Computational Modeling of Ternary Complex Formation and Targeted Protein Degradation Mediated by Molecular Glues

Jesus A Izaguirre,* Yujie Wu,* Zachary McDargh, Timothy Palpant, Asghar M Razavi, Fabio Trovato, Cheryl Koh, and Huafeng Xu*

Atommap Corporation, New York, NY 10017, USA

E-mail: jesus@atommap.com; yujie@atommap.com; huafeng@atommap.com

Abstract

We introduce GlueMap, a computational toolkit to discover and optimize molecular glues. GlueMap integrates structural and pharmacodynamic modeling, molecular dynamics (MD), and machine learning (ML) for modeling targeted protein degradation (TPD) mediated by molecular glues (MGs). GlueMap accurately models the structural ensemble of ternary complexes, predicts their thermodynamic stabilities, and evaluates downstream effects such as ubiquitination and degradation efficiency. We validated GlueMap's capabilities through two case studies: CRBN-dependent degradation of GSPT1 and DDB1-dependent degradation of CDK12. For CRBN-GSPT1, GlueMap successfully recovered known molecular glue degraders and identified novel candidates through prospective virtual screening, leveraging ternary complexes derived from extensive MD simulation. In the DDB1-CDK12 system, we observed strong correlations between ternary complex stability, quantified by binding free energy calculations, and degradation efficiency, though notable exceptions highlighted the importance of considering structural dynamics of the ternary complex in the context of the cullin ring

ligase. Through a supervised variational autoencoder (VAE) model combined with attention-based regression, GlueMap demonstrates high accuracy in predicting degradation potency from structural features. By integrating structural, thermodynamic, and ubiquitination metrics, GlueMap provides a multi-criteria approach to rational MG design, distinguishing it from traditional screening methods that rely primarily on docking scores. These results illustrate that GlueMap may provide mechanistic insights and accelerate molecular glue discovery and optimization across diverse protein targets.

Introduction

Molecular glues (MGs) are therapeutics that overcome the limitations of heterobifunctional degraders, particularly due to their smaller molecular weight and potentially superior physicochemical properties.[1] Most MG discovery efforts rely on phenotypic screening, followed by target identification and mechanism-of-action validation.[2,3]

The rational design of MG degraders follows two primary approaches: (1) predicting how candidate MGs modify protein surfaces to induce specific protein-protein interfaces (PPIs), or (2) predicting PPIs that form in solution and evaluating how MGs stabilize these ternary complexes. We follow this latter approach. For comprehensive resources: Huan *et al.*[4] provide an in-depth review of weak PPI identification (encounter complexes); Kozicka *et al.*[5] offer a comprehensive MG dataset; and Schreiber[6] discusses induced proximity degraders. Konstantinidou and Arkin[7] provide an extensive review of MGs discovered between 2014 and 2023.

Recent advancements in machine learning (ML) have transformed the field. Prael *et al.*[8] developed methods to predict CRBN-based molecular glues by screening 161 targets against up to 25,000 compounds. Monte Rosa scientists[9] expanded this approach by computationally mining the human proteome, predicting over 1,400 targets degradable by CRBN MGs. More recent developments suggest potential integration with advanced models like AlphaFold3[10]

or BOLTZ-1[11] to enhance predictive accuracy. A preprint[12] introduced an approach similar to these models for predicting MG-mediated ternary complexes.

Our approach utilizes physics-based simulations to model conformational landscapes and design MGs that can stabilize weak basal interactions.[13] The framework integrates multiple components, including machine learning models that demonstrate high accuracy in predicting degradation potency from structural features. Key elements include encounter and ternary complex prediction, molecular dynamics (MD) simulations for generating conformational landscapes (Markov State Models[14] and Non-Markovian models[15]), binding affinity predictions for binary and ternary complexes, ubiquitin-lysine distance predictions, and neural network models to derive experimental observables from MD data. These tools are integrated into GlueMap, a virtual screening framework, building on our prior work.[16,17] Alternative computational approaches that primarily rely on docking can serve as starting points for our modeling.[18]

Overview of GlueMap Workflow

In GlueMap (Fig. 1), we integrate protein structure prediction, protein-protein docking, and advanced simulations to predict structural ensembles of encounter complexes between POIs and effector proteins, as well as ternary complexes involving POIs, molecular glues, and effector proteins.

The workflow begins with protein-protein docking to identify plausible interaction surfaces. While we use MOE's protein-protein docking tool, alternative methods can also be employed. Following docking, all-atom MD simulations (1 to 2 μs) are conducted for each predicted PPI. These simulations refine the interaction interfaces using physics-based force fields, providing a more accurate representation of the molecular interactions (Fig. 1a).

Next, a PPI free energy landscape is constructed by integrating the MD simulation data using Markov State Modeling (MSM) and related techniques. This approach enables the identification and selection of metastable PPIs for further analysis (Fig. 1b). PPIs are

selected by interfacial pocket druggability and potential for ubiquitination.

We then perform small-molecule virtual screening against an ensemble of PPIs coming from the MSM. Pharmacophores are defined for each metastable PPI conformation to identify potential molecular glues, followed by rapid pharmacophore screening of virtual compound libraries, such as those available from Enamine (Fig. 1c).

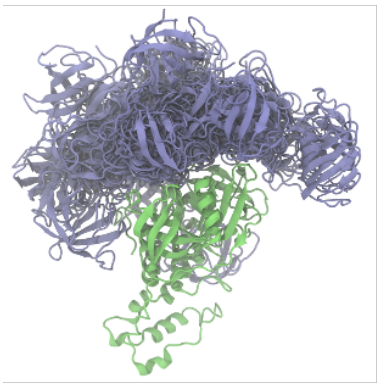
To score the screened compounds, short MD simulations (10 ns each) are conducted on candidates filtered using pharmacophores. Electrostatic and van der Waals interactions are calculated for each protein residue, enabling the identification of key binding residues and pockets. Based on these results, compounds are selected for longer MD simulations (200 ns each) (Fig. 1d,e). The final compound selection considers structural stability, thermodynamic stability—predicted by relative binding free energy (RBFE) calculations—and ubiquitination potential of the corresponding ternary complex to identify promising molecular glue candidates (Fig. 1f).

As demonstrated next by the results of applying GlueMap to various projects, the workflow should be tailored to meet the specific needs of each project. In some cases, only certain modules may be utilized, while in others, the generic steps outlined here may be replaced with more suitable alternatives. The modular design of GlueMap facilitates such customization.

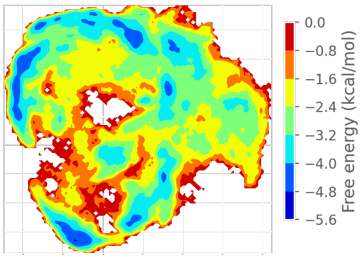
Results

Case Study: DDB1-dependent CDK12-Cyclin K degradation

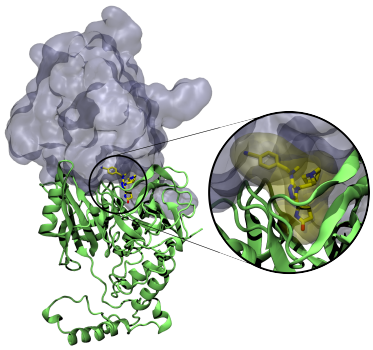
Kozicka *et al.*[5] report a rich dataset of molecular glues that degrade CDK12 and Cyclin K through DDB1. This dataset contains crystal structures, experimental binary and ternary affinities, and degradation results. We use this dataset to assess various capabilities of GlueMap.



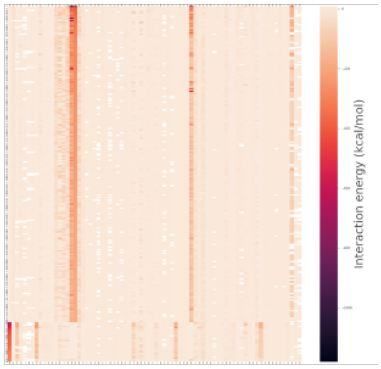
(a) Protein-Protein interface (PPI) docking followed by long MD simulations.



