PROTAC-Design-Evaluator (PRODE): An Advanced Method for In-Silico PROTAC Design

Ben Geoffrey A S, Deepak Agrawal,* Nagaraj M. Kulkarni, Rajappan Vetrivel, and Kishan Gurram
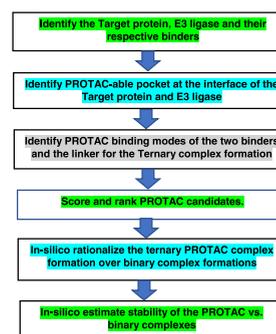
Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: PROTAC (proteolysis-targeting chimeras) is a rapidly evolving technology to target undruggable targets. The mechanism by which this happens is when a bifunctional molecule binds to a target protein and also brings an E3 ubiquitin ligase in proximity to trigger ubiquitination and degradation of the target protein. Yet, in-silico-driven approaches to design these heterobifunctional molecules that have the desired functional properties to induce proximity between the target protein and E3 ligase remain to be established. In this paper, we present a novel in-silico method for PROTAC design and to demonstrate the validity of our approach, we show that for a BRD4-VHL-PROTAC-mediated ternary complex known in the literature, we are able to reproduce the PROTAC binding mode, the structure of the ternary complex formed therein, and the free energy (ΔG) thermodynamics favoring ternary complexation through theoretical/computational methodologies. Further, we demonstrate the use of thermal titration molecule dynamics (TTMD) to differentiate the stability of PROTAC-mediated ternary complexes. We employ the proposed methodology to design a PROTAC for a new system of FGFR1-MDM2 to degrade the FGFR1 (fibroblast growth factor receptor 1) that is overexpressed in cancer. Our work presented here and named as PROTAC-Designer-Evaluator (PRODE) contributes to the growing literature of in-silico approaches to PROTAC design and evaluation by incorporating the latest in-silico methods and demonstrates advancement over previously published PROTAC in-silico literature.

The in-silico PROTAC design method (PRODE: PROTAC Design & Evaluator).

Identify the Target protein, E3 ligase and their respective binders

Identify PROTAC-able pocket at the interface of the Target protein and E3 ligase

Identify PROTAC binding modes of the two binders and the linker for the Ternary complex formation

Score and rank PROTAC candidates.

In-silico rationalize the ternary PROTAC complex formation over binary complex formations

In-silico estimate stability of the PROTAC vs. binary complexes

INTRODUCTION

PROTAC is a rapidly emerging technology for target protein degradation; yet, the approaches for in-silico-driven PROTAC design remain rather ad hoc and there is a need to establish methods to better rationalize PROTAC design in silico.[1−7] To rationalize PROTAC-mediated ternary complex formation, Drummond et al.[1] proposed the use of protein–protein docking and search of PROTAC-compatible protein–protein docked poses to rationalize PROTAC-mediated ternary complex formation. The development in the literature along the protein–protein docking approach to PROTAC ternary complex modeling recently culminated in the work of Mikhail et al.,[2] which concludes that the linker-compatible protein–protein docked pose corresponding to the protein of interest (POI) and the E3 ligase is the most suitable pose favoring the ternary complex formation. Further, co-operativity has been shown by Li et al.[3] as an important criterion for ternary complex formation. Co-operativity is defined as the ratio of binding constants corresponding to binary and ternary complex formation.[3] Specifically, for the PROTAC-mediated ternary complex to be formed, the ΔG_{bind} of binding corresponding to the ternary complex involving the protein of interest (POI)-PROTAC-E3 Ligase should be more negative than the ΔG_{bind} of binding corresponding to the binary complexes involving the protein of interest (POI)-PROTAC and PROTAC-E3 Ligase.[3] However, an in-silico

methodology to accurately predict the PROTAC binding mode and the PROTAC-mediated ternary complex structure has not been reported in the literature thus far. While Li et al. have employed ΔG_{bind} calculations retrospectively on already available PROTAC-mediated ternary complex structures available in RCSB to rationalize in silico the thermodynamics and kinetics of PROTAC-mediated ternary complex formation, the approach still cannot be applied in practice as the PROTAC-mediated ternary complex structure is not known in PROTAC design problems. We have developed here an in-silico methodology to predict the PROTAC binding mode and the ternary complex formed therein mediated by the PROTAC, overcoming this limitation. We demonstrate the ability of the proposed in-silico methodology to reproduce an experimentally known PROTAC binding mode and PROTAC-mediated ternary complex structure as in RCSB with PDB ID 8BDX for the BRD4-VHL system. This allows for ΔG_{bind} calculations developed in the PROTAC literature so far to be

A

used in practice, wherein the experimental PROTAC-mediated ternary complex structure is not known. Further, the stability of the PROTAC-mediated ternary complex is determined by the binding constant (K_d) associated with the ternary. As an additional in-silico component, we also employ the thermal titration molecular dynamics (TTMD) methodology developed by Pavan et al.[8] to differentiate the binding constants (K_d) associated with strong and weak binders to the PROTAC system to differentiate K_d values of different magnitudes associated with a more-stable and less-stable PROTAC-mediated ternary complex. Taken together, we believe our method presented here represents an important contribution to in-silico PROTAC design literature. Our PROTAC-Designer-Evaluator (PRODE) methodology, as presented in this paper, will help to reduce the time as well as costs of the PROTAC DMTA cycle and accelerate early-stage PROTAC drug discovery.

■ METHODOLOGY

Part A: In-silico methodology to reproduce the binding mode of PROTAC and the PROTAC-mediated ternary complex in the BRD4-VHL system corresponding to the PDB ID 8BDX. PROTAC 48 in PDB ID 8BDX tags Bromodomain-Containing Protein 4 (BRD4) with the E3 ligase von Hippel−Lindau (VHL) for degradation. In our previous work,[7] we proposed that the starting point required for in-silico PROTAC design is the knowledge of the binders for the target protein of interest (POI) that is required to be degraded and the binder for the E3 ligase found within the localization of the target protein. From PROTAC 48 found in PDB ID 8BDX, we extracted the binder for BRD4 and the binder for VHL, as well as the linker used to link these two binders. With these starting points of known binders to BRD4 and VHL, we proceed as follows to reproduce the experimentally known binding mode of PROTAC 48 and the structure of the PROTAC-mediated ternary complex.

Protein−Protein Docking to Obtain a Linker-Compatible Protein−Protein Pose of BRD4-VHL. The binders of BRD4 and VHL are first docked into their respective targets, and the binding mode of the binders associated with BRD4 and VHL is obtained. Next, the proteins BRD4 and VHL along with the binders in the respective binding mode identified are taken for protein−protein docking. The protein−protein docking was conducted using MEGADOCK 4.0. The docking procedure was executed as per the protocol given in the publication associated with MEGADOCK 4.0, which requires a definition of the receptor and ligand among the two proteins to be docked and the program generates docking poses (decoys) with an associated protein−protein interaction (PPI) score.[9] Among the top-scoring docking poses, the distance between the binders was analyzed and the top-scoring pose in which the binders to BRD4 and VHL are close enough in proximity to be connected by a linker was taken as the best protein−protein pose for PROTAC design.

