Article

Non-Markovian Dynamic Models Identify Non-Canonical KRAS-VHL Encounter Complex Conformations for Novel PROTAC Design

Yunrui Qiu, Rafal P. Wiewiora, Jesus A. Izaguirre, Huafeng Xu, Woody Sherman, Weiping Tang,* and Xuhui Huang*

Cite This: *JACS Au* 2024, 4, 3857−3868

Read Online

ACCESS | 📊 Metrics & More | 📰 Article Recommendations | 🆘 Supporting Information

ABSTRACT: Targeted protein degradation (TPD) is emerging as a promising therapeutic approach for cancer and other diseases, with an increasing number of programs demonstrating its efficacy in human clinical trials. One notable method for TPD is Proteolysis Targeting Chimeras (PROTACs) that selectively degrade a protein of interest (POI) through E3-ligase induced ubiquitination followed by proteasomal degradation. PROTACs utilize a warhead-linker-ligand architecture to bring the POI (bound to the warhead) and the E3 ligase (bound to the ligand) into proximity. The resulting non-native protein−protein interactions (PPIs) formed between the POI and E3 ligase lead to the formation of a stable ternary complex, enhancing cooperativity for TPD. A significant challenge in PROTAC design is the screening of the linkers to induce favorable non-native PPIs between POI and E3 ligase. Here,

Non-Markovian Dynamics Model

we present a physics-based computational protocol to predict noncanonical and metastable PPI interfaces between an E3 ligase and a given POI, aiding in the design of linkers to stabilize the ternary complex and enhance degradation. Specifically, we build the non-Markovian dynamic model using the Integrative Generalized Master equation (IGME) method from ∼1.5 ms all-atom molecular dynamics simulations of linker-less encounter complex, to systematically explore the inherent PPIs between the oncogene homologue protein and the von Hippel-Lindau E3 ligase. Our protocol revealed six metastable states each containing a different PPI interface. We selected three of these metastable states containing promising PPIs for linker design. Our selection criterion included thermodynamic and kinetic stabilities of PPIs and the accessibility between the solvent-exposed sites on the warheads and E3 ligand. One selected PPIs closely matches a recent cocrystal PPI interface structure induced by an experimentally designed PROTAC with potent degradation efficacy. We anticipate that our protocol has significant potential for widespread application in predicting metastable POI-ligase interfaces that can enable rational design of PROTACs.
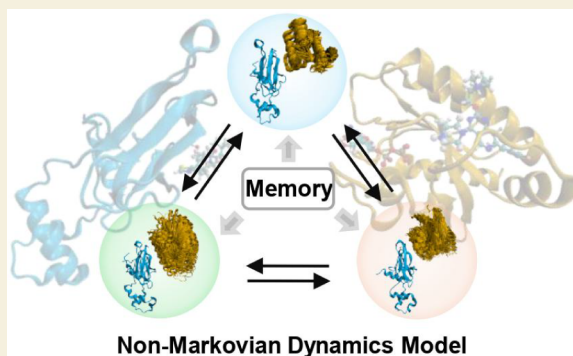
KEYWORDS: *non-Markovian dynamic models, Markov state models, protein−protein interface prediction, PROTAC, encounter complex*

1. INTRODUCTION

Small molecule heterobifunctional degraders, exemplified by proteolysis targeting chimeras (PROTACs), have the potential to transform drug discovery and therapeutic interventions by *degrading* proteins instead of *inhibiting* them.[1−4] Unlike the traditional small-molecule inhibitors that block the protein function through occupying the active or allosteric site of the protein of interest (POI), PROTACs can employ functional or nonfunctional binders to target the POI, inducing its degradation through a catalytic mechanism.[5,6] This approach provides opportunities to target many undruggable POIs that lack well-defined small molecule binding sites for functional blockade. A PROTAC comprises three distinct components: warhead, linker, and E3 ligand. With the warhead binding to the POI and E3 ligand binding to the E3 ligase, PROTAC facilitates the proximity between the POI and the E3 ubiquitin ligase, leading to the formation of a ternary complex. This complex could then trigger the ubiquitination of the POI,

marking it for degradation by the cellular proteasome machinery. Over the past two decades, significant effort has been dedicated to investigating and designing PROTACs.[7,8] However, the development of most PROTACs remains highly empirical, involving the time-consuming synthesis and screening of libraries with various linkers between the warhead and the E3 ligase ligand. This process aims to induce favorable non-native protein−protein interactions (PPIs) between the POI and E3 ligase.

Throughout the PROTAC-induced targeted protein degradation (TPD), establishing specific PPI between the POI and
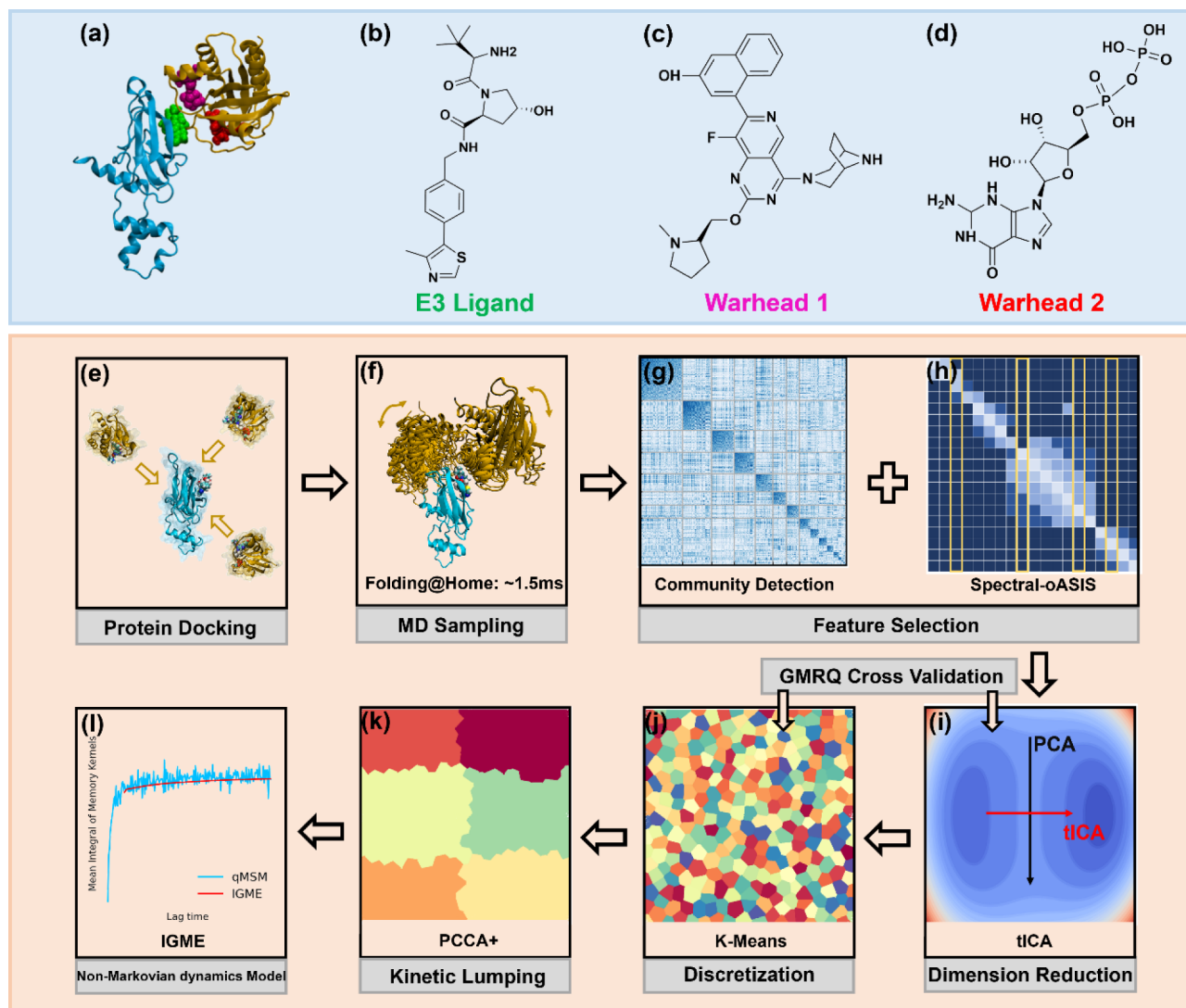
Figure 1. The KRAS-VHL encounter complex system (a-d) and the workflow of the construction of the non-Markovian IGME model (e-l). (a). The structure of the encounter complex from rigid protein docking, involving VHL (cyan) and KRAS (orange), along with the E3 ligand and two warheads. (b-d) Chemical structures of E3 ligand (green), warhead 1 (magenta) and warhead 2 (red). (e) Generate initial conformations for the encounter complex through rigid protein docking. (f) Perform extensive MD simulations using Folding@Home to explore the PPI interfaces of the encounter complex. (g-h) Utilize MoSAIC community detection and spectral-oASIS algorithms to extract essential pairwise distance features for representations of the PPI interfaces. (i) Identify the collective variables by tICA. (j) Cluster the projected MD conformations to microstates by K-Means algorithm. The hyperparameters for (i) and (j) are tuned through cross-validation based on the GMRQ score. (k) Lump the microstates to metastable macrostates by PCCA+ algorithm. (l) Model the transition dynamics between macrostates with IGME method.

