ACS OMEGA

Open Access

http://pubs.acs.org/journal/acsodf

Article

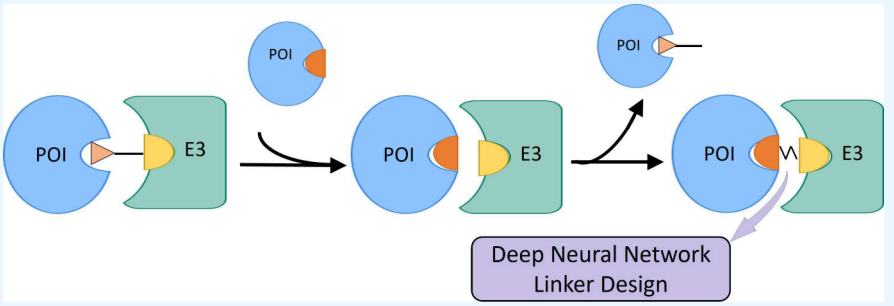
A Novel Rational PROTACs Design and Validation via AI-Driven Drug Design Approach

Cheng-Li Chou,[¶] Chieh-Te Lin,[¶] Chien-Ting Kao, and Chu-Chung Lin*

Cite This: *ACS Omega* 2024, 9, 38371−38384

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: The rational design of novel drug candidates presents a formidable challenge in modern drug discovery. Proteolysis-targeting chimeras (PROTACs) drug design is particularly demanding due to their limited crystal structure availability and design of a viable small molecule to bridge the protein of interest (POI) and ubiquitin−protein ligase (E3). An integrated approach that combines superimposition techniques and deep neural networks is demonstrated in this study to leverage the power of deep learning and structural biology to generate structurally diverse molecules with enhanced binding affinities. The superimposition technique ensures the congruence of initial and new protein−ligand pairs, which are evaluated via subsequent comprehensive screening using the root-mean-square deviation (RMSD), binding free energy (BFE), and buried solvent-accessible surface area (SASA). The final candidates are subjected to the incorporation of molecular dynamics (MD) and free energy perturbation (FEP) simulations to provide a quantitative evaluation of relative binding energies, reinforcing the efficacy and reliability of the generated molecules. The outcomes of the generated novel PROTACs molecules exhibit comparable structural attributes while demonstrating superior binding affinities within the binding pockets when contrasted with those of the established cocrystal ternary complexes. To enhance the generalizability of the workflow, we chose the ternary structure of the cellular inhibitor of apoptosis protein 1 (cIAP1) and Bruton's Tyrosine Kinase (BTK) for validating the chemical properties generated from the processes. The new linker molecules additionally showed superior affinity from the simulations. In summary, this methodology serves as an effective workflow to align computational predictions with current limitations, thereby introducing a novel paradigm in AI-driven drug design.

INTRODUCTION

Drug design faces several limitations and challenges that hinder the discovery and development of new therapeutics. These include the high cost and time associated with experimental screening methods, the limited availability of structural information for target molecules, and the inability to effectively address complex biological processes and interactions. Proteolysis-targeting chimeras (PROTACs) have recently been highlighted in the realm of therapeutic modalities for their potential to degrade targeted protein. PROTACs are heterobifunctional small molecules, which connect a target protein of interest (POI) and an E3 ubiquitin-protein ligase (E3) on each side via functional ligands.[1] The POI and E3 proteins are linked with an appropriate linker that conjugates the two ligands on an individual protein.[2] Degradation is initiated when a ternary complex is formed as POI-PROTACs-E3.[3] The ubiquitin-proteasome system subsequently seizes the targeted protein, facilitating its polyubiquitination, which is then followed by proteasomal degradation.[4] From a small molecule drug discovery point of view, designing a druggable PROTACs requires identifying the best spatial combination of the ternary complex and a thorough study of the structural and chemical properties of the POI and E3 ligase-complemented molecular dynamics.[5,6]

Multiple E3 ubiquitin ligases have been targeted for PROTACs development and represent promising chemical properties in drug discovery. One of the largest focuses with targeted protein degradation (TPD) is the members of the

Received: December 20, 2023
Revised: July 31, 2024
Accepted: August 15, 2024
Published: September 6, 2024

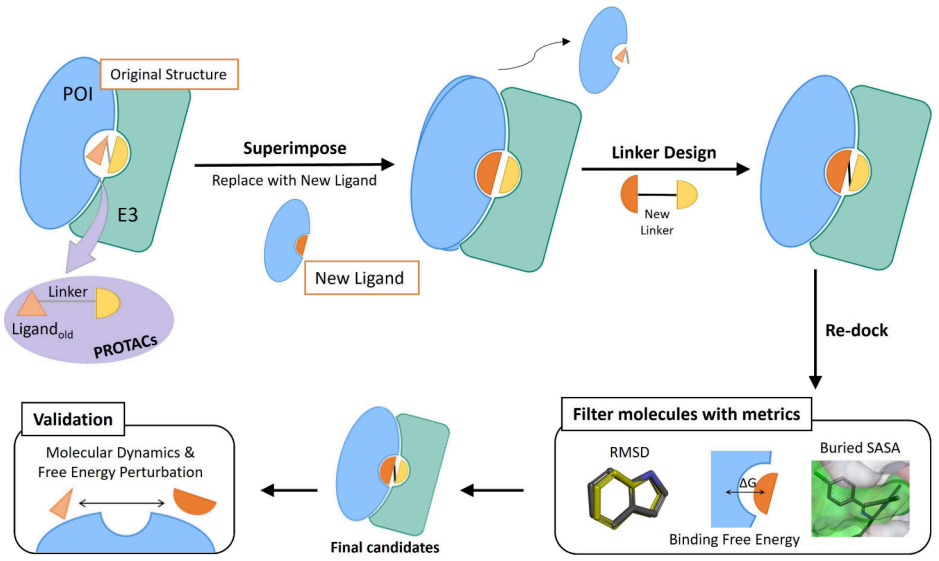


Figure 1. Schematic of the proposed workflow. The process includes (1) retrieving the ternary cocrystal structure of the POI-PROTACs-E3 ubiquitin ligase, (2) superimposing the POI and its linker with an identical POI attaching to a new ligand, (3) the new POI and E3 ubiquitin ligase are fed into the deep neural network so that the individual ligands are treated as fragments for linker design, (4) the generated novel molecules are filtered with multiple steps of chemical property screenings to remove the less druggable molecules, and (5) the final remaining candidates are validated with MD and FEP analysis to test the structural stability and dynamics of the generated molecules.

cullin RING ubiquitin ligase (CRL), which are modular protein assemblies centered around a cullin backbone.[7−9] Among the CRLs, the most commonly targeted ligases in clinical evaluation or approval are CRL2^{VHL} (hereafter called Von Hippel−Lindau tumor suppressor and abbreviated as VHL) and CRL4^{CRBN}. VHL is a substrate receptor that facilitates the specific recognition of target substrates to hijack ligases, along with other substrate receptors, contributing to the substrate specificity of CRLs. The concept of PROTACs involving the recruitment of E3 ubiquitin ligase and the POI has demonstrated successful applications, particularly in the case of VHL. This tumor suppressor plays a crucial role in oxygen sensing by targeting hypoxia-inducible factor α (HIF-α) subunits for ubiquitination and degradation. Hanzl et al.[10] have proven the innovative hybrid compounds, including PROTACs, effective in mapping mutations that confer resistance to degraders for a handful of E3 ligases.[11]

The application of PROTACs has shown success in the case of BRD4^{BD2} (hereinafter termed BRD4), which is a member of the bromodomain and extra terminal (BET) family. BRD4 is recognized for its involvement in orchestrating superenhancers and controlling the expression of oncogenes, thus making it a significant player in cancer biology.[12] Its significant POI associated with critical cancer pathways has been targeted for degradation. Recent studies showed PROTACs molecule structure of AT7[10] and MZ1[13] targeting on BRD4[14] to demonstrate successful BRD4 degradation through PROTACs that engage the VHL ligand on the BD2 domain. AT7 represents a potential small molecule inhibitor targeting BRD4, aiming to disrupt its interaction with acetylated histones and subsequently inhibit downstream oncogenic signaling pathways. The ternary complex formed by BRD4, AT7, and VHL represents a molecular assembly in which AT7 binds to BRD4, potentially influencing its activity, and VHL serves as a bridging molecule or a potential target for therapeutic intervention. However, the current ternary structure can be improved in its druggable potential and structural stability via state-of-the-art deep learning techniques.