PROTAC Design. The distance between the binders in the selected protein−protein docked pose is the key criterion for linker selection. Among the linkers available in PROTAC-DB[10] that match with the distance criteria for the PDB ID 8BDX, we used the linker for PROTAC 48 to connect the binders using Rdkit.

PROTAC Binding Mode Identification and Free Energy (ΔG) Calculations. *Binding Mode Identification.* For the PROTAC candidate obtained by connecting the

binders with the chosen linker, we generated multiple conformers of the PROTAC candidate using a torsional diffusion approach reported by Bowen et al.[11] The torsional diffusion approach is more efficient than traditional approaches to conformer generation as the torsional landscape of all possible conformers of the PROTAC is sampled more efficiently in the torsional diffusion approach as compared to traditional approaches.[11] The binding mode of the BRD4 and VHL binders in the selected protein−protein docked pose is chosen as the template, and an rdkit-based alignment of the generated conformers to the template is carried out. The objective is to obtain a conformer with the maximum alignment and minimum RMSD with the template, which retains all the interactions present in the binding mode of the binders of BRD4 and VHL and this is expected to be the binding mode associated with the PROTAC candidate.

Molecular Dynamics Simulation. Once the PROTAC binding mode is identified, the ΔG calculations associated with the binary and ternary complex as proposed by Li et al. are carried out to understand the co-operativity of ternary complexation. Before carrying out the MMPBSA-based binding free energy calculations, the stability of the PROTAC-mediated ternary complex was estimated using a 50 ns classical molecular dynamics simulation. In order to perform the molecular dynamics simulation, it is required to choose a force field. GROAMCS provides the option of using AMBER force fields for proteins, and the amber99sb-ildn force field was used for the protein system, while the ACPYPE AmberTools package was used to parameterize the built force field parameters for the ligand.[12,13] The PROTAC-mediated ternary complex was solvated in a cubic solvent box with the ternary complex system centered in the cubic box with at least 1.2 nm distance from the edge of the box and ions were added to physiological pH. The solvent model used was the TIP3P model. To mimic the physiological temperature and pressure, the system was heated and equilibrated to 310 K and 1 Bar in 100 ps NVT and NPT runs, which were executed with a time step of 1 fs while the temperature and pressure of the simulation were controlled using the Berendsen thermostat and barostat, respectively.[14,15] The production run was executed for 50 ns with a time step of 1 fs. The last 100 frames of the stable portion of the trajectory were taken for MMPBSA-based binding free energy calculations.

MMPBSA-Based Binding Free Energy Calculations. MMPBSA-based binding free energy calculations are based on the following equations[16]

$$\Delta G_{bind} = \Delta G_{complex} - \Delta G_{receptor} - \Delta G_{ligand}$$

where the ΔG for each case is estimated as

$$\Delta G = \Delta H - T\Delta S$$

The enthalpic contributions from bonded, nonbonded, and solvation energy components and the entropic contributions to ΔG were estimated using the tool gmx_mmpbsa.[16]

PROTAC-BRD4 and PROTAC-VHL form the two binary systems, while BRD4-PROTAC-VHL forms the two ternary systems. For MMPBSA calculations, the receptor of the ligand was defined as follows. For the PROTAC-BRD4 system, BRD4 was defined as the receptor and PROTAC as the ligand. For the PROTAC-VHL system, VHL was defined as the receptor and PROTAC as the ligand. For the BRD4-PROTAC-VHL system, BRD4-PROTAC was used as the receptor and VHL as the ligand. When the ΔG of Ternary(ΔG_{Ter}) is lower than the
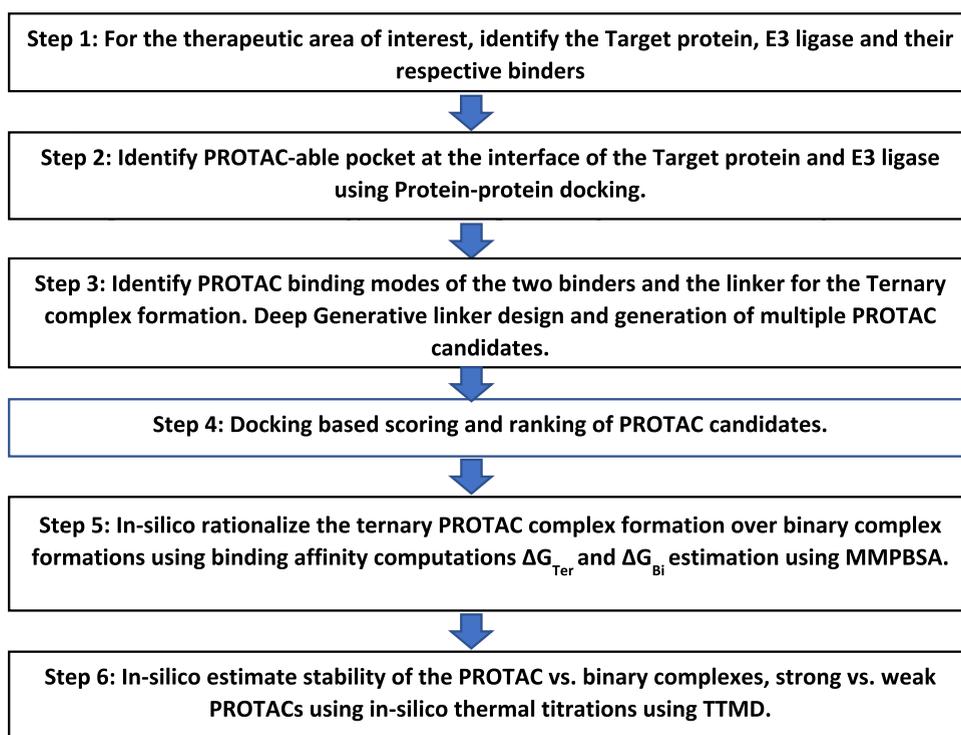
Step 1: For the therapeutic area of interest, identify the Target protein, E3 ligase and their respective binders

Step 2: Identify PROTAC-able pocket at the interface of the Target protein and E3 ligase using Protein-protein docking.

Step 3: Identify PROTAC binding modes of the two binders and the linker for the Ternary complex formation. Deep Generative linker design and generation of multiple PROTAC candidates.

Step 4: Docking based scoring and ranking of PROTAC candidates.

Step 5: In-silico rationalize the ternary PROTAC complex formation over binary complex formations using binding affinity computations ΔG_{Ter} and ΔG_{Bi} estimation using MMPBSA.

Step 6: In-silico estimate stability of the PROTAC vs. binary complexes, strong vs. weak PROTACs using in-silico thermal titrations using TTMD.