E3 ligase is critical.[9−12] Many degraders function by leveraging the stabilization of preexisting but weak PPIs between POIs and E3 ligases.[11−13] Additionally, both experiments and computational simulations reveal that PPIs of highly productive ternary complexes exhibit noticeable dynamical conformational heterogeneities, distinct from the static contacts found in crystal structures.[11,14−17] Previous biophysical and structural studies have also demonstrated that different PROTACs, even with the same warhead and E3 ligand but different linkers, can induce distinct PPIs in ternary complexes, leading to significant differences in degradation efficiency.[9,16−24] Therefore, investigating the complex and dynamic non-native PPIs between the POIs and E3 ligases is critical for understanding TPD mechanisms and guiding the rational design of novel PROTACs. An approach with great potential to explore all possible inherent PPIs between the POI and E3 ligase is to study the POI-E3 ligase encounter complex

without the linker.[10,25−27] The subsequent introduction of the linker to this encounter complex is akin to adding an additional geometric constraint.[10]

All-atom MD simulation offers a promising approach to reveal metastable and dynamical PPIs between the POI and E3 ligase.[10,11] It has been combined with enhanced sampling techniques to elucidate both the kinetic and thermodynamic properties of a PROTAC system.[28] However, simulating the formation of PPIs presents significant challenges due to the various ways in which the POI and E3 ligase can approach each other, as well as the conformational changes induced upon the formation of the encounter complex. The formation and conformational changes of encounter complex PPI interfaces often occur on milliseconds time scales, which exceed accessible length of the straightforward MD simulations for a system at the size of approximately 150,000 atoms. Adding to the complexity, there is a lack of dominant PPI, and all PPIs

may potentially serve as functional ones for PROTAC design.[10] Therefore, obtaining a comprehensive understanding of the conformational space of the encounter complex and identifying representative PPI interfaces, along with their equilibrium populations and transition rates between them, are challenging.

Markov State Models (MSMs) built from extensive MD simulations offer a potentially useful technique to address these challenges.[29−40] MSMs model dynamics through a series of Markovian jumps among conformational states at discrete lag times. MSMs also provide a rigorous pipeline to coarse-grain MD conformations into a few comprehensible states according to their dynamic metastability, facilitating the prediction of thermodynamic and kinetic properties associated with them. However, for MSMs to have predictive power, they must be constructed with a sufficiently long lag time to ensure that interstate transitions become Markovian, posing a major challenge as the lag time is constrained by the length of short MD simulations.,[38,39,41] To address this challenge, we recently developed an approach based on the Generalized Master Equation (GME), called the Integrative Generalized Master equation (IGME) method.[38] IGME captures non-Markovian dynamics by incorporating time-integrations of memory kernel functions, offering a promising approach to study PPIs in encounter complexes based on relatively limited MD simulation data.

In this study, we constructed an IGME model from 2,492 MD trajectories, with an average length of 605 ns (∼1.51 ms in total), to elucidate potential non-native PPIs between the oncogene homologue (KRAS) protein[42−45] and the von Hippel-Lindau (VHL) E3 ligase.[46,47] KRAS is the oncogene most frequently mutated in cancer,[42] and PROTAC-induced TPD is considered as a promising approach for treating KRAS-induced cancer.[43,45] We here simulated the formations and conformational changes of the encounter complex in the absence of the linker, but with KRAS bound to two different warheads and VHL bound to one ligand (Figure 1a−d). Using our simulation and dynamic modeling protocol, we revealed six metastable states characterized by distinct conformations of PPI interfaces and provided the corresponding thermodynamic and kinetic properties for each state. Based on the IGME model, we further evaluated additional structural properties of conformations within each state, such as the spatial proximity of the warhead and E3 ligand and the solvent-exposed sites of both. Consequently, we identified three metastable states that exhibit promising PPI interfaces for future linker design. Conformations from one of our predicted metastable states agree well with a recent ternary crystal structure[44] (with an average interface-RMSD of 5.42 ± 3.67 Å) involving a degrader of promising degradation efficiency. Our IGME modeling offers a systematic and efficient approach to legitimizing metastable PPIs in protein pairs, thereby facilitating rational PROTAC design.

2. RESULTS AND DISCUSSIONS

2.1. Elucidating the Dynamics of KRAS-VHL Encounter Complex Formation: IGME Outperforms MSM

We construct our IGME model from MD trajectories totaling ∼1.51 ms for studying the dynamics of the KRAS-VHL encounter complex formation (see Figure 1 panels e-l for our protocol). Specifically, to systematically explore the complete ensemble of PPI interface conformations, we employ rigid

protein docking to search the preferable PPIs from various approaching orientations, and then initiate unbiased MD simulations from these docking poses (see Figure 1e,f and Methods 4.1 for details). To build the 100-microstate MSM, we initially characterize the conformations of the encounter complex using all 25 330 internal pairwise distances between KRAS and VHL residues, and then employ the Molecular Systems Automated Identification of Cooperativity (MoSAIC) algorithm[48] and Spectral-oASIS algorithm[49] to identify 1,500 important distances as features for subsequent analysis (see Figures 1g,h and S2−S4). The implementation of these two algorithms ensures that the chosen distance features adequately represent various important collective motions around the PPI interfaces, while also capturing the slowest dynamics effectively (see Figures S2−S4 and Methods 4.2 for details). Subsequently, we apply the time-lagged independent component analysis (tICA)[32,50] with kinetic mapping[51] to project the encounter complex conformations onto five collective variables (CVs) (see Figure 1i) and then cluster them into 100 microstates via K-Means algorithm (see Figure 1j). The tICA-related hyperparameters and the number of microstates are optimized using cross-validation with the Generalized Matrix Rayleigh Quotient (GMRQ) score[52] (see Figure S5). More details about the construction and validation of microstate MSM are presented in Methods and Supporting Information.

To identify metastable PPIs of the KRAS-VHL encounter complex, we lump 100 microstates into six metastable macrostates using PCCA+[53,54] and build a 6-macrostate IGME model (see Figure 1k,l). Unlike Markovian MSMs, IGME utilizes the GME to evolve dynamics, considering the non-Markovian dynamics through time integrations of memory kernel functions. Given that the relaxation time of memory kernel functions is much shorter than the Markovian lag time for MSMs, IGME can model dynamics between a handful of metastable states with shorter segments of MD simulations compared to MSMs. As shown in Figure 2a, the integrations of memory kernels reach plateaus at around 50 ns, therefore accurate IGME models can be constructed at the memory kernel relaxation time $\tau_k > 50$ ns. An example of such an IGME model, constructed from MD simulation segments, each with the length of 150 ns ($\tau_k = 70ns$ and an additional segment of $L_{fit} = 80$ ns for fitting, see Methods 4.3 for details), is shown in Figure S11c. In sharp contrast, the MSM constructed with a much longer lag time of $\tau = 250$ ns still predicts significantly faster state-relaxation dynamics compared to the original MD simulations (Figure S11c). Furthermore, the root mean squared error (RMSE) of the MSMs' predicted dynamics is over an order of magnitude larger than that of the IGME models at different lag times (see Figure 2b and the Methods 4.3 for the details of the RMSE calculations). While IGME models consistently predict the slowest time scale and the mean first passage time (MFPT) across a wide range of lag times, MSMs always underestimate these values (see Figure 2c,d). We anticipate that to achieve comparable performance with IGME models, MSMs would require a significantly longer lag time, which is beyond the length of our MD simulations. As shown in Figure S11a, IGME models built with a wide range of hyperparameters: τ_k and L_{fit} robustly exhibit small RMSEs, i.e., below 0.1%. For the remaining sections, we choose the optimal IGME model as the one with the smallest RMSE value (constructed with $\tau_k = 70ns$ and $L_{fit} = 80ns$) to report the thermodynamic and kinetic properties of the PPI interfaces.
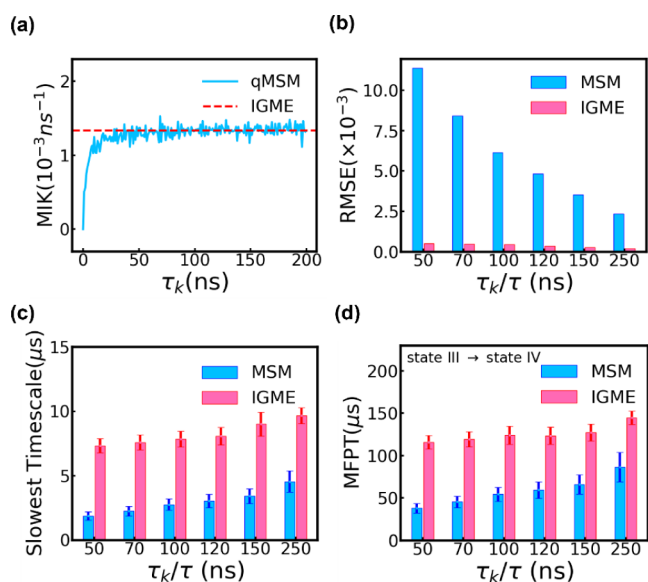
Figure 2. Non-Markovian IGME models outperform MSMs in elucidating the dynamics of the KRAS-VHL encounter complex formation. (a) Mean Integral of memory kernels (MIK) with different τ_k for six-states model calculated from quasi-MSM (qMSM) and IGME. (b) Root mean squared error (RMSE) of predicted transition probability matrices with respect to MD simulations. (c) Slowest implied time scale and (d) mean first passage time (MFPT) from State III to State IV, calculated from IGME models and MSMs constructed with various lag times. The error bars represent standard deviations estimated from 50 bootstraps of the data with replacement.