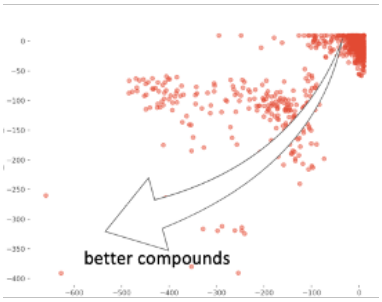
(b) Construction of a PPI free energy landscape.



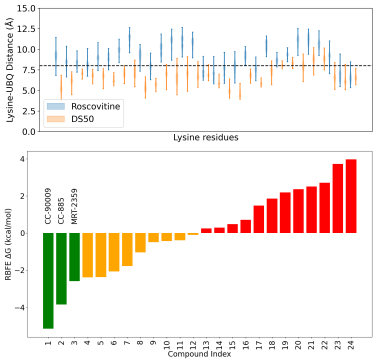
(c) Small molecule virtual screening against PPIs of interest from the conformational ensemble.



(d) Electrostatics and Van der Waals scoring.



(e) Initial Compound selection for longer MD simulations.



(f) Final compound selection.

Figure 1: Overview of the computational process for **GlueMap**: A. Protein-Protein interface (PPI) docking and simulations. B. PPI free energy landscape. C. Compound docking. D. Compound scoring. E. Initial compound selection. F. Final compound selection, using structural stability of the ligand in the pocket, ubiquitination profiles and ternary complex stabilities (binding free energies).

Prediction of ternary complex structures

We modeled ternary complexes between DDB1 and CDK12, mediated by analogs of the CDK12 inhibitor CR8. We start from docked CDK12-DDB1 poses, run HREMD (see Methods) starting from the resulting poses, cluster the conformations sampled in the simulations, and choose cluster centers of the highest populated clusters as our models. The resulting models are in good agreement with the crystal structures, which are *not* used as input to the simulations (Fig. 2).

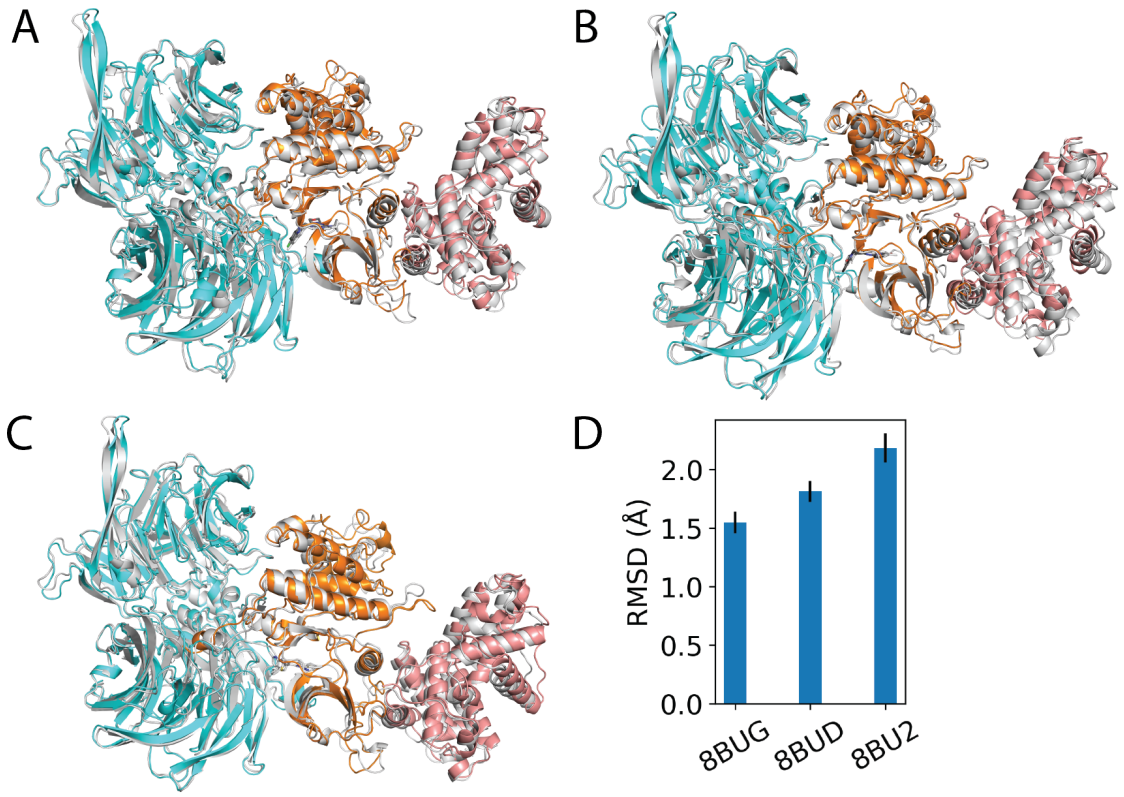


Figure 2: Comparison of modeled ternary complex structures with crystal structures for DDB1-CDK12-CYCK (A: PDB ID 8BU2; B: PDB ID 8BUD; C: PDB ID 8BUG). (D) The crystal structures belong to the highest-population conformational clusters identified through GlueMap. The top clusters had average E3 ligase-aligned ligand RMSDs of ∼2 Å.

Predicting stability of DDB1-MG-CDK12 ternary complexes using relative binding free energy calculations

Here, we applied relative binding free energy (RBFE) calculations[19–22] to predict the thermodynamic stabilities of the binary complexes of MG with CDK12 and of the ternary complexes (Fig. 3) of MG, DDB1 and CDK12, obtaining results in good agreement with experimental values (see Methods and Supplementary Information) for both.

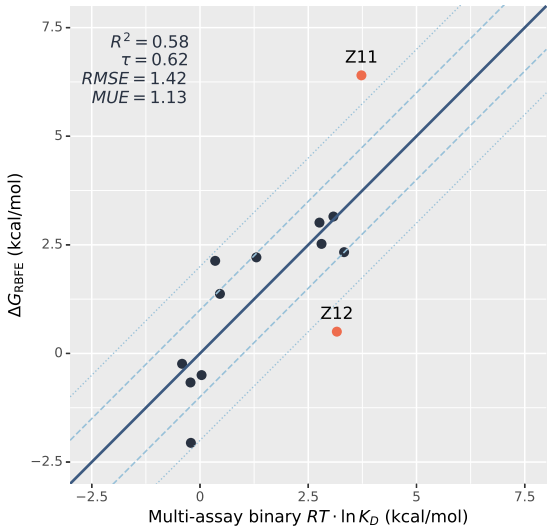
RBFE calculations consider the following set of binding reactions and alchemical transformations:

$$\begin{aligned} X + M_a &\xrightarrow{dG_X^a} X \cdot M_a \\ X + M_b &\xrightarrow{dG_X^b} X \cdot M_b \\ M_a &\xrightarrow{dG_{sol}^{a\to b}} M_b \\ X \cdot M_a &\xrightarrow{dG_X^{a\to b}} X \cdot M_b \end{aligned} \tag{1}$$

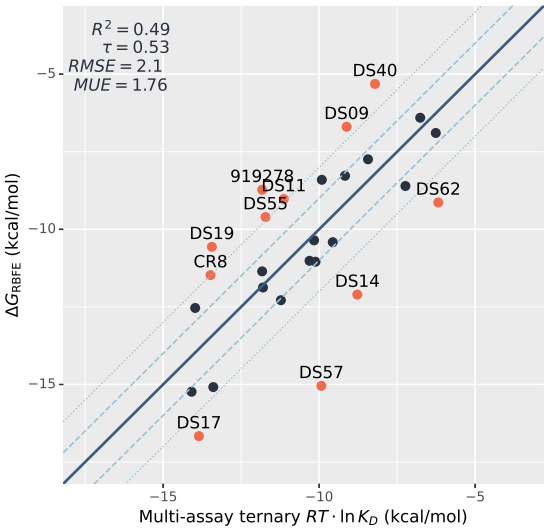
where $X = P, E, P \cdot E$ is the receptor and M_a and M_b are the two glue molecules to calculate the binding free energy difference for. The relative binding free energies can be calculated using the following equations:

$$\begin{aligned} ddG_X^{a\to b} &= dG_X^b - dG_X^a \\ &= dG_X^{a\to b} - dG_{sol}^{a\to b} \end{aligned} \tag{2}$$

Note that there are two binary RBFEs: $ddG_P^{a\to b}$ and $ddG_E^{a\to b}$ for P and E, respectively. Our workflow allows the user to choose to calculate either or both of them.



