

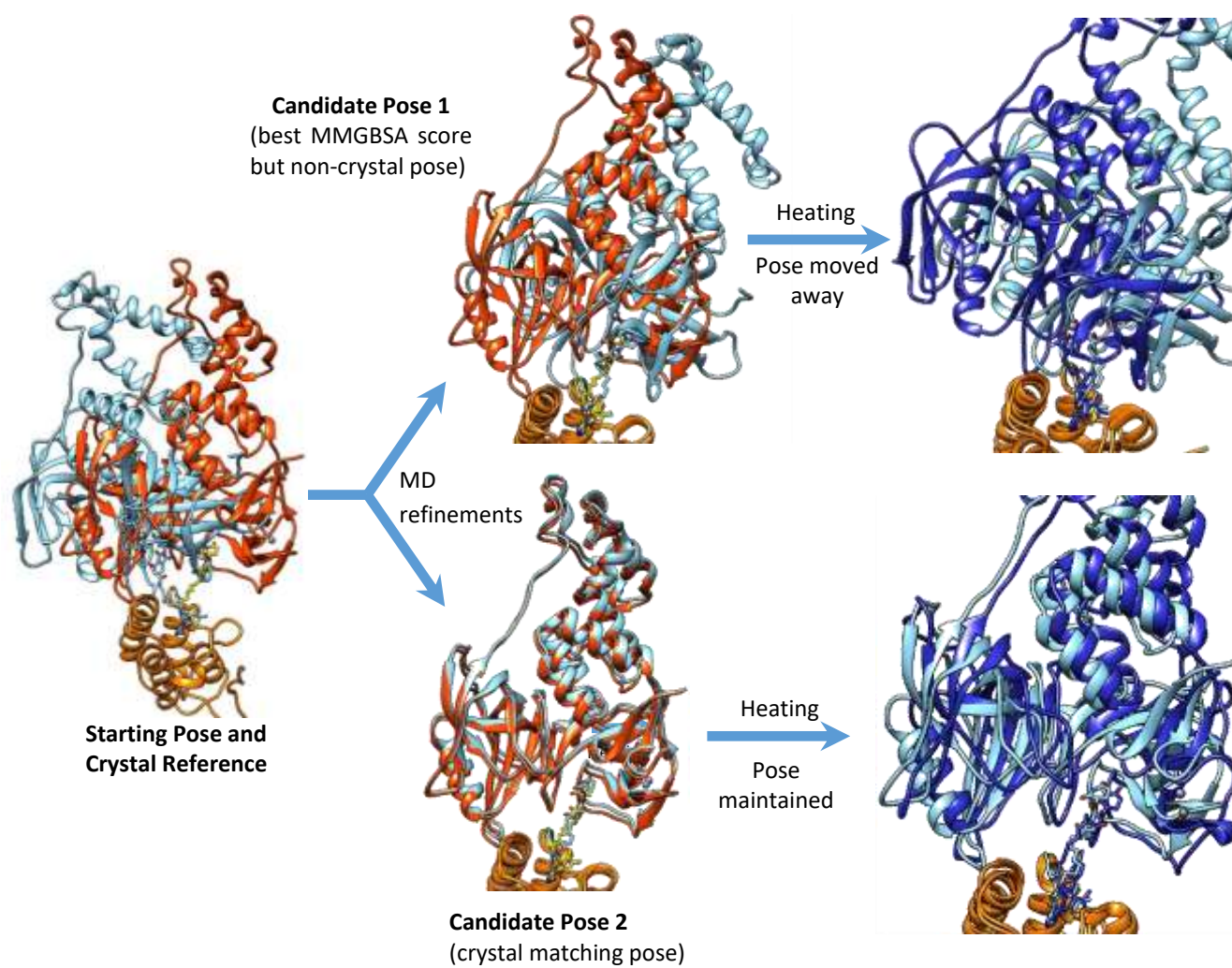
In Silico Modeling and Scoring of PROTAC-Mediated Ternary Complex Poses

Junzhuo Liao,¹ Xueqing Nie,¹ Ilona Christy Unarta,² Spencer S. Ericksen^{3*} and Weiping Tang^{1,2,3*}

Affiliations:

1. School of Pharmacy, University of Wisconsin-Madison, Madison, WI 53705, United States
2. Department of Chemistry, University of Wisconsin-Madison, Madison, WI 53706
3. Drug Development Core, UW Carbone Cancer Center, School of Medicine and Public Health, University of Wisconsin-Madison, 53705, United States

TOC graphic:



ABSTRACT

Proteolysis targeting chimeras (PROTACs) are bifunctional molecules that can induce the ubiquitination of targeted proteins via the formation of ternary complexes between an E3 ubiquitin ligase and a target protein. The poly-ubiquitinated target protein will be escorted to the proteasome for degradation. Rational design of PROTACs require knowledge of an accurate configuration of the PROTAC induced ternary complex. This study demonstrates that native ternary poses can be distinguished by scoring candidate poses based on the pose residence time. The scoring is essentially heat-and-dissociate trials of candidate poses sampled by MD and pre-ranked by the classic MM/GBSA method. It is practical, simple to use and self-intuitive, relying on the observation that the assumed more stable native crystal ternary poses maintained a longer residence time than non-native ones at both room and higher temperatures. A time score and temperature score were generated from multiple replicate trajectories. These scores were able to correctly identify the native pose from non-native ones in all the systems examined. The absolute numbers were comparable across different systems in all currently available VHL and CRBN-containing ternary crystal structures. Therefore, it is also possible to provide an empirical criteria for unresolved ternary structures that under the conditions of this study. If a ternary pose is stable up to over a certain threshold score, it is likely a native pose. The success of the method is in part attributed to the dynamic nature of the pose change analysis which naturally involves entropic effects, one that is intrinsically unavailable with faster static scoring methods that consider molecular mechanical energy only. Protein-protein binding entropy is much more significant than in protein-ligands binding. The success is also attributed to the fact that the protein structures themselves were all stable in the short heating trials.

1. INTRODUCTION

Since the discovery of cell-permeable small molecule ligands of E3 ubiquitin ligases, such as Von Hippel-Lindau (VHL) and cereblon (CRBN), for targeted protein degradation,^{1,2} proteolysis targeting chimeras (PROTACs) have become valuable tools for dissecting cell biology³ and also promising therapeutics⁴⁻⁷. PROTACs are bifunctional molecules having two ligands connected by a linker. One ligand binds to an E3 ubiquitin ligase, while the other ligand binds to the protein target. In the context of PROTACs, the “ligand” terminology usually only refers to the ligand warhead and not the whole small molecule. Throughout this report we use the terms “ligand warhead” to refer to part of the PROTAC, and simply PROTAC when mentioning the whole small molecule. The formation of target protein-PROTAC-E3 ligase ternary complex promotes ubiquitination of the protein target. Polyubiquitination marks the target protein for degradation in the proteasome. Unlike traditional small-molecule ligands that usually modulate target protein function in bound state, PROTACs only need to bind to the protein target transiently. Being able to degrade a target protein by transient association, PROTACs disrupt target function through a catalytic mechanism and may require lower dosing. The removal of the target protein may also achieve more sustained activity.

One of the pre-requisites for PROTAC to work is the formation of target protein-PROTAC-E3 ligase ternary complexes. After a suitable ligand is developed for the target protein, it needs to be linked to an E3 ligase ligand appropriately to promote a competent ternary binding arrangement for subsequent ubiquitination of the target protein. It is therefore important to characterize and understand how the ternary complexes are formed. Accurate knowledge of ternary complexes configurations is essential to structure-based PROTAC design and comparison of different PROTACs.

Computational approaches to PROTAC systems. In addition to obtaining a co-crystal structure from experiments, which is generally time-consuming, computational structure prediction methods have shown some potential for understanding the ternary complexes. A pioneering study in 2019 started with protein-protein docking as the first step, using the holo form protein+ligand of the ternary crystal. It is assumed that the two protein-ligand binary structures are known, and they are the same as in the ternary crystal. The docking is followed by independent PROTAC conformation sampling, and alignment of the ligand warheads from the sampled PROTAC onto the protein docking poses.⁸ Energy minimization of the aligned PROTAC was an important filter here. If the adjustment of PROTAC poses due to minimization was sufficiently small, an indication of less PROTAC linker strain generated during alignment, the whole ternary pose is predicted as a native crystal-like pose. A 40% hit rate could be achieved for obtaining crystal-like poses with the BRD4^{BD2}-PROTAC-VHL ternary system (PDB: 5T35) (**Figure 1** left, blue bar). However, it was also concluded that without prior knowledge of the crystal structures, it would be difficult to know which 40% are the correct poses given that there was no attempt to score the candidate poses. A 2020 update on this study used better filters and clustering to greatly improve the results for finding the native 5T35 pose.⁹ The largest cluster of ternary poses almost completely corresponded to the native pose (**Figure 1** left, green bar). However, the overall success rate for different systems is relatively low (**Figure 1**) and no other scoring was employed.

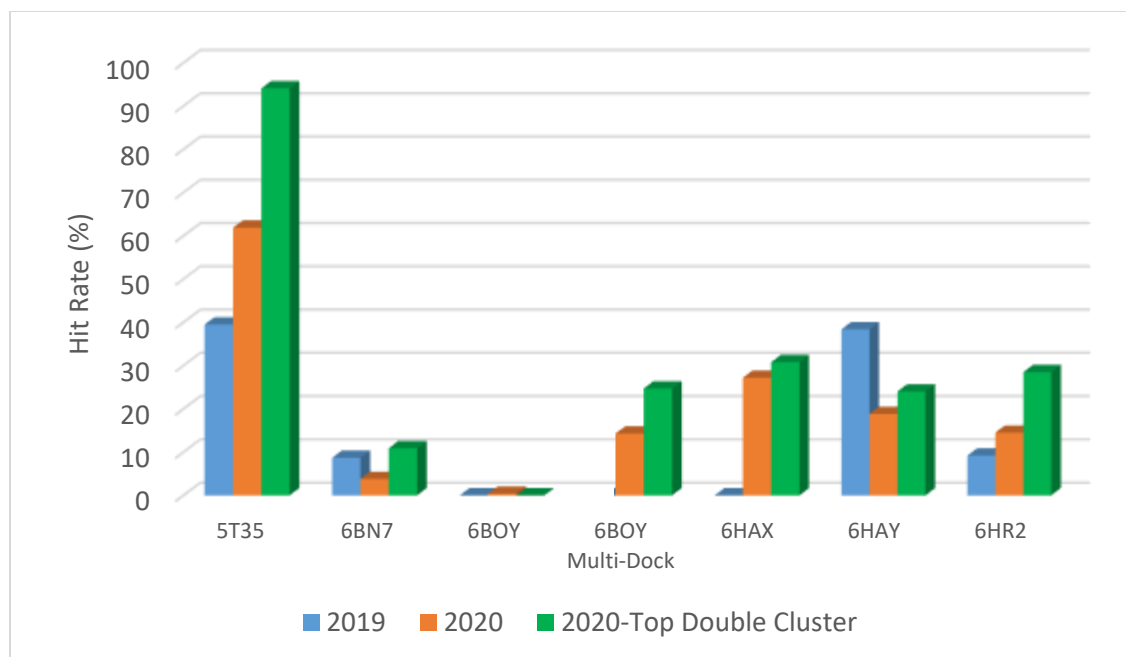


Figure 1. Previous search efforts from 2 studies combined, by Drummond et. al.

Another effort to determine ternary complex structures by Zaidman et. al used Rosetta-based modelling and provided improvements in accuracy and scope (**Figure 2**).¹⁰ Their method, named PRosettaC, involves two rounds of protein docking at the beginning. Later, another two rounds of Rosetta-based scoring is applied to the possible ternary complexes generated. The results are clustered with the population rank being the final scoring criteria.