2.2. Dynamic Heterogeneity of the Encounter Complex Associated with Diverse Metastable PPI Formation

With the optimal IGME model, we observe the PPI interfaces of encounter complex consist of diverse non-native interaction patterns and exhibit significant dynamical heterogeneities. As shown in Figure 3a, the free energy landscape of PPI interfaces, projected onto the top two CVs identified by tICA, reveal multiple free energy basins. Each basin is associated with distinct metastable macrostate, indicating the inherent flexibility and diversity for the formation of the PPI interfaces between KRAS and VHL. Our IGME model also shows that State VI is highly populated (72.71%), while the equilibrium populations of the other five states are all below 10% (Figure 3b). Strikingly, we observe significant different PPI interfaces formed and stabilized by diverse chemical interactions between different domains of KRAS and VHL in these six metastable states (see representative structures in Figure 3e). The transition rates between these metastable states are also predicted by our IGME model (Figure S12). To characterize different PPIs, we first illustrate the PPI patterns using the residue pairwise distance maps and analyzing the distance variance between contacting interface residues in KRAS and VHL. As shown in Figure S14 and S15, various PPIs display substantially different residue pairwise distance maps. Additional analysis of the contact frequency for each residue across PPI interfaces also indicates the heterogeneity of these PPIs (Figure S16). To further examine if there exists preference of specific nonbonded chemical interactions to stabilize these PPIs, we plotted the preferences of amino acid type and interactions for PPIs formed in different macrostates (see Figures S17 and S18). We observe that salt bridges and dipolar interactions are present in all PPIs, through the interactions between charged residues (e.g., Glu and Arg) and polar residues (e.g., Gln and His). Interestingly, PPIs in States I and IV exhibit additional hydrophobic interactions (e.g., via Leu and Val). These observations suggest that KRAS and VHL can form different non-native PPIs via diverse nonbonded interactions. These metastable non-native PPIs open new opportunities for PROTAC design.

Previous experimental and computational results have demonstrated that it is inadequate to solely rely on the crystal
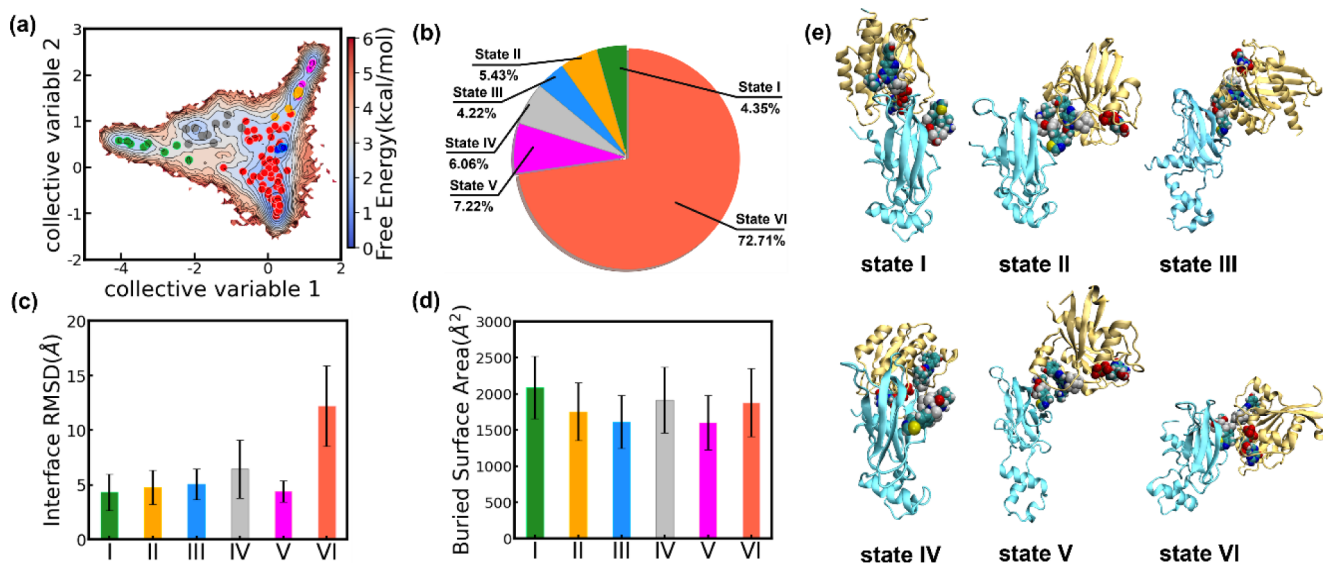

Figure 3. Interpretation of non-Markovian dynamics model. (a) The free energy landscape and distribution of states visualized on the top two tICA components. The free energy is estimated from the ultralong trajectory generated by running kinetic Monte Carlo with the microstate-MSM. Each point represents the center of a microstate, and its color corresponds to the macrostate label. (b) Stationary populations for macrostates predicted from the optimal IGME model. (c) The heterogeneity of each macrostate is visualized by calculating the interface-RMSD relative to the state center for all conformations within the state. (d) The buried area of PPI surfaces within each macrostate. (e) The representative conformations for each macrostate (selected from the microstates with the highest population).
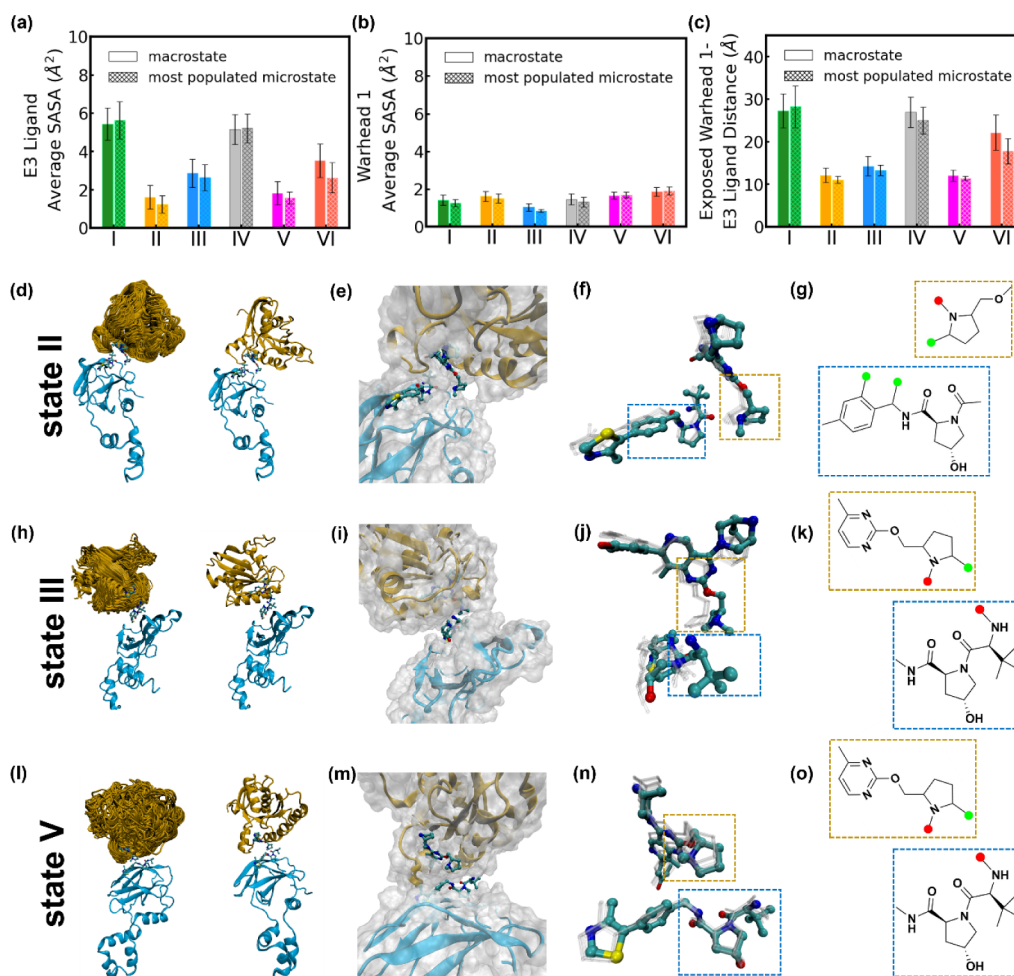
Figure 4. Protein−protein interfaces selected for linker filtering. (a-b). Average solvent accessible surface area (SASA) depicted for (a) E3 ligand and (b) warhead 1 molecules across all conformations within six macrostates and their respective most populated microstate. (c) The average pairwise distances between the exposed heavy E3 ligand atoms and warhead 1 atoms (top 50% SASA) are calculated across all conformations within six macrostates and their respective most populated microstate. Error bars represent standard deviations. Fifty randomly selected overlapping conformations and one representative single conformation of the PPI interface are visualized for State II (d-e), State III (h-i), and State V (i-m). The relative positions of the E3 ligand-warhead 1 and their partial chemical structures are displayed for State II (f-g), State III (j-k), and State V (n-o).