Figure 1. In-silico PROTAC design method (PRODE: PROTAC Design & Evaluator).

ΔG of binaries (ΔG_{Bi}), ternary complexation is favored according to Li et al.[3]

While PDB ID 8BDX was used as a validation of the proposed in-silico methodology, the proposed methodology was used to design a PROTAC for a novel system. Fibroblast growth factor receptor 1 (FGFR1) was chosen as the target protein of interest, and its known binder Erdafitinib was used in the PROTAC design. The appropriate E3 ligase MDM2 was chosen based on literature[17] and its binder Nutlin was used in the PROTAC design.

Part B: Thermal Titration Molecular Dynamics (TTMD) to Differentiate Ternary Complexation Ability for PROTAC. Further, PROTAC mediates the formation of the ternary complex with a different K_d (binding constant) value. The lower the value of K_d, the more stable the PROTAC-mediated ternary complex. Therefore, it is of value to differentiate, in silico, relatively low and high K_d values in the PROTAC DMTA cycle. We adopted a methodology developed by Pavan et al.[8] to differentiate strong and weak small molecule binders for the PROTAC system and differentiate the relatively high and low ternary K_d (binding constant) values associated with PDB IDs 8BDX and 8BDT[18] as a proof of concept of the adaptation of TTMD to the PROTAC-mediated ternary complex system. TTMD is based on the hypothesis that a weak binder has a less-stable binding mode at a higher temperature than a strong binder. Therefore, a short MD simulation was conducted; the interaction fingerprints were computed across the last 5 ns of the simulation; and Tanimoto similarity was computed between the interaction fingerprints across the first and the subsequent frames to understand the retainment of original interactions. It is expected that a more-stable PROTAC-mediated ternary complex having a lower K_d value will retain more of the original interaction. The MD simulation was carried out in

GROMACS[19] and the interaction fingerprint was carried out through PLIP.[20]

The in-silico PROTAC design method (PRODE referring to **PRO**TAC **D**esign & **E**valuator) is summarized as a flow diagram in Figure 1.

RESULTS AND DISCUSSION

Part A. The first part of the methodology involves reproducing in silico the binding mode of PROTAC 48 as

Table 1. Top-Scoring Poses and Linker Compatibility

pose ID	protein−protein interaction (PPI) score from MEGADOCK	linker compatibility
BRD4-VHL_Pose_1	4441	no
BRD4-VHL_Pose_2	4302	no
BRD4-VHL_Pose_3	4408	no
BRD4-VHL_Pose_4	4211	no
BRD4-VHL_Pose_5	4107	no
BRD4-VHL_Pose_6	3988	yes
BRD4-VHL_Pose_7	3971	no
BRD4-VHL_Pose_8	3853	no
BRD4-VHL_Pose_9	3757	no
BRD4-VHL_Pose_10	3691	no

found in PDB ID 8BDX and implementing protein−protein docking on BRD4-VHL to identify a linker-compatible protein−protein docked pose. In Table 1, top poses based on the highest protein−protein interaction scores and linker compatibility are shown.

The linker-compatible pose corresponding to Pose ID "BRD4-VHL_Pose_6" is shown in Figure 2. The binding pocket of VHL and BRD4 is orientated along the protein− protein interface, which enables the respective binders at the
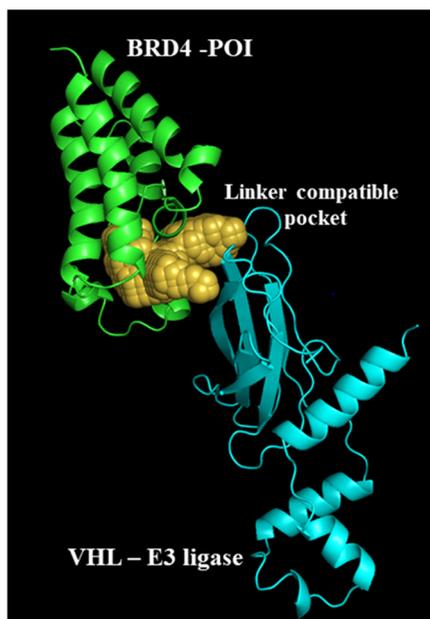
Figure 2. Linker-compatible pose corresponding to Pose ID — "BRD4-VHL_Pose_6".
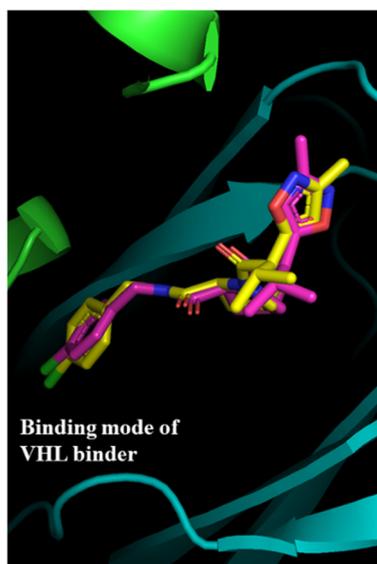

Figure 3. Superimposed docking pose with the crystallographic pose for the VHL binder.

pockets to be connected via a linker making them a linker-compatible pose for PROTAC design.

The next step in the methodology to reproduce the binding mode of PROTAC 48 in the BRD4-VHL system corresponding to the PDB ID 8BDX is to reproduce the binding mode of the individual binders in the respective pockets of BRD4 and VHL. The superimposition of the pose obtained from docking on the crystallographic pose is shown in Figure 3 for the VHL binder, where the crystallographic pose of the binder is depicted in yellow and the docked pose is shown in pink color.

Similarly, the binding mode of the BRD4 binder was also reproduced, and the superimposition of the crystallographic and docked pose obtained is shown in Figure 4.

The next step in the proposed methodology to reproduce the binding mode of the PROTAC is to find a linker of suitable
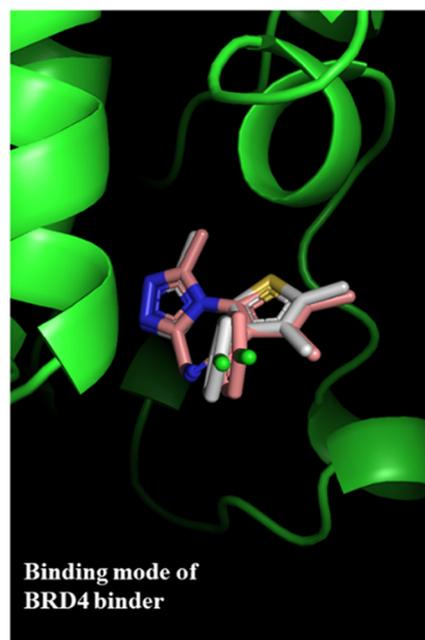

Figure 4. Superimposed docking pose with the crystallographic pose for the BRD4 binder.

