 4, as the default setting resulted in a trapped state whose connectivity between simulation

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 using 5 macrostates to coarse-grain the transition matrix.

4.11 Comparison of HREMD to SAXS experiment

We validated the HREMD-generated ensembles of Iso1 and Iso2 complexes by directly comparing to the experimental SAXS data. The theoretical SAXS profile was computed from each snapshot from the HREMD simulation trajectory using CRYSOL[78] available in a software package ATSAS.[79] The following CRYSOL 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 $idpflex$.[80] The ensemble-averaged theoretical SAXS profile was determined as below,

$$< I(q) >= \frac{1}{n}\sum_{i=1}^{n} I_i(q) \tag{10}$$

where n = 100,001 is the total number of frames in HREMD trajectory of Iso1 and Iso2 complex each. The ensemble-averaged theoretical SAXS profile was compared to experiment (Figure 6c) by minimizing chi-square (χ^2) given by,

$$\chi^2 = \frac{1}{(m-1)}\sum_{i=1}^{m}\left\{\frac{\left[< I_{expt}(q_i) > -(c < I_{calc}(q_i) > +b)\right]}{\sigma_{expt}(q_i)}\right\}^2 \tag{11}$$

where $< I_{expt}(q) >$ and $< I_{calc}(q) >$ 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 $I_{expt}(q)$.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

To study the impact of different bifunctional molecules 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 VHL-degrader-SMARCA2 simulations described above. We used targeted MD simulations (TMD)[81] to drive the activation of the CLR based on the active structure of a homologous E3 ligase, CRL-βTrCP (PDB ID 6TTU).[55] The full CRL-VHL system was built using PDB IDs 1LQB and 5N4W including VHL, ElonginB, ElonginC, Cullin2, and RBX1.[11,82] NEDD8 was placed near residue Lys689 of the CRL where neddylation occurs.

As the collective variable for TMD, we used residue-based RMSD of the last ~70 Cα atoms of the Cullin C-terminus (where neddylation and subsequent activation occur) of the Cullin1 from the 6TTU structure[55] as the reference state and modeled the Cullin2 from its inactive form in the 5N4W structure to this reference state. In addition, the Cα atoms of the entire NEDD8 protein from the 6TTU structure were also used as a reference structure during TMD. Residues 135 to 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/nm. The system was built using Amber force field 14 parameters with rectangular simulation box with a total number of atoms ~500K and ionic concentration of 0.120 M using KCl. The hydrogen mass repartitioning (HMR) was used to enable 4 fs timestep simulations. The TMD structure was then used to build the entire complex for CRL-VHL-Degrader-SMARCA2. 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.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

4.13 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 complex and generate a diverse set of putative closed conformations that place the E2-loaded ubiquitin close to lysine residues on SMARCA2. The results from the meta-eABF simulation are used to seed additional simulations for unbiased ensemble-scale sampling. Detailed description of the meta-eABF algorithm and its variants can be found elsewhere,[83–86] but for clarity we present a brief account here. Similar to adaptive biasing force methods (ABF), meta-eABF simulations also utilize adaptive free energy biasing force to enhance sampling along one or more collective variables (CV), but the practical implementation is different. 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 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 25-75 Å. During this simulation we saw multiple ring closing-

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 on Nvidia RTX 2080Ti GPU using OpenMM 7.5[87] interfaced with PLUMED 2.7.[88]

Acknowledgement

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.

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.

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.

Supporting Information Available

We make all experimental data used in this study available, including HDX-MS and a crystal structure of SMARCA2:ACBI1:VHL-Elongin C-Elongin B (PDB ID 7S4E). We also make available trajectory data for the conformational sampling of the crystal structures and the ternary complex formation simulations at
https://console.cloud.google.com/storage/browser/paperdata.
We have created a repository information about the format of the WES+HDX trajectory data, and source code needed to run WES+HDX at

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

https://github.com/stxinsite/degrader-ternary-complex-prediction.

References

(1) Wu, T.; Yoon, H.; Xiong, Y.; Dixon-Clarke, S. E.; Nowak, R. P.; Fischer, E. S. Targeted protein degradation as a powerful research tool in basic biology and drug target discovery. NAT STRUCT MOL BIOL **2020**, 27, 605–614.