structure of induced ternary complex to assess PROTAC performance. Instead, the dynamic behaviors of the ternary complex may exert a more influential role on degradation efficiency.[9,16−24] As our simulations of the encounter complex do not include the degrader linker, the encounter complex exhibits much greater heterogeneity among multiple protein domains. We next examine the structural heterogeneities within each metastable macrostate. The visualizations of multiple encounter complex conformations for each macrostate (see Figure S19) and the characterization of distance variance between contacting interface residues (see Figure S15) illustrate the high consistency of PPI interfaces in State I−V and significant flexibilities of interfaces in state VI. By further using the MD conformation located at the geometric center of each macrostate as the reference structure, we compute the interface-RMSD among all MD conformations within each of the six macrostates (Figure 3c). The interface-RMSD is calculated using the formula $\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - x_i^{ref})^2}$, where x_i represents the Cartesian coordinates of atoms in the interface residues (i.e., residues with an average minimal distance to the other protein less than 10 Å) after optimal

alignment, x_i^{ref} denotes the reference coordinates and N is the number of conformations. Except for the highest populated State VI, all other 5 macrostates display moderate interface-RMSD values which are comparable to those observed in the dynamical simulations of other PROTAC-induced ternary complexes.[11] This observation suggests that even in the absence of the degrader linker, KRAS and VHL can develop dynamic cooperativity during the formation of the encounter complex, resulting in various well-defined PPIs suitable as a baseline for linker design. Additionally, we evaluate the stability of the PPIs for each macrostates. Previous studies have suggested a correlation between buried surface areas (BSA) and experimentally measured binding affinity of the PPIs.[13] We quantify the BSAs of interfaces from different macrostates by subtracting the solvent-accessible surface area (SASA) of the two single proteins from the encounter complex. As shown in Figure 3d, our analysis shows that there are no noticeable differences in the BSA values of the six macrostates in our IGME model. This result is consistent with the equilibrium populations predicted by IGME, where States I−V exhibit comparable populations.

Conversely, although State VI has the largest overall equilibrium population, it encompasses diverse PPI binding modes with significantly shorter average lifetimes, making it unsuitable for linker design. We noticed that each metastable state spans a distinct region of varying size within the CV space (Figure S20a) and contains PPI interface binding modes with varying degrees of heterogeneity. We thus reported the average lifetime of microstates belonging to each metastable state as an indicator of the average lifetime for binding modes within each metastable state. Specifically, we fit the residence probability for each microstate at different lag times using the formula $p(t)$ $= ce^{kt}$ and define the microstate's lifetime as the time at which the residence probability decays $1/e$. As shown in Figure S20, we plotted the lifetime distribution projected onto the top two tICA CVs, and the average lifetime of microstates within each metastable state, respectively. Interestingly, although state VI has the largest overall equilibrium population, it contains diverse PPI binding modes, and these modes exhibit much shorter average lifetimes (Figure S20b). This observation suggests that State VI is a high-entropy state with diverse, short-lived PPIs that rapidly interconvert. Consequently, PPI interfaces from States I−V may serve as better candidates for further PROTAC design. In the current study, we did not find any metastable states with extremely low equilibrium populations. However, we recognize that the population predicted by the IGME model could be a valuable criterion for shortlisting PPIs interfaces. Specifically, we recommend that metastable states with low populations (e.g., < 1%) should not be prioritized for linker design.

Compared to simple rigid protein docking, our simulation and dynamic modeling protocol significantly enhances the refinement of PPI interface patterns and identifies the most metastable interfaces. Although the interfaces obtained from docking exhibit considerable diversity, many do not fall within the free energy basins when projected onto the top two CVs (see Figure S9). A detailed comparison of the structural differences between docking interfaces and those from metastable states revealed noticeable differences in their contact maps (see Figure S21). Furthermore, assigning docking interfaces to metastable states showed that while these interfaces span all six states, 43 out of 50 are predominantly associated with state VI (see Figure S9), which is unsuitable for subsequent linker design. This highlights the importance of dynamic modeling for accurate interface classification and the identification of the most metastable PPI interfaces.

2.3. Shortlisting Predicted PPIs Meeting Linker Constraints

The rational design of PROTAC linkers has been limited due to the challenges associated with predicting the preexisting PPIs between the POI and the E3 ligase. As our IGME model has characterized the equilibrium populations and kinetics of various metastable PPI interfaces across different macrostates, we next consider the geometries of interfaces and ligands within each macrostate to evaluate their potential for linker design. Throughout the MD simulations, we notice that the warheads and E3 ligand tightly bound to the protein pockets, with only ∼3.2% of trajectories showing them diffusing away from the binding site. We further filter out these conformations from postanalysis. Since the encounter complex exhibits varying degrees of conformational changes during the formation of different PPIs, warheads and E3 ligand expose different atoms and adopt different relative orientations

accordingly. To identify the exposed functional groups in the warheads and E3 ligand that could potentially be connected via a linker, we calculate the SASA for each of their atoms (Figure 4a,b). We identify exposed heavy atoms, defined as those with the top 50% SASA among all atoms, as having the linking potential. Furthermore, we measure the average pairwise distances between the exposed heavy atom pairs of the E3 ligand and the warheads to further assess the feasibility of linker design, as shown in Figures 4c and S22.

Figure 4a,b shows that the E3 ligand generally exhibits larger SASA compared to the warheads, primarily because of the shallow pocket of VHL.[55,56] The E3 ligand conformations from States I and IV are considerably more exposed than those from other states. This suggests that conformations from these two states may have multiple potential sites to be linked. However, upon examining the interactive profiles between KRAS and VHL in these two states (Figures 3e and S19), we find that developing any linker based on the conformations from these two states is impractical, as the warheads and E3 ligand are too far away from each other (between 25 and 30 Å, see Figure 4c). Previous studies have highlighted that linker length is one of the most crucial factors influencing the effectiveness of PROTACs, and excessively long linkers often lead to a reduction in potency. Generally, PROTACs with linkers containing more than 20 atoms have comparatively low potency.[18,25,57,58] Therefore, designing linkers based on the PPI interfaces from States I and IV poses significant challenges. Moreover, after evaluating the conformations from all states, we observe that warhead 2 (see its chemical structure in Figure 1d) consistently stays distant from the ligase ligand (Figure S22), suggesting challenges in developing a degrader using it.

Consequently, by excluding State VI due to its kinetically unstable PPI, and eliminating State I, State IV, and warhead 2 due to inappropriate distances between warheads and E3 ligand, the PPI interfaces from the remaining three states (II, III, and V) have the potential for further linker development between E3 ligand and warhead 1. We further visualize the PPI conformations and the relative positions of ligands for these three states. As shown in Figure 4d−o, the conformations within these three states maintain consistent interfaces while also exhibiting slight heterogeneities. In addition, the warhead and ligand approach each other at appropriate distances for adding the linker. As shown in Figure 4(f,j and n), we highlight the ligand and warhead atoms with the top 50% largest SASA using dashed boxes in their chemical structures. In these chemical structures, we have identified potential linking sites, marking them with colored dots based on their synthetic ease and frequency of use documented in the literature.[23,59−64] In particular, the red dots correspond to atomic sites that are commonly used in literature for attaching the linkers, while the green dots indicate sites that are less frequently used for this purpose. However, it is important to note that less frequently used attachment sites for VHL ligands may lead to highly effective degraders.[59] We anticipate that these selected conformations may aid in designing linkers that could further stabilize the naturally favorable PPIs.

Through our systematic analysis shown above, we found there are several crucial factors to consider when evaluating and selecting appropriate PPI interfaces for linker design. We summarize these factors and criteria in a schematic workflow, as illustrated in Figure S23. Specifically, quantifying their structural heterogeneity and average lifetimes for PPI binding modes helps in selecting states with long-lived PPI interfaces
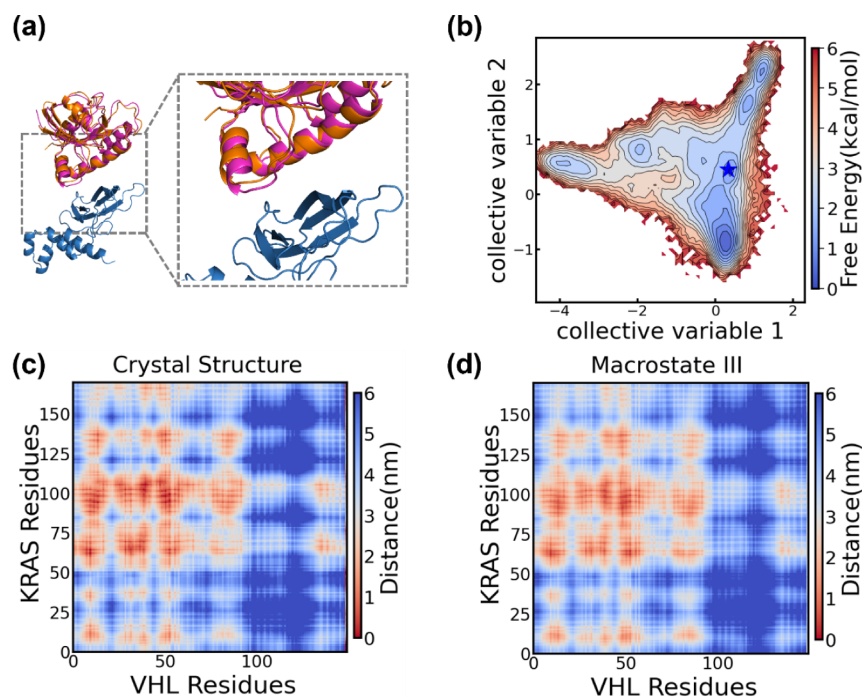
Figure 5. Comparison between computationally predicted PPI interfaces and the interface induced by the experimentally designed PROTAC. (a) Structural alignment between the crystal structure (magenta, PDB ID: 8QVU) and one PPI interface from most populated microstate in State III (orange). The interface with the smallest interface-RMSD (0.68 Å) is selected for visualization, and the alignment is based on the VHL protein. (b) Projection (blue star) of the crystal PPI interface of the ternary complex onto the top two CVs. (c) Pairwise distances between KRAS residues and VHL residues in the crystal structure of the ternary complex (PDB: 8QVU). (d) Averaged pairwise distances between KRAS residues and VHL residues across all conformations within macrostate III.