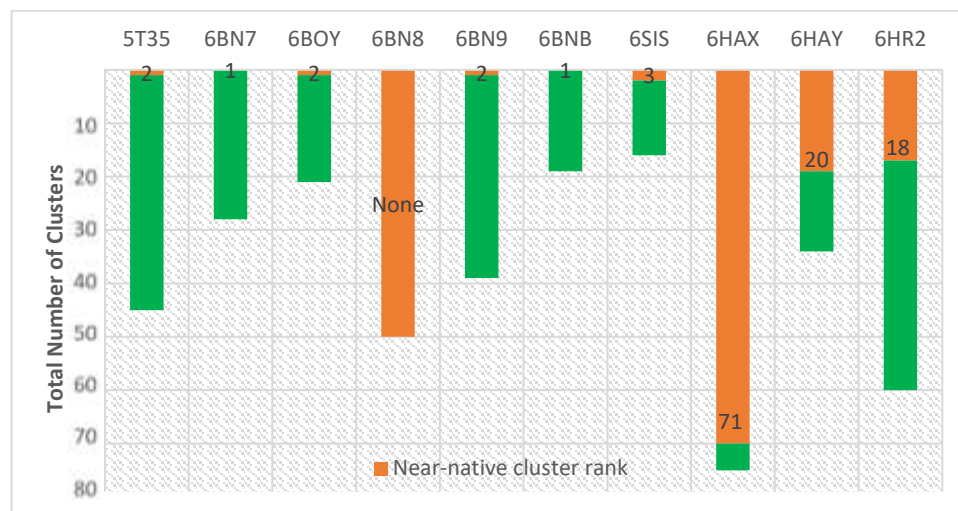


Figure 2. Reported results with the PRosettaC method. The population rank of the near-native cluster for each system is labeled. More green indicates a ranking closer to top, which is the ideal situation. More red indicates a ranking more deviated from top.

Indeed, these docking-based methods performed very well in terms of conformation space search and the native pose was almost always included in the results. However, the filtering or scoring so far have not been able to identify the native pose as the top in many systems. In an ideal situation, the native pose should be ranked at the top across many different systems. If this issue is resolved, then low hit rates (or low cluster population of the near-native pose) as seen above would not be a concern anymore.

In the present study, we focus on being able to distinguish native against non-native poses by our MD-based scoring method. The poses are generated by a selective local MD conformation sampling that found the native pose in the sampling results. The scoring is successfully demonstrated not only on our own conformation search results, but also on poses independently generated from other sources. Briefly, one can start with a classic MMGBSA enthalpy only prescore, and it is followed by heating-accelerated pose change scoring. It can be performed in a practical manner by medicinal chemists without extensive MD knowledge.

In our MD studies on PROTAC ternary systems, we have observed the native crystal ternary pose exist stably in our simulations (of up to ~100 ns long single trials), while non-native poses may or may not be stable during this time, a well-expected observation. We adapt this finding and convert the observed pose residence time into general scoring for ternary pose stability. It is a novel, effective and easy-to-use way of virtually evaluating the stability of each given ternary pose. To make the scoring method practical enough, the simulation wall time needs to be short enough. The method is essentially heating-accelerated pose change trials in MD. Our scoring was effective for all 4 of the systems we studied and the native pose was identified in all cases. It was also effective on systems we did not generate poses ourselves.

In a general process of $RL \rightleftharpoons R + L$, where R is receptor, L is ligand and RL is the bound complex, there are 2 effects of heating. 1) In simulation, it decreases the residence time of the bound complex. 2. On a macroscopic level, it increases the unbound-bound equilibrium constant $K = \frac{[R][L]}{[RL]}$ to favor the unbound state more. Dissociation of ligands have standard reaction enthalpy $\Delta H^\circ > 0$ and standard reaction entropy $\Delta S^\circ > 0$. From the Van 't Hoff equation as in (1) (assuming ΔH° and ΔS° do not change with temperature),

$$\ln K = -\frac{\Delta H^\circ - T\Delta S^\circ}{RT} = -\frac{\Delta H^\circ}{RT} + \frac{\Delta S^\circ}{R} \quad (1)$$

one can see the increase of K with temperature.

However, true observation of thermodynamic equilibrium can only be achieved with an ensemble of molecules, while simulation usually studies one complex. A ligand unbound from its pocket will be very unlikely to rebound in simulation, and when it does the unbound/bound population or time does not reflect any real K.^{11, 12} So here in this study it is by no means relying on the observation of equilibrium shift of our ternary poses.

Fortunately, the greatly decreased bound residence time can be easily observed. Heated temperatures have been demonstrated to accelerate protein unfolding simulations.¹³⁻¹⁶ While in our case we are not unfolding the protein but even avoiding it, the ternary pose change is clearly observed and accelerated. The protein structure itself is stable within our short MD trials, while ternary pose residence time was quite different for native vs non-native poses. At room temperature all candidate ternary poses are stable within the short tens of ns. With the pose residence time decreased to less than short tens of ns, one can easily compare different poses, as successfully demonstrated in this study.

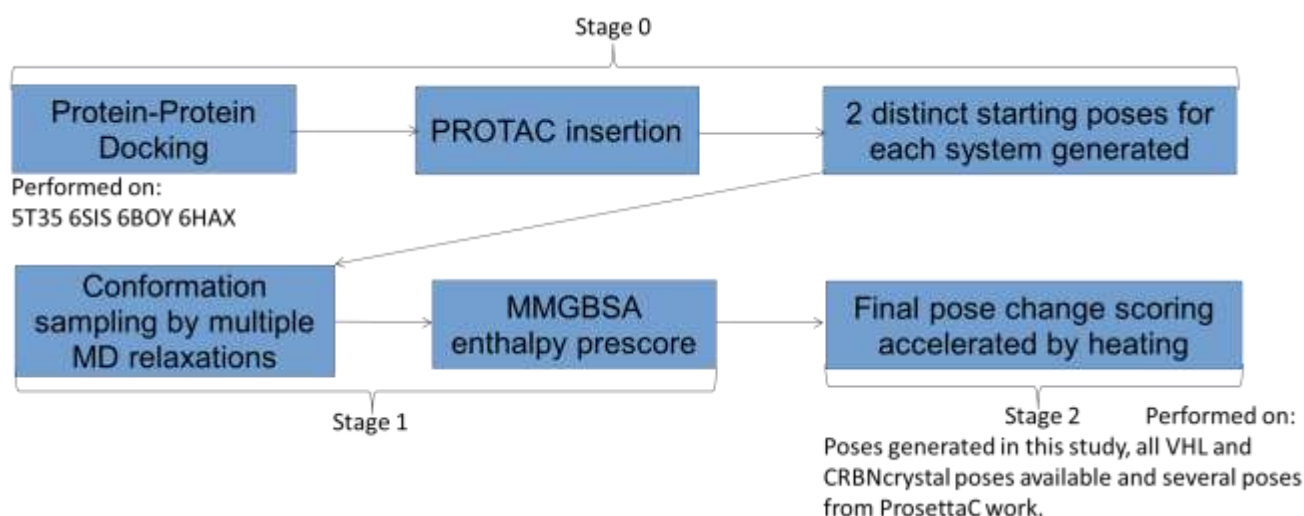
While there are other techniques¹⁷⁻²² capable of accelerating kinetics which potentially includes ternary complex pose changes, the heating is simplest for the user and indeed worked very well in this specific PROTAC study. Here we are only using heating for its increased kinetics. We are not concerned about the accuracy of any pathways or new structures in this stage. One may have worried about false negatives-i.e. a good pose changing away too soon and that would mean in the results the native pose would score lower than non-natives. However, we did not see such result. Notably, part of its success may be attributed to the fact that the protein structure themselves were all stable during our short trials. The short run times and not too high heating ceiling (410 K) may have also helped, keeping the water molecules all liquid.

Additionally, there are at least two more advantages of this method. 1) It is performed in explicit solvent, avoiding the introduction and choice of a new set of parameters needed for the commonly used GBSA measurements. 2) This method inherently includes any entropic effects to the evaluation of binding since one is looking at a dynamic

process being run multiple times, instead of static molecular-mechanical (MM) energy values only, when the very costly entropy calculations are not performed.

2. METHODS

The overall workflow (**Scheme 1**) of this study is as follows. We first demonstrate that it is possible to reach and maintain the native crystal pose (among other non-native poses) by conformational sampling from MD. Based on the score of the poses, the native one is ranked at the very top. Although MD relaxations are involved in stage 1, the scoring in stage 2 does not rely on a specific sampling method and can be used independently as long as there are poses generated.



Scheme 1. Overall workflow

2.1 Structure Preparation and Pretest

Protein docking. Protein-protein docking poses were first obtained using ClusPro,²³⁻²⁵ which is very accessible as a free online server and simple to use. For the design of most PROTACs, the two ligand warheads and their binding pockets are already known from crystal structures. For docking input, we used the holo form protein structures from independent protein-ligand binary crystals or the ternary crystal. The ligands were not included in the docking input since they are not recognized by ClusPro. Residues around each ligand pocket were defined as attractive regions to favor the known binding sites being closer together (**Table 1**). The docking results may still give some poses that had the two pockets too far for linker attachment, even with the defined attractions. These poses were excluded for further use. The top two ClusPro “balanced” scoring poses were then selected.

Table 1. Attraction residues defined in ClusPro. The residues were chosen around where the linker extends out. The smaller protein was defined as “receptor” because this would allow the larger protein to be relaxed more as the “ligand”. Cluspro gives both the backbone and sides chains some relaxation in the ligand, but the backbone is held rigid in the receptor.

System	Receptor	Ligand	Input PDB when independent binary crystal was used instead of cut-out from ternary
5T35 ²⁶	BRD4 ^{BD2} a-433, 385, 432	VHL c-67, 91, 69	4W9H (VHL-ligand), independent BRD4 ^{BD2} -ligand unavailable

6SIS ²⁷	BRD4 ^{BD2} a-433, 385, 432	VHL c-112, 109, 98, 90	4W9H (VHL-ligand), independent BRD4 ^{BD2} -ligand unavailable
6BOY ²⁸	BRD4 ^{BD1} c-87, 146, 140, 83	CRBN b-357, 386, 352	-
6HAX ²⁹	SMARCA2 ^{BD} a-1413, 1463, 1470	VHL c-98, 99, 108, 110	5NVX (VHL-ligand) and 6HAZ (SMARCA2 ^{BD} -ligand)

PROTAC insertion. The linkers were cut as indicated (**Figure 3**). To regenerate the linker in 3D, either conversion from its SMILES string by *omega2*^{30, 31} or visual building was used.

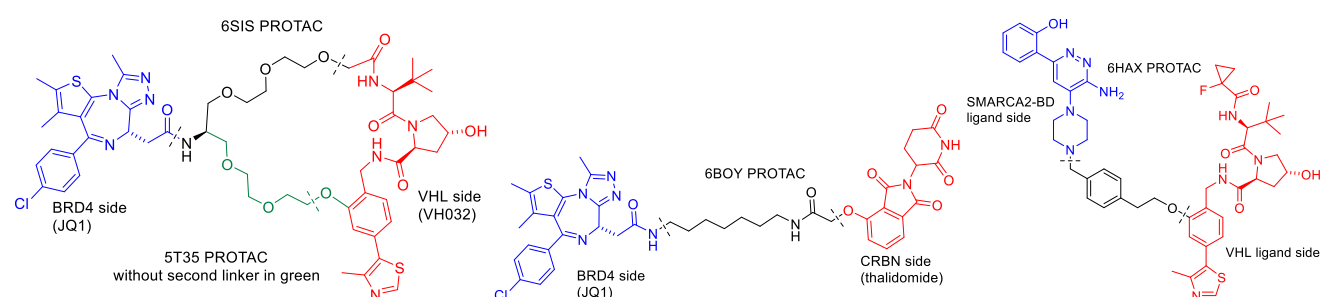


Figure 3. PROTAC chemical structures and cutting points during 3D model building.