(2) Schneider, M.; Radoux, C. J.; Hercules, A.; Ochoa, D.; Dunham, I.; Zalmas, L.-P.; Hessler, G.; Ruf, S.; Shanmugasundaram, V.; Hann, M. M.; Thomas, P. J.; Queisser, M. A.; Benowitz, A. B.; Brown, K.; Leach, A. R. The PROTACtable genome. NAT REV DRUG DISCOV **2021**, 1–9.

(3) Schapira, M.; Calabrese, M. F.; Bullock, A. N.; Crews, C. M. Targeted protein degradation: expanding the toolbox. NAT REV DRUG DISCOV **2019**, 18, 949–963.

(4) Coleman, K. G.; Crews, C. M. Proteolysis–Targeting Chimeras: Harnessing the Ubiquitin–Proteasome System to Induce Degradation of Specific Target Proteins. Annual Review of Cancer Biology **2017**, 2, 1–18.

(5) Matyskiela, M. E. et al. A Cereblon Modulator (CC-220) with Improved Degradation of Ikaros and Aiolos. Journal of Medicinal Chemistry **2018**, 61, 535–542.

(6) Chamberlain, P. P. et al. Structure of the human Cereblon–DDB1–lenalidomide complex reveals basis for responsiveness to thalidomide analogs. Nature Structural & Molecular Biology **2014**, 21, 803–809.

(7) Krönke, J. et al. Lenalidomide Causes Selective Degradation of IKZF1 and IKZF3 in Multiple Myeloma Cells. Science 343, 301–305.

(8) Ohoka, N. et al. In Vivo Knockdown of Pathogenic Proteins via Specific and Nongenetic Inhibitor of Apoptosis Protein (IAP)-dependent Protein Erasers (SNIPERs)*. Journal of Biological Chemistry **2017**, 292, 4556–4570.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(9) Wei, J. et al. Harnessing the E3 Ligase KEAP1 for Targeted Protein Degradation. Journal of the American Chemical Society **2021**, 143, 15073–15083.

(10) Rodriguez-Gonzalez, A.; Cyrus, K.; Salcius, M.; Kim, K.; Crews, C. M.; Deshaies, R. J.; Sakamoto, K. M. Targeting steroid hormone receptors for ubiquitination and degradation in breast and prostate cancer. Oncogene **2008**, 27, 7201–7211.

(11) Hon, W.-C.; Wilson, M. I.; Harlos, K.; Claridge, T. D.; Schofield, C. J.; Pugh, C. W.; Maxwell, P. H.; Ratcliffe, P. J.; Stuart, D. I.; Jones, E. Y. Structural basis for the recognition of hydroxyproline in HIF-1α by pVHL. Nature **2002**, 417, 975–978.

(12) Sakamoto, K. M.; Kim, K. B.; Kumagai, A.; Mercurio, F.; Crews, C. M.; Deshaies, R. J. Protacs: Chimeric molecules that target proteins to the Skp1–Cullin–F box complex for ubiquitination and degradation. Proceedings of the National Academy of Sciences **2001**, 98, 8554–8559.

(13) Roy, M. J.; Winkler, S.; Hughes, S. J.; Whitworth, C.; Galant, M.; Farnaby, W. l.; Rumpel, K.; Ciulli, A. SPR-Measured Dissociation Kinetics of PROTAC Ternary Complexes Influence Target Degradation Rate. ACS CHEM BIOL **2019**, 14, 361–368.

(14) Hughes, S.; Ciulli, A. Molecular recognition of ternary complexes: a new dimension in the structure-guided design of chemical degraders. ESSAYS BIOCHEM **2017**, 61, 505–516.

(15) Zorba, A. et al. Delineating the role of cooperativity in the design of potent PROTACs for BTK. Proceedings of the National Academy of Sciences **2018**, 115, 201803662.