and consistent binding modes. States with interface-RMSD values larger than 10 Å should be carefully considered, given that the typical structural heterogeneity of the POI−PROTAC-E3 ligase ternary complex is moderate.[11] And states with large interface-RMSD usually serve as connectors for diverse, short-lived PPI binding modes that rapidly interconvert. Subsequently, the equilibrium populations of states and the BSA of configurations within states can serve as references for PPI binding affinity. States with very low populations (e.g., < 1%) and very small BSAs (e.g., < 500 Å^2)[13] should be used cautiously for linker design. Then we could employ SASA to identify solvent-exposed heavy atoms on the warhead or ligand as potential linking sites. Targeting these exposed sites with a linker can largely retain the inherent binding modes of the warhead and ligand. The identification process can be highly system-dependent, as variations in pocket shapes and PPI binding modes significantly influence the conformations of the warhead and ligand. Meanwhile, it is also essential to integrate chemical synthesis knowledge during this step. Additionally, the distance between selected linking sites offers guidance for determining the appropriate linker length. PPI interfaces with linking sites that are too far apart should be discarded, as typical linkers span 5 to 15 carbon atoms in length.[25]

2.4. Our Predicted PPI Interface (State III) Agrees with the Structure Induced by an Experimentally Designed PROTAC

A recent experimental study by Johannes et al. successfully designed and completed the preclinical validation of a single small molecule degrader, targeting KRAS and related mutant cancer proteins with VHL E3 ligase.[44] In this study, the authors reported a cocrystal structure of the degrader in complex with KRAS and VHL at a resolution of 2.2 Å (PDB:

8QVU), as shown in Figure 5a. We find that the PPI interface in this ternary cocrystal structure is structurally similar to the most populated microstate from our State III (Figures 5a,b and S19). Upon further examination of the pairwise distance map of the crystal PPI interface, we observe a high degree of consistency with the pairwise distance map of the ensemble of interfaces within macrostate III (Figure 5c,d), where salt bridges and dipolar interactions stabilize the PPI. The interface-RMSDs between the crystal structure and the ensemble of interfaces from macrostate III and its corresponding most populated microstate are as small as 5.42 ± 3.67 Å and 3.76 ± 2.37 Å, respectively. The interface with the smallest interface-RMSD, visualized in Figure 5a, has a value of just 0.68 Å and shares an identical contact map with the crystal structure. Furthermore, the BSA of crystal structure is 1,556 Å^2, which is also consistent with State III (1,612 ± 367 Å^2) and its most populated microstate (1,587 ± 206 Å^2). This agreement between the PPI interfaces in State III with the experimental crystal structure provides compelling validation of the predictions from our IGME model. In contrast, among all interfaces generated from protein docking, we found only one that is closest to state III and exhibited the smallest interface-RMSD with the crystal structure. However, the PPI interfaces in this docking structure still show substantial differences in the contact map (with a Frobenius norm of 4.82 between the docking and crystal structures) and a significant interface RMSD of 8.28 Å compared to the crystal structure (see Figure S24), indicating that our MD simulations and IGME modeling significantly improve the accuracy of the initial docking structures.

Probing the dynamical interactions and identifying non-native PPIs between protein pairs without degrader poses

significant challenges for various experimental methods. The weak binding affinity of the encounter complex makes it challenging for the structure biology approaches like X-ray crystallography.[65−67] While NMR spectroscopy or hydrogen—deuterium exchange mass spectrometry can detect the interactions, their time-resolution is limited.[11,67] Recently, it has also been shown that data-driven machine-learning approaches such as AlphaFold and AlphaFold-Multimer face challenges in accurately predicting non-native PPIs, particularly when the interface area is limited.[68,69] Our results demonstrate that non-native PPIs in encounter complex could be systematically and accurately predicted in atomistic detail by integrating parallel short unbiased MD simulations with non-Markovian dynamics modeling (i.e., IGME). Our models provide advantages of revealing both non-native PPIs and their dynamic heterogeneities simultaneously, thereby offering ensembles of metastable PPI interfaces for subsequent high-throughput linker design. In the future, the transition rates between macrostates predicted by our IGME model could be valuable in aiding the design of flexible linkers to stabilize two distinct metastable PPI interfaces that interconvert rapidly. Therefore, we anticipate that IGME holds significant potential for generalization in future PROTAC discovery.

3. CONCLUSIONS

PROTAC-induced TPD is regarded as one of the most promising approaches for small molecule-based drug discovery. However, the rational design of PROTACs remains challenging due to factors such as the large size of the multiprotein system and the complex, dynamic protein interactions. In this study, we present a physics-based approach to identify the complete ensemble of intrinsic and dynamic PPI interfaces between KRAS and VHL proteins by investigating the linker-less encounter complex. Specifically, we demonstrate that our IGME model, a non-Markovian dynamics model, constructed from extensive MD simulations (∼1.5 ms), is able to elucidate the metastable states of PPI interfaces and accurately predict their associated thermodynamic and kinetic properties. We show that IGME models significantly outperform MSMs in predicting slow dynamics associated with the encounter complex formation between KRAS and VHL. The six metastable states characterized by our IGME model represent distinct PPI interfaces of the encounter complex. Upon evaluating the stabilities and geometries of the PPI interfaces in each state, we narrowed down to three states (State II, III, and V) with promising PPI interfaces for future PROTAC linker design. The interfaces from the selected metastable states are primarily maintained by electrostatic interactions and display local dynamic heterogeneity, serving as a good basis for linker docking. We validate our theoretical predictions by showing that one of our selected PPI interfaces (State III) is highly consistent with a recent cocrystal structure containing the PPI induced by an experimentally validated PROTAC for the KRAS-VHL system. We anticipate that our predicted PPI interfaces for the KRAS-VHL system will provide valuable insights for future linker design. We believe that the rigorous foundations of this strategy, grounded in physical simulations and statistical thermodynamics, will lead to broad applicability across diverse systems, facilitating more efficient designs of efficacious PROTACs.

4. METHODS

4.1. All-Atom MD Simulation Setup for KRAS-VHL Encounter Complex

The conformation of VHL protein (PDB ID: 1VCB)[70] and KRAS protein (PDB ID: 7RPZ)[71] are respectively derived from crystal structures identified through X-ray diffraction. We first utilize the PyRosetta docking package[72] to generate sets of initial KRAS-VHL encounter complex structures. The rigid-body docking is performed while restraining the distance between the linker attachment atoms of warhead 1 and the E3 ligand as 20 Å, ensuring the formation of interfaces appropriate for subsequent design. Subsequently, we employ the K-Means algorithm to categorize the obtained docking structures into 50 clusters based on their root-mean-square deviations. The structure nearest to each cluster center is chosen as the initial structure for the subsequent MD simulations. We then protonate and solvate the initial poses in cubic boxes with explicit TIP3P water[73] and add counterions to maintain the neutrality of the system. Next, we employ the OpenMM package[74] to conduct all-atom simulations across Folding@Home,[75] with the in-house parametrized force field for the small molecules (i.e., ligand and warheads) and the AMBER ff14SB force field[76] for the proteins. The final obtained data set used for postanalysis consists of 2,492 trajectories, totaling 1.51 ms of aggregate simulation time, with an average trajectory length of 605 ns. Please refer to Supporting Information Sections 1 and 2 and for more details of system setup and all-atom MD simulations.

4.2. Construction and Validation for Microstate-MSM

Following our proposed pipeline in Figure 1, we construct the microstate-MSM to study the inherent PPIs between KRAS and VHL protein. The detailed procedures are described below:

(1). Classify the collective motions at PPI interfaces via MoSAIC:[48] we initially construct representations for PPI interfaces by utilizing the internal pairwise distances between KRAS residues (170 residues) and VHL residues (149 residues), leading to 25 330 pairwise distance features (see Figure S2a). Subsequently, we apply the MoSAIC algorithm[48] to cluster these features into 27 communities, with approximately 10% of the features filtered out as unimportant noise (see Figure S2b and more details in Supporting Information Section 3.1). Through visualization of the features within each community, we further exclude 11 communities associated with collective conformational changes unrelated to PPI interfaces (see Figure S3), resulting in 16 communities encompassing a total of 14 402 features (see Figure S4 and more details in Supporting Information Section 3.1).