Table 2. PROTAC Conformers with Maximum Alignment on the Basis of RMSD and Retention of Interactions

ID	RMSD	SF-CNN Protein-PROTAC interaction score
PROTAC_CRYSTAL_Pose	0.000	9.099
PROTAC_pose_3031	2.084	8.519
PROTAC_pose_3915	2.484	7.490
PROTAC_pose_1002	2.680	6.747
PROTAC_pose_3110	2.852	6.628
PROTAC_pose_2875	2.882	7.348
PROTAC_pose_725	2.994	7.102
PROTAC_pose_3184	3.464	6.625
PROTAC_pose_963	3.466	6.618
PROTAC_pose_2079	3.487	7.709
PROTAC_pose_327	3.542	7.686
PROTAC_pose_1278	3.563	6.778
PROTAC_pose_281	3.764	5.854

length to link the binders of BRD4 and VHL in their respective binding modes such that the PROTAC molecule retains all the key interactions present in the individual binders of BRD4 and VHL, respectively. For the BRD4-VHL system, the linker from PDB ID 8BDX was adopted and rdkit was used to connect the binders with the linker and make the PROTAC molecule. 4000 conformers of this PROTAC candidate were generated using the torsional diffusion methodology as per Bowen et al.[11] The binding mode of the BRD4 and VHL binders individually is taken as template, and rdkit was used to obtain the transformation matrix required to align the generated PRTOAC conformers to the template. The transformation matrix encodes the translation and rotations required to align PROTAC conformers to the binding mode of the individual binders of BRD4 and VHL, respectively, to identify the PROTAC conformer, which has maximum alignment to the binding mode of BRD4 and VHL binders such that the PROTAC molecule retains all of the interactions present in the individual binders of BRD4 and VHL. The results for the
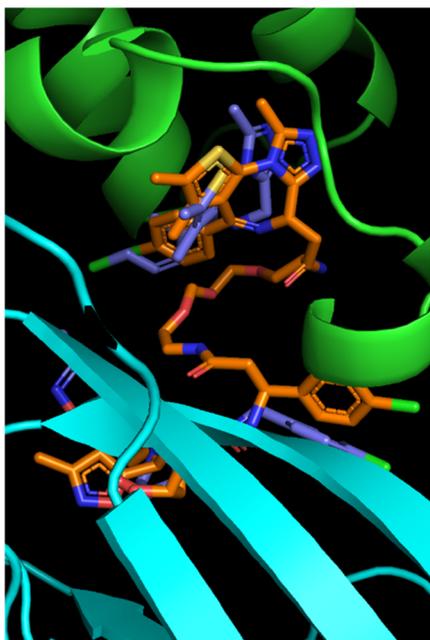
Figure 5. Alignment of PROTAC_Pose_3031 with the binding mode of the individual binders of BRD4 and VHL.

PROTAC conformers with maximum alignment quantified by the RMSD and retainment of interactions quantified by the SF-CNN score are provided in Table 2.

The SF-CNN Protein-PROTAC interaction score comes from a deep learning model[21] trained on the PDBbind data set and is meant to provide a quantifiable score to rank the interaction between protein residues and small organic ligands. We adopted this model to generate a quantifiable interaction score between the PROTAC molecule and its pocket residues.

The first row contains the crystallographic pose of PROTAC in PDB 8BDX, which obtains a maximum score of 9.099. Among the PROTAC conformers generated through our approach and aligned with the template of the pose of the individual binders of BRD4 and VHL, the 3031st conformer with the PROTAC pose ID, PROTAC_pose_3031, has a maximum alignment with the crystallographic pose of RMSD 2.084 and has a maximum interaction retained as quantified by the SF-CNN Protein-PROTAC interaction score.

The alignment of PROTAC_Pose_3031 with the binding mode of the individual binders of BRD4 and VHL is shown in Figure 5.

Similarly, the alignment of the binding mode of PROTAC_pose_3031 with the crystallographic pose of the PROTAC as found in PDB entry 8BDX is shown in Figure 6.

Following the identification of the binding mode of PROTAC, the stability of the PROTAC-mediated ternary complex formed therein was evaluated using classical molecular dynamics simulation. The backbone root mean square deviation (RMSD) stabilization and the stabilization of the radius of gyration (RoG) graphs shown in Figures 7 and 8 indicate the stability and compactness of the PROTAC-mediated ternary complex, respectively.

Frames from 49 to 50 ns were used to perform the MMPBSA-based ΔG calculations as proposed by Li et al. to rationalize in silico the thermodynamics and kinetics of ternary complex formation. The ΔG calculations are shown in Table 3.

From the data in Table 3, it can be inferred that since $\Delta G_{TER} < \Delta G_{BI}$, ternary complexation is thermodynamically favored.

PROTAC Design for FGFR1. Following the approach detailed above, we next designed a PROTAC for a novel system. Fibroblast growth factor receptor 1 (FGFR1) is overexpressed in cancer cells and is targeted for inhibition in colorectal cancer. We use the known binder of FGFR1, which
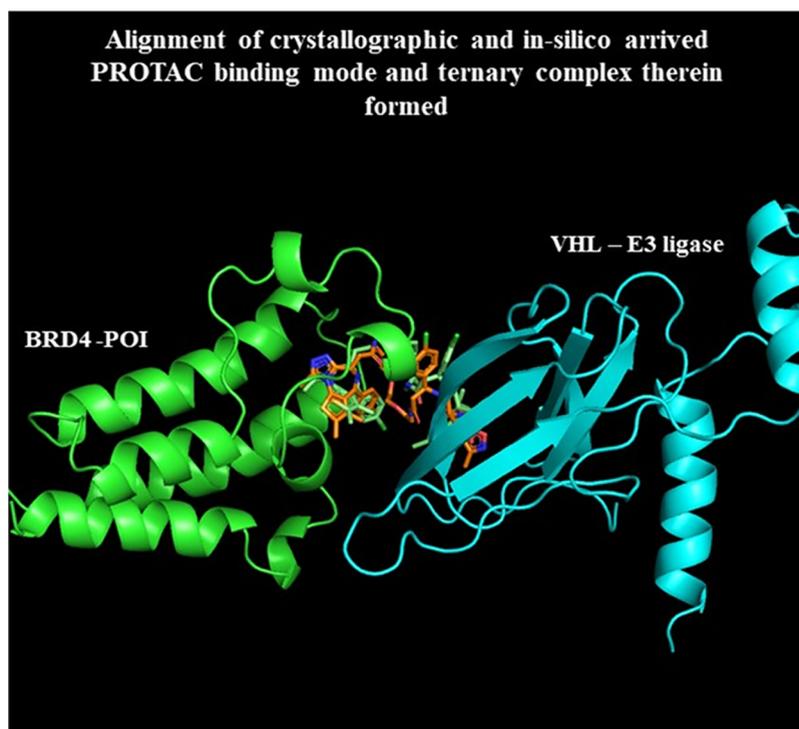