Current studies leverage the deep neural network in PROTACs drug discovery for the rapid simulation and generation of novel structures. The technique demonstrates the feasibility of screening more druggable targets.[15,16] Graph neural networks (GNN) have emerged as a prominent technology for their intrinsic capability to learn specific task characteristics via graph representations while simultaneously retaining vital atom and bond interactions within the molecular structure.[17,18] By skillfully aggregating information from surrounding atoms encoded via the atom feature vector, the GNN effectively retains the representation of each atom with the option to feed chemistry laws to the network. The network employs recursive encoding to capture the connected bond feature vector in addition to atom information through messages passing across the molecular graphs. The ensuing readout operation consolidates these representations, atom-bond interactions, and incoherence.[19−21] Notably, the GNN models have demonstrated their superiority or comparability to traditional descriptor-based models in predicting chemical properties.[22,23]

Further, the utilization of a gated graph neural network (GGNN) has proven to surpass conventional molecular graph generation methods in deep generative models, showcasing its practicality in structure formation for drug design.[17,24,25] Currently, multiple studies rely on two-dimensional SMILES-based representations embedded in the feature vector to generate novel molecules.[26−29] However, these approaches overlook the crucial aspects of molecular structures in three-dimensional space, leading to significant deviation from the initial structure design. DeLinker, inspired by Liu et al., represents the first endeavor to incorporate GGNN in linker design, with the emphasis on retaining three-dimensional structural information and generating novel small molecules from the input fragments.[27,30] Nonetheless, the lack of effective methods for refining the generated molecules and the absence of comprehensive validation in molecular conformations remain evident in these studies.

https://doi.org/10.1021/acsomega.3c10183
ACS Omega 2024, 9, 38371−38384

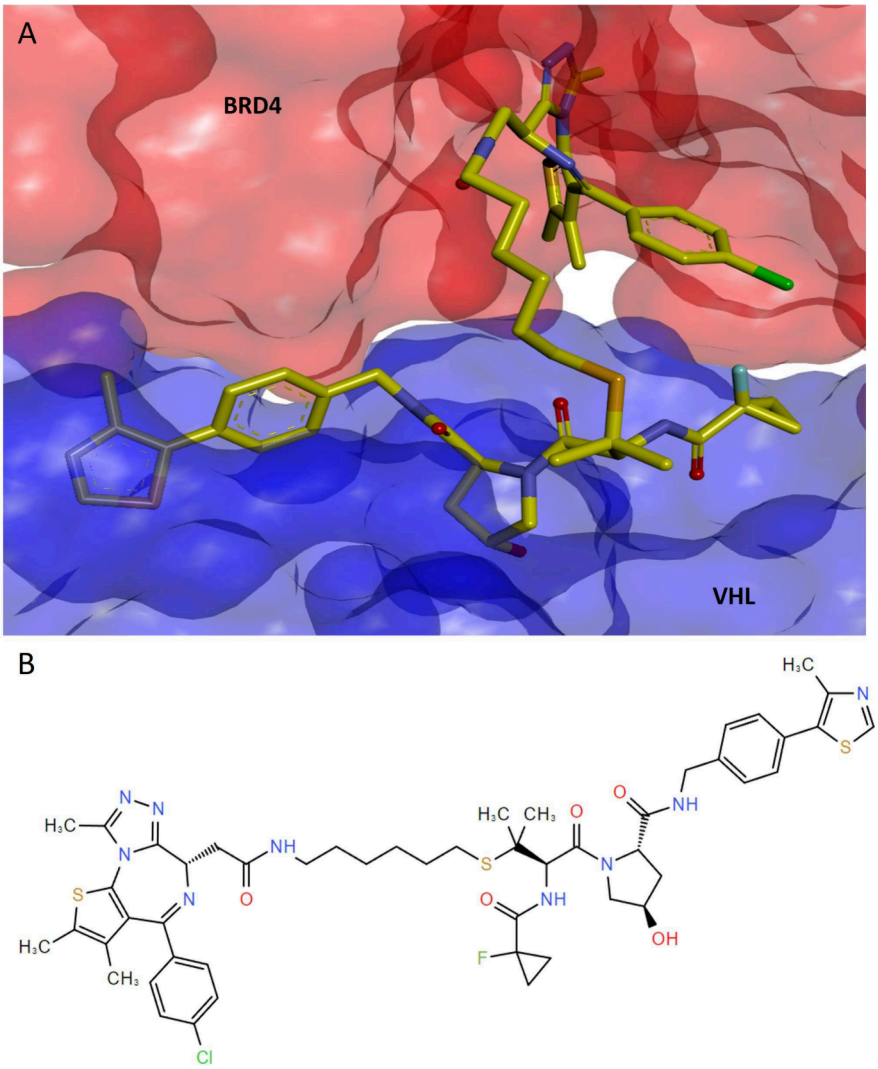


Figure 2. BRD4-AT7-VHL ternary structure and the chemical structural formula of AT7 are depicted in the following manner. (A) The spatial information on 7znt structure. The proteins of BRD4 and VHL are labeled by red and blue labels. (B) The chemical structural formula illustration of the AT7 molecule shows its ligand linkage to the E3 ubiquitin ligase and POI.

To overcome the limitation of sparse PROTACs ternary structure and adopt the emerging computational networks, this study demonstrates the first methodology to integrate the superimposition of a POI-ligand pair and deep neural network in PROTACs linker generation. First, we superimpose the new POI-ligand pair to substitute the POI-AT7 pair. The new POI-ligand pairs, BRD4-MS417 and BRD4-Compound 18 (C18) complexes, are verified via molecular docking, indicating that these molecules reside in the same binding pocket as the initial BRD4-AT7 pair. Further, the adopted deep neural network, AIMLinker,[31] generates novel molecules to bridge the POI and E3 ligands with consideration of the structural 3D information. The generated molecules are subjected to molecular docking analysis to extract the molecules with better chemical properties. This docking assessment includes the calculation of root-mean-square deviation (RMSD), evaluation of relative binding free energy (BFE), and buried solvent-accessible surface area (SASA). Lastly, the molecular dynamics (MD) and free energy perturbation (FEP) simulations are applied to measure the robustness and effectiveness of the candidate molecules. This comprehensive pipeline establishes a pioneering approach, wherein overcoming the current limited POI-PROTACs-E3 ternary structure and implementing state-of-the-art deep learning are required to advance the domain of drug discovery. Additionaly, we validate the pipeline with another POI-linker-E3 pair to show the reliability.

METHODS

We provide the pipeline of our proposed methodology in Figure 1. The original ternary structure of POI-PROTACs-E3 ubiquitin ligase is retrieved from Protein Data Bank (PDB).[32] A crucial aspect of TPD involves the binding of ligands to the POI and the E3 ubiquitin ligase on each end and a PROTACs molecule that enhances affinity and leads to the formation of a ternary structure. We then superimpose the POI and its associated linker with a POI variant, which includes a new ligand for further PROTACs design. Next, the new POI and initial E3 ubiquitin ligase are treated as input data for the deep neural network as fragments, providing the initial anchors to design novel PROTACs linkers. The produced novel molecules from the network form new ternary structures with diverse binding chemical properties. We filtered the novel ternary complexes with buried SASA, BFE, and RMSD to exclude the molecules with less viability and less potential to be

https://doi.org/10.1021/acsomega.3c10183
ACS Omega 2024, 9, 38371−38384

druggable. Lastly, we validate the final candidates with MD and FEP analyses to verify the robustness of the generated molecules.

Superimposition. The plastic binding of the PROTACs between the ligase and the substrate exhibits diverse conformations influenced by both linker length and relative position. The PROTACs molecule AT7 stands out for its remarkable selectivity properties with degradation activity (DC_{50}) of approximately 20.8 nM and BFE of −30.94 kcal/mol, and its ternary cocrystal structure, BRD4-AT7-VHL (PDB: 7znt), has been elucidated by Hanzl et al. and deposited in the PDB.[10] This 7znt ternary complex, integrating structural, biochemical, and cellular aspects, represents a neodegrader-mediated PROTACs structure. To visualize the relative spatial arrangement of the BRD4-AT7-VHL complex, Figure 2A displays BRD4 and VHL structures via Discovery Studio Visualizer (DSV),[33] with AT7 acting as the bridging molecule connecting E3 ubiquitin ligase VHL and target protein ligase BRD4. Figure 2B illustrates the chemical structural formula of AT7.

The core of our methodology involves a robust structural alignment and superimposition procedure to seamlessly integrate the new protein−ligand complex within the binding pocket of interest. We leverage the relatively larger availability of POI-ligand structure with better chemical properties in the same binding pocket as BRD4-AT7-VHL. We identify BRD4-C18 (PDB: 6vix) and BRD4-MS417 (PDB: 6duv) as two superimposition POI-ligand targets, which contain better or comparable inhibition activity.[34] Since the chosen POI-ligand targets lie in the same binding pocket as the AT ternary structure, the superimposition pose shows a structurally reasonable length between BRD4 and VHL for PROTACs linker design and is favorable to be druggable. The RMSD tool in BIOVIA Discovery Studio (DS) was used to perform the superimposition process. During the alignment process, it is not solely based on the protein backbone; it also encompasses side-chain conformations and ligand orientations.

We retrieve the BRD4-ligand structure from Faivre et al.,[35] where the BRD4 along with its ligand chemical structure is released to logically select the reasonable begin atom when designing the PROTACs linker molecule. Next, this BRD4-ligand is replaced with the new BRD4 attaching to a new ligand, and we ensure the protein backbone alignment to form a new POI-ligand pair. We then identify the starting atoms for designing linker molecules on the new BRD4-ligand pair. Lastly, we superimpose the new POI-ligand pair to form a configuration of separate POI-ligand and E3-ligand molecules.

PROTACs Generation. In preparation for linker generation, we employ the crystal pose of the BRD4-AT7-VHL complex and the corresponding binding moieties to design the PROTACs linker. The input data for the deep neural network are two fragments comprising individual ligands from BRD4 and VHL, respectively. The original linker moiety from the cocrystal structure is removed. We extracted the AT7 molecule from the 7znt structure. Notably, the fragment of the BRD4 ligand, established as a potent inhibitor by Filippakopoulos et al.,[36] is visually depicted in Figure 2B. The VHL ligand is characterized by a structure akin to the ligand released by Frost et al.[37] The closest unlinked atoms are predefined as the anchors for parameter calculations for the network. The processed fragments provide essential 3D spatial data, namely, the relative angle and distance in the free space. The information is calculated by the predetermined and unchanging nature of the cocrystal structure obtained from the PDB. Subsequently, the network utilizes this spatial information along with the two fragments as input parameters. The details about the multimodal neural network are described in the following sections.

Architecture. The multimodal encoder-decoder network, AIMLinker, presents a data-driven deep learning network that seamlessly integrates the generation and design of novel structural linkers between input fragments. This is followed by a postprocessing step to refine the predicted structures. This work adopts a network for PROTACs linker generation. Built upon insights from prior works by Imrie et al.[30] and Liu et al.,[27] AIMLinker capitalizes on unlinked fragments, leveraging their relative spatial position and orientation data to produce linker structures. AIMLinker is fine-tuned through training and modifications to meet the unique requirements of producing PROTACs molecules, subsequently undergoing postprocessing to enhance the drug-like potential of the generated molecules.