(2). Select the features capturing slow dynamics by Spectral-oASIS:[49] We apply the spectral-oASIS algorithm[49] for the second round of feature selection, through which 1,500 features are automatically identified. These features are shown to effectively capture the top three slowest dynamic modes. (see Figure S2c and more details in Supporting Information Section 3.2)

(3). Reduce dimensionality by tICA:[32,50] We employ tICA with kinetic mapping[51] to linearly construct five collective variables (CVs) from 1,500 features. The MD conformations are projected on these CVs and further clustered into 100 microstates using the K-Means algorithm. The optimal hyper-parameters (i.e., number of CVs, tICA relaxation time and the number of microstates) are determined by cross-validations with the generalized matrix Rayleigh quotient (GMRQ) score[52] (see Figure S5 and Supporting Information Section 4).

(4). Validate the microstate MSM: Based on the 100 microstates model, we further conduct Implied Time Scale (ITS) analysis and Chapman-Kolmogorov (CK) test and validate the Markovian lag time for microstates-MSM is 200 ns (see Figure S6 and more details in Supporting Information Section 4). Additionally, we validated that the microstates are well-connected and its transition probability matrix does not contain any disconnected components (see Figures S7 and S8).

4.3. IGME Modeling of Encounter Complex

To identify metastable states of the PPI interface, facilitate the interpretation, and acquire the associated thermodynamic and kinetic properties, we employ our recently developed IGME method[38] to construct a model comprising only six representative states. We first utilize the PCCA+ algorithm[53,54] to lump the 100 microstates into 6 macrostates, given the largest time scale gap is between the fifth and sixth transition modes (see Figure S6a). We performed a crisp state lumping by assigning each microstate to the macrostate with the highest membership value. Subsequently, the IGME is applied to accurately model the transition dynamics between macrostates, accounting for non-Markovian dynamics through the time-integration of memory kernel functions. Specifically, IGME precisely describes the evolution of the transition probabilities matrices (TPMs) with the lag time longer than memory relaxation time τ_k by $T(t \geq \tau_k) = A\hat{T}^t$, where matrices A and $\hat{T}$ are estimated from simulations (Supporting Information Section 5).

To determine τ_k, we utilize two approaches: one employs our previous quasi-MSM technique,[41] which computes the memory kernel matrix at various times using the greedy algorithm with discretized GME, the other approach involves applying IGME to approximate the time-integrated memory kernel (see more details in Supporting Information Section 5). The mean integral memory kernel (MIK), defined as $MIK(t) = \frac{1}{N}\sqrt{\sum_{i,j=1}^{N}\left(\int_0^t K_{ij}(\tau)d\tau\right)^2}$, computed from two approaches are well consistent and the memory relaxation time τ_k is decided as $50ns$ (Figure 2a).

To build the optimal non-Markovian dynamics model, we employ multiple sets of TPMs with different lag time range $\{T^{MD}(\tau_k + n\Delta t)\}_{n=0}^{L_{fit}}$ to estimate the matrices A and $\hat{T}$ using least-squares fitting with a Lagrangian approach.[77] The optimal range, parametrized by τ_k and L_{fit} are decided by time-averaged root mean squared error (RMSE) with respect to MD simulations. After a systematic scan, we ultimately identify the optimal fitting range: $\tau_k = 70ns$ and $L_{fit} = 80ns$ (see Figure S11a). The optimal IGME model, along with its predicted thermodynamic and kinetic properties for the metastable states, is visualized as a transition network (Figure S12 and more details in Supporting Information Section 5).

■ ASSOCIATED CONTENT

Data Availability Statement

The MD simulation trajectories have been deposited in Open Science Framework (10.17605/OSF.IO/MERS6). The force field parameters for the two warhead molecules and one ligand molecule are provided as supplementary data files.

ⓈⒾ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacsau.4c00503.

> Detailed descriptions of setups for protein−protein docking and all-atom MD simulations, principles and procedures for constructing and validating MSM and IGME models; Supplementary Figures section: Additional results and figures that complement the main text (PDF)
>
> Summary of forcefield parameters (ZIP)

■ AUTHOR INFORMATION

Corresponding Authors

Weiping Tang − *Lachman Institute for Pharmaceutical Development, School of Pharmacy, University of Wisconsin-Madison, Madison, Wisconsin 53705, United States;* ⓞ orcid.org/0000-0002-0039-3196; Email: weiping.tang@wisc.edu

Xuhui Huang − *Department of Chemistry, Theoretical Chemistry Institute, University of Wisconsin-Madison, Madison, Wisconsin 53706, United States; Data Science Institute, University of Wisconsin-Madison, Madison, Wisconsin 53706, United States;* ⓞ orcid.org/0000-0002-7119-9358; Email: xhuang@chem.wisc.edu

Authors

Yunrui Qiu − *Department of Chemistry, Theoretical Chemistry Institute, University of Wisconsin-Madison, Madison, Wisconsin 53706, United States; Data Science Institute, University of Wisconsin-Madison, Madison, Wisconsin 53706, United States;* ⓞ orcid.org/0009-0003-2200-0490

Rafal P. Wiewiora − *Psivant Therapeutics, Boston, Massachusetts 02210, United States*

Jesus A. Izaguirre − *Atommap Corporation, NY, New York 10013, United States*

Huafeng Xu − *Atommap Corporation, NY, New York 10013, United States;* ⓞ orcid.org/0000-0001-5447-0452

Woody Sherman − *Psivant Therapeutics, Boston, Massachusetts 02210, United States*

Complete contact information is available at:
https://pubs.acs.org/10.1021/jacsau.4c00503

■ REFERENCES

(1) Schneekloth, J. S., Jr.; Fonseca, F. N.; Koldobskiy, M.; Mandal, A.; Deshaies, R.; Sakamoto, K.; Crews, C. M. Chemical Genetic Control of Protein Levels: Selective in Vivo Targeted Degradation. *J. Am. Chem. Soc.* **2004**, *126* (12), 3748−3754.

(2) Zhou, P. Targeted protein degradation. *Curr. Opin. Chem. Biol.* **2005**, *9* (1), 51−55.

(3) Raina, K.; Crews, C. M. Chemical Inducers of Targeted Protein Degradation*. *J. Biol. Chem.* **2010**, *285* (15), 11057−11060.

(4) Békés, M.; Langley, D. R.; Crews, C. M. PROTAC targeted protein degraders: The past is prologue. *Nat. Rev. Drug Discovery* **2022**, *21* (3), 181−200.

(5) Lai, A. C.; Crews, C. M. Induced protein degradation: An emerging drug discovery paradigm. *Nat. Rev. Drug Discovery* **2017**, *16* (2), 101−114.

(6) Liu, Z.; Hu, M.; Yang, Y.; Du, C.; Zhou, H.; Liu, C.; Chen, Y.; Fan, L.; Ma, H.; Gong, Y.; et al. An overview of PROTACs: A promising drug discovery paradigm. *Mol. Biomed.* **2022**, *3* (1), 46.

(7) 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. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (15), 8554−8559.

(8) Gao, X.; Burris, H. A., III; Vuky, J.; Dreicer, R.; Sartor, A. O.; Sternberg, C. N.; Percent, I. J.; Hussain, M. H.; Rezazadeh Kalebasty,

A.; Shen, J. Phase 1/2 study of ARV-110, an androgen receptor (AR) PROTAC degrader, in metastatic castration-resistant prostate cancer (mCRPC). *J. Clin. Oncol.* **2022**, *40* (6_suppl), 17.

(9) Chamberlain, P. P.; Hamann, L. G. Development of targeted protein degradation therapeutics. *Nat. Chem. Biol.* **2019**, *15* (10), 937−944.

(10) Mostofian, B.; Martin, H.-J.; Razavi, A.; Patel, S.; Allen, B.; Sherman, W.; Izaguirre, J. A. Targeted Protein Degradation: Advances, Challenges, and Prospects for Computational Methods. *J. Chem. Inf. Model.* **2023**, *63* (17), 5408−5432.

(11) Dixon, T.; MacPherson, D.; Mostofian, B.; Dauzhenka, T.; Lotz, S.; McGee, D.; Shechter, S.; Shrestha, U. R.; Wiewiora, R.; McDargh, Z. A.; et al. Predicting the structural basis of targeted protein degradation by integrating molecular dynamics simulations with structural mass spectrometry. *Nat. Commun.* **2022**, *13* (1), 5884.

(12) Cowan, A. D.; Ciulli, A. Driving E3 Ligase Substrate Specificity for Targeted Protein Degradation: Lessons from Nature and the Laboratory. *Annu. Rev. Biochem.* **2022**, *91*, 295−319.

(13) Rui, H.; Ashton, K. S.; Min, J.; Wang, C.; Potts, P. R. Protein−protein interfaces in molecular glue-induced ternary complexes: Classification, characterization, and prediction. *RSC Chem. Biol.* **2023**, *4* (3), 192−215.

(14) Daurio, N. A.; Zhou, H.; Chen, Y.; Sheth, P. R.; Imbriglio, J. E.; McLaren, D. G.; Tawa, P.; Rachdaoui, N.; Previs, M. J.; Kasumov, T.; O'Neil, J.; Previs, S. F. Examining Targeted Protein Degradation from Physiological and Analytical Perspectives: Enabling Translation between Cells and Subjects. *ACS Chem. Biol.* **2020**, *15* (10), 2623−2635.

(15) Li, W.; Zhang, J.; Guo, L.; Wang, Q. Importance of Three-Body Problems and Protein−Protein Interactions in Proteolysis-Targeting Chimera Modeling: Insights from Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2022**, *62* (3), 523−532.