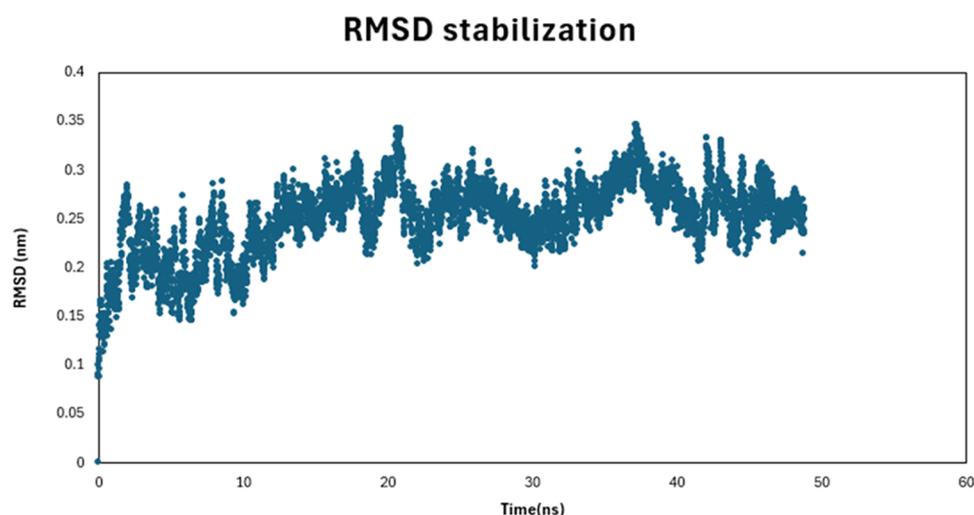**Figure 6.** Alignment of PROTAC_Pose_3031 with the crystallographic pose of PROTAC as found in PDB ID 8BDX.

RMSD stabilization

Figure 7. RMSD stabilization indicates the stability of the ternary complex.

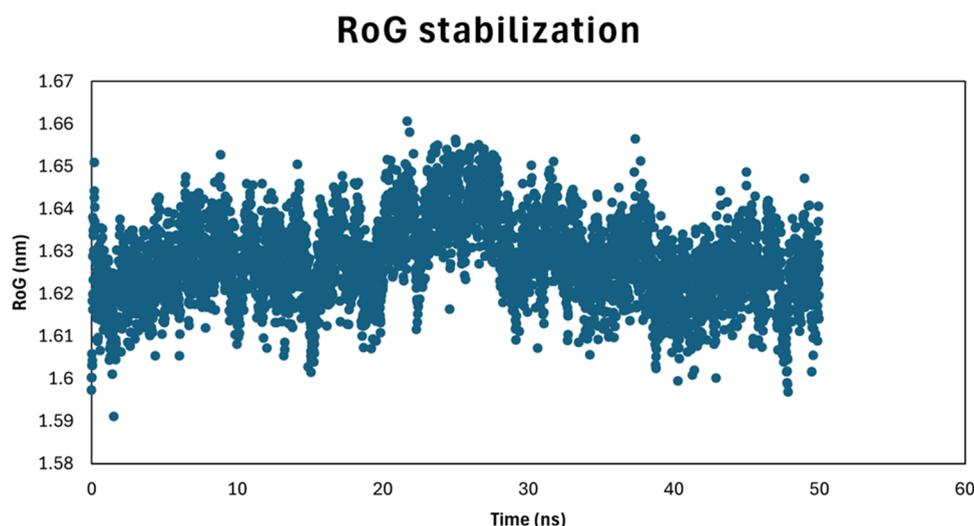## RoG stabilization

Figure 8. RoG stabilization indicates compactness of the ternary complex.

Table 3. ΔG Calculations to Rationalize In-Silico Ternary Complex Favorability

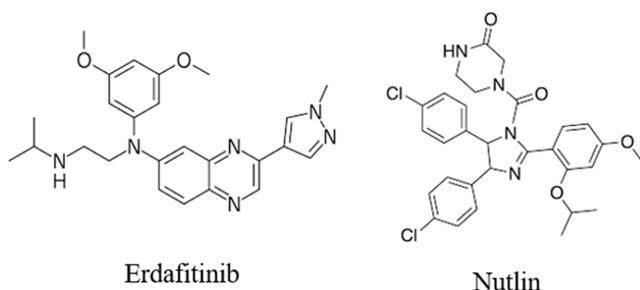| complex	ΔH	$-T\Delta S$	$\Delta G = \Delta H - T\Delta S\Delta G_{BI}$ of binary kcal/mol	$\Delta G = \Delta H - T\Delta S\Delta G_{TER}$ of ternary kcal/mol
BRD4-PROTAC	−48.17	8.78	−39.39	
BRD4-PROTAC-VHL	−101.82	13.53		−88.29
VHL-PROTAC	−36.92	7.58	−29.34	

Figure 9. Structures of erdafitinib and nutlin.

is a marketed drug by the name Erdafitinib in the PROTAC design. The E3 ligase Mouse Double Minute 2 (MDM2) was

chosen based on literature[17] and its binder nutlin was used in the PROTAC design.

The structures of erdafitinib and nutlin are shown in Figure 9.