The generation process involves iterative edge generation and atom addition, with a focus on 14 permitted atom types. Utilizing a breadth-first approach, the model generates molecules while adhering to basic atomic valency rules through a masking step. Additionally, it is flexible to specify the number of atoms between anchors, thus maximizing variation in linker molecules while ensuring validity of the generated structures. The network is also capable of determining the quantity of generated molecules, followed by a postprocessing stage that removes chemically unsound, duplicate, and illogical structures. This process serves as the additional step to remove the unwanted molecules due to the nature of the noisy output from the network.

The architecture utilizes a standard GGNN to process input fragments into a graph representation. Within this structure, each atom and bond corresponds to a node and edge, respectively, labeled as a and k. The graph information is channeled through the GGNN core encoder structure, facilitating the seamless integration and continuous refinement of the hidden states associated with nodes and edges. This dynamic integration occurs iteratively throughout the learning process, enabling the network to progressively enhance its understanding of the structural relationships within the molecular entities.

Initiating with node expansion, each node holds a random hidden state, z_v, extracted from a d-dimensional normal distribution, where l signifies the number of features of the hidden state. The expansion nodes are subsequently denoted as h_v, accompanied by structural information derived from the SoftMax output of a learned mapping function, represented as y. This mapping function can be substituted with other functions to associate hidden state z_v with diverse atom types, influenced by the selected linker length. Next, the iterative process of edge selection, edge labeling, and node updating generates new molecules. The initial node u considers the length and spatial information on the neighborhood node v, configured by the initial input fragment anchors. The graph representation of the connected node is repeatedly added to the chain until the expansion nodes are all chained without further formation. The subsequent stages involve assessing the potential for edge formation based on the basic valency constraint and updating nodes through GGNN. The constructed feature vector $\Phi^t_{u,v}$ with subsequent node v is denoted as $h_v \sim y(z^t_u)$. The feature vector at time point t considers the adjacent node:

https://doi.org/10.1021/acsomega.3c10183
ACS Omega 2024, 9, 38371−38384

$$\Phi_{u,v}^{t} = [h_u^t, h_v^t, d_{u,v}, H^0, H^t, I]$$

where h_u^t and h_v^t are the hidden states with the distance $d_{v,u}$ in the graph representation. H^0 and H^t indicate the local and global information on the nodes at the time point t, respectively. The length and relative angle of the anchors in the 3D structural information are stored as I. The probability distribution of the candidate edge is represented as

$$p(u \leftrightarrow v|\Phi_{u,v}^{t}) = p(k|\Phi_{u,v}^{t}, u \leftrightarrow v)\cdot p(u \leftrightarrow v|\Phi_{u,v}^{t})$$

where the edge k can be single, double, or triple bonds subject to the basin valency constraints.

Upon completion of the iteration process, the network produces the largest intact molecule while discarding unconnected nodes. Notably, the stereochemistry information on the generated molecules is not provided during the generative process, necessitating postprocessing steps to screen the predicted molecules and test their robustness. In essence, the model emerges as a comprehensive and versatile approach for designing PROTACs molecules, fusing deep learning and graph-based methodologies to achieve tailored structural linkers.

Model Training. A training data set consisting of a conventional ZINC data set[38,39] and PROTAC-DB[40] is employed for training within a variational autoencoder (VAE) framework. The data set includes 160,491 molecules, with 157,221 molecules from ZINC and 3270 PROTACs-specific molecules from PROTAC-DB. The ZINC data set contains chemically complex compounds to emulate the ligand buried in the large POI-PROTACs-E3 ternary structure, while the PROTAC-DB data set encompassed all available compounds. Each molecule is divided into two fragments and a linker segment, ensuring that the linker has a minimum of three heavy atoms while retaining the possibility of generating intact ring structures.

The data set is split into 90% for training and 10% for validation utilizing 10-fold cross-validation to mitigate overfitting. The hyperparameters are tuned for optimal performance. The model focuses on fragment-molecule pairs, aiming to reconstruct a linked molecule Y from input fragments Y and a latent code z, which represents the learned mapping. The latent code was constrained to a low-dimensional vector to enforce effective learning from input fragments and subsequent regeneration. The loss function combined a reconstruction term and a Kullback–Leibler regularization loss, with an encoder representing the Gaussian probability density of node positions and a decoder aiming to reconstruct input molecules from the fragments. The Kullback–Leibler regularization term managed the divergence between predicted molecular spatial distributions and the probability vector derived from the linked molecule. This approach introduced variations to the traditional VAE loss, as previously suggested by Yeung et al.[41] The training process focused on enhancing the network's ability to predict atom types and reconstruct targeted molecules.

Postprocessing. The model generates a library of 2D chemical structures for PROTACs molecules, which undergo postprocessing to eliminate undesirable targets. The postprocessing involves several filters integrated into the model, addressing issues arising from the graph computational process and linked substructures. The first filter removes duplicate predictions, resulting in a unique molecule set. Additionally, nonlinker substructures, where two fragments do not form a compound with the linker moiety, are filtered out. Another filter identifies unfavorable substructures that are unsuitable for chemical synthesis or as druggable targets such as acid halides, disulfide bonds, and small cyclic rings with double bonds. This library is further refined by assessing the feasibility of newly generated substructures as potential drug leads.

Moreover, molecules violating Bredt's Rule,[42] containing specific bridged-ring structures with a carbon–carbon double bond at a bridgehead atom, are excluded from the target pool. These filtration steps effectively reduce unwanted molecules, optimizing computational resources and simulation time. The systematic utilization of these filters ensures that the resulting molecule pool consists of candidates with promising drug-like properties, enhancing the quality and applicability of the generated novel PROTACs molecules.

Molecular Properties Filtration. *Molecular Docking.* In order to select potential molecular candidates, it is essential to construct 3D protein–protein interaction poses and 3D conformations of the generated molecules. The cocrystal structure of BRD4-AT7-VHL is available in the PDB and the simulated spatial conformations were retrieved via DSV (Table S1). Additionally, the 2D chemical structures of postprocessed molecules were converted into 3D PROTACs conformations through DSV. To validate the robustness of our docking methodology, the reference compound AT7 was subjected to the same process, transforming it into a series of 3D conformations. The energy minimization techniques were subsequently applied to optimize these 3D forms of molecules.[43]

Following the preparation of protein–protein interaction poses and 3D conformations for PROTAC candidates, the docking analysis employs CDOCKER program[44] embedded in DS to retrieve optimal PROTACs binding poses. This was achieved by configuring the binding pocket as a grid, enabling each 3D PROTACs to bind freely to our superimposed BRD4-VHL structure. The binding site sphere was defined as having an 8 Å radius surrounding the entire PROTACs binding pocket. During the docking procedure, aspects such as the binding energy, biochemical properties, and entropy are comprehensively considered. For each PROTACs candidate, 10 docking poses were generated and passed to subsequent RMSD analysis.

RMSD Calculation. The RMSD introduced by Bell et al.[45] was employed as a quantitative tool to measure the structural similarity between corresponding atoms within two molecules:

$$\mathrm{RMSD}(v, w) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(v_i - w_i)^2}$$

where v and w denoted as the atoms in AT7 and generated molecule, receptively. Notably, n represents the atom count in the ligand, while v_i signifies the Euclidean distance between the i-th pair of the corresponding atom w_i. By using the RMSD tool within DS, the superimposing and calculating RMSD values allows for evaluating the structural similarity level of the generated linker molecules against the cocrystal structure reference compound in the PDB.

Binding Free Energy Simulation. To access the binding affinity between the protein and ligand, we employed the *Calculation Binding Energies* protocol in DS, coupled with the *Generalized Born with Molecular Volume* implicit solvent model[46] for predictive analysis. The BFE between a protein and a ligand within a solvent environment was determined through the following formulation:

https://doi.org/10.1021/acsomega.3c10183
ACS Omega 2024, 9, 38371−38384

$$\Delta G_{\mathrm{BFE}} = G_c - (G_l + G_p)$$

where G_c indicates the total free energy associated with the protein−ligand complex, whereas G_l and G_p correspond to the respective total free energies of the separated ligand and protein in the solvent.

To further enhance the rigor of our investigation, we incorporated spatial constraints by imposing a fixed spatial position for the BRD4-VHL interactions. We afford the generated bridging molecules freedom of movement to allow a more accurate assessment of their interaction potential. Through this ΔG_{BFE} analysis and consideration of spatial constraints, we provided a reliable evaluation of the validity and potential interaction capability of the generated molecules within the protein−ligand complex context.

Buried SASA Calculation. We measured the SASA of the ligand within the protein binding pocket via the *Analyze Ligand Poses* protocol to provide valuable insights into the interaction dynamics between the ligand and the receptor. Specifically, we calculated the $\mathrm{SASA_{ligand\text{-}pocket}}$, which corresponds to the SASA of the ligand when it was situated within the protein binding pocket. Furthermore, we determined the ligand SASA in isolation, representing $\mathrm{SASA_{ligand}}$ without the presence of the receptor. This isolated SASA measurement served as a baseline reference for the ligand surface area when not engaged with the protein.

The $\mathrm{SASA_{buried}}$ ratio was then computed as a means to quantitatively assess the extent of burial or burial efficiency of the ligand within the binding pocket.[47] It is calculated as

$$\mathrm{SASA_{buried}}\ (\%) = \left(1 - \frac{\mathrm{SASA_{ligand\text{-}pocket}}}{\mathrm{SASA_{ligand}}}\right) \times 100\%$$

where $\mathrm{SASA_{ligand\text{-}pocket}}$ represents the ligand SASA value in the binding pocket, and $\mathrm{SASA_{ligand}}$ is the SASA of the ligand without the receptor. The resulting $\mathrm{SASA_{buried}}$ value provided a percentage representation of the ligand surface area that was shielded or "buried" upon interaction with the receptor.