(16) Schiemer, J. et al. Snapshots and ensembles of BTK and cIAP1 protein degrader ternary complexes. NAT CHEM BIOL **2021**, 17, 152–160.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(17) Nowak, R. P.; DeAngelo, S. L.; Buckley, D.; He, Z.; Donovan, K. A.; An, J.; Safaee, N.; Jedrychowski, M. P.; Ponthier, C. M.; Ishoey, M.; Zhang, T.; Mancias, J. D.; Gray, N. S.; Bradner, E. S., J. E. Fischer Plasticity in binding confers selectivity in ligand-induced protein degradation. Nature Chemical Biology **2018**, 14, 706–714.

(18) Farnaby, W.; Koegl, M.; Roy, M. J.; Whitworth, C.; Diers, E.; Trainor, N.; Zollman, D.; Steurer, S.; Karolyi-Oezguer, J.; Riedmueller, C., et al. BAF complex vulnerabilities in cancer demonstrated via structure-based PROTAC design. Nature chemical biology **2019**, 15, 672–680.

(19) Huang, H.-T.; Dobrovolsky, D.; Paulk, J.; Yang, G.; Weisberg, E. L.; Doctor, Z. M.; Buckley, D. L.; Cho, J.-H.; Ko, E.; Jang, J., et al. A chemoproteomic approach to query the degradable kinome using a multi-kinase degrader. Cell chemical biology **2018**, 25, 88–99.

(20) Bondeson, D. P.; Smith, B. E.; Burslem, G. M.; Buhimschi, A. D.; Hines, J.; Jaime-Figueroa, S.; Wang, J.; Hamman, B. D.; Ishchenko, A.; Crews, C. M. Lessons in PROTAC design from selective degradation with a promiscuous warhead. Cell chemical biology **2018**, 25, 78–87.

(21) Ward, C. C.; Kleinman, J. I.; Brittain, S. M.; Lee, P. S.; Chung, C. Y. S.; Kim, K.; Petri, Y.; Thomas, J. R.; Tallarico, J. A.; McKenna, J. M., et al. Covalent ligand screening uncovers a RNF4 E3 ligase recruiter for targeted protein degradation applications. ACS chemical biology **2019**, 14, 2430–2440.

(22) Zengerle, M.; Chan, K.-H.; Ciulli, A. Selective Small Molecule Induced Degradation of the BET Bromodomain Protein BRD4. ACS CHEM BIOL **2015**, 10, 1770–1777.

(23) Gadd, M. S.; Testa, A.; Lucas, X.; Chan, K.-H.; Chen, W.; Lamont, D. J.; Zẽngerle, M.; Ciulli, A. Structural basis of PROTAC cooperative recognition for selective protein degradation. NAT CHEM BIOL **2017**, 13, 514–521.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(24) Farnaby, W. et al. BAF complex vulnerabilities in cancer demonstrated via structure-based PROTAC design. NAT CHEM BIOL **2019**, 15, 672–680.

(25) Testa, A.; Hughes, S. J.; Lucas, X.; Wright, J. E.; Ciulli, A. Structure-Based Design of a Macrocyclic PROTAC. Angewandte Chemie International Edition **2020**, 59, 1727–1734.

(26) Zaidman, D.; Prilusky, J.; London, N. PRosettaC: Rosetta Based Modeling of PROTAC Mediated Ternary Complexes. J CHEM INF MODEL **2020**, 60, 4894–4903.

(27) Bai, N.; Kirubakaran, P.; Karanicolas, J. Rationalizing PROTAC-mediated ternary complex formation using Rosetta. J. Chem. Inf. Model. **2021**, 61, 1368–1382.

(28) 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 **2020**, 60, 5234–5254.

(29) Shaheer, M.; Singh, R.; Sobhia, M. E. Protein degradation: a novel computational approach to design protein degrader probes for main protease of SARS-CoV-2. J BIOMOL STRUCT DYN **2021**, 1–13.

(30) 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 **2020**, 60, 5234–5254.

(31) Huber, G. A.; Kim, S. Weighted-ensemble Brownian dynamics simulations for protein association reactions. BIOPHYS J **1996**, 70, 97–110.

(32) Zhang, B. W.; Jasnow, D.; Zuckerman, D. M. The "weighted ensemble" path

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

sampling method is statistically exact for a broad class of stochastic processes and binning procedures. J CHEM PHYS **2010**, 132, 054107.