Using Chimera,³² in each of the two selected protein dock poses, placing the ligand warheads into their pockets were based on the known crystal of each. Since the PROTAC was not in the docking stage, some minor translational adjustment of the protein (mainly perpendicular to the protein contact plane) was applied visually to provide just enough space for the PROTAC linker to avoid clashes with the 2 proteins structures. The linker was then attached visually. To remove distorted bond lengths and angles of the visually attached linker, we proceeded first without considering the proteins. To this end, 100 steps steepest descent + 10 steps conjugate gradient minimization was performed 10 times on the linker. When the proteins are included, another 100 steps steepest descent + 10 steps conjugate gradient minimization steps were performed 10 times. Only linker atoms and on one occasion a single protein side chain that is too close to the linker were allowed to move. More detailed step-by-step model building procedures which were performed in Chimera are provided in the SI.

Preparation for MD. AM1-BCC charges were assigned to the PROTACs.³³ In *tleap*, the *gaff2* forcefield was used for the PROTAC. The *ff19SB* force field was used for proteins.³⁴ An octahedral box of OPC water extending at least 10 Å from solute atoms were generated.³⁵ Na⁺ and Cl⁻ counterions were added to neutralize the system.³⁶ Hydrogen-mass-repartitioning (HMR) assigning solute hydrogens a mass of 3.024 was used to enable a 0.004 ps timestep, or 0.002 ps at higher temperatures.³⁷

In our pretest stage when the newer *ff19SB* was not yet released, *gaff* for PROTAC, *ff14SB* for proteins,³⁸ and TIP3P water was used.³⁹ There were no differences in the results with either setup. All final results reported here are with the newer *ff19SB* setup.

For implicit solvent simulations (6BOY) and MMGBSA energy calculations, *mbondi2* radii set and GB^{OBC2} (*igb*=5) model was used.⁴⁰ HMR was not applied.

Pretest. One advantage of MD search is that only poses belonging to a local energy minimum will remain stable, thus eliminating the majority of exhaustive poses. Though at the onset of this study, there were very few MD simulations on PROTAC systems. To first examine the capability of MD simulation being able to recapitulate and

stabilize a PROTAC ternary system at an experimental crystal pose, VHL-PROTAC-BRD4 (PDB: 5T35) was used for a pretest starting with a single GTX 1070Ti graphics card on a local workstation. After a short routine minimization and equilibration stage, 110 ns of MD relaxation at 310 K was applied to the two constructed starting poses (**Table 1** row 2, when using cut-out binary crystals from known ternary), each given 5 replicates. Another control run was performed using the crystal ternary as the starting pose to examine pose stability and for subsequent energy scoring.

2.2 Conformation Sampling

The sampling was explorative and the number of runs is expected to be increased accordingly based on the scoring results. Based on our pretest experience, hardware limits and literature reports on multiple short runs⁴¹ being better than a single long run, we decided to start with at least 10 relatively short runs for each starting pose. The GPU version of AMBER was used in all simulations reported in this study.⁴²⁻⁴⁵

Table 2. Sampling amount for production runs

System	No. of starting poses	Simulation time for each single replicate	Actual replicates for each start pose
5T35	2	50 ns	x10
6BOY	2	20 ns (GB ^{OBC2} (igb=5))	x30
6SIS	2	50 ns	x10 for pose 1, x20 for pose 2
6HAX	1+1	50 ns	x20

For each system, we wanted to observe the pose with the most favored MMGBSA prescore stabilized in the last 10 ns occurring from at least two trajectories. Starting from 10 (or 30 for implicit solvent 6BOY), we increased the number of replicates initiated from each starting pose if the lowest MMGBSA prescore trajectory only showed up once. This is to best mimic the situation without prior knowledge of the crystal pose. Additionally, the final heating rescore later in this study can be the guide to further increase sampling both in terms of starting pose numbers and simulation trials, if no pose has reached a >17.5 time and >380 temperature score, empirically drawn from examining all the currently available CRBN and VHL ternary crystals.

Explicit solvent MD trials. A short minimization and equilibration stage was performed proceeding the production stage, with 5,000 steps of steepest descent minimization (ntmin=2) with heavy atoms restrained, followed by MD equilibration in the NPT ensemble (ntb=2, ntp=1, Montecarlo barostat, barostat=2, taup=0.5, Langevin thermostat, ntt=3, gamma_ln=2) at 310 K of 150,000 steps (0.001 ps timestep), gradually releasing the heavy atom restraints. Complete force evaluation was applied (ntf=1, ntc=1). The minimization and equilibration conditions were initially adapted from the works of the Rizzo group⁴⁶ and gradually modified throughout the course of this study. The production stage of 50 ns was run in the NVT ensemble (Langevin thermostat, ntt=3, gamma_ln=0.01). The long-distance cutoff was set to 8 Å. The SHAKE algorithm was used to restrict hydrogen bond length changes, and no forces are calculated for bonds with hydrogens (ntc=2, ntf=2). The timestep was set to 0.004 ps, and snapshots were recorded every 0.1 ns.

Implicit solvent MD trials. This applies to the 6BOY search. GB^{OBC2} was used(igb=5, ntb=0, ntp=0). After allowing hydrogens to fully minimize while heavy atoms restrained at 5 kcal/mol/Å for 5,000 steps steepest descent (ntmin=2), CRBN and its ligand warhead were held mostly rigid by weight restraints of 10 kcal/mol/Å on all atoms. BRD4 was held by internal distance restraints of all C- α atoms to a single C- α based on the starting pose. The BRD4 ligand warhead was held in its pocket by distance restraints. Based on the starting pose atom distances, the distance restraints were a parabola well generated in *cptraj* 1 Å wide and extending up linearly outside the parabolic well region. The force constant for restraint was 10 kcal/mol/Å. MD equilibration at 310 K for 250,000 steps (0.001 ps

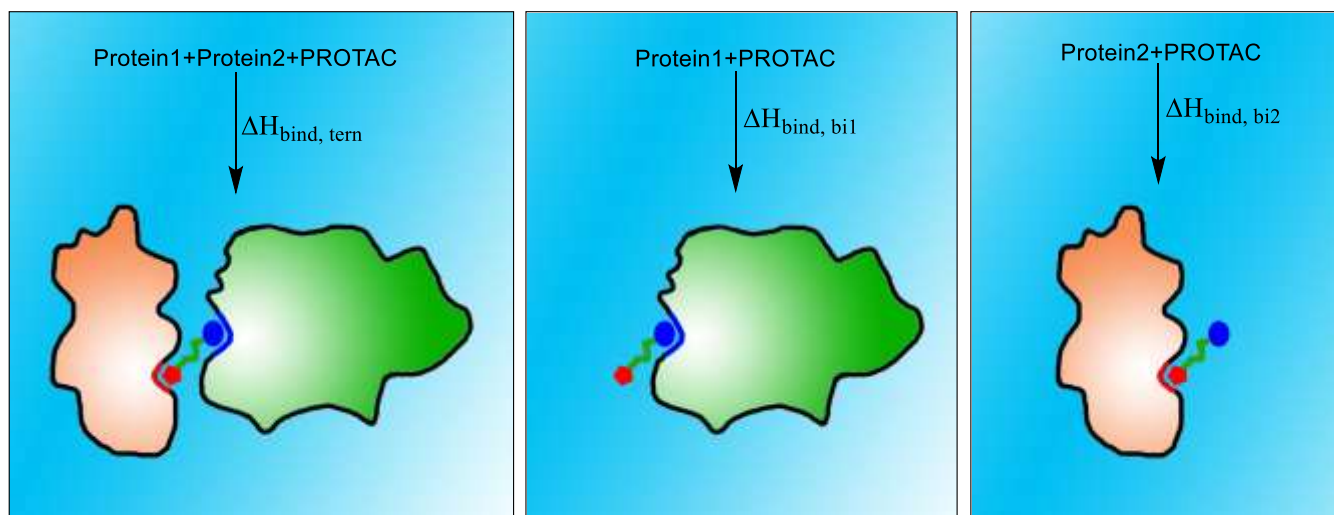
timestep) with complete force evaluation was applied (ntf=1, ntc=1). The production stage was run for 20 ns. The SHAKE algorithm was used to restrict hydrogen bond length changes, and no forces are calculated for bonds with hydrogens (ntc=2, ntf=2). The Langevin thermostat (ntt=3, gamma_ln=2) was used. No long-distance cut-off was applied. The timestep was set to 0.002 ps, and snapshots were recorded every 0.1 ns.

Combined clustering of frames. For each system, all frames from the last 10 ns of each MD run were combined and clustered. For each cluster, one representative pose with the lowest cumulative distance to all other frames in cluster was extracted. A cluster may span 1 or more trajectories of final 10 ns. When more than one trajectory corresponded to a cluster, only the 10 ns from the trajectory with the lowest enthalpy was used to rank this cluster/pose. The top 5 poses extracted from the top 5 clusters based on MMGBSA energy of the trajectory was used for rescoring by heating. Using the best consecutive 10 ns is the unified selection criteria we have established for prescoring, which also streamlines the process to perform with many trajectories. Clusters that are significantly smaller than 10 ns indicate pose instability. Their corresponding trajectories generally had poor MMGBSA energies and they were therefore eliminated in this stage. Even occasionally when a trajectory with a poorly formed cluster was within the top 5 in energy (i.e. **Figure 6**, 5T35 1.1 and **Figure S8**, 6HAX 1.20), we maintained the same prescoring criteria.

In *cpptraj*, the DBScan algorithm was used.⁴⁷ Prior to clustering, in each trajectory the smaller BRD4 (or BD) was aligned to its crystal. Coordinates of the other protein's C- α (excluding the region bound to the deleted anchor proteins) and its ligand warhead heavy atoms were measured for clustering. Aligning the smaller protein helped make the movements and differences more prominent in the analysis. The Epsilon was set to 1.5 Å and minpoints set to 1. No sieving was used so all frames were used to cluster.

2.3 MMGBSA Prescore

We take advantage of the well-established and commonly applied MMGBSA energy calculation method as the first round of scoring,⁴⁸⁻⁵⁰ which are usually for ligand-receptor binary systems. We expand its application to the PROTAC ternary system here.



Scheme 2. In-aqueous binding energy scheme for calculating ternary binding energy $\Delta H_{\text{bind,tern}}$ and cooperativity binding energy $\Delta\Delta H$ in equations (2)-(4).

$$\Delta H_{\text{bind,tern}} = H_{\text{complex}} - H_{\text{protein1}} - H_{\text{PROTAC}} - H_{\text{protein2}} \quad (2)$$

$$\Delta H_{\text{bind,coop}}(\Delta\Delta H) = \Delta H_{\text{bind,tern}} - \Delta H_{\text{bind,bi1}} - \Delta H_{\text{bind,bi2}} \quad (3)$$

$$\begin{aligned}
&= (H_{\text{complex}} - H_{\text{protein1}} - H_{\text{PROTAC}} - H_{\text{protein2}}) - (H_{\text{bi1}} - H_{\text{protein1}} - H_{\text{PROTAC}}) - (H_{\text{bi2}} - H_{\text{protein2}} \\
&\quad - H_{\text{PROTAC}}) \\
&= H_{\text{complex}} + H_{\text{PROTAC}} - H_{\text{bi1}} - H_{\text{bi2}} \quad (4)
\end{aligned}$$

Using the GB^{OBC2} (igb=5) model with mbondi2 radii,⁵¹ the last 10 ns of each trajectory was subject to MM/GBSA scoring of ternary binding energies, as in eq. (2). With MMPBSA.py,⁵² for each single trajectory the program was run 3 times, each time having the PROTAC, protein1 or protein2 defined as the ligand, and the remaining two components as the receptor. The required parameter-topology files for each run were stripped from that of the complex by *cpptraj*.