(a) The predicted binding free energies for binary complexes of molecular glues binding to CDK12.



(b) The predicted binding free energies for ternary complexes of molecular glues binding to DDB1-CDK12.

Figure 3: The predicted binding free energies for binary and ternary complexes compared to the experimental values for DDB1-CDK12.

Prediction of ubiquitination and degradation from DDB1-MG-CDK12 ternary complex ensembles

We extended our MD simulations to include the ubiquitin transfer complex, or cullin ring ligase (CRL), enabling us to explore how different molecular glues influence the structural ensembles and the accessibility of ubiquitin to lysine residues on CDK12. These simulations were performed by combining models of the DDB1-MG-CDK12 ternary complex with the crystal structure of the DDB1-Cul4A-Rbx1 complex (PDB ID 2HYE)[23] to obtain models of the full CRL complex induced by each MG.

Experimental data reveal that ternary complex stability correlates well with degradation efficiency for the DDB1-MG-CDK12 system,[5] but there are notable outliers. For instance, we compared two MGs, Roscovitine and DS50. Although the ternary complexes they form exhibit similar thermal stability, Roscovitine is a significantly stronger degrader than DS50 (Fig. 4). This discrepancy arises from differences in the structural landscape of the ternary complexes induced by the two MGs. HREMD simulations of the ternary complexes induced

by these MGs show that DS50 positions potential ubiquitination sites (lysine residues) closer to the distal end of the CRL complex (Fig. 4a), where ubiquitin would be loaded. Here, we use the C-terminus of Rbx1 (also located at the distal end of the CRL complex) as a proxy for ubiquitin to assess the ubiquitination potential of each MG. Distances from the ubiquitination sites to Rbx1 were assessed by aligning our simulated ternary complexes to 10 frames selected from 100 ns of MD simulations of the full CRL complex induced by CR8.

To quantify this effect, we defined a ubiquitination score for each molecule as the total number of lysine residues located within a predefined distance cutoff from the C-terminus of Rbx1. Using a cutoff of 8 Å, the distribution of ubiquitination scores is shifted to higher values for Roscovitine compared to DS50 (Fig. 4b).

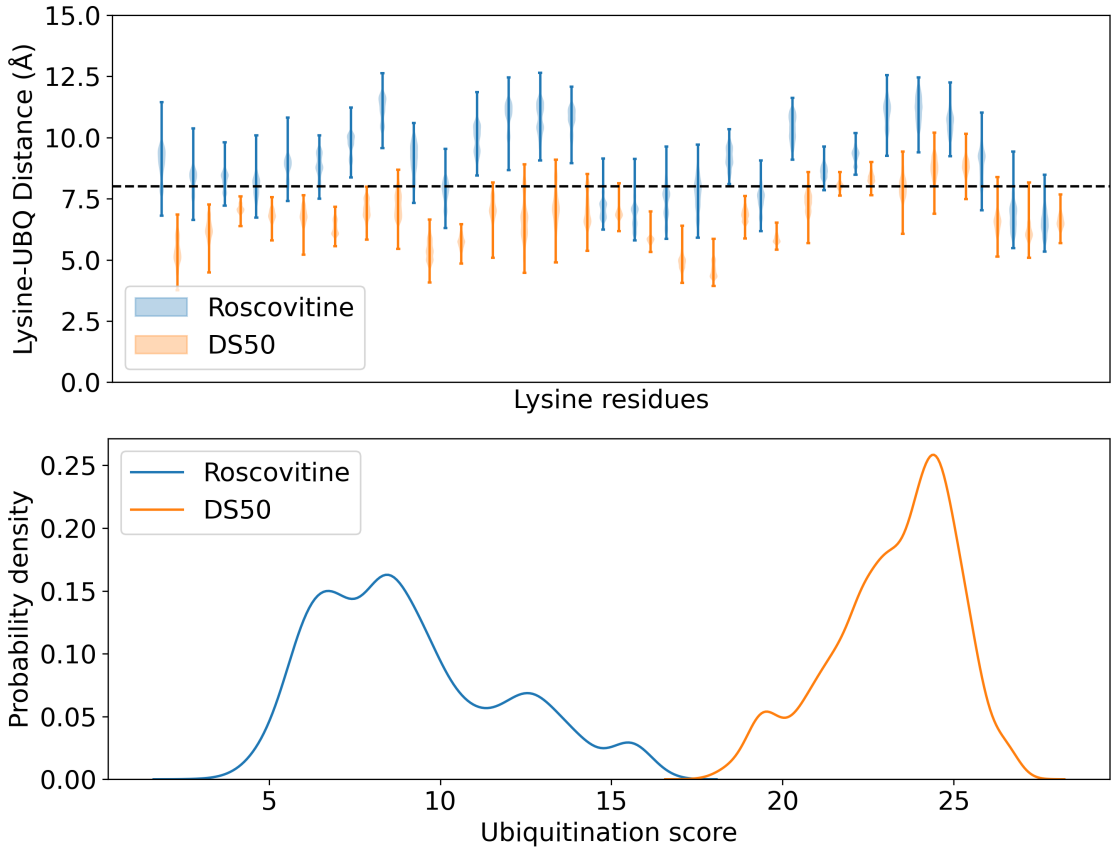


Figure 4: (a) Lysine accessibility and (b) ubiquitination scores for Roscovitine and DS50. The ubiquitination score represents the number of lysine residues located within 8 Å of the C-terminal residue of Rbx1.

Building on this observation, we developed a machine learning model to predict degra-

dation efficiency based on the structural features of the ternary complex induced by a given MG. The model incorporates inter-chain distances and distances between putative ubiquitination sites and the C-terminus of ubiquitin loaded onto the E2 ligase of the CRL complex (Fig. 9). Details of the model are provided in the Methods section.

Our model is a supervised variational autoencoder (VAE) that uses a randomly selected ensemble of frames from HREMD simulations of MG-induced ternary complexes to predict their $pDC50$. The model was trained on HREMD simulations of five MGs and tested on a separate set of five MGs not included in the training data. It accurately predicted the potency of the test set MGs, achieving an R^2 value of 0.910.

One advantage of using a supervised VAE is its ability to create a structured latent space, where learned features correlate with degradation potency. As shown in Fig. 5B, embeddings from the test set simulations reveal that more potent MGs tend to sample lower values of Feature 1, corresponding to the lower region of the scatter plot.