The approach detailed above was followed to design the PROTAC. The linker-compatible protein−protein docked pose between FGFR1 and MDM2 was obtained. The individual binding mode of the binders of FGFR1 and MDM2 was obtained. 13 Angstrom was the distance between the binders in their respective binding modes in the linker-compatible FGFR1-MDM2 docked pose. Based on the distance needed to link the two binders in their respective binding modes, 4 PEG linkers of lengths varying around 13 Angstrom were chosen with linker IDs 8, 7, 184, and 373 available in PROTAC-DB with the following hyperlinks:
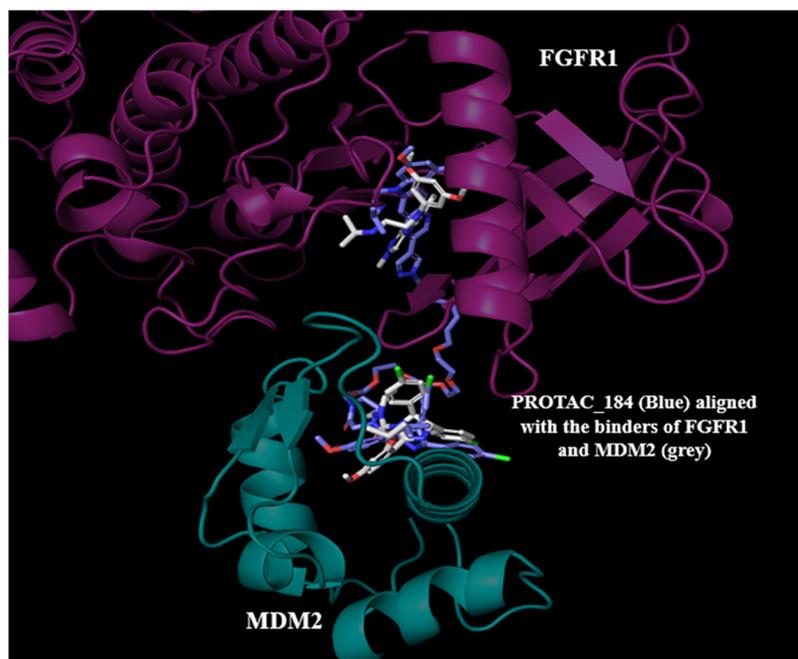
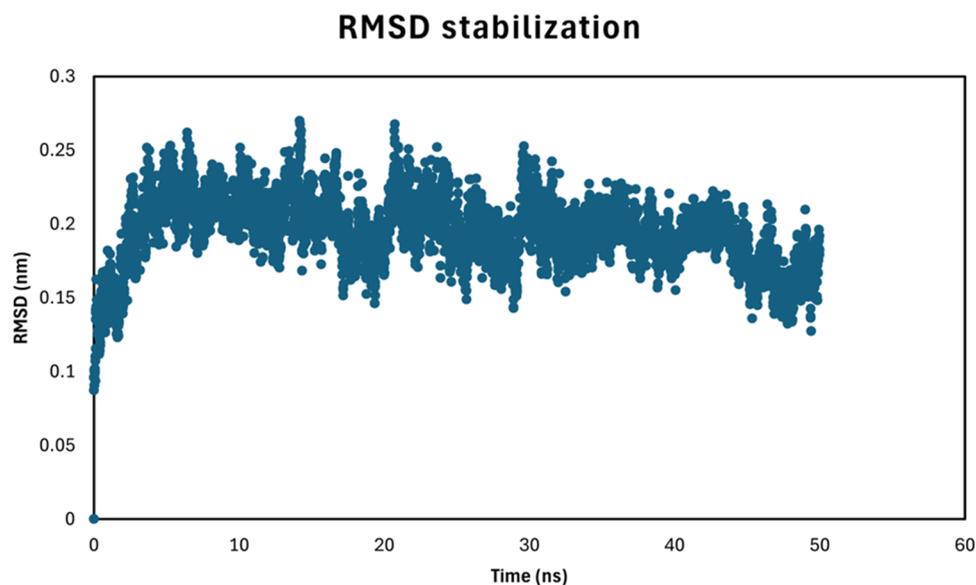**Figure 10.** Superimposed pose of PROTAC_184 on the binders of FGFR1 and MDM2.

Figure 11. RMSD stabilization indicates the stability of the ternary complex.

Table 4. ΔG Calculations to Rationalize the In-Silico Ternary Complex Favorability for FGFR1

complex	ΔH	$-T\Delta S$	$\Delta G = \Delta H - T\Delta S\Delta G_{BI}$ of binary kcal/mol	$\Delta G = \Delta H - T\Delta S\Delta G_{TER}$ of ternary kcal/mol
FGFR1-PROTAC	−51.64	8.64	−43.3	
FGFR1-PROTAC-MDM2	−98.4	15.11		−83.29
MDM2-PROTAC	−39.14	7.8	−31.34	

Linker_8: http://cadd.zju.edu.cn/protacdb/compound/dataset=linker&id=8

Linker_7: http://cadd.zju.edu.cn/protacdb/compound/dataset=linker&id=7

Linker_184: http://cadd.zju.edu.cn/protacdb/compound/dataset=linker&id=184

Linker_373: http://cadd.zju.edu.cn/protacdb/compound/dataset=linker&id=373

PROTAC candidates were generated out of the different linkers. 4,000 conformers were generated for each PROTAC candidate and were aligned to the template of binders of FGFR1 and MDM2 in their individual binding modes to check for the PROTAC candidate with maximum alignment indicated by a minimum RMSD and maximum retained interactions present in the binding mode of FGFR1 and MDM2 binders. It was found that the PROTAC candidate

Table 5. TTMD Results Comparing In Silico the Ternary Complexation Stability of PDB ID: 8BDX vs PDB ID: 8BDT

TTMD	temperature			
	300 K		450 K	
	K_D (Ternary) = 0.006 μM (strong binder) PDB ID: 8BDT	K_D (Ternary) = 0.043 μM (weak binder) PDB ID: 8BDX	K_D (Ternary) = 0.006 μM (strong binder) PDB ID: 8BDT	K_D (Ternary) = 0.043 μM (weak binder) PDB ID: 8BDX	remarks
average interactions retained	0.843	0.866	0.768	0.644	strong binder retains significantly more interactions at high temperature than the weak binder.

Table 6. TTMD Results Comparing In Silico the Ternary Complexation Stability of PDB ID: 7JTP vs PDB ID: 7JTO

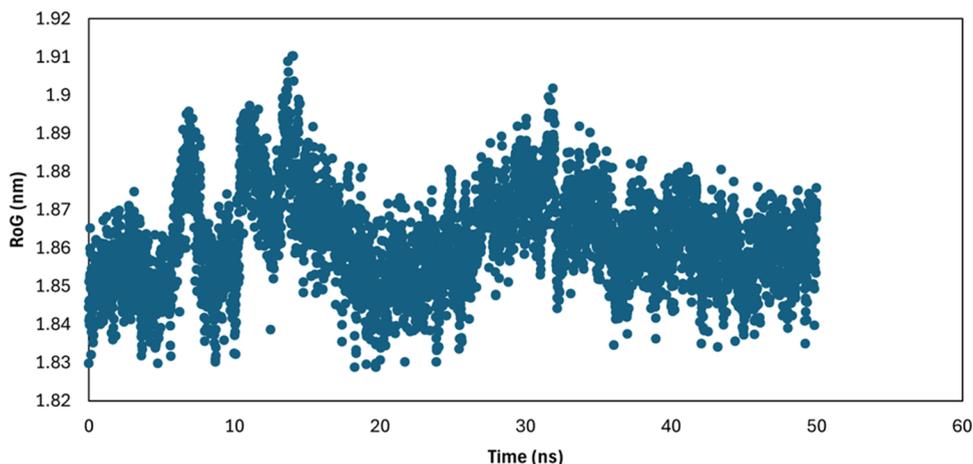| TTMD	temperature			
	300 K		450 K	
	K_D (Ternary) = 0.005 μM (strong binder) PDB ID: 7JTP	K_D (Ternary) = 0.520 μM (weak binder) PDB ID: 7JTO	K_D (Ternary) = 0.005 μM (strong binder) PDB ID: 7JTP	K_D (Ternary) = 0.520 μM (weak binder) PDB ID: 7JTO	remarks
average interactions retained	0.919	0.783	0.792	0.706	interactions retained are higher for the strong binder than the weak binder.

RoG stabilization

Figure 12. RoG stabilization indicates the compactness of the ternary complex.