(33) Zwier, M. C.; Adelman, J. L.; Kaus, J. W.; Pratt, A. J.; Wong, K. F.; Rego, N. B.; Suarez, E.; Lettieri, S.; Wang, D. W.; Grabe, M.; Zuckerman, D. M.; Chong, L. T. WESTPA: An Interoperable, Highly Scalable Software Package for Weighted Ensemble Simulation and Analysis. J CHEM THEORY COMPUT **2015**, 11, 800–809.

(34) Abdul-Wahid, B.; Feng, H.; Rajan, D.; Costaouec, R.; Darve, E.; Thain, D.; Izaguirre, J. A. AWE-WQ: Fast-Forwarding Molecular Dynamics Using the Accelerated Weighted Ensemble. Journal of Chemical Information and Modeling **2014**, 54, 3033–3043.

(35) Feng, H.; Costaouec, R.; Darve, E.; Izaguirre, J. A. A comparison of weighted ensemble and Markov state model methodologies. The Journal of Chemical Physics **2015**, 142, 214113.

(36) Donyapour, N.; Roussey, N. M.; Dickson, A. REVO: Resampling of ensembles by variation optimization. J. Chem. Phys. **2019**, 150.

(37) Lotz, S. D.; Dickson, A. Wepy: A Flexible Software Framework for Simulating Rare Events with Weighted Ensemble Resampling. ACS Omega **2020**, 5, 31608–31623.

(38) Husic, B. E.; Pande, V. S. Markov state models: From an art to a science. Journal of the American Chemical Society **2018**, 140, 2386–2396.

(39) Liu, X.; Zhang, X.; Lv, D.; Yuan, Y.; Zheng, G.; Zhou, D. Assays and technologies for developing proteolysis targeting chimera degraders. Future Medicinal Chemistry **2020**, 12, 1155–1179.

(40) Deller, M. C.; Kong, L.; Rupp, B. Protein stability: A crystallographer's perspective. Acta Crystallogr F Struct Biol Commun **2016**, 72, 72–95.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(41) Dagbay, K. B.; Bolik-Coulon, N.; Savinov, S. N.; Hardy, J. A. Caspase-6 Undergoes a Distinct Helix-Strand Interconversion upon Substrate Binding*. J BIOL CHEM **2017**, 292, 4885–4897.

(42) Dagbay, K. B.; Hardy, J. A. Multiple proteolytic events in caspase-6 self-activation impact conformations of discrete structural regions. Proceedings of the National Academy of Sciences of the United States of America **2017**, 114, E7977–E7986.

(43) MacPherson, D. J.; Mills, C. L.; Ondrechen, M. J.; Hardy, J. A. Tri-arginine exosite patch of caspase-6 recruits substrates for hydrolysis. J BIOL CHEM **2019**, 294, 71–88.

(44) Douglass, E. F.; Miller, C. J.; Sparer, G.; Shapiro, H.; Spiegel, D. A. A Comprehensive Mathematical Model for Three-Body Binding Equilibria. Journal of the American Chemical Society **2013**, 135, 6092–6099.

(45) Kochert, B. A.; Iacob, R. E.; Wales, T. E.; Makriyannis, A.; Engen, J. R. Hydrogen-Deuterium Exchange Mass Spectrometry to Study Protein Complexes. Methods in Molecular Biology **2018**, 1764, 153–171.

(46) Wales, T. E.; Engen, J. R. Hydrogen exchange mass spectrometry for the analysis of protein dynamics. MASS SPECTROM REV **2006**, 25, 158–170.

(47) Gallagher, E. S.; Hudgens, J. W. Mapping Protein-Ligand Interactions with Proteolytic Fragmentation, Hydrogen/Deuterium Exchange-Mass Spectrometry. Methods in Enzymology **2016**, 566.

(48) Saglam, A. S.; Chong, L. T. Protein–protein binding pathways and calculations of rate constants using fully-continuous, explicit-solvent simulations. Chemical Science **2018**, 10, 2360–2372.