(16) Gadd, M. S.; Testa, A.; Lucas, X.; Chan, K.-H.; Chen, W.; Lamont, D. J.; Zengerle, M.; Ciulli, A. Structural basis of PROTAC cooperative recognition for selective protein degradation. *Nat. Chem. Biol.* **2017**, *13* (5), 514−521.

(17) Farnaby, W.; Koegl, M.; Roy, M. J.; Whitworth, C.; Diers, E.; Trainor, N.; Zollman, D.; Steurer, S.; Karolyi-Oezguer, J.; Riedmueller, C.; Gmaschitz, T.; Wachter, J.; Dank, C.; Galant, M.; Sharps, B.; Rumpel, K.; Traxler, E.; Gerstberger, T.; Schnitzer, R.; Petermann, O.; Greb, P.; Weinstabl, H.; Bader, G.; Zoephel, A.; Weiss-Puxbaum, A.; Ehrenhöfer-Wölfer, K.; Wöhrle, S.; Boehmelt, G.; Rinnenthal, J.; Arnhof, H.; Wiechens, N.; Wu, M.-Y.; Owen-Hughes, T.; Ettmayer, P.; Pearson, M.; McConnell, D. B.; Ciulli, A. BAF complex vulnerabilities in cancer demonstrated via structure-based PROTAC design. *Nat. Chem. Biol.* **2019**, *15* (7), 672−680.

(18) Cyrus, K.; Wehenkel, M.; Choi, E.-Y.; Han, H.-J.; Lee, H.; Swanson, H.; Kim, K.-B. Impact of linker length on the activity of PROTACs. *Mol. BioSyst.* **2011**, *7* (2), 359−364.

(19) Weerakoon, D.; Carbajo, R. J.; De Maria, L.; Tyrchan, C.; Zhao, H. Impact of PROTAC Linker Plasticity on the Solution Conformations and Dissociation of the Ternary Complex. *J. Chem. Inf. Model.* **2022**, *62* (2), 340−349.

(20) Paiva, S.-L.; Crews, C. M. Targeted protein degradation: Elements of PROTAC design. *Curr. Opin. Chem. Biol.* **2019**, *50*, 111−119.

(21) Chan, K.-H.; Zengerle, M.; Testa, A.; Ciulli, A. Impact of Target Warhead and Linkage Vector on Inducing Protein Degradation: Comparison of Bromodomain and Extra-Terminal (BET) Degraders Derived from Triazolodiazepine (JQ1) and Tetrahydroquinoline (I-BET726) BET Inhibitor Scaffolds. *J. Med. Chem.* **2018**, *61* (2), 504−513.

(22) Roy, M. J.; Winkler, S.; Hughes, S. J.; Whitworth, C.; Galant, M.; Farnaby, W.; Rumpel, K.; Ciulli, A. SPR-Measured Dissociation Kinetics of PROTAC Ternary Complexes Influence Target Degradation Rate. *ACS Chem. Biol.* **2019**, *14* (3), 361−368.

(23) Smith, B. E.; Wang, S. L.; Jaime-Figueroa, S.; Harbin, A.; Wang, J.; Hamman, B. D.; Crews, C. M. Differential PROTAC substrate specificity dictated by orientation of recruited E3 ligase. *Nat. Commun.* **2019**, *10* (1), 131.

(24) 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, J. E.; Fischer, E. S. Plasticity in binding confers selectivity in ligand-induced protein degradation. *Nat. Chem. Biol.* **2018**, *14* (7), 706−714.

(25) Dong, Y.; Ma, T.; Xu, T.; Feng, Z.; Li, Y.; Song, L.; Yao, X.; Ashby, C. R., Jr.; Hao, G.-F. Characteristic roadmap of linker governs the rational design of PROTACs. *Acta Pharm. Sin. B* **2024**.

(26) 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* (10), 5234−5254.

(27) Wu, J.; Wang, W.; Leung, C.-H. Computational strategies for PROTAC drug discovery. *Acta Mater. Med.* **2023**, *2* (1), 42−53.

(28) Tang, R.; Wang, Z.; Xiang, S.; Wang, L.; Yu, Y.; Wang, Q.; Deng, Q.; Hou, T.; Sun, H. Uncovering the Kinetic Characteristics and Degradation Preference of PROTAC Systems with Advanced Theoretical Analyses. *JACS Au* **2023**, *3* (6), 1775−1789.

(29) Wan, H.; Voelz, V. A. Adaptive Markov state model estimation using short reseeding trajectories. *J. Chem. Phys.* **2020**, *152* (2), 024103.

(30) Konovalov, K. A.; Unarta, I. C.; Cao, S.; Goonetilleke, E. C.; Huang, X. Markov State Models to Study the Functional Dynamics of Proteins in the Wake of Machine Learning. *JACS Au* **2021**, *1* (9), 1330−1341.

(31) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. Markov models of molecular kinetics: Generation and validation. *J. Chem. Phys.* **2011**, *134* (17), 174105.

(32) Pérez-Hernández, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noé, F. Identification of slow molecular order parameters for Markov model construction. *J. Chem. Phys.* **2013**, *139* (1), 015102.

(33) Mardt, A.; Pasquali, L.; Wu, H.; Noé, F. VAMPnets for deep learning of molecular kinetics. *Nat. Commun.* **2018**, *9* (1), 5.

(34) Qiu, Y.; O'Connor, M. S.; Xue, M.; Liu, B.; Huang, X. An Efficient Path Classification Algorithm Based on Variational Autoencoder to Identify Metastable Path Channels for Complex Conformational Changes. *J. Chem. Theory Comput.* **2023**, *19* (14), 4728−4742.

(35) Liu, B.; Qiu, Y.; Goonetilleke, E. C.; Huang, X. Kinetic network models to study molecular self-assembly in the wake of machine learning. *MRS Bull.* **2022**, *47* (9), 958−966.

(36) Konovalov, K. A.; Wu, C.-G.; Qiu, Y.; Balakrishnan, V. K.; Parihar, P. S.; O'Connor, M. S.; Xing, Y.; Huang, X. Disease mutations and phosphorylation alter the allosteric pathways involved in autoinhibition of protein phosphatase 2A. *J. Chem. Phys.* **2023**, *158* (21), 215101.

(37) Liu, B.; Xue, M.; Qiu, Y.; Konovalov, K. A.; O'Connor, M. S.; Huang, X. GraphVAMPnets for uncovering slow collective variables of self-assembly dynamics. *J. Chem. Phys.* **2023**, *159* (9), 094901.

(38) Cao, S.; Qiu, Y.; Kalin, M. L.; Huang, X. Integrative generalized master equation: A method to study long-timescale biomolecular dynamics via the integrals of memory kernels. *J. Chem. Phys.* **2023**, *159* (13), 134106.

(39) Yik, A. K.-H.; Qiu, Y.; Unarta, I. C.; Cao, S.; Huang, X. A Step-by-Step Guide on How to Construct Quasi-Markov State Models to Study Functional Conformational Changes of Biological Macromolecules *A Practical Guide to Recent Advances in Multiscale Modeling and Simulation of Biomolecules* Wang, Y.; Zhou, R. Eds.; AIP Publishing LLC: 2023.

(40) Jiang, H.; Sheong, F. K.; Zhu, L.; Gao, X.; Bernauer, J.; Huang, X. Markov State Models Reveal a Two-Step Mechanism of miRNA Loading into the Human Argonaute Protein: Selective Binding followed by Structural Re-arrangement. *PloS Comput. Biol.* **2015**, *11* (7), No. e1004404.

(41) Cao, S.; Montoya-Castillo, A.; Wang, W.; Markland, T. E.; Huang, X. On the advantages of exploiting memory in Markov state

models for biomolecular dynamics. *J. Chem. Phys.* **2020**, *153* (1), 014105.

(42) Huang, L.; Guo, Z.; Wang, F.; Fu, L. KRAS mutation: From undruggable to druggable in cancer. *Signal Transduction Targeted Ther.* **2021**, *6* (1), 386.

(43) Bond, M. J.; Chu, L.; Nalawansha, D. A.; Li, K.; Crews, C. M. Targeted Degradation of Oncogenic KRASG12C by VHL-Recruiting PROTACs. *ACS Cent. Sci.* **2020**, *6* (8), 1367−1375.

(44) Popow, J.; Farnaby, W.; Gollner, A.; Kofink, C.; Fischer, G.; Wurm, M.; Zollman, D.; Wijaya, A.; Mischerikow, N.; Hasenoehrl, C.; et al. Targeting cancer with small molecule pan-KRAS degraders. *Science* **2024**, *385* (6715), 1338−1347.

(45) Tolcher, A. W.; Wu, X.; Zhang, J. S.; Barve, M. A.; Gabrail, N. Y.; Sommerhalder, D.; Rodriguez Rivera, I. I.; Wilks, S.; Wu, Y.; Liu, R.; et al. A phase 1, first-in-human, open-label, dose escalation and dose expansion study of TST005 in patients with locally advanced or metastatic solid tumors. *J. Clin. Oncol.* **2023**, *41* (4_suppl), 10−1200.

(46) 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* (57), 7201−7211.

(47) Hon, W.-C.; Wilson, M. I.; Harlos, K.; Claridge, T. D. W.; 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* (6892), 975−978.