Table 7. TTMD Results for FGFR1 Binder in the PROTAC_184-Mediated Ternary Complex

	temperature ramp						remarks
	300 K		400 K		450 K		
TTMD	PROTAC_184	FGFR1 binder	PROTAC_184	FGFR1 binder	PROTAC_184	FGFR1 binder	PRTOAC_184 retains most interactions of the FGFR1 binder at higher temperature.
average interactions retained	0.814	0.831	0.783	0.791	0.711	0.777	

Table 8. TTMD Results for MDM2 Binder in the PROTAC_184-Mediated Ternary Complex

	temperature ramp						remarks
	300 K		400 K		450 K		
	PROTAC_184	MDM2 binder	PROTAC_184	MDM2 binder	PROTAC_184	MDM2 binder	PRTOAC_184 retains most of the relevant interactions of the MDM2 binder at higher temperature.
average interactions retained	0.813	0.821	0.734	0.715	0.597	0.623	

made with Linker_184 had a minimum RMSD and maximum retainment of interactions. The corresponding superimposed pose of the PROTAC_184 on the binders of FGFR1 and MDM2 is shown in Figure 10.

After the identification of the PROTAC binding mode, the stability of the ternary complex formed therein was estimated by using a classical molecular dynamics simulation. The backbone root mean square deviation (RMSD) stabilization and the stabilization of the radius of gyration (RoG) graphs shown in Figures 11 and 12 indicate the stability and compactness of the PROTAC-mediated ternary complex, respectively.

Next, ΔG calculations as proposed by Li et al. were carried out to rationalize the in-silico ternary complex thermodynamic favorability, and the results are tabulated in Table 4.

From the above calculations, ternary co-operativity and ternary complexation thermodynamic favorability are concluded positively.

Part B. In this section, we adopt the TTMD methodology developed by Pavan et al. to differentiate the ternary complex stability mediated by two PROTACs with a relatively higher and lower K_d value for the PROTAC-mediated ternary complex. We execute a 5 ns classical MD simulation of the PROTAC-mediated ternary complex of PDB 8BDX (PRO-

TAC_48) and PDB 8BDT (**PROTAC_51**), which have two different K_d values across a temperature ramp of low (300 K) and high (450 K). A 50 kJ/mol restraint on the backbone of the protein was applied to ensure the backbone integrity at 450 K while the ligand was freely excited at 450 K. The interaction fingerprints were computed across the generated trajectories and the Tanimoto similarity of interaction fingerprints of the first and the subsequent frames of 5 ns of the simulation was computed to determine the interactions retained. It was expected that a more-stable complex will have more interactions retained at a higher temperature than a less-stable complex. The results obtained for PDB entries 8BDX and 8BDT are tabulated in Table 5.

The average number of interactions retained is defined as the average of Tanimoto similarity scores calculated between the first and the subsequent frames of the 5 ns simulation. As is evident from Table 5, PDB 8BDT, which has a lower K_d value, has more interactions retained at a higher temperature.

To further establish confidence in the use of the TTMD method for the PROTAC system, we carried out the approach for another system taken from the literature and found that the trend was repeated for that system. As shown in Table 6, the strong binder with a higher ternary K_d value has a higher

number of interactions retained at a higher temperature compared to the weak binder.

Having validated the adaptation for use of the TTMD methodology for the PROTAC system, we used the TTMD methodology to generate TTMD profiles to understand whether the PROTAC_184 designed for the FGFR1-MDM2 system retained the interactions found in the individual binders of FGFR1 and MDM2 across an increasing temperature ramp. Table 7 shows the TTMD results for the FGFR1 binder in PROTAC_184. As we can see, the decline in interactions is comparatively small, as would be expected of a strong binder.

Similarly, Table 8 shows the TTMD results for the MDM2 binder in PROTAC_184.

From the generated trajectories, it is observed that PROTAC_184 retained the key hydrogen bond interactions (ASP641, ALA564) of the FGFR1 binder at higher temperatures. Similarly, it was found that PRTOAC_184 retained the hydrophobic interactions (29LEU, 32LEU, 36ILE, 66PHE, and 68VAL) of the MDM2 binder at higher temperatures. This is further quantitatively captured in the TTMD profile generated and is indicative of the PROTAC mediating a stable ternary complex formation.

CONCLUSIONS, LIMITATIONS, AND FUTURE DIRECTION

The binding mode of the PROTAC and the ternary complex formed therein as in PDB ID 8BDX was reproduced using the in-silico methodology proposed in our work. The proposed methodology of the PROTAC-Designer-Evaluator (PRODE) is as follows. A linker-compatible protein–protein docked pose is selected from among the low-energy poses obtained from carrying out protein–protein docking between the protein of interest (POI) and the E3 ligase. The binding mode of the individual binders of the protein of interest and E3 ligase is obtained, and based on the distance to link them, linkers are chosen for PROTAC design. A torsional diffusion approach is used to generate the multiple conformers of the PROTAC candidates that are aligned to the individual binders of the protein of interest and E3 ligase. A low-energy PROTAC conformer with the minimum RMSD and retaining the most interactions as in the individual binders is chosen as the PROTAC binding mode, and the resulting ternary complex is used for carrying out further advanced MD-based free energy calculations to determine the thermodynamic favorability of ternary complexation. Further, we demonstrate the use of thermal titration molecular dynamics to differentiate the PROTAC's ability to mediate a stable ternary complex. The proposed methodology was used to design a PROTAC candidate for FGFR1 that is overexpressed in cancer and is a target for colorectal cancer. While the PROTAC-Designer-Evaluator (PRODE) approach focuses on PROTAC's ability to mediate ternary complex formation, ADMET evaluations for the designed PROTAC candidate are still outside the scope of the PROTAC-Designer-Evaluator (PRODE). Rules such as Lipinski's rule for small molecules are yet to be formulated for the design of orally bioavailable PROTAC and since there are rules that are derived out of large amounts of experimental data, we have to incorporate such ADMET evaluations in PROTAC-Designer-Evaluator (PRODE) in the future even as the data and understanding around PROTACs grow. However, the present work addresses certain research gaps and contributes to the growing literature on in-silico PROTAC design.

ASSOCIATED CONTENT

Data Availability Statement

The data and software packages used are as follows:[1] For the BRD4-VHL system used for validation, we obtained the structures from RCSB protein data bank with PDB IDs 8BDX (https://www.rcsb.org/structure/8BDX) and 8BDT (https://www.rcsb.org/structure/8BDX).[2] For the FGFR1-MDM2 system used to design the PROTAC, we obtained the structures from PDB IDs 4RWL (https://www.rcsb.org/structure/4RWL) (for FGFR1) and 1RV1 (https://www.rcsb.org/structure/1rv1) (for MDM2).[3] For the known binders of FGFR1 and MDM2, we used DrugBank (https://go.drugbank.com/drugs/DB12147) and PROTAC-DB 2.0 (http://cadd.zju.edu.cn/protacdb/downloads), respectively.[4] For PROTAC pocket identification and PROTAC binding mode prediction, we used the following software: MEGA-DOCK for protein−protein docking (https://github.com/akiyamalab/MEGADOCK); CAVIAR for Pocket identification (https://github.com/jr-marchand/caviar); and rdkit for chem-informatics tasks (https://github.com/rdkit/rdkit).[5] For molecular dynamics simulations involving ternary complex modeling, TTMD, and free energy calculations, we used GROMACS 2023 software (https://github.com/gromacs/gromacs).[6] For MMPBSA-based free energy calculations, we used GMX_MMPBSA software (https://github.com/Valdes-Tresanco-MS/gmx_MMPBSA).[7] For interaction fingerprinting in TTMD, we used PLIP (https://github.com/pharmai/plip).