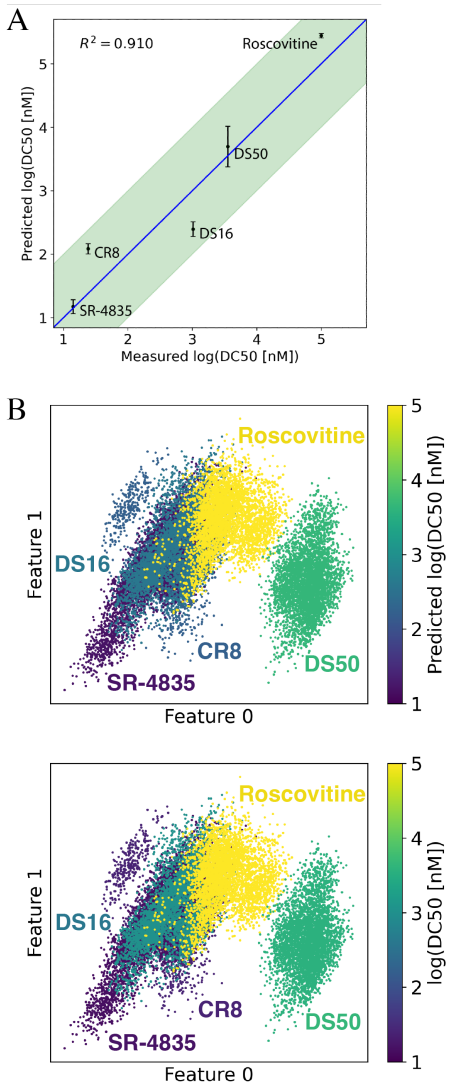


Figure 5: Results of the supervised VAE model. (A) The model accurately predicts pDC_{50} for the test set of molecular glues excluded from training. (B) The structured latent space of the model differentiates molecules based on their pDC_{50}, with more potent MGs sampling lower values of Feature 1.

Case Study: CRBN-dependent GSPT1 degradation

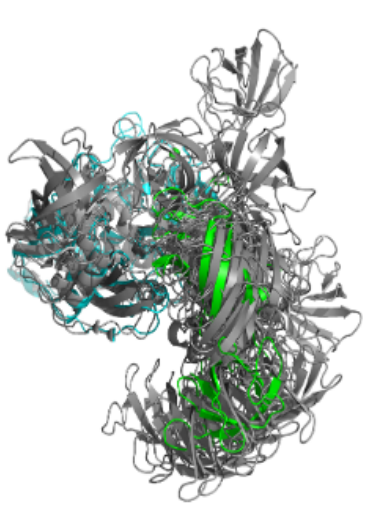
We applied GlueMap to identify CRBN-recruiting molecular glue (MG) degraders of GSPT1. GSPT1, a translation termination factor,[24] plays a critical role in various cancers, including acute myeloid leukemia (AML) and MYC-driven solid tumors. Targeting GSPT1 for degradation has emerged as a promising therapeutic strategy, particularly through MGs that recruit CRBN to facilitate GSPT1 degradation.[25] Degrading GSPT1 disrupts protein translation, triggering cell apoptosis via activation of the integrated stress response. Preclinical studies have demonstrated potent anti-tumor activity from GSPT1 degradation in AML and MYC-driven cancers, with several lead compounds progressing to clinical trials.[26] Additionally, GSPT1 degradation induces TP53-independent cell death, making it a viable strategy for cancers resistant to p53-dependent therapies.[27] Notably, GSPT1 degraders spare normal hematopoietic stem cells,[27] reinforcing their potential for selective cancer therapy. Selective degraders like compound 9q have shown strong antiproliferative effects against leukemia cell lines, underscoring their promise for cancer treatment.[28]

Prediction of ternary complex structure

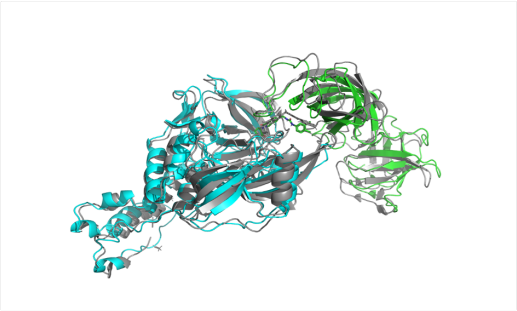
We first validated GlueMap by modeling the ternary complex between CRBN, the known MG CC-885, and GSPT1. GlueMap can accurately predict the ternary complex structures of GSPT1-MG-CRBN, as demonstrated by the excellent agreement with the crystal structure (PDB ID 5HXB[29]) of the ternary complex formed by GSPT1, CC-885, and CRBN (Fig. 6).

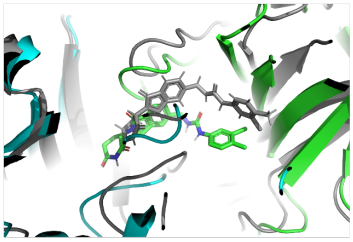
Prospective virtual screening using simulated CRBN-GSPT1 encounter complex structures

For prospective virtual screening, we employed a simulated encounter complex (EC) structure of CRBN and GSPT1 derived from extensive MD simulations of 40 protein-protein interaction (PPI) poses (see Methods for details). The selected EC structure demonstrated favorable stability metrics, including a GBVI energy of -47 kcal/mol and a buried surface

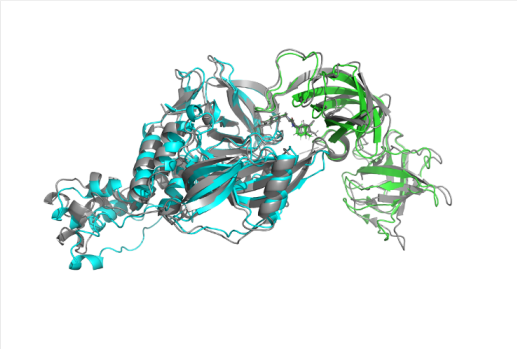


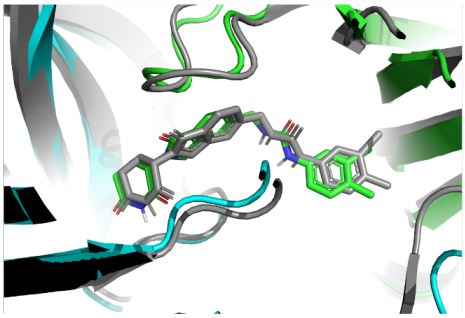
(a) Protein-protein poses derived from docking GSPT1 and CRBN.





(b) Initial ternary complex model of GSPT1 – CC-885 – CRBN after compound docking and scoring. I-RMSD = 8.0 Å; ligand RMSD = 2.1 Å.





(c) Refined ternary complex model after long MD simulations (100 ns x 2 replicas). I-RMSD = 3.5 Å; ligand RMSD = 1.0 Å.

Figure 6: Accurate prediction of the GSPT1 – CC-885 – CRBN ternary complex using GlueMap.

area of 1670 Å^2, comparable to the crystal structure (−57 kcal/mol and 1314 Å^2, respectively; PDB ID: 6XK9). This EC configuration presented a well-defined binding pocket near CRBN's CC-885 site, making it suitable for screening. Other simulated ECs showed similar stabilities but were excluded due to either narrow binding pockets or GSPT1-CRBN interactions far from the CC-885 site.

A library of over 11,000 CRBN-targeted compounds was docked to this EC using pharmacophore constraints of conserved MG interactions with CRBN, derived from the crystal structure (6XK9). We prioritized 15 compounds with docking scores $S < -8$ and specific interactions with both CRBN and GSPT1. Following initial docking, stability was assessed using GlueMap's complex stability workflow. Compounds were ranked based on ligand RMSD and RMSF values compared to reference MD simulations of 6XK9.

Three compounds, MG1, MG2, and MG3, emerged as promising candidates, exhibiting root-mean-square deviation (RMSD) and root-mean-square fluctuation (RMSF) values below 2.4 Å and 0.6 Å, respectively. Detailed interaction analysis revealed that MG1 and MG2 formed stabilizing hydrogen bonds with key residues on GSPT1 (K475) and CRBN (E377), whereas MG3 lacked such interactions, explaining its instability. Fig. 7 highlights the docking and stability results for these candidates.

This virtual screening campaign demonstrates GlueMap's ability to predict novel molecular glues through a combination of docking, dynamic stability analysis, and thermodynamic evaluations. It further validates the utility of simulated encounter complexes in expanding the scope of molecular glue discovery.