This detailed SASA analysis contributed to the ligand−receptor interaction landscape. The $\mathrm{SASA_{ligand\text{-}pocket}}$, $\mathrm{SASA_{ligand}}$, and $\mathrm{SASA_{buried}}$ ratios collectively offered valuable quantitative metrics to characterize the binding mode and binding efficiency of ligands within the binding pocket of the protein. This approach enhanced our ability to discern the structural aspects of ligand−protein interactions to facilitate the verification of the chemical properties of the generated molecules.

Molecular Dynamics Simulation. MD simulates the motion of atoms and molecules according to classical mechanics, offering insights into the conformational changes and the flexibility of biomolecular systems. It allows various binding modes and the prediction of ligand−protein interactions with a high accuracy. By doing MD simulations, we can capture dynamic events such as protein−ligand binding and unbinding processes, enabling the characterization of binding kinetics and thermodynamics.

The binding affinities of the eight selected PROTACs (AT7 crystal pose, C18−02−0014, C18−02−0019, C18−02−0055, C18−02−0130, MS417−01−0019, MS417−01−0157, and MS417−01−0164) with BRD4-VHL were further evaluated via 10 ns MD simulations perturbation. All MD simulations were conducted using the DS with the CHARMM force field,[48] which was used to obtain the energy-minimized conformations of all the processed complexes. Next, the generated complexes were solvated with an explicit periodic boundary model with added TIP3P water molecules to the cubic simulation boxes.[49] The solvated complexes were followed by neutralization using sodium and chloride ions.

Each BRD4-PROTACs-VHL complex was initially minimized to remove the unfavorable contacts for 200 steps by using the steepest descent method and subjected to 20 ps of heating from 50 to 300 K. Subsequently, a 50 ps equilibrium run was performed. Finally, periodic boundary dynamics simulations of 10 ns were conducted for the production step in a constant-temperature, constant-pressure ensemble at 1 atm and 300 K. All covalent bonds involving hydrogen were constrained during the simulations using the SHAKE algorithm.[50] The particle mesh Ewald method was used to treat the long-range electrostatic interactions.[51] The output trajectory files were saved every 2 ps from a 10 ns period and used for subsequent structural stability analysis, encompassing binding free energy, ligand nonbonded interactions, and the protein Radius of Gyration (Rg). The number of ligand nonbonded interactions was calculated between the PROTACs molecule and the BRD4-VHL. In the MD simulations of proteins, Rg is often used to assess the compactness of a protein, which represents the root-mean-square distance of all atoms in a molecule from their common center of mass. The calculated Rg value was subtracted from the initial Rg value (starting structure).

Free Energy Perturbation Simulation. FEP complements MD simulation by systematically calculating the ligand structure perturbation and the corresponding free energy differences, providing insights into the energetics of ligand binding. This approach facilitates rational drug design by enabling the prediction of relative binding affinities for a range of ligand analogs, thereby guiding scaffold hopping and optimization efforts.

We employed FEP simulations to measure the binding affinities of the generated molecules with the reference cocrystal structure complex. FEP simulations provided ligand−protein interactions and quantified the associated binding energies, thereby enabling insights into molecular recognition and phase transitions. To facilitate accurate prediction and characterization, we selected the best two distinct PROTACs configurations of each generated structure with relatively better chemical properties of conformation with BRD4-VHL. Notably, the ternary structure was free of constraints to emulate the natural in vivo conditions. By using the *CHARMM Relative and FEP calculations* protocol in DS, we conducted FEP simulations employing a lambda window protocol. Specifically, we employed a range of 20 lambda windows (0.0, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.85, 0.9, 0.91, 0.93, 0.95, 0.96, 0.97, 0.98, 0.99, and 1.0) throughout the FEP schedule. Each simulation run encompassed a duration of 8 ns, and we performed three independent FEP schedule runs to ensure the robustness and reproducibility of our results.

The core objective of our FEP simulations was the computation of the relative BFE between two ligand states, denoted as l_0 and l_1. This quantification was expressed as $\Delta\Delta G_{\mathrm{binding}}^{l_0 \to l_1}$, capturing the energetic difference associated with the transformation from ligand state l_0 to l_1 in the binding complex:

$$\Delta\Delta G_{\mathrm{binding}}^{l_0 \to l_1} = \Delta G_c^{l_0 \to l_1} - \Delta G_l^{l_0 \to l_1}$$

https://doi.org/10.1021/acsomega.3c10183
ACS Omega 2024, 9, 38371−38384

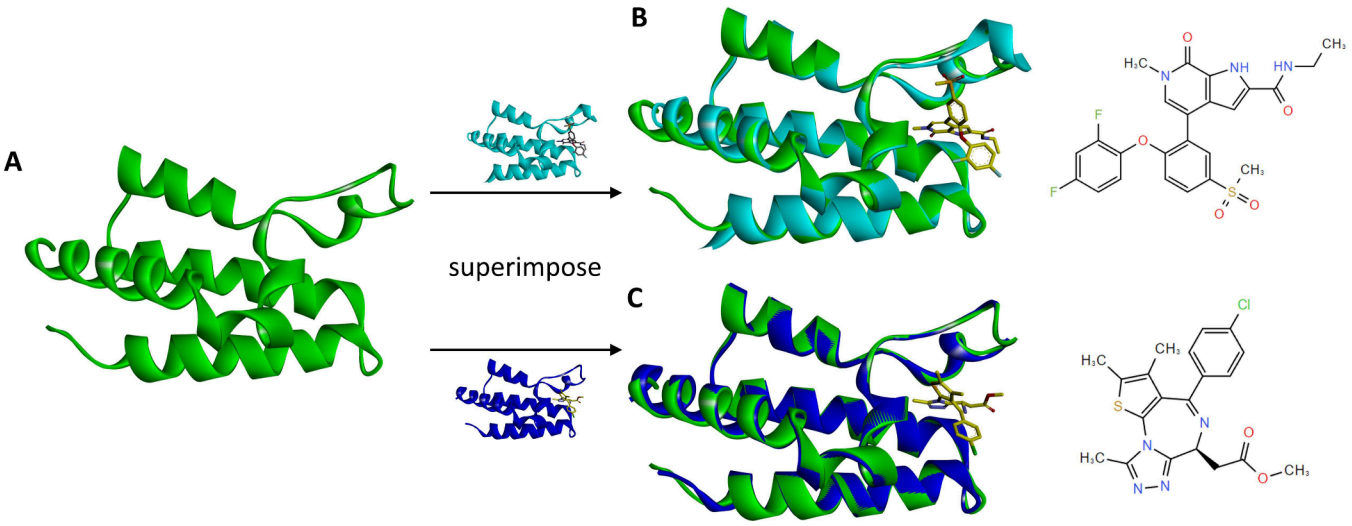


Figure 3. Schematic of the superimposition process of BRD4-ligand pairs. (A) The initial structure of BRD4 is labeled in green. We superimpose the reference complex with a new BRD4-ligand pair. (B) The superimposition structure of BRD4-C18 is labeled in light blue and the C18 ligand chemical formula is shown on the far right. (C) MS417 and its BRD4 structures are depicted as dark blue and superimposed with the reference complex. The congruence shows a high structural similarity of the new and old POI-ligand pairs. Note that the protein backbone similarity is the focus compared to the ligand congruent level.

The formula is defined as the difference between the free energy change in the complex ($\Delta G_c^{l_0 \to l_1}$) and the energy change of the ligand in solution ($\Delta G_l^{l_0 \to l_1}$). This formulation enabled us to evaluate and compare the binding affinities of different ligand states within the BRD4-PROTACs-VHL complex. By leveraging the FEP simulation, we provided the thermodynamics of ligand−protein interactions to test the rational design of potent ligands with improved binding affinities.

RESULTS

We demonstrate a first-in-class methodology of integrating POI-ligand pair superimposition and designing novel PROTACs linkers via deep neural networks. The following results first show the superimposition of the structural similarity to validate that the POI-ligand pair is effectively overlaying on the protein backbone and leaving the ligand part to be the target designing PROTACs molecules. Next, we show the performance of generating molecules via a deep neural network. The model provides novel PROTACs molecules that are favorable to be druggable. Further, the docking process includes RMSD, BFE, and buried SASA calculations to subsequently validate and screen the generated molecules to retain better structures with better chemical properties as the final candidates. Finally, MD and FEP analyses are adopted to compare the robustness of the candidates with the AT7 crystal structure.

Superimposition. In Figure 3, the application of the superimposition technique is demonstrated to show the extent of structural congruence between the POI and its respective ligand. The visualization is facilitated by DSV to employ distinct color coding and highlight the matched structural elements. Specifically, the original BRD4 (labeled in green) in Figure 3A serves as the reference configuration, showing the structural alignment of the new protein−ligand pair. The superimposed structure includes new BRD4 with C18 and MS417, respectively. Figure 3B showcases the alignment of BRD4-C18 (labeled in light blue) accompanied by the chemical formula of C18 presented on the adjacent side. The new structure of the BRD4-C18 pair shows high congruence with the initial BRD4 structure. Subsequently, Figure 3C illustrates the alignment between the reference pair and BRD4-MS417, which is labeled in dark blue. Both POI-ligand pairs exhibit high structural congruence with the reference complex.

PROTACs Generation. We utilize a deep neural network architecture to take the ligands on BRD4 and VHL as input, leading to the generation of novel PROTACs molecules through a deep GGNN architecture. The model operation entails the utilization of unlinked fragments located on either side of the ligands, with predefined anchor points offering crucial spatial angle and distance information, thus serving as essential parameters for the model configuration. The individual fragments from BRD4-C18 and BRD4-MS417 are inputted, enabling the model to harness its acquired knowledge from the training data set to produce novel linker molecules with precision and ingenuity.