(48) Diez, G.; Nagel, D.; Stock, G. Correlation-Based Feature Selection to Identify Functional Dynamics in Proteins. *J. Chem. Theory Comput.* **2022**, *18* (8), 5079−5088.

(49) Litzinger, F.; Boninsegna, L.; Wu, H.; Nüske, F.; Patel, R.; Baraniuk, R.; Noé, F.; Clementi, C. Rapid Calculation of Molecular Kinetics Using Compressed Sensing. *J. Chem. Theory Comput.* **2018**, *14* (5), 2771−2783.

(50) Noé, F.; Nüske, F. A Variational Approach to Modeling Slow Processes in Stochastic Dynamical Systems. *Multiscale Model. Simul.* **2013**, *11* (2), 635−655.

(51) Noé, F.; Clementi, C. Kinetic Distance and Kinetic Maps from Molecular Dynamics Simulation. *J. Chem. Theory Comput.* **2015**, *11* (10), 5002−5011.

(52) McGibbon, R. T.; Pande, V. S. Variational cross-validation of slow dynamical modes in molecular kinetics. *J. Chem. Phys.* **2015**, *142* (12), 124105.

(53) Deuflhard, P.; Weber, M. Robust Perron cluster analysis in conformation dynamics. *Linear Algebra Appl.* **2005**, *398*, 161−184.

(54) Röblitz, S.; Weber, M. Fuzzy spectral clustering by PCCA+: Application to Markov state models and data classification. *Adv. Data Anal. Classi.* **2013**, *7* (2), 147−179.

(55) Galdeano, C.; Gadd, M. S.; Soares, P.; Scaffidi, S.; Van Molle, I.; Birced, I.; Hewitt, S.; Dias, D. M.; Ciulli, A. Structure-Guided Design and Optimization of Small Molecules Targeting the Protein−Protein Interaction between the von Hippel−Lindau (VHL) E3 Ubiquitin Ligase and the Hypoxia Inducible Factor (HIF) Alpha Subunit with in Vitro Nanomolar Affinities. *J. Med. Chem.* **2014**, *57* (20), 8657−8663.

(56) Diehl, C. J.; Ciulli, A. Discovery of small molecule ligands for the von Hippel-Lindau (VHL) E3 ligase and their use as inhibitors and PROTAC degraders. *Chem. Soc. Rev.* **2022**, *51* (19), 8216−8257.

(57) Bemis, T. A.; La Clair, J. J.; Burkart, M. D. Unraveling the Role of Linker Design in Proteolysis Targeting Chimeras. *J. Med. Chem.* **2021**, *64* (12), 8042−8052.

(58) Troup, R. I.; Fallan, C.; Baud, M. G. J. Current strategies for the design of PROTAC linkers: A critical review. *Explor. Targeted Anti-Tumor Ther.* **2020**, *1* (5), 273−312.

(59) Setia, N.; Almuqdadi, H. T. A.; Abid, M. Journey of Von Hippel-Lindau (VHL) E3 ligase in PROTACs design: From VHL ligands to VHL-based degraders. *Eur. J. Med. Chem.* **2024**, *265*, 116041.

(60) Zoppi, V.; Hughes, S. J.; Maniaci, C.; Testa, A.; Gmaschitz, T.; Wieshofer, C.; Koegl, M.; Riching, K. M.; Daniels, D. L.; Spallarossa, A.; Ciulli, A. Iterative Design and Optimization of Initially Inactive

Proteolysis Targeting Chimeras (PROTACs) Identify VZ185 as a Potent, Fast, and Selective von Hippel−Lindau (VHL) Based Dual Degrader Probe of BRD9 and BRD7. *J. Med. Chem.* **2019**, *62* (2), 699−726.

(61) Disch, J. S.; Duffy, J. M.; Lee, E. C. Y.; Gikunju, D.; Chan, B.; Levin, B.; Monteiro, M. I.; Talcott, S. A.; Lau, A. C.; Zhou, F.; Kozhushnyan, A.; Westlund, N. E.; Mullins, P. B.; Yu, Y.; von Rechenberg, M.; Zhang, J.; Arnautova, Y. A.; Liu, Y.; Zhang, Y.; McRiner, A. J.; Keefe, A. D.; Kohlmann, A.; Clark, M. A.; Cuozzo, J. W.; Huguet, C.; Arora, S. Bispecific Estrogen Receptor α Degraders Incorporating Novel Binders Identified Using DNA-Encoded Chemical Library Screening. *J. Med. Chem.* **2021**, *64* (8), 5049−5066.

(62) Han, X.; Zhao, L.; Xiang, W.; Qin, C.; Miao, B.; Xu, T.; Wang, M.; Yang, C.-Y.; Chinnaswamy, K.; Stuckey, J.; Wang, S. Discovery of Highly Potent and Efficient PROTAC Degraders of Androgen Receptor (AR) by Employing Weak Binding Affinity VHL E3 Ligase Ligands. *J. Med. Chem.* **2019**, *62* (24), 11218−11231.

(63) Shah, R. R.; De Vita, E.; Sathyamurthi, P. S.; Conole, D.; Zhang, X.; Fellows, E.; Dickinson, E. R.; Fleites, C. M.; Queisser, M. A.; Harling, J. D.; Tate, E. W. Structure-Guided Design and Optimization of Covalent VHL-Targeted Sulfonyl Fluoride PRO-TACs. *J. Med. Chem.* **2024**, *67* (6), 4641−4654.

(64) Tovell, H.; Testa, A.; Maniaci, C.; Zhou, H.; Prescott, A. R.; Macartney, T.; Ciulli, A.; Alessi, D. R. Rapid and Reversible Knockdown of Endogenously Tagged Endosomal Proteins via an Optimized HaloPROTAC Degrader. *ACS Chem. Biol.* **2019**, *14* (5), 882−892.

(65) Słabicki, M.; Kozicka, Z.; Petzold, G.; Li, Y.-D.; Manojkumar, M.; Bunker, R. D.; Donovan, K. A.; Sievers, Q. L.; Koeppel, J.; Suchyta, D.; Sperling, A. S.; Fink, E. C.; Gasser, J. A.; Wang, L. R.; Corsello, S. M.; Sellar, R. S.; Jan, M.; Gillingham, D.; Scholl, C.; Fröhling, S.; Golub, T. R.; Fischer, E. S.; Thomä, N. H.; Ebert, B. L. The CDK inhibitor CR8 acts as a molecular glue degrader that depletes cyclin K. *Nature* **2020**, *585* (7824), 293−297.

(66) Simonetta, K. R.; Taygerly, J.; Boyle, K.; Basham, S. E.; Padovani, C.; Lou, Y.; Cummins, T. J.; Yung, S. L.; von Soly, S. K.; Kayser, F.; et al. Prospective discovery of small molecule enhancers of an E3 ligase-substrate interaction. *Nat. Commun.* **2019**, *10* (1), 1402.

(67) Vaynberg, J.; Qin, J. Weak protein−protein interactions as probed by NMR spectroscopy. *Trends Biotechnol.* **2006**, *24* (1), 22−27.

(68) Pereira, G. P.; Gouzien, C.; Souza, P. C. T.; Martin, J. Challenges in predicting PROTAC-mediated Protein-Protein Inter-faces with AlphaFold. *bioRxiv* **2024**, 2024−03.

(69) Yin, R.; Feng, B. Y.; Varshney, A.; Pierce, B. G. Benchmarking AlphaFold for protein complex modeling reveals accuracy determi-nants. *Protein Sci.* **2022**, *31* (8), No. e4379.

(70) Stebbins, C. E.; Kaelin, W. G., Jr.; Pavletich, N. P. Structure of the VHL-EloninginC-EloninginB Complex: Implications for VHL Tumor Suppressor Function. *Science* **1999**, *284* (5413), 455−461.

(71) Wang, X.; Allen, S.; Blake, J. F.; Bowcut, V.; Briere, D. M.; Calinisan, A.; Dahlke, J. R.; Fell, J. B.; Fischer, J. P.; Gunn, R. J.; Hallin, J.; Laguer, J.; Lawson, J. D.; Medwid, J.; Newhouse, B.; Nguyen, P.; O'Leary, J. M.; Olson, P.; Pajk, S.; Rahbaek, L.; Rodriguez, M.; Smith, C. R.; Tang, T. P.; Thomas, N. C.; Vanderpool, D.; Vigers, G. P.; Christensen, J. G.; Marx, M. A. Identification of MRTX1133, a Noncovalent, Potent, and Selective KRASG12D Inhibitor. *J. Med. Chem.* **2022**, *65* (4), 3123−3133.

(72) Chaudhury, S.; Lyskov, S.; Gray, J. J. PyRosetta: A script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* **2010**, *26* (5), 689−691.

(73) Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A* **2001**, *105* (43), 9954−9960.

(74) 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.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PloS Comput. Biol.* **2017**, *13* (7), No. e1005659.

(75) Voelz, V. A.; Pande, V. S.; Bowman, G. R. Folding@home: Achievements from over 20 years of citizen science herald the exascale era. *Biophys. J.* **2023**, *122* (14), 2852−2863.

(76) 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* (8), 3696−3713.

(77) Wu, Y.; Cao, S.; Qiu, Y.; Huang, X. Tutorial on how to build non-Markovian dynamic models from molecular dynamics simulations for studying protein conformational changes. *J. Chem. Phys.* **2024**, *160* (12), 121501.