Recovery of previously identified molecular glues using GlueMap

To validate GlueMap's capabilities in identifying known molecular glues, we retrospectively screened a virtual library of cereblon binders comprising approximately 11,000 compounds sourced from Enamine. This library was augmented with three known molecular glues: CC-90009, CC-885, and MRT-2359. Using the crystal structure 6XK9 as the basis, we followed

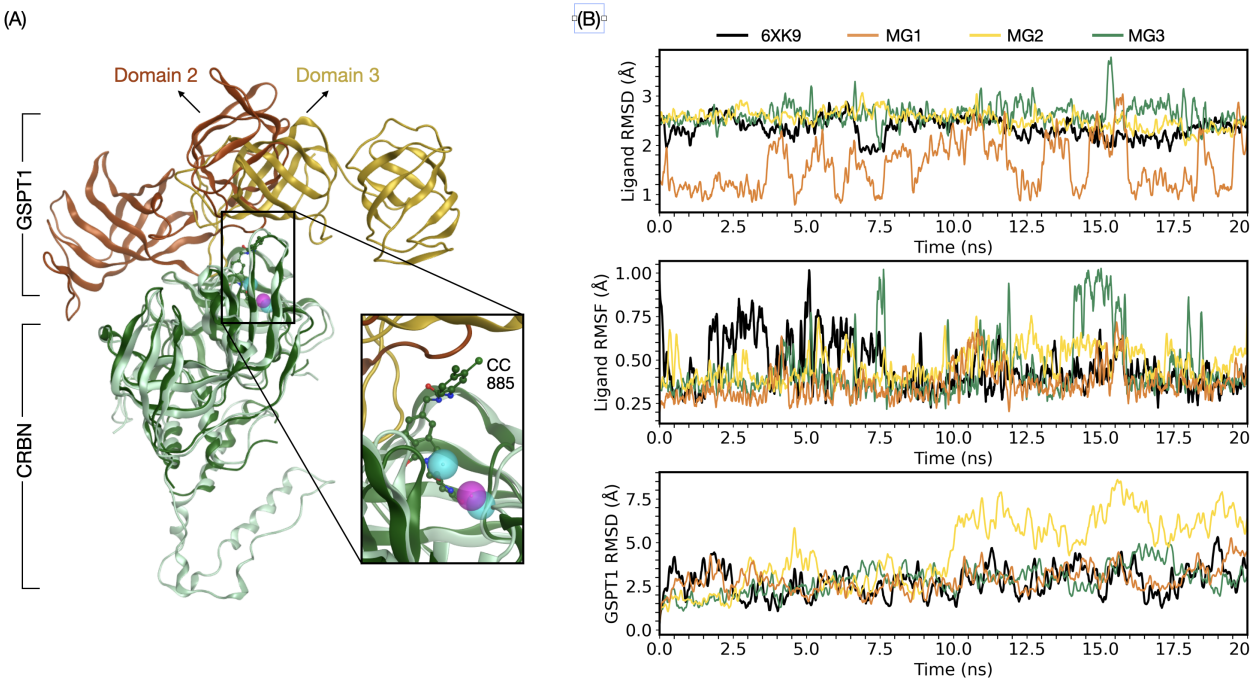


Figure 7: (A) The CRBN-GSPT1 encounter complex used for virtual screening is shown in light green (CRBN) and brown (GSPT1), compared with the co-crystal structure 5HXB. The structures were superimposed on CRBN Cα atoms. In the inset, the three pharmacophore features used for docking are highlighted as transparent spheres and residues 145 to 155 are not shown. (B) Results of the complex stability workflow: colored lines in three panels indicate the hit compounds MG1 to MG3 found in our VS. The black line indicates the reference time series calculated from the 6XK9 complex stability workflow.

the GlueMap workflow outlined in Fig. 1.

Initial filtering based on docking scores and predicted interactions with GSPT1 narrowed the library to 100 candidates. These candidates were further refined using long molecular dynamics (MD) simulations, with stability assessed via ligand root-mean-square fluctuation (RMSF). Of the 100 compounds, 24 exhibited high stability and were subjected to relative binding free energy (RBFE) calculations. Among the screened compounds, the three known molecular glues—CC-90009, CC-885, and MRT-2359—ranked among the top-performing molecules, as shown in Fig. 8.

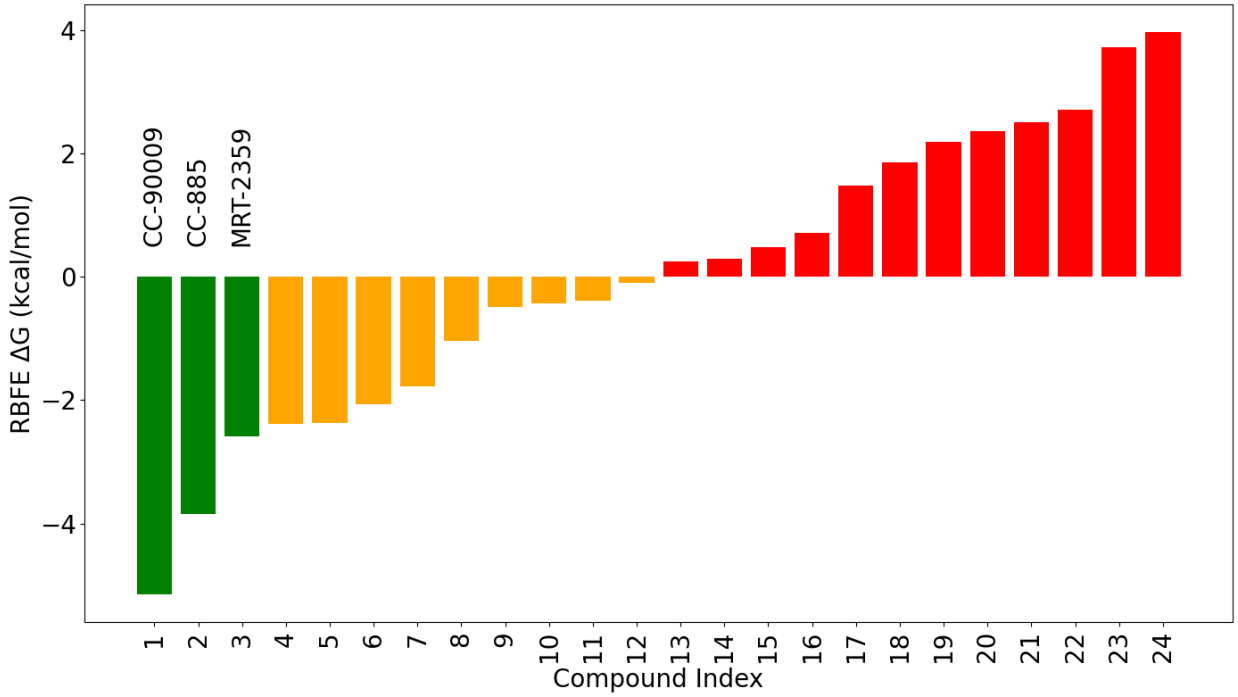


Figure 8: Molecular glues binding to GSPT1-CRBN complex. The three lowest values, corresponding to CC-90009, CC-885, and MRT-2359, are highlighted in green. Values less than zero are shown in orange, while values greater than zero are in red.

These results underscore GlueMap's ability to reliably recover previously identified molecular glues by integrating docking, MD simulation stability, and RBFE predictions, thereby offering a robust and comprehensive evaluation framework. This methodology extends prior work[30] by incorporating enhanced thermodynamic and dynamic stability assessments alongside traditional docking metrics.

Methods

Molecular Dynamics simulations

Forty structures generated from protein-protein docking of GSPT1 and CRBN were used as starting points for molecular dynamics (MD) simulations. The starting structures were prepared using the MOE software package from the Chemical Computing Group. To optimize simulation efficiency, the large protein DDB1 was removed, reducing the system size.