We employ a controlled fragment range during molecule generation, affording the neural network the flexibility to craft structures ranging from linear to ring-like architectures. To ensure the production of viable drug leads, we systematically refine these outputs through a postprocessing regimen consisting of two filters. The first filter is designed to remove redundancy and nonlinker configurations, ensuring that the fragments effectively combine to form a unified compound via the linker structure. The resulting subset of molecules from this filter embodies uniqueness and novelty while falling short of possessing the characteristics necessary for drug candidacy. The second filter targets substructures incompatible with drug-like molecules, thereby eliminating elements unsuitable for our intended application. Upon the completion of these post-process procedures, the network generates a final set of 220 and 224 molecules for C18 and MS417, respectively. The generated molecules are novel and valid PROTACs structures that possess the capability of being druggable and potential leading targets.

Docking Process. We utilize CDOCKER as a docking and validation tool to facilitate the evaluation of the generated molecules in comparison with the established AT7 structure.

https://doi.org/10.1021/acsomega.3c10183
ACS Omega 2024, 9, 38371−38384

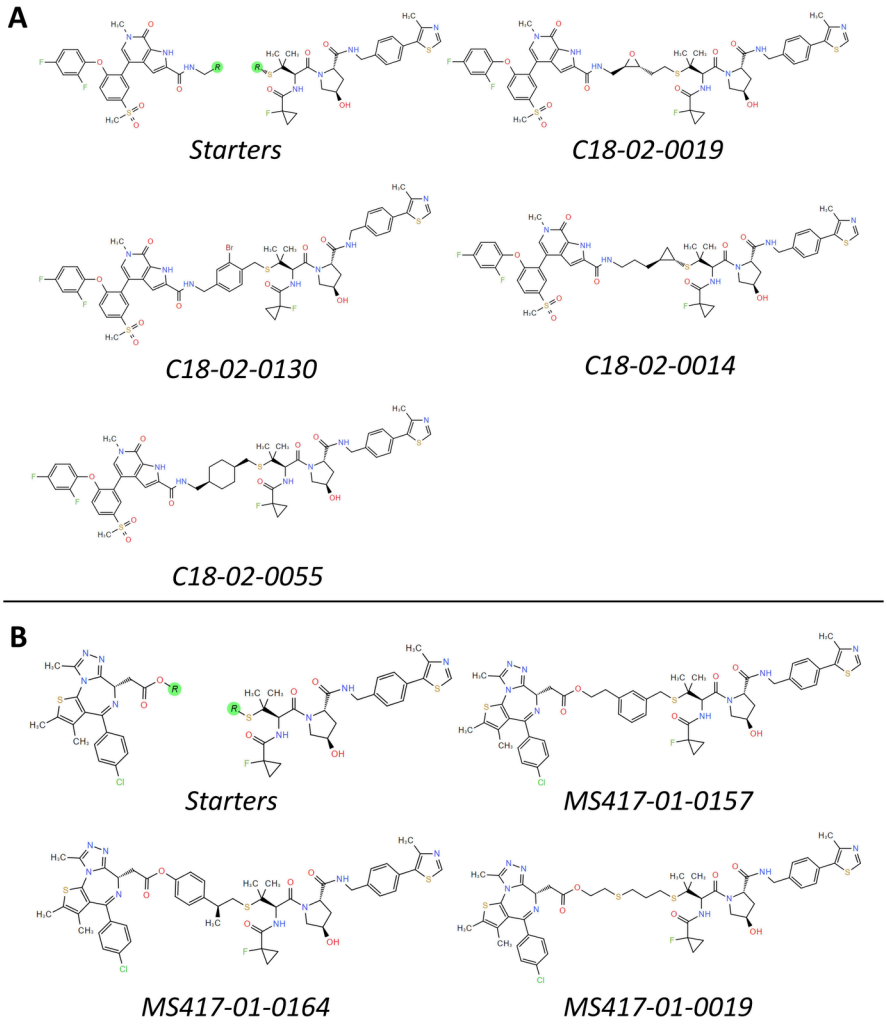


Figure 4. Superimposition process replaces the protein backbone and its ligand on BRD4 while retaining the identical ligand on VHL. In A, the MS417 PROTACs molecule without a linker structure is shown as the input fragments, and the potent initial atom for designing novel linker molecules is denoted as R*. The input fragments of C18 are shown in part B with the individual ligands on POI and E3. The labeling method is applied to the potential atom on both BRD4 and VHL ligands. The R* serves as the anchor to the deep neural network and provides spatial information about the relative spatial position of the input fragments. The generated molecules of the C18 and MS417 groups are separately shown in parts A and B, respectively. Note that these final candidate molecules were screened with RMSD, buried SASA, and BFE calculations.

The redocking procedure of the compound within the binding pockets of both BRD4 and VHL further solidifies the role of CDOCKER as a benchmarking instrument for estimating the quality of the generated molecules. To ensure accuracy, the free energy of AT7 is constrained to enable the retrieval of the nearest docking pose and binding affinity, as evidenced in the 7znt crystal structure. The cohorts of 220 and 224 molecules generated through the deep neural network is evaluated. Adhering to established protocols and congruent with biological interactions, each molecule is allowed a maximum of 10 binding poses to align with best practices in emulating binding interactions within the binding pocket. We select an RMSD threshold to ensure molecules exhibiting superior values are selectively retained and attain potential drug-like entities.

The outcome of the screening process is visually presented in Figure 4A and B, where ligands from POI and E3 are individually extracted and designated as input fragments. These molecules are the final candidates to be further simulated by FEP calculations. The atoms potentially to be the initial linking positions are marked with an R* and are designated as anchors, imparting spatial information vital to the network generation process. The chemical composition of the generated molecules is presented alongside the reference compound, AT7, illustrated in Figure 2B, providing a comprehensive visualization of the molecular attributes.

RMSD and BFE Calculations. Table 1 presents an overview of the RMSD and ΔG_{BFE} values to demonstrate the structural and thermodynamic congruence between the reference and generated molecules. Within the context of C18 molecule generation, C18−02−0019 shows an RMSD of 1.20 Å, surpassing its group in terms of structural congruence. An evaluation of the MS417 molecules reflects that MS417−01−0157 notably shows a better structural alignment of 0.69 Å, outperforming both the MS417−01−0164 and MS417−01−0019 molecules.

In the evaluation of ΔG_{BFE}, C18−02−0019 shows a value of −43.71, which is significantly better than that of the reference molecule. A lower BFE indicates a more favorable binding affinity, an attribute that reinforces the better chemical properties of C18−02−0019. Meanwhile, within the MS417

https://doi.org/10.1021/acsomega.3c10183
ACS Omega 2024, 9, 38371−38384

Table 1. Docking Performance of the Crystal Structure of BRD4-AT7 and the Generated Molecules on BRD4-C18 and BRD4-MS417[a]

Linker molecules	RMSD (Å)	ΔG_{BFE} (kcal/mol)	$SASA_{buried}$ (%)
AT7 (reference)	–	−24.44	81.77
C18−02−0014	1.22	−37.60	84.75
C18−02−0019	1.20[b]	−43.71	84.74
C18−02−0055	1.26	−35.77	84.08
C18−02−0130	1.23	−36.54	84.97
MS417−01−0019	0.76	−44.98	80.75
MS417−01−0157	0.69[b]	−35.29	80.21
MS417−01−0164	0.74	−46.68	81.73

[a]We individually calculate the RMSD, BFE, and $SASA_{buried}$ values on each final candidate molecule. [b]The lowest RMSD value of the group.

cohort, MS417−01−0164 stands out in BFE evaluation with slightly better BFE than that of C18−02−0019.

In the $SASA_{buried}$ analysis, the C18 group shows a consistently higher buried percentage than AT7. This suggests a stronger binding affinity within the binding pocket compared with the reference molecule. However, the MS417 group shows a comparable or lower value than AT7.

Buried SASA Calculations. In order to establish the viability of the buried SASA ratio as a substantive screening metric for the assessment of the generated molecules, the interplay between the RMSD and the buried SASA ratio is employed. The study conducted by Mukherjee et al. underscores the augmentation of SASA upon protein−protein binding and consequently leads to an expansion in ligand buried extent.[52] We infer an increase in the binding affinity and structural congruence as a consequence of the binding process. We apply the buried SASA ratio within both the C18 and MS417 groups to collectively measure groups of the C18 and MS417 molecules. The relationship as depicted in Figure 5, aligning with the established findings demonstrates a negative slope between RMSD and the buried SASA ratio.

The relationship is further supported by the r-square value, which serves as an indicator of the fitting quality. In particular, the C18 group exhibits an enhanced r-square value to denote a more uniformly distributed set of values, thus bolstering the correlation. This alignment between the theoretical concepts and computational simulations converges to support the robust rationale underlying the adoption of the buried SASA ratio for molecular screening. We further provide a detailed calculation of the individual values in Table S4. These values include the buried SASA ratio, $SASA_{ligand\text{-}pocket}$, and $SASA_{ligand}$ of the reference AT7 structure and each final candidate molecule.

MD and FEP Simulations. We show the relative binding free energy results from each simulation in Table 3. The ternary complexes were subjected to 10 ns MD simulations and calculation of the BFE. The ternary complex formed by C18−02−0014 presents the lowest calculated values, suggesting that this conformation has the most stable complexes among the C18 group. In the MS417 group, MS417−01−0157 shows a better binding free energy in the group.

Figure 6 shows the POI-PROTACs-E3 ternary complexes of reference AT7 compound and the final candidates of C18 and MS417 groups before and after 10 ns MD simulation. The ternary complexes can freely rotate and move during the simulation to reach stable conformation within certain time frames. The snapshots were taken at the beginning and final time points of the simulations.

In Table 2, we demonstrate the substantial number of nonbonded interactions and ΔRg to validate the stability of the generated molecular structures. C18−02−0130 has the greatest number of nonbonded interactions and the lowest ΔRg value in the C-18 group. Both values are significantly better than those of the reference AT7 molecule. In the MS417 group, MS417−01−0157 has a greater number of nonbonded interactions, and MS417−01−0157 shows the lowest ΔRg. Note that the number of nonbonded interactions and ΔRg consistently show better molecular stability in both groups except MS417−01−0019.