In addition to ternary binding energy, cooperativity binding energy is a common parameter for PROTAC ternary binding systems.^{28, 53} The ternary binding energy may be very favorable if a single ligand warhead happens to be very strong, but it does not completely reflect the stability of the ternary system. Therefore, we want to separately subtract all PROTAC-single protein binding energies from the total ternary binding energy as a second measure, $\Delta\Delta H$, as in eq. (3) or simplified as eq. (4). This is the same energy as defined and measured by Roy and Ciulli et al. by SPR.⁵³ When comparing poses for ranking, both ΔH and $\Delta\Delta H$ were considered by simply summing up of the two.

2.4 Final Rescore by Heating-Accelerated Pose Change Trials (HAPOC)

Heating runs. To perform our heating-accelerated pose change (HAPOC) analysis for PROTAC ternary poses, the clustering and MMGBSA served as the first round of scoring (**Table 3**). For each system, a minimum of top 5 poses based on their corresponding H score are then taken for the final rescoring.

Each extracted candidate pose was used as the starting frame in the heating. Only the two proteins and PROTAC were retained in the pose. Any pose from any method of sampling can be fed into this step. When starting directly from the known crystal, no conformation search or prescoring was needed.

Before heating, all poses were reassigned AM1-BCC partial charges for the PROTAC and prepared with the same *tleap* workflow as for the MD relaxations. HMR was applied. A short minimization and equilibration stage was performed proceeding the actual heating stage, with 5,000 steps of steepest descent minimization (ntmin=2) with heavy atoms restrained, followed by MD equilibration in the NPT ensemble (Langevin thermostat, gamma_ln=2, Montecarlo barostat, ntb=2, ntp=1, barostat=2, taup=0.5) at 310 K of 150,000 steps (0.001 ps timestep), gradually releasing the heavy atom restraints. The production stage of 50 ns was run in the NPT ensemble (Langevin thermostat, ntt=3, gamma_ln=0.01, Montecarlo barostat, ntb=2, ntp=1, barostat=2, taup=0.5). The long-distance cutoff was set to 8 Å. The SHAKE algorithm was used to restrict hydrogen bond length changes, and no forces are calculated for bonds with hydrogens (ntc=2, ntf=2). The timestep was set to 0.002 ps, and snapshots recorded every 0.05 ns.

Generating scores from the trajectories. With the smaller protein (BRD4 or BD) and its ligand warhead aligned to the starting pose, the RMSD of the other side to its start was measured as an indicator of movement and stability. One could either measure the core regions of the other protein or the other ligand warhead, where they moved in correlation with each other. We measured the other ligand warhead. One example is shown in **Figure 4**.

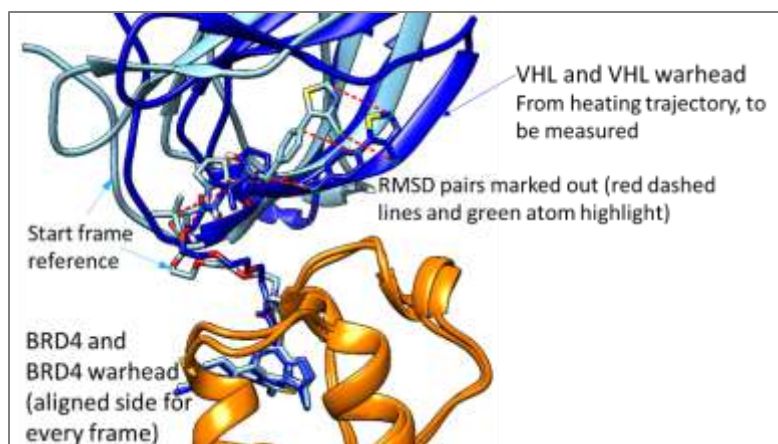


Figure 4. One example taken from 5T35 heating rescore trails. Representative heavy atoms on the unaligned ligand warhead (dark blue) were measured to the candidate pose (start) in terms of per-frame RMSD.

To convert the RMSDs into a final score indicating ternary pose stability, we evaluated it in two ways, and generated a time score (τ score) and temperature score (T score). The τ score represents the simulation time (in ns) of the moving PROTAC end being within 3.5 Å of the starting pose. With 500 frames in total for 25 ns, the total number of frames that had a 10-frame moving average of < 3.5 Å RMSD to the start are counted and divided by 20. For the T score, when this same moving average RMSD increased to ≥ 4 Å, the thermostat temperature is recorded. If the RMSD falls back to moving average < 3 Å later in the trajectory, the thermostat temperature is recorded again when the next time it crosses to ≥ 4 Å. If by the end of the heating scheme, the RMSD has never crossed over 4 Å, or fell back to less than 3 Å and have not crossed over 4 Å yet, then a 420 thermostat temperature is recorded for this instance. For each trajectory, all recorded thermostat temperatures are averaged to give a single T score. The 10 T scores from the 10 replicates are then averaged again, as well as the 10 τ scores. There are two reasons why the heating is not performed at temperatures higher than 410K. 1) The non-native poses rarely reached 410K and still remained bound. Only some crystal poses could be heated to 410 K without the pose change occurring. The difference is already obvious. 2) The increased likelihood of protein unfolding at high temperature may tend to make all poses look like negative ones. When actually performing the RMSD measurements, in addition to the indicated region in **Figure 4**, the aligned protein, unaligned protein, aligned ligand warhead and linker were all plotted as control measures to monitor potential abnormalities and other signs of instability, such as protein unfolding and ligand dissociation from pocket. Fortunately, these events were rarely seen, which indicated that the situation was not complicated by these factors.

Evaluating poses generated independently from other sources. Ternary docking poses from other sources can also be fed into the heating step for final rescoring. To demonstrate independent usage, systems and poses generated and reported by the PROsettaC study^{10, 54} were taken for rescoring with our heating-accelerated pose change method. In the reported study, ternary poses were clustered, and their scoring criteria was cluster population rank. Of the 10 PDB systems studied with the reported method in literature, two systems had the near-native pose as the top cluster, and four systems had it ranked 2 or 3 (**Figure 2**). Both MMGBSA binding energy and heating scores were evaluated for their top poses generated for 6BOY and 6HAY, each representing a good but not perfect, and poorer but not worst system. For the non-native poses, a single pose was chosen from each cluster. For the native pose control, 6BOY poses from the found near-native cluster (no.2) were compared to the crystal native pose and the one closest to the crystal was used. In 6HAY the near-native cluster (no.20) was only a singleton and was used directly. The native crystal itself as the starting point was also used in both systems. The τ score and T score were generated in the same way as previously described.

To obtain MMGBSA binding energies of these independent poses, a short minimization and equilibration stage was first performed, which is the same as in the explicit solvent conformation sampling and heating rescore steps. The production stage was run at 310 K for 20 ns with the Langevin thermostat ($\gamma_{ln}=2$) in the NVT ensemble

with snapshots collected every 0.1 ns (25,000 steps). HMR was applied to enable a 0.004 ps timestep. This procedure was done at least 3 times for each pose.

Since we are scoring a given pose instead of looking for new poses, we had to make sure the frames were pertaining to that pose in terms of RMSD to start with BD or BRD4 aligned for every frame. The generated trajectories were carefully examined to use either the 0-10, 5-15 or 10-20 ns portion with the lowest CRBN or VHL warhead RMSD. At the same time it had to satisfy the condition that this consecutive 10 ns portion had the RMSD $< 3.5 \text{ \AA}$ in 90 or more frames out of 100. If no portion satisfied, the trajectory was not used. Ultimately each pose had three trajectories maintaining around the local minima of itself.

2.5 Wall Time Considerations.

For each of the 25 ns heating runs under the conditions reported here, the smaller VHL systems will take about 3-4 hours, and the larger CRBN systems will take about 6-7 hours on an NVIDIA GeForce RTX 2080Ti GPU. This is well suited for parallel running.

3. Results and Discussion

3.1 Pretest Results

We are delighted to find that two of each of the five runs converged onto the crystal pose. In the control run the crystal pose was stable over 110 ns.

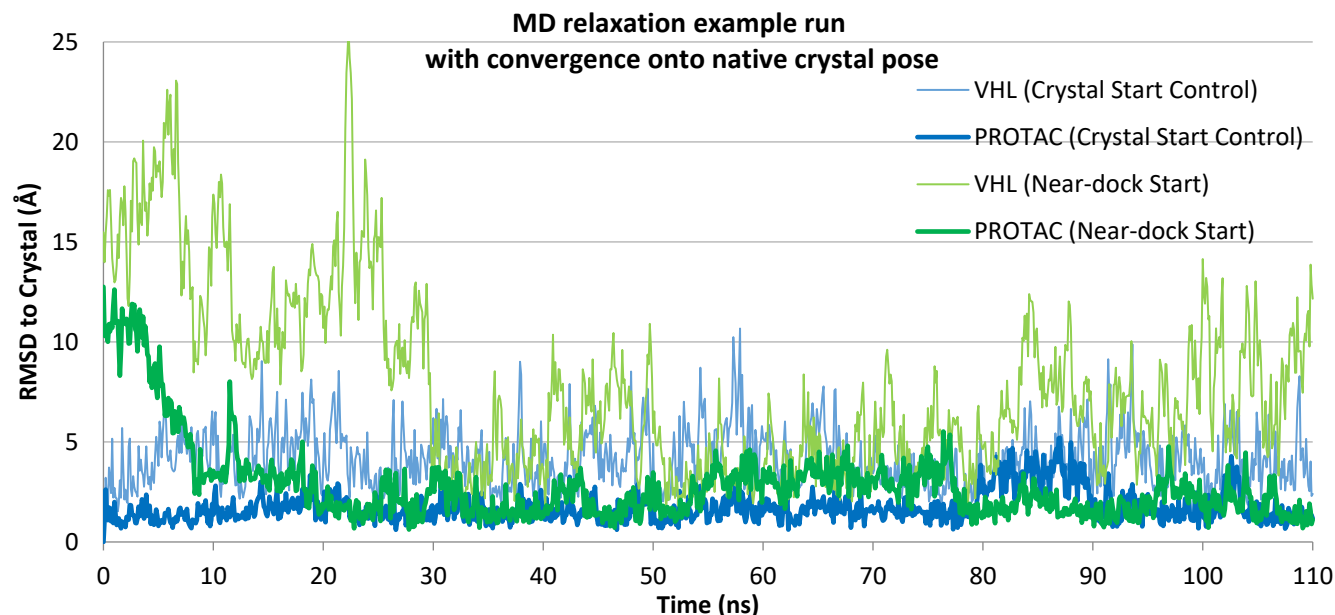


Figure 5. MD relaxation (1 example trial, 2-MD3 in **Table 3**) reaching convergence onto native crystal pose from built near-dock starting pose in the VHL-PROTAC-BRD4 (PDB: 5T35) system. The BRD4 C- α were aligned to the crystal pose for every frame. PROTAC plotted based on VHL ligand warhead side. VHL was plotted based on its C- α .