(49) Jubb, H. C.; Higueruelo, A. P.; Ochoa-Montaño, B.; Pitt, W. R.; Ascher, D. B.; Blundell, T. L. Arpeggio: A Web Server for Calculating and Visualising Inter-

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

atomic Interactions in Protein Structures. Journal of Molecular Biology **2017**, 429, 365–371.

(50) Eron, S. Finding a way out of the labyrinth: degrader-induced ternary complex modeling. July 14, 2021 *The Protein Society 35th Anniversary Symposium*, July 7-9, 12-14, 2021.

(51) Zhang, M. M.; Beno, B. R.; Huang, R. Y.-C.; Adhikari, J.; Deyanova, E. G.; Li, J.; Chen, G.; Gross, M. L. 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. **2019**, 91, 15709–15717.

(52) Méndez, R.; Leplae, R.; De Maria, L.; Wodak, S. J. Assessment of blind predictions of protein–protein interactions: Current status of docking methods. PROTEINS **2003**, 52, 51–67.

(53) Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernández, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.-H.; Noé, F. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. Journal of Chemical Theory and Computation **2015**, 11, 5525–5542.

(54) Molgedey, L.; Schuster, H. G. Separation of a mixture of independent signals using time delayed correlations. Phys. Rev. Lett. **1994**, 72, 3634–3637.

(55) Baek, K.; Krist, D. T.; Prabu, J. R.; Hill, S.; Klügel, M.; Neumaier, L.-M.; von Gronau, S.; Kleiger, G.; Schulman, B. A. NEDD8 nucleates a multivalent cullin–RING–UBE2D ubiquitin ligation assembly. Nature **2020**, 578, 461–466.

(56) Dagbay, K. B.; Hardy, J. A. Multiple proteolytic events in caspase-6 self-activation impact conformations of discrete structural regions. P NATL ACAD SCI USA **2017**, 114, E7977–E7986.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(57) Wales, T. E.; Fadgen, K. E.; Gerhardt, G. C.; Engen, J. R. High-Speed and High-Resolution UPLC Separation at Zero Degrees Celsius. AANAL BIOANAL CHEM **2008**, 80, 6815–6820.

(58) Winn, M. D. et al. Overview of the CCP4 suite and current developments. Acta Crystallographica Section D: Biological Crystallography **2011**, 67, 235–242.

(59) Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. Acta Crystallographica Section D: Biological Crystallography **2004**, 60, 2126–2132.

(60) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. Application of Maximum Likelihood Refinement. Refinement of Protein structures, Proceedings of Daresbury Study Weekend

(61) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. Refinement of Macromolecular Structures by the Maximum-Likelihood method. Acta Crystallographica Section D: Biological Crystallography **1997**, 53, 240–255.

(62) Pannu, N. J.; Murshudov, G. N.; Dodson, E. J.; Read, R. A. Incorporation of Prior Phase Information Strengthen Maximum-Likelihood Structure Refinement. Acta Crystallographica Section D: Biological Crystallography **1998**, 54, 1285–1294.

(63) Murshudov, G. N.; Lebedev, A.; Vagin, A. A.; Wilson, K. S.; Dodson, E. J. Efficient anisotropic refinement of Macromolecular structures using FFT. Acta Crystallographica Section D: Biological Crystallography **1999**, 55, 247–255.

(64) Winn, M.; Isupov, M.; Murshudov, G. N. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. Acta Crystallographica Section D: Biological Crystallography **2001**, 57, 122–133.

(65) Steiner, R.; Lebedev, A.; Murshudov, G. N. Fisher's information matrix in maximum likelihood molecular refinement. Acta Crystallographica Section D: Biological Crystallography **2003**, 59, 2114–2124.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(66) Winn, M.; Murshudov, G. N.; Papiz, M. Z. Macromolecular TLS refinement in REFMAC at moderate resolutions. Methods in Enzymology **2003**, 374, 300–321.

(67) Skubak, P.; Murshudov, G. N.; Pannu, N. S. Direct incorporation of experimental phase information in model refinement. Acta Crystallographica Section D: Biological Crystallography **2004**, 60, 2196–2201.

(68) Vagin, A. A.; Steiner, R. S.; Lebedev, A. A.; Potterton, L.; McNicholas, S.; Long, F.; Murshudov, G. N. REFMAC5 dictionary: organisation of prior chemical knowledge and guidelines for its use. Acta Crystallographica Section D: Biological Crystallography **2004**, 60, 2284–2295.