For each initial protein-protein interaction (PPI) structure, two replicas were simulated for 2 microseconds each, resulting in a total of 80 simulations. MD simulations were performed using OpenMM with hydrogen mass re-partitioning and a 4 fs integration time step. Harmonic restraints were applied to maintain the structural integrity of a Zn ion in the CRBN protein. Systems were solvated with a 9 Å buffer and neutralized to an ionic concentration of 100,mM NaCl, resulting in systems of approximately 170,000 atoms.

The Amber14 force field was applied for proteins, while parameters for the compound CC-885 were derived using internal and commercially available parameterization tools. The small molecule parameters are provided as a zip file in the Supporting Information.

Encounter Complex clustering

To analyze encounter complexes from MD simulations of CRBN and GSPT1, we utilized a neural network autoencoder in combination with time-lagged independent component analysis (tICA).[31] The input to the autoencoder was an inter-chain distance matrix representing residues at the CRBN-GSPT1 interface. Interface residues were defined as those whose Cα atoms came within 20 Å of each other at any point during the MD simulations.

The autoencoder compressed the interfacial distance matrix into a six-dimensional latent representation, which reconstructed the original matrix through a decoder network. This latent space captured over 95% of the variance in a held-out test set comprising 20% of MD data (approximately 12,000 frames).

The interfacial distances were projected into the latent space, and tICA was used with a lag time of 5 ns to identify the system's slowest dynamical modes. Features learned through tICA were scaled using the kinetic map of reference,[32] and frames were clustered into five groups using Ward's method, implemented with the scikit-learn AgglomerativeClustering module.[33]

Compound library for virtual screening

A CRBN-targeted library of over 11,000 compounds was curated from the Enamine database in SDF format. The library was processed in two steps: 1. Using MOE's `sdwash` utility,[34] salts and small components were removed, protomers at pH 7 were enumerated, and hydrogens were added. 2. Using MOE's `sdfilter` utility,[34] protomers with a predicted frequency $> 20\%$ were retained.

The final library contained 12,532 compounds, enriched with highly populated protomers. The following commands were used:

```
sdwash library.sdf -sdf:V3000 -addH -pH 7.0 -component -o library-w.sdf
sdfilter library-w.sdf -sdf:V3000 -numfield %C 20+ -o final-library.sdf
```

Prospective virtual screening workflow and analysis

The prospective virtual screening workflow began with pharmacophore-guided docking in the CRBN region. Based on structural analysis of co-crystal structures 5HXB and 6XK9, three pharmacophore features were defined: two hydrogen bond acceptor sites for interactions between the glutarimide of CC-885 (5HXB) or CC-90009 (6XK9) and CRBN residues H378 and W380, and one hydrogen bond donor feature for the interaction between CC-90009 and H378 (Fig. 7).

The processed ligand library was docked against the CRBN-GSPT1 structure using MOE with rigid receptor docking and pharmacophore constraints.[34] For each compound, up to two

conformations were saved, and standard values were used for all docking parameters.

Docked poses were assessed by: 1. S Score: Calculated by MOE, with more negative scores indicating better binding. 2. Interaction Profiles: Obtained using MOE's PLIF module to confirm interactions with CRBN and GSPT1.

Docking stability was evaluated with an in-house workflow consisting of 10 cycles of 20 ns MD simulations. Systems were solvated in a rectangular box with a 12 Å buffer and a NaCl concentration of 150 mM. Heavy atoms were restrained during gradual heating, and simulations were run in the NPT ensemble using Amber's `pmemd.cuda` module with hydrogen mass re-partitioning and a 4 fs integration time step.

To quantify stability, the fraction of time that the ligand RMSD ($\langle f_{RMSD} \rangle$) remained below 5 Å relative to the minimized docked pose was calculated. Complexes with $\langle f_{RMSD} \rangle < 0.5$ were filtered out at any stage. Additional analyses were performed on stable complexes and included ligand RMSF and GSPT1 RMSD.

Hamiltonian Replica Exchange Molecular Dynamics (HREMD)

HREMD simulations[16,35] were performed in Amber for these crystal structures of DDB1-CDK12-Cyclin K: 6TD3, 8BU1, 8BU2, 8BU5, 8BU9, 8BUD, 8BUG, 8BUN, 8BUR, and 8BUT.[5] Each system was prepared and energy-minimized using MOE, solvated with a 10 Å buffer, and neutralized with Na and Cl ions. Simulations were run for 50 ns per system with 32 replicas spanning 298.15 K to 400 K logarithmically. A time step of 2 fs was used. Proteins were parameterized using the Amber ff14SB force field,[36] and small molecules were parameterized using commercially available software. The small molecule parameters are provided as a zip file in the Supporting Information.

Supervised representation learning

Using HREMD simulations, we trained a supervised variational autoencoder (VAE) model to predict the pDC_{50} of molecular glues based on the free energy landscape of DDB1-molecular

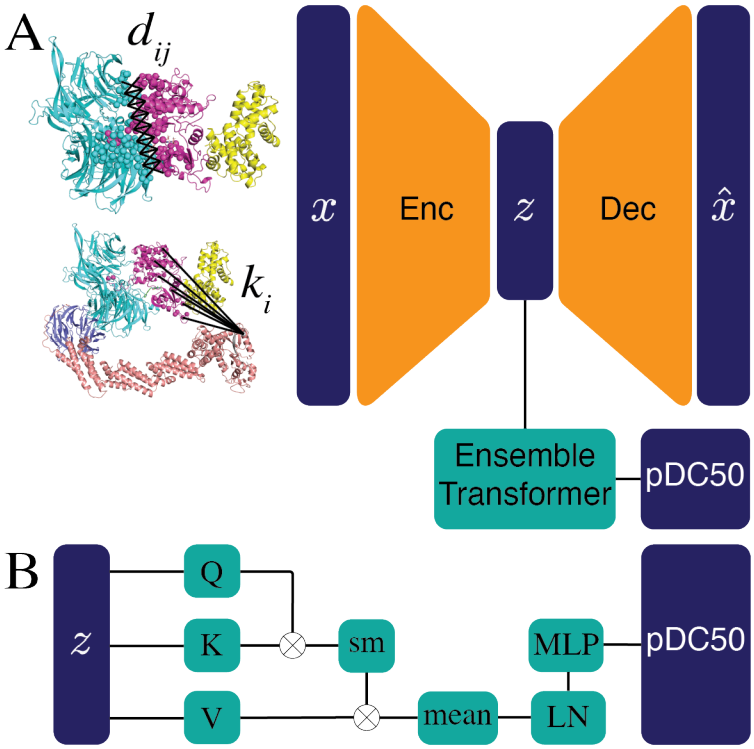


Figure 9: (A) Schematic of the supervised VAE model. Simulation frames are featurized using inter-chain residue distances (d_{ij}) and lysine-ubiquitin distances (k_i). (B) Regression module for pDC_{50} prediction using self-attention. Abbreviations: SM (softmax), LN (layer normalization), MLP (multi-layer perceptron).

glue-CDK12-Cyclin K complexes. The VAE was combined with an attention-based regression module to structure the latent space for degradation potency prediction.

Simulations were featurized by combining inter-chain Cα distances between DDB1 and CDK12 with lysine-ubiquitin distances in the CRL complex. Simulations of ternary complexes were aligned with CRL poses to include lysine-ubiquitin distances where full CRL data were unavailable. These features formed a frame-wise vector for HREMD simulations, which served as input for the model.

Deriving binary and ternary binding constants from experimental assays

We determine the equilibrium binding constants by simultaneously fitting data from binary binding assays and ternary complex formation assays for CDK12 and DDB1 (Fig. 10).