We show the FEP method to calculate the relative binding affinity of the protein−ligand interaction. Table 3 shows that the binding affinity exhibited by the final candidates in comparison to the crystal structure is superior, except MS417−01−0019. The relative binding energy appears particularly pronounced within C18−02−0130 and MS417−01−0164 to reflect the establishment of robust POI-PROTACs-E3 ternary complexes of the generated molecules. Among the final candidates, C18−02−0130 possesses the best relative binding affinity with an average value of −1.91 kcal/mol. This result supports the previous docking measurement and further reflects the potency of adopting the state-of-the-art deep neural networks for the systematic generation of novel drug-like molecules.

Validation of Generalizability. In order to validate the robustness and generalizability of the study, we chose the ternary structure of the cellular inhibitor of apoptosis protein 1 (cIAP1) and Bruton's Tyrosine Kinase (BTK) (PDB: 6w7o)

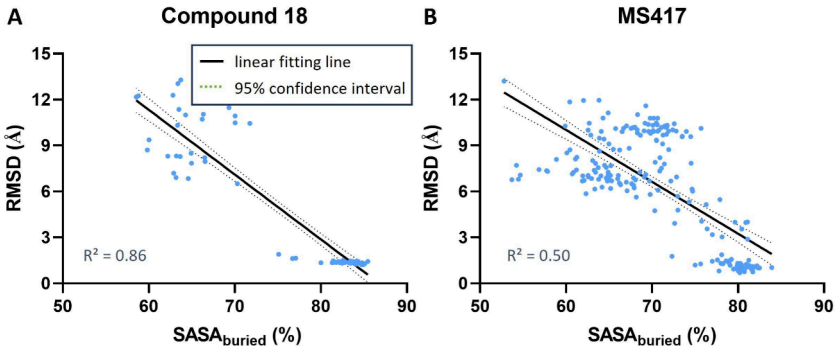


Figure 5. The relationship between RMSD and $SASA_{buried}$. The linear fitting curve shows a negative slope. This finding coincides with Mukherjee et al.'s finding that both SASA and protein−protein binding affinity increase to provide a high extent of structural similarity. The correlation slopes of compound 18 and MS417 are shown in A and B, respectively.

https://doi.org/10.1021/acsomega.3c10183
ACS Omega 2024, 9, 38371−38384

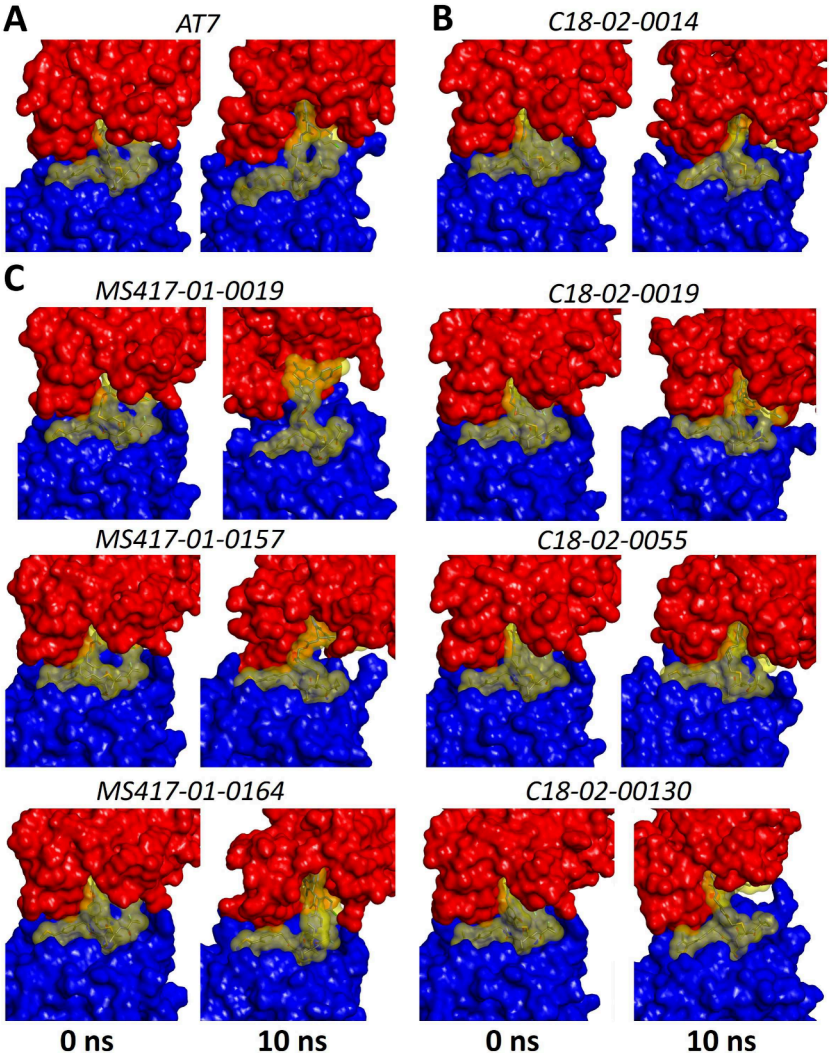


Figure 6. Spatial conformation movement of the reference AT7 compound and the final candidates of C18 and MS417 groups before and after 10 ns MD simulation. BRD4 and VHL are labeled in red and blue, respectively, and the generated PROTACs are labeled in yellow. The reference conformations of AT7 are shown in A. The generated novel PROTACs of C18 and MS417 conformations are shown in B and C, respectively.

Table 2. MD Simulation Trajectory Analysis for the Number of Interactions and ΔRg of AT7 and the Generated Molecules on BRD4-C18 and BRD4-MS417

Compound name	Number of nonbonded interactions	ΔRg (Å)
AT7 (reference)	28.68 ± 3.67	0.88 ± 0.26
C18−02−0014	35.18 ± 3.76	0.62 ± 0.39
C18−02−0019	32.07 ± 4.76	0.55 ± 0.27
C18−02−0055	33.03 ± 3.63	0.83 ± 0.41
C18−02−0130	36.73 ± 3.75	0.13 ± 0.29
MS417−01−0019	26.53 ± 4.59	1.06 ± 0.40
MS417−01−0157	38.95 ± 4.87	0.88 ± 0.56
MS417−01−0164	30.75 ± 3.72	0.46 ± 0.34

Table 3. Averaged MD and FEP Simulation Results between AT7 and the Generated Molecules[a]

Compound name	MD ($\Delta G_{binding}$ (kcal/mol))	FEP ($\Delta\Delta G_{binding}$ (kcal/mol))
C18−02−0014	−69.83 ± 7.09	−0.40 ± 3.96
C18−02−0019	−58.03 ± 6.91	−0.86 ± 2.31
C18−02−0055	−69.03 ± 7.39	−0.90 ± 1.07
C18−02−0130	−63.24 ± 7.18	−1.91 ± 1.51
MS417−01−0019	−47.67 ± 6.94	0.14 ± 0.61
MS417−01−0157	−63.14 ± 13.25	−1.41 ± 1.78
MS417−01−0164	−59.89 ± 8.78	−1.50 ± 2.56

[a]We evaluated the FEP experiment in triplicate independent runs and retrieved the average and error range.

and repeated the above processes for measuring their chemical properties generated from each step.[53−55] There were three molecules selected as the final candidates. Table S6 presents the RMSD and ΔG_{BFE} results, highlighting that all three novel molecules exhibit a significantly higher BFE than the reference molecule (TL7). Notably, the $SASA_{buried}$ values for all three molecules are marginally lower than that of TL7, indicating slight differences in the extent of burial ratio upon binding. The comparable MD simulation trajectory analysis results are detailed in Table S7, providing further insights into the stability and conformational behavior of the novel molecules. Additionally, the averaged MD and FEP simulation results are also summarized in Table S8 to demonstrate that the newly generated molecules possess comparable binding characteristics. When benchmarked against the BRD4-linker-CRBN ternary structure, all three novel molecules exhibit reasonable binding energy, suggesting that they maintain effective

https://doi.org/10.1021/acsomega.3c10183
ACS Omega 2024, 9, 38371−38384

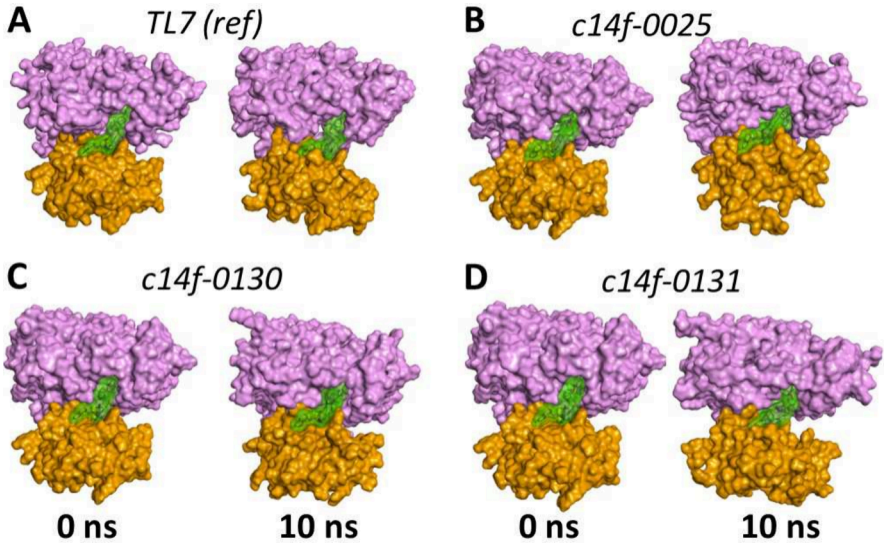


Figure 7. Trajectory movement of the validated ternary structure acquired from the 10 ns MD simulation. The pink and orange conformations are BTK and cIAP1, respectively. The green structures are the novel linker molecules generated by the deep neural network. A is the reference structure of TL7, and B−D are the structures of the generated novel PROTACs.

interactions within the binding pocket. This indicates that the new molecules could potentially serve as viable alternatives or enhancements to TL7 in the context of this biochemical framework. Figure 7 shows the ternary structure movement at the 10 ns MD simulation. The complexes were allowed to move freely during the simulation to achieve the most stable conformation and reach the minimum energy state. The snapshots were taken at the starting and ending time points. The validation results substantially support our proposed workflow and demonstrate the generalizability of the process.