(69) 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 **2017**, 50, 1545–1553.

(70) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J CHEM THEORY COMPUT **2015**, 11, 3696–3713, PMID: 26574453.

(71) Pracht, P.; Bohle, F.; Grimme, S. Automated exploration of the low-energy chemical space with fast quantum chemical methods. Phys. Chem. Chem. Phys. **2020**, 22, 7169–7192.

(72) Grimme, S. Exploration of Chemical Compound, Conformer, and Reaction Space with Meta-Dynamics Simulations Based on Tight-Binding Quantum Chemical Calculations. J. Chem. Theory Comput. **2019**, 15, 2847–2862.

(73) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. J. Chem. Theory Comput. **2019**, 15, 1652–1671.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(74) Gray, J. J.; Moughon, S.; Wang, C.; Schueler-Furman, O.; Kuhlman, B.; Rohl, C. A.; Baker, D. Protein–Protein Docking with Simultaneous Optimization of Rigid-body Displacement and Side-chain Conformations. J. Mol. Biol. **2003**, 331, 281–299.

(75) Marze, N. A.; Roy Burman, S. S.; Sheffler, W.; Gray, J. J. Efficient flexible backbone protein–protein docking for challenging targets. Bioinformatics **2018**, 34, 3461–3469.

(76) Bussi, G. Hamiltonian replica exchange in GROMACS: a flexible implementation. MOL PHYS **2014**, 112, 379–384.

(77) Wang, L.; Friesner, R. A.; Berne, B. J. Replica exchange with solute scaling: a more efficient version of replica exchange with solute tempering (REST2). J PHYS CHEM B **2011**, 115, 9431–9438.

(78) Svergun, D.; Barberato, C.; Koch, M. H. J. *CRYSOL* – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. J APPL CRYSTALLOGR **1995**, 28, 768–773.

(79) Manalastas-Cantos, K.; Konarev, P. V.; Hajizadeh, N. R.; Kikhney, A. G.; Petoukhov, M. V.; Molodenskiy, D. S.; Panjkovich, A.; Mertens, H. D. T.; Gruzinov, A.; Borges, C.; Jeffries, C. M.; Svergun, D. I.; Franke, D. *ATSAS 3.0*: expanded functionality and new tools for small-angle scattering data analysis. J APPL CRYSTALLOGR **2021**, 54, 343–355.

(80) 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. Journal of Open Source Software **2018**, 3.

(81) 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 computational biology **2006**, 2, e134.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.26.461830; this version posted September 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(82) Edmondson, S. D.; Yang, B.; Fallan, C. Proteolysis Targeting Chimeras (PROTACs) in 'Beyond Rule-of-Five' Chemical Space: Recent Progress and Future Challenges. BIOORG MED CHEM LETT **2019**, 29, 1555–1564.

(83) Comer, J.; Gumbart, J. C.; Hénin, J.; Lelièvre, T.; Pohorille, A.; Chipot, C. The adaptive biasing force method: Everything you always wanted to know but were afraid to ask. The Journal of Physical Chemistry B **2015**, 119, 1129–1151.

(84) Lesage, A.; Lelievre, T.; Stoltz, G.; Hénin, J. Smoothed biasing forces yield unbiased free energies with the extended-system adaptive biasing force method. The Journal of Physical Chemistry B **2017**, 121, 3676–3685.

(85) Fu, H.; Zhang, H.; Chen, H.; Shao, X.; Chipot, C.; Cai, W. Zooming across the free-energy landscape: shaving barriers, and flooding valleys. The journal of physical chemistry letters **2018**, 9, 4738–4745.

(86) Fu, H.; Shao, X.; Cai, W.; Chipot, C. Taming rugged free energy landscapes using an average force. Accounts of chemical research **2019**, 52, 3254–3264.

(87) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D., et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. PLoS computational biology **2017**, 13, e1005659.

(88) Bonomi, M. Promoting transparency and reproducibility in enhanced molecular simulations. Nature methods **2019**, 16, 670–673.