The glue-mediated formation of ternary complexes is modeled using the chemical reactions shown in Eq. 3:

$$
\begin{aligned}
&P + E \xrightleftharpoons{K_{PE}} P \cdot E \\
&P + M \xrightleftharpoons{K_{PM}} P \cdot M \\
&P \cdot E + M \xrightleftharpoons{K_{PE,M}} P \cdot M \cdot E \\
&P \cdot M + E \xrightleftharpoons{K_{PM,E}} P \cdot M \cdot E
\end{aligned} \tag{3}
$$

In this model, the molecular glue (M) is assumed to bind only to protein P and not to protein E. For the DDB1-CDK12 system, the molecular glue specifically binds to CDK12 (P) but not to DDB1 (E).

The concentrations of the molecular species are determined by solving the following

equations:

$$
\begin{aligned}
K_{PE} &= \frac{[P \cdot E]}{[P][E]} \\
K_{PM} &= \frac{[P \cdot M]}{[P][M]} \\
K_{PE,M} &= \frac{[P \cdot M \cdot E]}{[P \cdot E][M]} \\
[P]_0 &= [P] + [P \cdot E] + [P \cdot M] + [P \cdot M \cdot E] \\
[E]_0 &= [E] + [P \cdot E] + [P \cdot M \cdot E] \\
[M]_0 &= [M] + [P \cdot M] + [P \cdot M \cdot E]
\end{aligned} \tag{4}
$$

where $[X]_0$ is the initial concentration of molecule $X = P, E, M$.

In the binding assays (binary and ternary), the initial concentrations $[P]_0$ and $[E]_0$ are fixed, and the concentration $[M]_0$ is titrated. The equilibrium concentrations are thus functions of $[M]_0$.

In the TR-FRET assay of ternary complex formation, the signal is taken to be

$$
S([M]_0) = \gamma_{P \cdot E}[P \cdot E]([M]_0) + \gamma_{P \cdot M \cdot E}[P \cdot M \cdot E]([M]_0) \tag{5}
$$

The binary binding between P and M is characterized by a competitive binding assay, where M competes against a tracer T in binding to P:

$$
\begin{aligned}
P + M &\overset{K_{PM}}{\rightleftharpoons} P \cdot M \\
P + T &\overset{K_{PT}}{\rightleftharpoons} P \cdot T
\end{aligned} \tag{6}
$$

The signal is taken to be

$$
R([M]_0) = \gamma_T[T]([M]_0) + \gamma_{P \cdot T}[P \cdot T]([M]_0) \tag{7}
$$

Given the dose responses in TR-FRET and competitive binding assays, we fit the reaction constants K_{PE}, K_{PM}, $K_{PE,M}$, $K_{PM,E}$, K_{PT} and the signal-to-concentration ratios $\gamma_{P.E}$, $\gamma_{P.M.E}$, γ_T, $\gamma_{P.T}$. The objective function to minimize is

$$\begin{aligned} F(\vec{K},\vec{\gamma}) &= \sum_{M\in\text{TR-FRET}} \sum_{\{[M]_0\}} \left(S([M]_0) - S_{e,[M]_0}\right)^2 \\ &+ \sum_{M\in\text{binary}} \sum_{\{[M]_0\}} \left(R([M]_0) - R_{e,[M]_0}\right)^2 \\ &+ \eta_{P\cdot E} \sum_{M\in\text{TR-FRET}} \left(\gamma_{P\cdot E,M} - \gamma_{P\cdot E}\right)^2 \\ &+ \eta_T \sum_{M\in\text{binary}} \left(\gamma_{T,M} - \gamma_T\right)^2 + \eta_{P\cdot T} \sum_{M\in\text{binary}} \left(\gamma_{P\cdot T,M} - \gamma_{P\cdot T}\right)^2 \end{aligned} \tag{8}$$

with the constraint of thermodynamic cycle closure:

$$K_{PE}K_{PE,M} = K_{PM}K_{PM,E} \ \ \forall M \tag{9}$$

In Eq. 8, $S_{e,[M]_0}$ is the experimental readout of TR-FRET signal at concentration $[M]_0$ for M, and $R_{e,[M]_0}$ is the experimental readout of tracer-competition assay at concentration $[M]_0$ for M.

The binary $K_{D,b}$ and ternary $K_{D,t}$ values are summarized in the Excel spreadsheet in the Supplementary Information.

Relative Binding Free Energy calculations

To evaluate the thermodynamic stability of ternary and binary complexes involving molecular glues, we employed a robust workflow for calculating relative binding free energies (RBFEs). This method is based on well-established thermodynamic cycles (Eqs. 1 and 2) and incorporates state-of-the-art molecular dynamics (MD) and free energy analysis techniques. The workflow is fully automated, progressing from the initial molecular design to the generation of final results and visualizations.

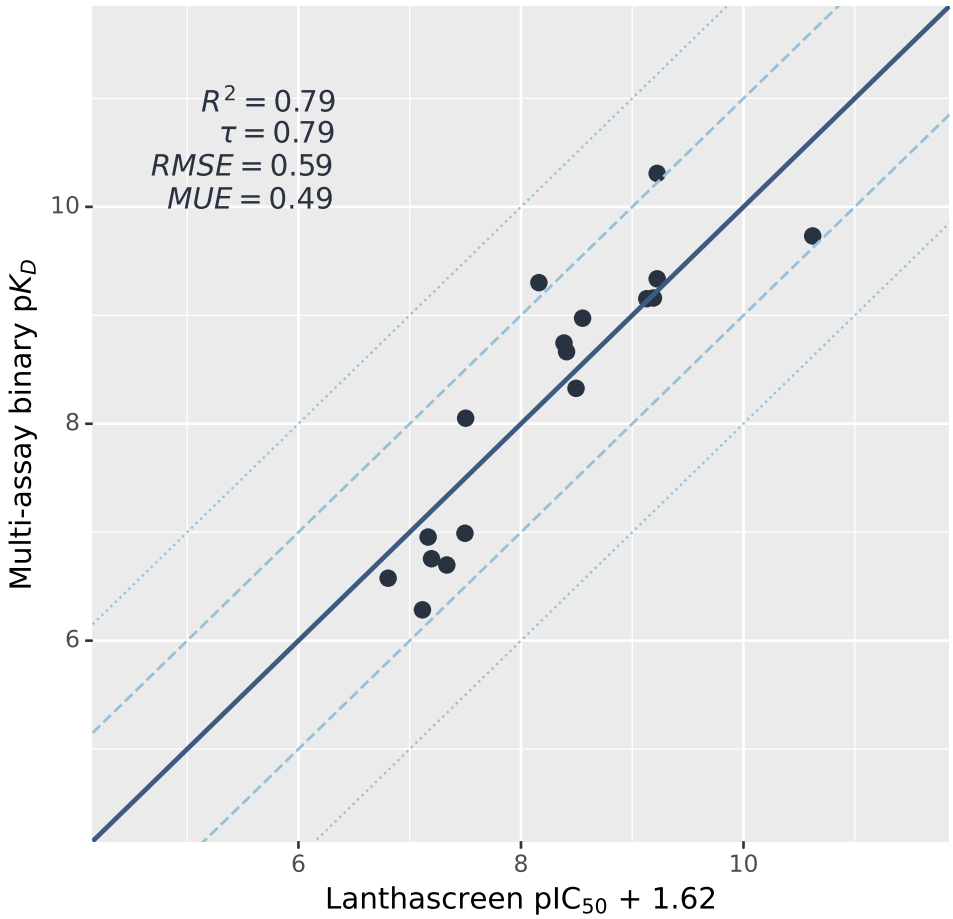


Figure 10: Comparing the binary $K_{\mathrm{D,b}}$ and ternary $K_{\mathrm{D,t}}$ values determined by the global fit of chemical reaction models to the literature values of IC_{50} and EC_{50} determined from phenomenological models.