The last 10 ns of the 10 total replicates + 1 control run was measured for MM/GBSA ternary binding ΔH and cooperativity binding $\Delta\Delta H$. The 4 runs that converged onto the crystal pose, along with the control crystal start run, had the 5 highest cooperativity binding $\Delta\Delta H$ and occupied 5 out of the top 6 scores for ternary binding ΔH (**Table 3**). This encouraging result suggested that a local MD refinement is potentially capable of finding stabilized ternary poses, and a classic MM/GBSA binding score may be able to rank the native pose as the most energetically favorable one. Therefore, we expanded our scope of prediction to more ternary systems, including ones that failed from previously reported methods. Additionally, we also used independent ligand-bound protein structures to better mimic the situation where no ternary crystal was available. The use of independent binary structures was not demonstrated in any of the previous work on PROTAC ternary pose prediction published previously.^{8-10, 55} During the preparation of this manuscript, the first method starting from independent binary structures and also apo protein structures was reported for the modeling of PROTAC-mediated ternary complex poses.⁵³

Table 3. For the pretest using 5T35, four out of ten MD relaxations found and stabilized at the native crystal pose and the corresponding MMGBSA enthalpy energy calculations give a very good ranking. Error margin shown in s.e.m. All units in kcal/mol. 2-MD3 is the example trial shown in **Figure 5**.

Near-Docked Complex and Trial No.	Native state convergence	Ternary binding enthalpy ($\Delta H_{\text{bind, tern}}$)	rank	Coop. enthalpy of ternary complex ($\Delta\Delta H$)	rank
crystal starting pose	control	-117.5 ± 0.6	5.5	-25.3 ± 0.8	4.5
1-MD1	YES	-119.1 ± 1.1	5	-26.4 ± 1.5	4
1-MD4	YES	-119.3 ± 0.9	4	-32.1 ± 1.2	2

2-MD3	YES	-122.1 \pm 0.9	3	-30.2 \pm 1.5	3
2-MD4	YES	-122.3 \pm 0.8	2	-32.7 \pm 1.1	1
1-MD2	No	-112.6 \pm 0.8	8	-14.5 \pm 1.1	9
1-MD3	No	-105.3 \pm 0.7	10	-11.7 \pm 0.7	10
1-MD5	No	-127.4 \pm 0.8	1	-21.1 \pm 1.1	7
2-MD1	No	-113.9 \pm 0.8	7	-21.7 \pm 0.9	6
2-MD2	No	-105.8 \pm 1.0	9	-15.7 \pm 1.2	8
2-MD5	No	-115.7 \pm 0.9	6	-22.4 \pm 1.1	5

3.2 Sampling, Clustering and MMGBSA Prescoring Results

The sampled frames from the MD trajectories were analyzed by MMGBSA and clustering in the same stage. With the clustering, one can easily see 1) whether a run has mostly or fully converged onto a single pose, and 2) if convergence to a single pose is reoccurring from independent starts. It quickly provides an overall picture without the need to visualize every trajectory. **Figure 6** shows clustering analysis from 5T35. Clustering for 6SIS, 6BOY, 6HAX are shown in SI, **Figures S6-S8**. In the graphs, the trials are denoted as numbers x.y, where x indicates the starting pose number and y indicates the MD trial number for that start pose.

The searches for 5T35 and 6SIS involving VHL from an independent crystal found the native ternary relatively easily. Based on the criteria of having the lowest MMGBSA scoring trajectory showing up more than once, 6SIS was given an additional 10 runs for the second starting pose.

6BOY being larger than the rest was observed to move slower than desired when first ran in explicit solvent. To speed up sampling, implicit solvent was used. Even though the nominal simulation speed may not always be faster for larger systems due to not employing a cutoff while PME (particle mesh Ewald method) is employed for explicit solvent, the conformation search in implicit solvent is still faster due to the lack of solvent viscosity.^{56, 57} With the increased flexibility, we need to do the following to restrain internal degrees of freedom. 1) CRBN was held rigid at all atoms. 2) PROTAC warheads were held in the binding pocket. 3) BRD4 was internally restrained at C- α . If the restraints were released, the MD relaxations did not reach the crystal pose within the search we did. This also meant that only the cut-out binary from 6BOY ternary found the crystal. The available independent binary of human CRBN-ligand (i.e. PDB: 4TZ4) was still somewhat different than the 6BOY CRBN. It could not be held rigid for the purpose of searching for the 6BOY pose.

While 6HAX with independent binary starting from docking also didn't reach the crystal pose within the search we did, the cut-out binary with start pose artificially modified from the ternary did. In the first 10 runs the lowest enthalpy score trajectory (not comparing the independent binary starts from docking) showed up twice, but we still gave it another 10 runs just to sample more because the lowest enthalpy trajectory moved away from the clustered pose in the final few ns (SI **Figure S8** run 1.5, visually confirmed). In the next 10 runs, this same pose showed up in 2 stabilized trajectories (**Figure S8**, runs 1.17, 1.18), and no more runs were done. It was also the crystal pose. Poses found from both starting versions (last row, **Table 2**) were later jointly compared with their MMGBSA scores. Since two poses from the independent binary starting version had an even lower enthalpy score than the found native, they were good candidates to test our final rescoring.

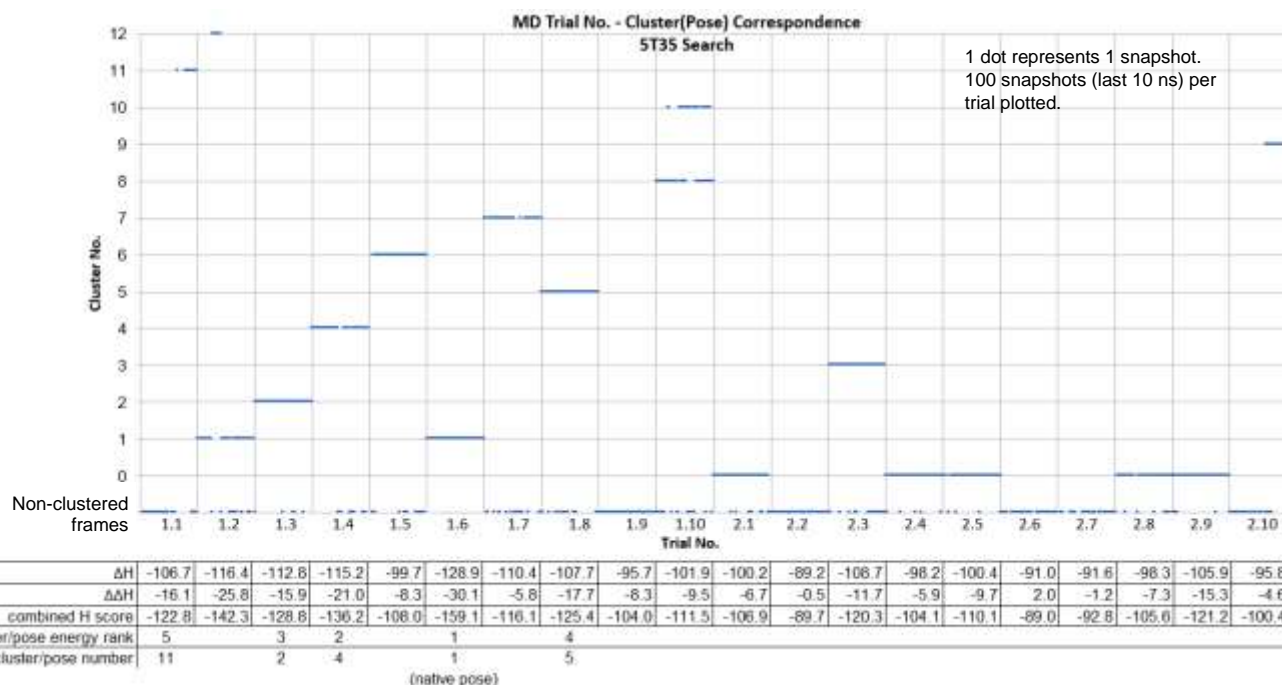


Figure 6. Clustering analysis of the last 10 ns from all 20 runs for 5T35 search. 1.x are trials starting from pose 1 and 2.x are trials starting from pose 2. The corresponding MMGBSA energy for each run is shown below. For every cluster, if it only spanned a single trajectory (e.g. cluster no.2), its representative pose is corresponded to the MMGBSA H score of this trajectory (-128.8). If more than one trajectory converged onto that cluster (e.g. cluster no.0), the most favorable H score of the two or more corresponding trajectories will be used (-121.2 from trial 2.9).

Table 4. MMGBSA prescoring results. 6BOY and 6HAX were two of the four systems that found lower enthalpy score poses. Statistics calculated from the last 10 ns of the trajectory (or best scoring trajectory if more than 1) corresponding to each candidate pose.

a. 5T35

candidate pose # to be rescored with heating	ΔH	$\Delta\Delta H$	combined score ($\Delta H + \Delta\Delta H$)
1 (found native pose)	-128.9 ± 0.8	-30.1 ± 0.6	-159.1 ± 1.3
4	-115.2 ± 0.8	-21.0 ± 0.5	-136.2 ± 1.2
2	-112.8 ± 0.7	-15.9 ± 0.5	-128.8 ± 1.1
5	-107.7 ± 0.7	-17.7 ± 0.5	-125.4 ± 1.1
11	-106.7 ± 1.1	-16.1 ± 0.8	-122.8 ± 1.8

b. 6SIS

candidate pose # to be rescored with heating	ΔH	$\Delta\Delta H$	combined score ($\Delta H + \Delta\Delta H$)
3 (found native pose)	-120.9 ± 0.7	-30.4 ± 0.5	-151.3 ± 1.1
1	-112.3 ± 0.8	-15.6 ± 0.7	-127.9 ± 1.4
4	-116.8 ± 0.9	-6.9 ± 0.5	-123.8 ± 1.3
11	-112.1 ± 0.7	-8.2 ± 0.6	-120.3 ± 1.1
9	-107.0 ± 0.7	-12.9 ± 0.5	-119.9 ± 1.1

c. 6BOY

candidate pose # to be rescored with heating	ΔH	$\Delta\Delta H$	combined score ($\Delta H + \Delta\Delta H$)
0	-89.2 ± 0.5	-22.9 ± 0.5	-112.1 ± 1.0
3 (found native pose)	-84.6 ± 0.5	-16.1 ± 0.3	-100.7 ± 0.7
5	-83.4 ± 0.6	-15.8 ± 0.5	-99.1 ± 1.0
4	-84.9 ± 0.5	-12.5 ± 0.3	-97.4 ± 0.7
2	-83.1 ± 0.5	-10.1 ± 0.4	-93.1 ± 0.8

d. 6HAX

candidate pose # to be rescored with heating	ΔH	$\Delta\Delta H$	combined score ($\Delta H + \Delta\Delta H$)
<u>ind.</u> bi. + dock.0	-106.3 ± 0.9	-12.5 ± 0.7	-118.8 ± 1.6
<u>ind.</u> bi. + dock.1	-102.0 ± 0.7	-12.5 ± 0.5	-114.5 ± 1.1
0 (found native pose)	-102.2 ± 0.9	-10.1 ± 0.7	-112.3 ± 1.5
4	-92.5 ± 0.6	-12.5 ± 0.5	-105.0 ± 1.1
1	-94.4 ± 0.5	-7.5 ± 0.3	-101.9 ± 0.6
2	-97.9 ± 0.5	-3.8 ± 0.2	-101.7 ± 0.2
11	-91.3 ± 1.0	-9.6 ± 0.8	-100.9 ± 1.7

While MMGBSA enthalpies gave 5T35 a good ranking by placing the crystal pose corresponding trajectory at the top, it failed to do the same for 6BOY. An RMSD-energy funnel plot (**Figure S9**) further corroborated that there were other configurations with lower binding enthalpy than the found crystal corresponding pose.