DISCUSSION

This work shows that the integration of superimposition techniques and deep neural networks in PROTACs drug design offers multifaceted benefits that significantly advance the field of AI-driven drug design. The utilization of superimposition provides a crucial bridge between the available crystallographic structures and the broader design space. By enabling the replacement of the initial POI and its linker with new protein−ligand pairs that share the same binding pocket, superimposition bypasses the limitations posed by a sparsity of crystallographic data for the specific POI-PROTACs-E3 ternary complex. This is especially relevant in the realm of PROTACs, where the intricate nature of these molecules often poses challenges in elucidating their precise binding modes due to their dynamic and adaptable nature.

The deep neural network further introduces a layer of complexity into the drug design process. By leveraging advanced graph learning, the neural network excels in capturing intricate spatial and structural relationships inherent in ligands and proteins. The ability to generate novel PROTACs linker molecules based on learned patterns from the training data set reinforces the potential of adopting a state-of-the-art deep neural network as a powerful tool for molecular design. During the process, we demonstrate a careful selection of fragments on each side of the ligands, employing predefined anchors to impart spatial angle and distance information. The generated molecules are substantially screened by a postprocessing step to ensure their viability. The resultant molecules exhibit success rates comparable to those of other advanced machine learning methods, affirming the effectiveness of our approach.

The utility of chemical property screening using RMSD, BFE, and buried SASA simulations is particularly significant to retaining the potent novel molecules to become druggable. The RMSD values provide a quantitative assessment of the structural similarity between the generated molecules and the reference compounds. A lower RMSD indicates higher congruence between the two structures. The BFE calculation provides insight into whether the molecule has a stronger affinity to the protein and the capability to bridge the protein−protein interaction. The results show that the final candidates consistently exhibit stronger binding affinity. Additionally, the SASA parameter offers insights into the extent of surface burial upon protein−ligand interaction, serving as a relevant metric for molecular screening. Our study establishes a discernible relationship between the RMSD, BFE, and buried SASA ratio, providing a robust rationale for employing the potent screening criterion. Furthermore, MD and FEP simulations exemplify the true essence of molecular interactions, offering a comprehensive view of the binding affinity between POI-PROTACs-E3. The simulation of calculating the relative binding affinity interactions demonstrates the predictive power of our approach. The better relative binding affinity observed in certain molecules, including C18−02−0055, C18−02−0130, and MS417−01−0164 underlines their suitability for drug-like applications. These findings substantiate the potency of the integrated approach, wherein the neural network generates molecules that exhibit superior chemical properties and binding affinities in validating the viability of the generated PROTACs linker molecules.

The limitation of the study pertains to the quality and availability of crystallographic structures for the initial and new protein−ligand pairs. The assumption that the binding pockets of the new proteins and ligands are identical with those of the initial complex may not hold in all cases. In some cases, the binding pocket conformations are subject to inherent variability influenced by factors such as protein flexibility, the solvent environment, and even ligand-induced conformational changes. These dynamic aspects introduce a layer of

https://doi.org/10.1021/acsomega.3c10183
ACS Omega 2024, 9, 38371−38384

complexity that can potentially compromise the applicability of the superimposition process. Variations in the binding pocket conformation and side chain orientations could lead to inaccuracies in the superimposition process, potentially affecting the subsequent analysis and interpretation of screening results. The other limitation is the time scale of the simulations, which may not fully capture the complex dynamics of the system. Classical MD simulations of nanosecond scale, as utilized in this study, may not provide a sufficient potential energy surface to represent the system behavior and various structural-dynamic properties. Recent studies may incorporate a few millisecond time frames for sufficient time to allow for molecular dynamics. Incorporating accelerated sampling approaches such as replica exchange or metadynamics could enhance the coverage of conformational space and provide a more comprehensive understanding of the system's dynamics and interactions.

CONCLUSION

This study demonstrates a first-in-class integration of superimposition techniques and deep neural networks that holds significant promise in advancing PROTACs drug design. This innovative approach offers a systematic and efficient means of generating novel molecules by harnessing the capabilities of deep learning, while ensuring structural congruence through superimposition. The successful alignment of initial and new protein−ligand pairs enables comprehensive screening using metrics such as the RMSD, BFE, and buried SASA simulations.

In the future, the convergence of cutting-edge deep learning methodologies and structural biology tools holds the potential to reshape the landscape of drug discovery. Further refinement of the integrated approach could be beneficial in the design of PROTACs molecules and other therapeutics. By harnessing the strengths of deep learning and structural alignment simulations, this approach offers a promising avenue for accelerating drug development, optimizing lead compounds, and ultimately contributing to the advancement of precision medicine.

ASSOCIATED CONTENT

Data Availability Statement

All data mentioned in this study are publicly available in the ZINC data set, PROTAC-DB, and PDB. We retrieved the training and validation data from the above databanks. All the data we applied can be found in the Supporting Information. The code and remaining data details can be found at (https://github.com/AnHorn/AIMLinker).

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.3c10183.

Information on the deep neural network implementation, details on the data preparation, and validation results; Neural network hyperparameter settings in Table S1; individual SASA values to the reference compound in Table S2; final candidates selection process in Table S3; FEP simulation on the three runs in Table S4; averaged relative BFE in Table S5; docking performance on the validation complexes in Table S6; MD analysis on the validation complexes in Table S7; MD and FEP results on the validation complexes in Table S8; validation of docking alignment in Figure S1; substructures of postprocess filters in Figure S2; 2D chemical formula structures of TL7 and the novel generated molecules in Figure S3 (PDF)

Additional figure data and software information (ZIP)

AUTHOR INFORMATION

Corresponding Author

Chu-Chung Lin − *AnHorn Medicines Co., Ltd., Taipei 115202, Taiwan*; Email: cclin@anhornmed.com

Authors

Cheng-Li Chou − *AnHorn Medicines Co., Ltd., Taipei 115202, Taiwan*

Chieh-Te Lin − *Department of Biomedical Engineering, University of California Davis, Davis, California 95616, United States*; orcid.org/0000-0001-5262-1050

Chien-Ting Kao − *AnHorn Medicines Co., Ltd., Taipei 115202, Taiwan*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.3c10183

Author Contributions

¶C.-L.C. and C.-T.L. contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank the National Center for High-performance Computing (NCHC) for providing computational and storage resources. We would like to extend our gratitude to our funders who have made this research possible. We deeply appreciate your commitment to research and innovation, and we look forward to continuing our collaborative efforts to make a positive impact on society.

REFERENCES

(1) 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*, 8554−8559.

(2) Bai, N.; Miller, S. A.; Andrianov, G. V.; Yates, M.; Kirubakaran, P.; Karanicolas, J. Rationalizing PROTAC-mediated ternary complex formation using Rosetta. *J. Chem. Inf. Model.* **2021**, *61*, 1368−1382.

(3) Sun, X.; Gao, H.; Yang, Y.; He, M.; Wu, Y.; Song, Y.; Tong, Y.; Rao, Y. PROTACs: great opportunities for academia and industry. *Sig. Transduct. Target Ther.* **2019**, *4*, 64.

(4) Finley, D. Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annual review of biochemistry* **2009**, *78*, 477−513.

(5) Cecchini, C.; Tardy, S.; Ceserani, V.; Theurillat, J.-P.; Scapozza, L. Exploring the ubiquitin-proteasome system (UPS) through PROTAC technology. *Chimia* **2020**, *74*, 274−274.

(6) Westermaier, Y.; Barril, X.; Scapozza, L. Virtual screening: an in silico tool for interlacing the chemical universe with the proteome. *Methods* **2015**, *71*, 44−57.

(7) Petroski, M. D.; Deshaies, R. J. Function and regulation of cullin−RING ubiquitin ligases. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 9−20.

(8) Harper, J. W.; Schulman, B. A. Cullin-RING ubiquitin ligase regulatory circuits: a quarter century beyond the F-box hypothesis. *Annual review of biochemistry* **2021**, *90*, 403−429.

(9) Kramer, L. T.; Zhang, X. Expanding the landscape of E3 ligases for targeted protein degradation. *Current Research in Chemical Biology* **2022**, *2*, 100020.

(10) Hanzl, A.; Casement, R.; Imrichova, H.; Hughes, S. J.; Barone, E.; Testa, A.; Bauer, S.; Wright, J.; Brand, M.; Ciulli, A.; et al.