AUTHOR INFORMATION

Corresponding Author

Deepak Agrawal − *Sravathi AI Technology Pvt. Ltd., Bengaluru, Karnataka 560099, India;* orcid.org/0009-0005-1722-0864; Email: deepak.a@sravathi.ai

Authors

Ben Geoffrey A S − *Sravathi AI Technology Pvt. Ltd., Bengaluru, Karnataka 560099, India*

Nagaraj M. Kulkarni − *Sravathi AI Technology Pvt. Ltd., Bengaluru, Karnataka 560099, India*

Rajappan Vetrivel − *Sravathi AI Technology Pvt. Ltd., Bengaluru, Karnataka 560099, India*

Kishan Gurram − *Sravathi AI Technology Pvt. Ltd., Bengaluru, Karnataka 560099, India*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acsomega.3c07318

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

REFERENCES

(1) Drummond, M. L.; Williams, C. I. In silico modeling of PROTAC-mediated ternary complexes: validation and application. *J. Chem. Inf. Model.* 2019, 59 (4), 1634−1644.

(2) Ignatov, M.; Jindal, A.; Kotelnikov, S.; Beglov, D.; Posternak, G.; Tang, X.; Maisonneuve, P.; et al. High Accuracy Prediction of

PROTAC Complex Structures. *J. Am. Chem. Soc.* **2023**, *145*, 7123−7135, DOI: 10.1021/jacs.2c09387.

(3) Li, W.; Zhang, J.; Guo, L.; et al. 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*, 523−532, DOI: 10.1021/acs.jcim.1c01150.

(4) Weng, G.; Li, D.; Kang, Y.; Hou, T. Integrative modeling of PROTAC-mediated ternary complexes. *J. Med. Chem.* **2021**, *64* (21), 16271−16281.

(5) 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* (no. 3), 1368−1382.

(6) Liao, J.; Nie, X.; Unarta, I. C.; Ericksen, S. S.; Tang, W. In silico modeling and scoring of PROTAC-mediated ternary complex poses. *J. Med. Chem.* **2022**, *65* (no. 8), 6116−6132.

(7) Ben Geoffrey, A. S.; Kulkarni, N. M.; Agrawal, D.; Vetrivel, R.; Gurram, K. A New In-Silico Approach for PROTAC Design and Quantitative Rationalization of PROTAC mediated Ternary Complex Formation. *bioRxiv* **2022**, No. 2022.07.11.499663, DOI: 10.1101/2022.07.11.499663.

(8) Pavan, M.; Menin, S.; Bassani, D.; Sturlese, M.; Moro, S. Qualitative Estimation of Protein−Ligand Complex Stability through Thermal Titration Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2022**, *62* (no. 22), 5715−5728.

(9) Ohue, M.; Shimoda, T.; Suzuki, S.; Matsuzaki, Y.; Ishida, T.; Akiyama, Y. MEGADOCK 4.0: an ultra−high-performance protein−protein docking software for heterogeneous supercomputers. *Bioinformatics* **2014**, *30* (no. 22), 3281−3283.

(10) Weng, G.; Cai, X.; Cao, D.; Du, H.; Shen, C.; Deng, Y.; He, Q.; Yang, B.; Li, D.; Hou, T. PROTAC-DB 2.0: an updated database of PROTACs. *Nucleic Acids Res.* **2023**, *51* (no. D1), D1367−D1372.

(11) Bowen, J.; Corso, G.; Chang, J.; Barzilay, R.; Jaakkola, T. Torsional diffusion for molecular conformer generation. 2022. arXiv:2206.01729. arXiv.org e-Print archive. https://arxiv.org/abs/2206.01729.

(12) Case, D. A.; Aktulga, H. M.; Belfon, K.; Cerutti, D. S.; Cisneros, G. A.; Cruzeiro, V. W. D.; Forouzesh, N.; et al. AmberTools. *J. Chem. Inf. Model.* **2023**, *63* (no. 20), 6183−6191.

(13) Sousa da Silva, A. W.; Vranken, W. F. ACPYPE-Antechamber python parser interface. *BMC Res. Notes* **2012**, *5*, 1−8.

(14) van Gunsteren, W. F.; Berendsen, H. J. C. Algorithms for macromolecular dynamics and constraint dynamics. *Mol. Phys.* **1977**, *34* (no. 5), 1311−1327.

(15) Hünenberger, P. H.Thermostat Algorithms for Molecular Dynamics Simulations. In *Advanced Computer Simulation: Approaches for Soft Matter Sciences I*; Springer, 2005; pp 105−149.

(16) Valdés-Tresanco, M. S.; Valdés-Tresanco, M. E.; Valiente, P. A.; Moreno, E. gmx_MMPBSA: a new tool to perform end-state free energy calculations with GROMACS. *J. Chem. Theory Comput.* **2021**, *17* (no. 10), 6281−6291.

(17) Hines, J.; Lartigue, S.; Dong, H.; Qian, Y.; Crews, C. M. MDM2-recruiting PROTAC offers superior, synergistic antiproliferative activity via simultaneous degradation of BRD4 and stabilization of p53. *Cancer Res.* **2019**, *79* (no. 1), 251−262.

(18) Krieger, J.; Sorrell, F. J.; Wegener, A. A.; Leuthner, B.; Machrouhi-Porcher, F.; Hecht, M.; Leibrock, E. M.; et al. Systematic Potency and Property Assessment of VHL Ligands and Implications on PROTAC Design. *ChemMedChem.* **2023**, No. e202200615, DOI: 10.1002/cmdc.202200615.

(19) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: fast, flexible, and free. *J. Comput. Chem.* **2005**, *26* (no. 16), 1701−1718.

(20) Salentin, S.; Schreiber, S.; Haupt, V. J.; Adasme, M. F.; Schroeder, M. PLIP: fully automated protein−ligand interaction profiler. *Nucleic Acids Res.* **2015**, *43* (no. W1), W443−W447.

(21) Wang, Y.; Wei, Z.; Xi, L. Sfcnn: a novel scoring function based on 3D convolutional neural network for accurate and stable protein−ligand affinity prediction. *BMC Bioinf.* **2022**, *23* (no. 1), 222.