It is known that protein-protein bindings have large entropy contributions.⁵⁸ Entropy calculations were done on the 6BOY system (**Table S2**) on a different set of earlier results, and it did improve the rankings. In these earlier results the same pose (**Table S2**, trajectories 0.4 and 0.8) as in **Table 4c** pose0 was found, and it was lower in binding enthalpy than the found native pose. However, the entropy was extremely costly to calculate. The entropy calculation was by *nmode* in MMGBSA.py with a subset of frames. A pruned protein was used by partially adapting the method reported by Ryde et. al.⁵⁹ While adding the entropy improved the binding score of the found native poses somewhat, at the same time the size of the system made the calculations only marginally practical to perform, even with pruning.

Since entropy is essentially related to the probability of an energy configuration being reached, direct accurate calculation of entropy can only be obtained macroscopically via eq. (1). To work around this and focus on a single complex, thermodynamic integration (TI) and normal mode vibrational analysis (as in *nmode*) are some of the few approximations available. Deriving from the Gibbs equation, at constant temperature the entropic term disappears as in basic equation (5), where V is volume and P is pressure, a simplest case for TI.

$$dG = VdP \quad (5)$$

Practically, in TI for ligand binding free energy calculations, the situation is a bit more complicated and pressure is not the quantity being perturbed. Instead, a gradual alchemical perturbation is performed from ligand A to ligand B and the potential energy changes is integrated to obtain ΔG .^{60,61} Nmode analysis obtains absolute vibrational entropy via

$$S_{vib} = -R \ln \left(1 - e^{-\frac{h\nu_0}{kT}} \right) + \frac{N_A h \nu_0 e^{-\frac{h\nu_0}{kT}}}{T \left(1 - e^{-\frac{h\nu_0}{kT}} \right)} \quad (6)$$

where ν_0 is the frequency of each normal mode of the solute.⁶² Nonetheless, they are still limited in various ways. TI needs to be a slow process for small changes between two states, which is not suitable for large conformation changes. Calculating of normal mode vibrational frequency is extremely computationally demanding with a large margin of error, and it is more accurate for gas phase than solution phase.⁶³ So it is not surprising that entropy calculations are often neglected for in-silico drug discovery.

Whether due to inaccuracy in the MMGBSA itself or lack of entropy calculations, merely relying on the MMGBSA enthalpy score is not enough to identify the native ternary pose from non-native in some systems. But it is useful as a prescore since the native pose is still generally favored. **Tables 4a-d** show the top scoring poses. The full MMGBSA enthalpy results are listed at the bottom of the clustering results, **Figure 6** bottom, and **Figure S6-S8** bottom.

3.3 Final Rescoring by Heating-Accelerated Pose Change Trials (HAPOC)

One way of running heating studies for examining pose stabilities was simply running the system at a given higher temperature (e.g. 350K or 400K), which was attempted first. One inconvenience was that the stable pose should still be mostly bound at the chosen temperature, while non-stable poses mostly moved away, all within the short tens of ns. This would then involve testing a series of elevated temperatures (e.g. 310, 350, 380 K each etc.) with different MD trials, and this optimal temperature is expected to be different for each PROTAC ternary system. If directly evaluating the results from the same series of elevated temperatures across all ternary systems, this will help skip the optimal temperature determination step, and total residence time across the whole temperature series can then be a score for each pose. However, weighing the scores for different temperatures should be included, which would introduce an extra parameter. So eventually, we decided to cover the whole temperature range in 1

MD trial by heating the system from 310K to 410K, as presented. Since the residence time fluctuates among runs, multiple repeats were attempted. With only 10 short replicates for each pose, it was sufficient for ranking the poses and making the results of native vs non-native statistically significant in most systems.

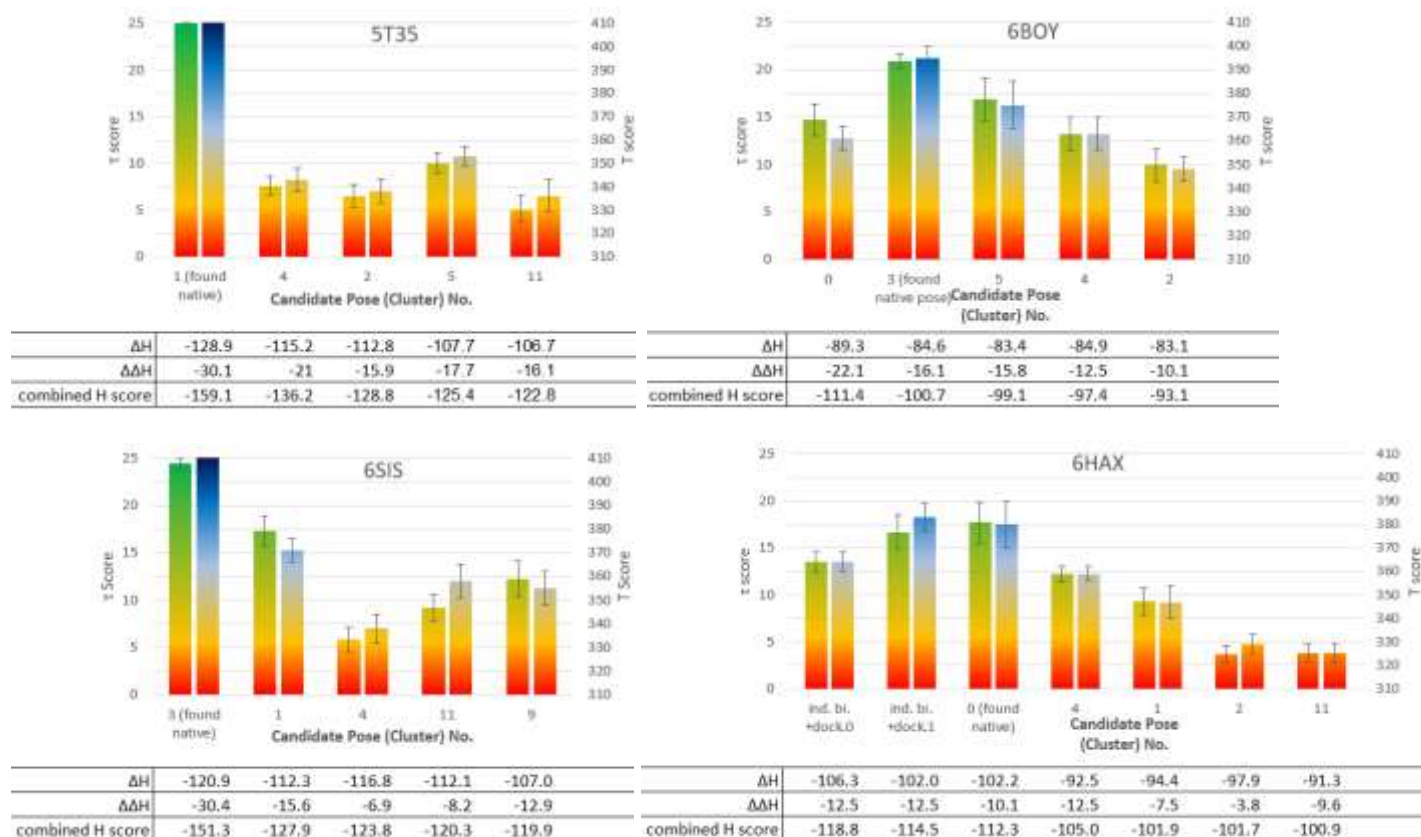


Figure 7. Heating rescore results of top MMGBSA scoring poses. Left bar shows time score (τ), right bar shows temperature score (T). Error bars show s.e.m. Corresponding MMGBSA energy scores for each pose shown at bottom of each graph, the same values as in **Tables 4a-d**.

Both the τ and T scores gave very similar results as expected, with the τ score marginally better. The T score was included so that we can estimate at what temperatures the pose changes are occurring. Even though the heated simulation trials already returned good results, we still wanted to briefly investigate some behaviors of the OPC water at this somewhat elevated temperature by looking at its density. From one of the heating trials, the system density over temperature was plotted and compared to the density of real-world water over temperature. Their trends matched very well (**Figure S10**). At temperatures above the real-world boiling temperature, the density continued the same smooth trend with no surges, indicating the water still solvating the solute.

Another more important concern is protein structure stability during the heating. If the protein structure itself is changing during the heating trials, results collected in those scenarios may include false negatives. One of the easily observable signs for the change of protein structure during heating is whether or not the RMSD of the aligned protein is increasing (to $> 3 \text{ \AA}$). Modifications in those scenarios may include protein structure restraints during the heating, however that is not within the scope of this current study. Fortunately, the protein structures were stable without restraints for all the poses we tested.

We then seek to provide a method to address one of the most important questions on the PROTAC-ternary pose prediction. In the absence of a known ternary crystal structure, how do we know the correct ternary pose has been reached? A relative score is insufficient because it may vary across different systems like the MMGBSA scores in **Table 4**, even if the relative score is accurate within each ternary system. Results of the scoring by heating-accelerated pose change within the same ternary system and across different systems indicated that the four crystal ternary poses were all observed to be stable up to at least 17.5 in terms of the τ score and 380 in terms of the T score, while non-native poses generally were not. We then decided to measure the τ and T scores for all the available VHL and CRBN-containing ternary crystal structures.⁶⁴

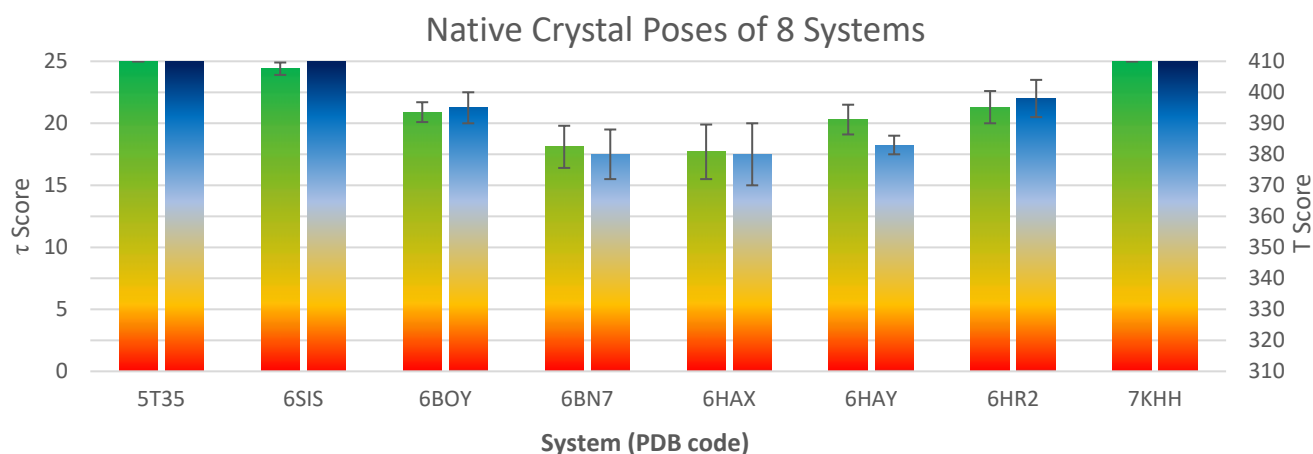


Figure 8. Scores of all resolved crystal structures. 5T35, 6SIS, 6BOY, and 6HAX scores from found native pose in our study. 6BN7, 6HAY, 6HR2 and 7KHH scores measured from crystal PDB pose. Error bars show s.e.m.