Functional E3 ligase hotspots and resistance mechanisms to small-molecule degraders. *Nat. Chem. Biol.* **2023**, *19*, 323−333.
(11) Hollingsworth, S.; Johnson, S.; Khakbaz, P.; Meng, Y.; Mouchlis, V.; Pierce, O.; Prytkova, V.; Vik, E.; Weiss, D.; Shanmugasundaram, V. The rise of targeting chimeras (TACs): next-generation medicines that preempt cellular events. *Med. Chem. Res.* **2023**, *32*, 1294.
(12) Donati, B.; Lorenzini, E.; Ciarrocchi, A. BRD4 and Cancer: going beyond transcriptional regulation. *Mol. Cancer* **2018**, *17*, 164.
(13) 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.
(14) 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*, 514−521.
(15) Vamathevan, J.; Clark, D.; Czodrowski, P.; Dunham, I.; Ferran, E.; Lee, G.; Li, B.; Madabhushi, A.; Shah, P.; Spitzer, M.; et al. Applications of machine learning in drug discovery and development. *Nat. Rev. Drug Discovery* **2019**, *18*, 463−477.
(16) Zheng, S.; Tan, Y.; Wang, Z.; Li, C.; Zhang, Z.; Sang, X.; Chen, H.; Yang, Y. Accelerated rational PROTAC design via deep learning and molecular simulations. *Nature Machine Intelligence* **2022**, *4*, 739−748.
(17) Chen, H.; Engkvist, O.; Wang, Y.; Olivecrona, M.; Blaschke, T. The rise of deep learning in drug discovery. *Drug discovery today* **2018**, *23*, 1241−1250.
(18) Zhou, J.; Cui, G.; Hu, S.; Zhang, Z.; Yang, C.; Liu, Z.; Wang, L.; Li, C.; Sun, M. Graph neural networks: A review of methods and applications. *AI open* **2020**, *1*, 57−81.
(19) Jiang, D.; Wu, Z.; Hsieh, C.-Y.; Chen, G.; Liao, B.; Wang, Z.; Shen, C.; Cao, D.; Wu, J.; Hou, T. Could graph neural networks learn better molecular representation for drug discovery? A comparison study of descriptor-based and graph-based models. *J. Cheminform.* **2021**, *13*, 1−23.
(20) Sun, M.; Zhao, S.; Gilvary, C.; Elemento, O.; Zhou, J.; Wang, F. Graph convolutional networks for computational drug development and discovery. *Briefings in bioinformatics* **2020**, *21*, 919−935.
(21) Flam-Shepherd, D.; Wu, T. C.; Friederich, P.; Aspuru-Guzik, A. Neural message passing on high order paths. *Machine Learning: Science and Technology* **2021**, *2*, No. 045009.
(22) Li, J.; Cai, D.; He, X. Learning graph-level representation for drug discovery. *arXiv*, September 12, 2017. DOI: 10.48550/arXiv.1709.03741.
(23) Withnall, M.; Lindelöf, E.; Engkvist, O.; Chen, H. Building attention and edge message passing neural networks for bioactivity and physical−chemical property prediction. *J. Cheminform.* **2020**, *12*, 1−18.
(24) Mercado, R.; Rastemo, T.; Lindelöf, E.; Klambauer, G.; Engkvist, O.; Chen, H.; Bjerrum, E. J. Graph networks for molecular design. *Mach. Learn.: Sci. Technol.* **2021**, *2*, No. 025023.
(25) Guo, J.; Knuth, F.; Margreitter, C.; Janet, J. P.; Papadopoulos, K.; Engkvist, O.; Patronov, A. Link-INVENT: generative linker design with reinforcement learning. *Digital Discovery* **2023**, *2*, 392.
(26) Simonovsky, M.; Komodakis, N. Graphvae: Towards generation of small graphs using variational autoencoders. In *Artificial Neural Networks and Machine Learning−ICANN 2018: 27th International Conference on Artificial Neural Networks*, Rhodes, Greece, October 4−7, 2018; Proceedings, Part I 27; pp 412−422.
(27) Liu, Q.; Allamanis, M.; Brockschmidt, M.; Gaunt, A. Constrained graph variational autoencoders for molecule design. *Adv. Neur. Inform. Process. Syst.* 2018, *31*. arXiv:1805.09076.
(28) Yang, Y.; Zheng, S.; Su, S.; Zhao, C.; Xu, J.; Chen, H. SyntaLinker: automatic fragment linking with deep conditional transformer neural networks. *Chemical science* **2020**, *11*, 8312−8322.
(29) Zheng, S.; Lei, Z.; Ai, H.; Chen, H.; Deng, D.; Yang, Y. Deep scaffold hopping with multimodal transformer neural networks. *J. Cheminform.* **2021**, *13*, 1−15.
(30) Imrie, F.; Bradley, A. R.; van der Schaar, M.; Deane, C. M. Deep generative models for 3D linker design. *J. Chem. Inf. Model.* **2020**, *60*, 1983−1995.
(31) Kao, C.-T.; Lin, C.-T.; Chou, C.-L.; Lin, C.-C. Fragment Linker Prediction Using the Deep Encoder-Decoder Network for PROTACs Drug Design. *J. Chem. Inf. Model.* **2023**, *63*, 2918.
(32) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The protein data bank. *Nucleic acids research* **2000**, *28*, 235−242.
(33) Biovia, D. S. *Discovery Studio Visualizer*, v21.1.0.20298; Dassault Systèmes: San Diego, 2021.
(34) Sheppard, G. S.; Wang, L.; Fidanze, S. D.; Hasvold, L. A.; Liu, D.; Pratt, J. K.; Park, C. H.; Longenecker, K.; Qiu, W.; Torrent, M.; et al. Discovery of N-Ethyl-4-[2-(4-fluoro-2, 6-dimethyl-phenoxy)-5-(1-hydroxy-1-methyl-ethyl) phenyl]-6-methyl-7-oxo-1 H-pyrrolo [2, 3-c] pyridine-2-carboxamide (ABBV-744), a BET Bromodomain Inhibitor with Selectivity for the Second Bromodomain. *J. Med. Chem.* **2020**, *63*, 5585−5623.
(35) Faivre, E. J.; McDaniel, K. F.; Albert, D. H.; Mantena, S. R.; Plotnik, J. P.; Wilcox, D.; Zhang, L.; Bui, M. H.; Sheppard, G. S.; Wang, L.; et al. Selective inhibition of the BD2 bromodomain of BET proteins in prostate cancer. *Nature* **2020**, *578*, 306−310.
(36) Filippakopoulos, P.; Qi, J.; Picaud, S.; Shen, Y.; Smith, W. B.; Fedorov, O.; Morse, E. M.; Keates, T.; Hickman, T. T.; Felletar, I.; et al. Selective inhibition of BET bromodomains. *Nature* **2010**, *468*, 1067−1073.
(37) Frost, J.; Galdeano, C.; Soares, P.; Gadd, M. S.; Grzes, K. M.; Ellis, L.; Epemolu, O.; Shimamura, S.; Bantscheff, M.; Grandi, P.; et al. Potent and selective chemical probe of hypoxic signalling downstream of HIF-α hydroxylation via VHL inhibition. *Nat. Commun.* **2016**, *7*, 13312.
(38) Irwin, J. J.; Shoichet, B. K. ZINC- a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **2005**, *45*, 177−182.
(39) Irwin, J. J.; Tang, K. G.; Young, J.; Dandarchuluun, C.; Wong, B. R.; Khurelbaatar, M.; Moroz, Y. S.; Mayfield, J.; Sayle, R. A. ZINC20—a free ultralarge-scale chemical database for ligand discovery. *J. Chem. Inf. Model.* **2020**, *60*, 6065−6073.
(40) Weng, G.; Shen, C.; Cao, D.; Gao, J.; Dong, X.; He, Q.; Yang, B.; Li, D.; Wu, J.; Hou, T. PROTAC-DB: an online database of PROTACs. *Nucleic acids research* **2021**, *49*, D1381−D1387.
(41) Yeung, S.; Kannan, A.; Dauphin, Y.; Fei-Fei, L. Tackling over-pruning in variational autoencoders. *arXiv*, June 9, 2017. DOI: 10.48550/arXiv.1706.03643.
(42) Fawcett, F. S. Bredt's rule of double bonds in atomic-bridged-ring structures. *Chem. Rev.* **1950**, *47*, 219−274.
(43) Hahn, M. Receptor surface models. 1. Definition and construction. *J. Med. Chem.* **1995**, *38*, 2080−2090.
(44) Wu, G.; Robertson, D. H.; Brooks, C. L., III; Vieth, M. Detailed analysis of grid-based molecular docking: A case study of CDOCKER—A CHARMm-based MD docking algorithm. *Journal of computational chemistry* **2003**, *24*, 1549−1562.
(45) Bell, E. W.; Zhang, Y. DockRMSD: an open-source tool for atom mapping and RMSD calculation of symmetric molecules through graph isomorphism. *J. Cheminform.* **2019**, *11*, 1−9.
(46) Lee, M. S.; Salsbury, F. R., Jr; Brooks, C. L., III Novel generalized Born methods. *J. Chem. Phys.* **2002**, *116*, 10606−10614.
(47) Singh, E.; Jha, R. K.; Khan, R. J.; Kumar, A.; Jain, M.; Muthukumaran, J.; Singh, A. K. A computational essential dynamics approach to investigate structural influences of ligand binding on Papain like protease from SARS-CoV-2. *Computational Biology and Chemistry* **2022**, *99*, 107721.
(48) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S. a.; Karplus, M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *Journal of computational chemistry* **1983**, *4*, 187−217.
(49) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.

https://doi.org/10.1021/acsomega.3c10183
ACS Omega 2024, 9, 38371−38384

(50) Elber, R.; Ruymgaart, A. P.; Hess, B. SHAKE parallelization. *European Physical Journal Special Topics* **2011**, *200*, 211−223.
(51) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N log (N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(52) Mukherjee, S.; Bahadur, R. P. An account of solvent accessibility in protein-RNA recognition. *Sci. Rep.* **2018**, *8*, 10546.
(53) Qi, Y.; Xia, P. Cellular inhibitor of apoptosis protein-1 (cIAP1) plays a critical role in β-cell survival under endoplasmic reticulum stress: promoting ubiquitination and degradation of C/EBP homologous protein (CHOP). *J. Biol. Chem.* **2012**, *287*, 32236−32245.
(54) Pal Singh, S.; Dammeijer, F.; Hendriks, R. W. Role of Bruton's tyrosine kinase in B cells and malignancies. *Mol. Cancer* **2018**, *17*, 57.
(55) Schiemer, J.; Horst, R.; Meng, Y.; Montgomery, J. I.; Xu, Y.; Feng, X.; Borzilleri, K.; Uccello, D. P.; Leverett, C.; Brown, S.; et al. Snapshots and ensembles of BTK and cIAP1 protein degrader ternary complexes. *Nat. Chem. Biol.* **2021**, *17*, 152−160.

https://doi.org/10.1021/acsomega.3c10183
ACS Omega 2024, 9, 38371−38384