All the different crystal poses were observed to be stable with at least a 17.5 τ score and 380 T score (**Figure 8**). From these results we provide an empirical criteria. For the evaluation of a given PROTAC-mediated ternary pose using the heating-accelerated pose change (HAPOC) method and conditions in this study, the pose that is stable to at least 17.5 τ score and 380 T score over an average of 10 or more runs is very likely the correct native pose. The non-native poses generally moved away much sooner. One can constantly search for more candidate poses for heating reevaluation until this empirical criteria is reached. Run conditions are described in the methods part and input scripts along with all structures are available on Zenodo⁶⁵.

3.4 Testing on independently generated poses

The conformation search used in this study only covers a limited scope and is far from a global search. The process of reaching the native within a short amount of time is quite dependent on docking to get a good starting point, though partial success was obtained even with independent binary crystals as the inputs. The final HAPOC scoring is exceptionally successful in distinguishing native from non-native ternary poses and can be used independently from our MD-based pose sampling. Just like using the crystal coordinates as inputs as demonstrated above, PROsetta poses were evaluated.^{10, 54}

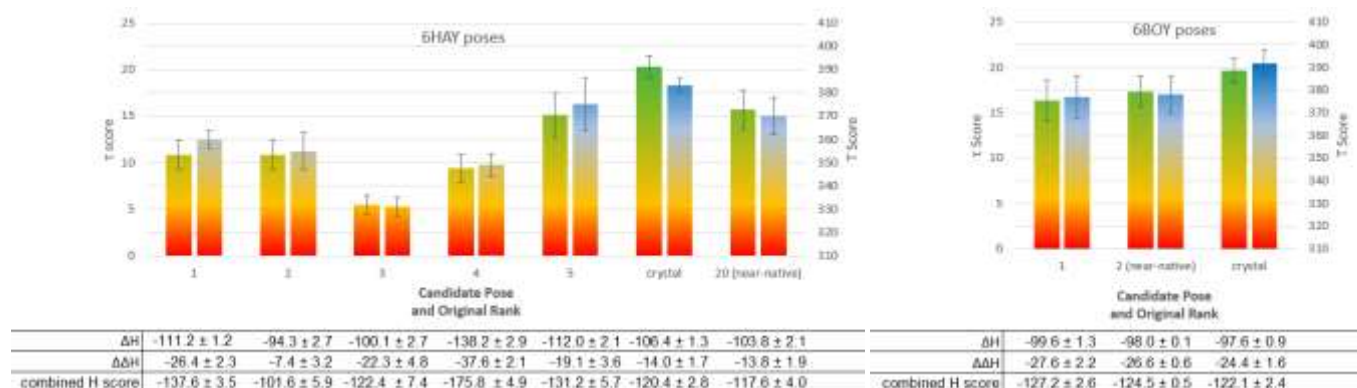


Figure 9. HAPOC rescore of top candidate poses from PROsettaC. Pose no. indicate cluster population rank. Error bars of τ and T score show s.e.m. over 10 trajectories. Error bars of MMGBSA enthalpies show s.e.m. from the 3 averages of 3 trajectories.

In the 310 K trajectories for MMGBSA evaluations, for 6BOY pose 1, the BRD4 ligand warhead was observed to partially dissociate from its pocket in some runs unexpectedly. This unexpected dissociation was not observed for pose 2 or the crystal one. It could be possible that in a poor pose, strain is pulling the ligand out of its pocket (while remaining interactions are stronger than the ligand-BRD4 interactions). Even though this observation by itself should be sufficient to consider 6BOY pose 1 as a poor, non-native one, the MMGBSA energies from runs with the ligand properly bound were still collected for comparing with the native and near-native.

Originally, the near-native pose for 6BOY was ranked 2nd (out of 22), and 6HAY was ranked 20th (out of 35). Applying MMGBSA enthalpy calculations to the poses, we observed some improvement, though some of these top poses may simply be enthalpically favored ones. In such case, it will not be surprising that they will rank ahead of the native pose based on static energy techniques to distinguish favorable poses. The HAPOC method shown here mostly corrected the pose ranks. From the results based on HAPOC (**Figure 9**), one can see that all of the non-native poses originally ranked at the top scored less favorable than the near-native pose based on the τ score, excluding two statistical ties not distinguished within 10 runs. The native crystal pose clearly scored the best. We noticed the scoring difference between the near-native and the crystal. Even though the structure of their near-native poses had a reasonable C- α RMSD to the crystal pose, there were actually less protein-protein and protein-PROTAC contacts in the near-native pose (**Table 5**). This seemingly subtle difference is well reflected in the HAPOC scoring as expected. A pose isn't ideal enough if the side chains are not within proper intermolecular contacts. The crystal structure is one example of having good contacts.

Table 5. Number of contact pairs in crystal structures and found natives, defined by a Van der Waals radii overlap of > -0.4 Å.

Molecule Pair	CRBN-BRD4	BRD4-PROTAC	PROTAC-CRBN
6BOY Crystal	84	61	83
6BOY near-native (pose 2, Figure 9)	65	29	69
6BOY found native from own study with MD search	63	44	61
Molecule Pair	VHL-BD	VHL-PROTAC	PROTAC-BD
6HAY Crystal	45	85	125
6HAY near-native (pose 20, Figure 9)	43	84	103

4. Conclusion

The prediction of an accurate ternary pose is essential for structure-based PROTAC design when the structure of the ternary complex is not available. This study is intended to predict PROTAC ternary structures *in silico* in the absence of a known ternary co-crystal structure. We have developed a novel way to score the PROTAC-mediated ternary complexes by running heating-accelerated pose change trials of candidate poses with explicit solvent MD, and measure its average pose-residence time. It is highly successful in the systems available for testing. While the candidate poses in this study were generated by a selective local MD conformation search, the scoring is not limited to MD generated poses. Instead, they can be used independently as we have successfully demonstrated. The results generated from this study represent a substantial improvement over the limited number of previous attempts on PROTAC ternary complex predictions, which do not provide scoring or only indirectly rely on enthalpy-based static scoring techniques.

Our method presented here considered both enthalpic and entropic effects. Entropy is especially important when protein-protein interactions are involved. It therefore provides a much more complete picture based on the success in all the systems we examined here. The process is relatively straightforward, self-intuitive, and easy to use for medicinal chemists without extensive knowledge in MD. Even with repeated heating replicates, the calculations are much faster than normal mode entropy calculations for systems of this size with a huge interface. While MMGBSA enthalpy is still used for prescoring, the final HAPOC rescore is run in explicit solvent without involving a choice of another set of GB parameters.

ASSOCIATED CONTENT

Supporting Information.

Step-by-step model building procedures with UCSF Chimera.

Additional data and results.

Provided in Zenodo database.

Input structures and scripts used throughout the whole study, as well as output files and results. doi.org/10.5281/zenodo.5645771

AUTHOR INFORMATION

Corresponding Authors

Spencer S. Ericksen

E-mail: ssericksen@wisc.edu

Weiping Tang

E-mail: weiping.tang@wisc.edu

ACKNOWLEDGMENT

We would like to thank Dr. Xuhui Huang from the University of Wisconsin-Madison for valuable discussions and inputs. Support for this research was provided by the University of Wisconsin - Madison Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation through a UW2020 award.

With the exception of the pretest results generated on a workstation with a single GPU, the majority of the computationally demanding work was performed using the computing resources and assistance of the UW-Madison Center for High Throughput Computing (CHTC) in the Department of Computer Sciences. The CHTC is supported by UW-Madison, the Advanced Computing Initiative, the Wisconsin Alumni Research Foundation, the Wisconsin Institutes for Discovery, and the National Science Foundation, and is an active member of the Open Science Grid, which is supported by the National Science Foundation and the U.S. Department of Energy's Office of Science. Molecular graphics images were produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco with support from NIH P41-GM103311.

REFERENCES

1. Bondeson, D. P.; Mares, A.; Smith, I. E. D.; Ko, E.; Campos, S.; Miah, A. H.; Mulholland, K. E.; Routly, N.; Buckley, D. L.; Gustafson, J. L.; Zinn, N.; Grandi, P.; Shimamura, S.; Bergamini, G.; Faelth-Savitski, M.; Bantscheff, M.; Cox, C.; Gordon, D. A.; Willard, R. R.; Flanagan, J. J.; Casillas, L. N.; Votta, B. J.; den Besten, W.; Famm, K.; Kruidenier, L.; Carter, P. S.; Harling, J. D.; Churcher, I.; Crews, C. M., Catalytic in vivo protein knockdown by small-molecule PROTACs. *Nat. Chem. Biol.* **2015**, *11* (8), 611-617.
2. Winter, G. E.; Buckley, D. L.; Paulk, J.; Roberts, J. M.; Souza, A.; Dhe-Paganon, S.; Bradner, J. E., Phthalimide conjugation as a strategy for in vivo target protein degradation. *Science* **2015**, *348* (6241), 1376.
3. Burslem, G. M.; Crews, C. M., Proteolysis-Targeting Chimeras as Therapeutics and Tools for Biological Discovery. *Cell* **2020**, *181* (1), 102-114.
4. Salami, J.; Crews, C. M., Waste disposal—An attractive strategy for cancer therapy. *Science* **2017**, *355* (6330), 1163.
5. Lai, A. C.; Crews, C. M., Induced protein degradation: an emerging drug discovery paradigm. *Nat. Rev. Drug Discovery* **2017**, *16* (2), 101-114.
6. Wan, Y.; Yan, C.; Gao, H.; Liu, T., Small-molecule PROTACs: novel agents for cancer therapy. *Future Med. Chem.* **2020**, *12* (10), 915-938.
7. Konstantinidou, M.; Li, J.; Zhang, B.; Wang, Z.; Shaabani, S.; Ter Brake, F.; Essa, K.; Dömling, A., PROTACs— a game-changing technology. *Expert Opin. Drug Discov.* **2019**, *14* (12), 1255-1268.
8. 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.
9. Drummond, M. L.; Henry, A.; Li, H.; Williams, C. I., Improved Accuracy for Modeling PROTAC-Mediated Ternary Complex Formation and Targeted Protein Degradation via New In Silico Methodologies. *J. Chem. Inf. Model.* **2020**, *60* (10), 5234-5254.
10. Zaidman, D.; Prilusky, J.; London, N., PROsettaC: Rosetta Based Modeling of PROTAC Mediated Ternary Complexes. *J. Chem. Inf. Model.* **2020**.
11. Araki, M.; Matsumoto, S.; Bekker, G.-J.; Isaka, Y.; Sagae, Y.; Kamiya, N.; Okuno, Y., Exploring ligand binding pathways on proteins using hypersound-accelerated molecular dynamics. *Nat. Commun.* **2021**, *12* (1), 2793.
12. Salmaso, V.; Moro, S., Bridging Molecular Docking to Molecular Dynamics in Exploring Ligand-Protein Recognition Process: An Overview. *Front. Pharmacol.* **2018**, *9* (923).
13. Mayor, U.; Johnson, C. M.; Daggett, V.; Fersht, A. R., Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. *Proc. Natl. Acad. Sci.* **2000**, *97* (25), 13518.
14. Mayor, U.; Gyuosh, N. R.; Johnson, C. M.; Grossmann, J. G.; Sato, S.; Jas, G. S.; Freund, S. M. V.; Alonso, D. O. V.; Daggett, V.; Fersht, A. R., The complete folding pathway of a protein from nanoseconds to microseconds. *Nature* **2003**, *421* (6925), 863-867.
15. Rocco, A. G.; Mollica, L.; Ricchiuto, P.; Baptista, A. M.; Gianazza, E.; Eberini, I., Characterization of the Protein Unfolding Processes Induced by Urea and Temperature. *Biophys. J.* **2008**, *94* (6), 2241-2251.
16. Settanni, G.; Fersht, A. R., High Temperature Unfolding Simulations of the TRPZ1 Peptide. *Biophys. J.* **2008**, *94* (11), 4444-4453.
17. Pierce, L. C. T.; Salomon-Ferrer, R.; Augusto F de Oliveira, C.; McCammon, J. A.; Walker, R. C., Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. *J. Chem. Theory Comput.* **2012**, *8* (9), 2997-3002.
18. Tiwary, P.; Limongelli, V.; Salvalaglio, M.; Parrinello, M., Kinetics of protein–ligand unbinding: Predicting pathways, rates, and rate-limiting steps. *Proc. Natl. Acad. Sci.* **2015**, *112* (5), E386.
19. Mollica, L.; Decherchi, S.; Zia, S. R.; Gaspari, R.; Cavalli, A.; Rocchia, W., Kinetics of protein-ligand unbinding via smoothed potential molecular dynamics simulations. *Scientific Reports* **2015**, *5* (1), 11539.

20. Miao, Y.; Feher, V. A.; McCammon, J. A., Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *J. Chem. Theory Comput.* **2015**, *11* (8), 3584-3595.
21. Deb, I.; Frank, A. T., Accelerating Rare Dissociative Processes in Biomolecules Using Selectively Scaled MD Simulations. *J. Chem. Theory Comput.* **2019**, *15* (11), 5817-5828.
22. Wolf, S.; Lickert, B.; Bray, S.; Stock, G., Multisecond ligand dissociation dynamics from atomistic simulations. *Nat. Commun.* **2020**, *11* (1), 2918.
23. Kozakov, D.; Hall, D. R.; Xia, B.; Porter, K. A.; Padhorny, D.; Yueh, C.; Beglov, D.; Vajda, S., The ClusPro web server for protein-protein docking. *Nat. Protoc.* **2017**, *12* (2), 255-278.
24. Vajda, S.; Yueh, C.; Beglov, D.; Bohnuud, T.; Mottarella, S. E.; Xia, B.; Hall, D. R.; Kozakov, D., New additions to the ClusPro server motivated by CAPRI. *Proteins: Struct., Funct., Bioinf.* **2017**, *85* (3), 435-444.
25. Kozakov, D.; Beglov, D.; Bohnuud, T.; Mottarella, S. E.; Xia, B.; Hall, D. R.; Vajda, S., How good is automated protein docking? *Proteins: Struct., Funct., Bioinf.* **2013**, *81* (12), 2159-2166.
26. Gadd, M. S.; Testa, A.; Lucas, X.; Chan, K.-H.; Chen, W.; Lamont, D. J.; Zengerle, M.; Ciulli, A., Structural basis of PROTAC cooperative recognition for selective protein degradation. *Nat. Chem. Biol.* **2017**, *13* (5), 514-521.
27. Testa, A.; Hughes, S. J.; Lucas, X.; Wright, J. E.; Ciulli, A., Structure-Based Design of a Macrocyclic PROTAC. *Angew. Chem. Int. Ed.* **2020**, *59* (4), 1727-1734.
28. Nowak, R. P.; DeAngelo, S. L.; Buckley, D.; He, Z.; Donovan, K. A.; An, J.; Safaee, N.; Jedrychowski, M. P.; Ponthier, C. M.; Ishoey, M.; Zhang, T.; Mancias, J. D.; Gray, N. S.; Bradner, J. E.; Fischer, E. S., Plasticity in binding confers selectivity in ligand-induced protein degradation. *Nat. Chem. Biol.* **2018**, *14* (7), 706-714.
29. Farnaby, W.; Koegl, M.; Roy, M. J.; Whitworth, C.; Diers, E.; Trainor, N.; Zollman, D.; Steurer, S.; Karolyi-Oezguer, J.; Riedmueller, C.; Gmaschitz, T.; Wachter, J.; Dank, C.; Galant, M.; Sharps, B.; Rumpel, K.; Traxler, E.; Gerstberger, T.; Schnitzer, R.; Petermann, O.; Greb, P.; Weinstabl, H.; Bader, G.; Zoephel, A.; Weiss-Puxbaum, A.; Ehrenhöfer-Wölfer, K.; Wöhrle, S.; Boehmelt, G.; Rinnenthal, J.; Arnhof, H.; Wiechens, N.; Wu, M.-Y.; Owen-Hughes, T.; Ettmayer, P.; Pearson, M.; McConnell, D. B.; Ciulli, A., BAF complex vulnerabilities in cancer demonstrated via structure-based PROTAC design. *Nat. Chem. Biol.* **2019**, *15* (7), 672-680.
30. Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T., Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **2010**, *50* (4), 572-584.
31. OMEGA 4.1.1.1: OpenEye Scientific Software, Santa Fe, NM. <http://www.eyesopen.com>.
32. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605-12.
33. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A., Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Model.* **2006**, *25* (2), 247-260.
34. Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguet, L.; Huang, H.; Miguels, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; Simmerling, C., ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput.* **2020**, *16* (1), 528-552.
35. Izadi, S.; Anandakrishnan, R.; Onufriev, A. V., Building Water Models: A Different Approach. *The Journal of Physical Chemistry Letters* **2014**, *5* (21), 3863-3871.
36. Joung, I. S.; Cheatham, T. E., Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem.* **2008**, *112* (30), 9020-9041.
37. Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E., Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J. Chem. Theory Comput.* **2015**, *11* (4), 1864-1874.
38. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696-3713.

39. 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* (2), 926-935.
40. Onufriev, A.; Bashford, D.; Case, D. A., Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Struct., Funct., Bioinf.* **2004**, *55* (2), 383-394.
41. Knapp, B.; Ospina, L.; Deane, C. M., Avoiding False Positive Conclusions in Molecular Simulation: The Importance of Replicas. *J. Chem. Theory Comput.* **2018**, *14* (12), 6127-6138.
42. Le Grand, S.; Götz, A. W.; Walker, R. C., SPFP: Speed without compromise—A mixed precision model for GPU accelerated molecular dynamics simulations. *Comput. Phys. Commun.* **2013**, *184* (2), 374-380.
43. Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C., Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9* (9), 3878-3888.
44. Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C., Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **2012**, *8* (5), 1542-1555.
45. D.A. Case, K. B., I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro,; T.A. Darden, R. E. D., G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R Harris, S. Izadi, S.A. Izzi, mailov, K. K., A. Kovalenko, R. Krasny, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu,; T. Luchko, R. L., V. Man, K.M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, A. Onufriev, F.; Pan, S. P., R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, N.R.; Skrynnikov, J. S., J. Swails, R.C. Walker, J. Wang, L. Wilson, R.M. Wolf, X. Wu, Y. Xiong, Y. Xue,; Kollman, D. M. Y. a. P. A. *AMBER 2020*, University of California, San Francisco., 2020.
46. Carrascal, N.; Rizzo, R. C., Calculation of binding free energies for non-zinc chelating pyrimidine dicarboxamide inhibitors with MMP-13. *Bioorg. Med. Chem. Lett.* **2009**, *19* (1), 47-50.
47. Roe, D. R.; Cheatham, T. E., PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9* (7), 3084-3095.
48. Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T., Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112* (16), 6127-6129.
49. Hou, T.; Wang, J.; Li, Y.; Wang, W., Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2011**, *51* (1), 69-82.
50. Wang, E.; Sun, H.; Wang, J.; Wang, Z.; Liu, H.; Zhang, J. Z. H.; Hou, T., End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. *Chem. Rev. (Washington, DC, U. S.)* **2019**, *119* (16), 9478-9508.
51. Onufriev, A.; Case, D. A.; Bashford, D., Effective Born radii in the generalized Born approximation: the importance of being perfect. *J. Comput. Chem.* **2002**, *23* (14), 1297-304.
52. Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E., MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **2012**, *8* (9), 3314-3321.
53. Roy, M. J.; Winkler, S.; Hughes, S. J.; Whitworth, C.; Galant, M.; Farnaby, W.; Rumpel, K.; Ciulli, A., SPR-Measured Dissociation Kinetics of PROTAC Ternary Complexes Influence Target Degradation Rate. *ACS Chem. Biol.* **2019**, *14* (3), 361-368.
54. Nir, Z. D. L., Results files of PROsettaC for the following PDBs: 5T35, 6BOY, 6BN7, 6BN8, 6BN9, 6BNB, 6HAX, 6HAY, 6HR2, 6SIS. 2020.
55. Weng, G.; Li, D.; Kang, Y.; Hou, T., Integrative Modeling of PROTAC-Mediated Ternary Complexes. *J. Med. Chem.* **2021**.
56. Izadi, S.; Anandakrishnan, R.; Onufriev, A. V., Implicit Solvent Model for Million-Atom Atomistic Simulations: Insights into the Organization of 30-nm Chromatin Fiber. *J. Chem. Theory Comput.* **2016**, *12* (12), 5946-5959.

57. Nguyen, H.; Maier, J.; Huang, H.; Perrone, V.; Simmerling, C., Folding Simulations for Proteins with Diverse Topologies Are Accessible in Days with a Physics-Based Force Field and Implicit Solvent. *J. Am. Chem. Soc.* **2014**, *136* (40), 13959-13962.
58. Sun, Z.; Yan, Y. N.; Yang, M.; Zhang, J. Z. H., Interaction entropy for protein-protein binding. *J. Chem. Phys.* **2017**, *146* (12), 124124.
59. Genheden, S.; Kuhn, O.; Mikulskis, P.; Hoffmann, D.; Ryde, U., The Normal-Mode Entropy in the MM/GBSA Method: Effect of System Truncation, Buffer Region, and Dielectric Constant. *J. Chem. Inf. Model.* **2012**, *52* (8), 2079-2088.
60. Lawrenz, M.; Baron, R.; McCammon, J. A., Independent-Trajectories Thermodynamic-Integration Free-Energy Changes for Biomolecular Systems: Determinants of H5N1 Avian Influenza Virus Neuraminidase Inhibition by Peramivir. *J. Chem. Theory Comput.* **2009**, *5* (4), 1106-1116.
61. Steinbrecher, T.; Case, D. A.; Labahn, A., Free energy calculations on the binding of novel thiolactomycin derivatives to E. coli fatty acid synthase I. *Bioorg. Med. Chem.* **2012**, *20* (11), 3446-3453.
62. McQuarrie, D. A., *Statistical thermodynamics*. Harper and Row: New York, 1973.
63. Genheden, S.; Ryde, U., The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* **2015**, *10* (5), 449-461.
64. Dragovich, P. S.; Pillow, T. H.; Blake, R. A.; Sadowsky, J. D.; Adaligil, E.; Adhikari, P.; Chen, J.; Corr, N.; dela Cruz-Chuh, J.; Del Rosario, G.; Fullerton, A.; Hartman, S. J.; Jiang, F.; Kaufman, S.; Kleinheinz, T.; Kozak, K. R.; Liu, L.; Lu, Y.; Mulvihill, M. M.; Murray, J. M.; O'Donohue, A.; Rowntree, R. K.; Sawyer, W. S.; Staben, L. R.; Wai, J.; Wang, J.; Wei, B.; Wei, W.; Xu, Z.; Yao, H.; Yu, S.-F.; Zhang, D.; Zhang, H.; Zhang, S.; Zhao, Y.; Zhou, H.; Zhu, X., Antibody-Mediated Delivery of Chimeric BRD4 Degradators. Part 2: Improvement of In Vitro Antiproliferation Activity and In Vivo Antitumor Efficacy. *J. Med. Chem.* **2021**, *64* (5), 2576-2607.
65. doi.org/10.5281/zenodo.5645771.