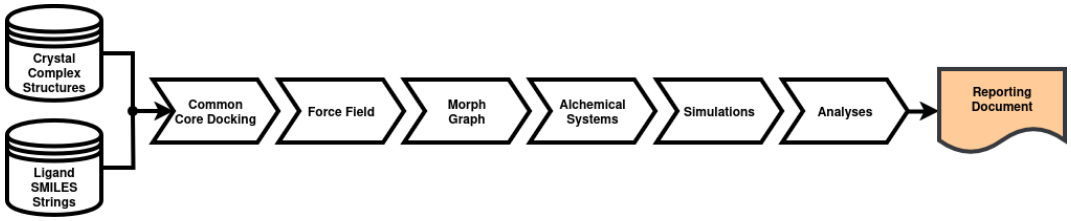


Figure 11: A flowchart showing the overall steps of Atommap's RBFE workflow.

RBFE workflow overview

The overall RBFE workflow is summarized in Fig. 11 and involves the following steps

1. Docking and structure preparation: Design molecules are docked into the binding sites of binary and ternary complexes using a common core docking strategy. This approach ensures alignment of conserved substructures among the congeneric series of compounds while allowing flexibility for unique substructures to optimize binding interactions. Experimental or computationally derived ternary complex structures serve as references for docking.

2. Force Field Parameterization: Parameters for the molecular glue candidates are generated on-the-fly using a proprietary force field generator. These parameters ensure compatibility with the Amber14 force field used for the protein components.

3. Atom Matching and Morph Graph Construction: Atom mappings are computed between pairs of molecules to quantify structural similarity. These mappings are used to construct a morph graph, where nodes represent molecules and edges define transformations between them. The graph is optimized to maximize similarity while maintaining connectivity across all nodes.

4. System Preparation: For each edge in the morph graph, three alchemical systems are constructed: the molecule in water, the molecule in the binary complex, and the molecule in the ternary complex. Binary complexes are generated by removing one receptor from the ternary complex. Systems are solvated with explicit water and ions to achieve a physiological salt concentration (0.15 M).

5. Simulation Protocol: Systems undergo energy minimization followed by relaxation through short MD simulations. Replica Exchange Molecular Dynamics (REMD) simulations are performed with optional REST2[37] enhanced sampling to ensure thorough exploration of the conformational space.

6. Free Energy Calculations: Free energy differences are computed for each edge in the morph graph using the Multistate Bennett Acceptance Ratio (MBAR) or Bennett Acceptance Ratio (BAR) methods. Results are aggregated and analyzed using the DiffNet framework to provide maximum likelihood estimates of the RBFEs for ternary and binary complexes.

7. Results and Reporting: The computed RBFEs are visualized and summarized in an Excel report, including detailed plots and numerical values. This facilitates comparison between compounds and aids in prioritizing candidates for further experimental validation.

The described RBFE workflow provides high accuracy by incorporating enhanced sampling techniques and optimized atom mappings; it is also scalable due to its elastic and automated nature; and it provides structural and interaction data, in addition to free energy values, to offer a deeper understanding of the factors driving molecular glue efficacy.

Discussion

GlueMap effectively models the structural ensemble of ternary complexes, predicts their thermodynamic stabilities, and evaluates downstream effects such as ubiquitination and degradation efficiency of molecular glues (MGs). By combining structural and thermodynamic metrics with ubiquitination potential, GlueMap enables a multi-criteria approach to rational MG design. This study and prior work[16,17] demonstrate the ability of computational modeling to predict the conformational landscapes of ternary complexes with high atomistic accuracy. These landscapes are instrumental in identifying molecular glues that stabilize specific ternary complexes with desirable properties, such as enhanced ubiquitination potential and stability.

The case studies of CRBN-GSPT1 and DDB1-CDK12 illustrate GlueMap's utility. For CRBN-GSPT1, GlueMap successfully recovered known MG degraders and identified novel

candidates through prospective virtual screening using a simulated PPI. These findings emphasize the value of incorporating encounter complex simulations into the virtual screening pipeline, enabling the identification of binding poses conducive to ternary complex formation.

In the DDB1-CDK12 system, the correlation between ternary complex stability and degradation efficiency and between predicted binding free energies and measured ternary complex stability highlighted GlueMap's predictive power. Disparity between Roscovitine and DS50 degradation efficiencies, despite similar thermal stabilities, underscore the importance of structural features such as lysine-ubiquitin proximity. GlueMap's ability to model these differences enables it to fine-tune MGs in optimizing degradation efficiency.

The application of machine learning further enhances GlueMap's predictive capabilities. The supervised variational autoencoder (VAE) model demonstrated high accuracy in predicting degradation potency from structural features, offering a powerful tool for prioritizing MG candidates.

Future developments could extend GlueMap's application to additional proteins, including membrane proteins, and a wider range of E3 ligases, broadening its application scope. Incorporating experimental feedback and refining its machine learning models, for example by fine-tuning models like AlphaFold3 or BOLTZ-1 might further enhance its predictive accuracy and effectiveness in discovering novel molecular glues.

Conclusion

In this work, we demonstrated GlueMap's versatility and predictive accuracy in the discovery and prioritization of molecular glues, using the case studies of CRBN-dependent degradation of GSPT1 and DDB1-dependent degradation of CDK12. GlueMap may help identify novel molecular glue degraders, provide mechanistic insight into their activity, and enable their rational design and optimization. Our work underscores the potential of computational modeling and design to accelerate the development of molecular glues across a broad range

of protein targets and E3 ligases.

Supporting Information Available

The supplementary information contains an Excel spreadsheet of the binary $K_{D,b}$ and ternary $K_{D,t}$ values fit to experimental data, used for comparison to binding free energy calculations. Also included are zip files with the small molecule parameters used in reported simulations of CDK12-DDB1 and GSPT1-CRBN complexes.

References

(1) Ting, P. Y. et al. A molecular glue degrader of the WIZ transcription factor for fetal hemoglobin induction. *Science* **2024**, *385*, 91–99.

(2) Ng, A.; Offensperger, F.; Cisneros, J. A.; Scholes, N. S.; Malik, M.; Villanti, L.; Rukavina, A.; Ferrada, E.; Hannich, J. T.; Koren, A.; Kubicek, S.; Superti-Furga, G.; Winter, G. E. Discovery of Molecular Glue Degraders via Isogenic Morphological Profiling. *ACS Chemical Biology* **2023**, *18*, 2464–2473, PMID: 38098458.

(3) Perry, M.; Munier, C.; Neves, J.; Gunnarsson, A.; Edfeldt, F.; Landrieu, I.; others Rational discovery of molecular glues for the glucocorticoid receptor – 14-3-3 protein-protein interaction starting with application of traditional hit-finding methods. *ChemRxiv* **2024**, This content is a preprint and has not been peer-reviewed.

(4) Rui, H.; Ashton, K. S.; Min, J.; Wang, C.; Potts, P. R. Protein–protein interfaces in molecular glue-induced ternary complexes: classification, characterization, and prediction. *RSC Chemical Biology* **2023**, *4*, 192–215.

(5) Kozicka, Z.; Thomä, N. H. Haven't got a glue: Protein surface variation for the design of molecular glue degraders. *Cell Chemical Biology* **2021**, *28*, 1032–1047.

(6) Schreiber, S. L. Molecular glues and bifunctional compounds: Therapeutic modalities based on induced proximity. *Cell Chemical Biology* **2024**, S2451945624002058.

(7) Konstantinidou, M.; Arkin, M. R. Molecular glues for protein-protein interactions: Progressing toward a new dream. *Cell Chemical Biology* **2024**, *31*, 1064–1088, Publisher: Elsevier.

(8) Prael, F. J.; Cox, J.; Sturm, N.; Kutchukian, P.; Forrester, W. C.; Michaud, G.; Blank, J.; Shen, L.; Rodríguez-Pérez, R. Machine learning proteochemometric models for Cereblon glue activity predictions. *Artificial Intelligence in the Life Sciences* **2024**, *6*, 100100.

(9) Petzold, G. et al. Mining the CRBN Target Space Redefines Rules for Molecular Glue-induced Neosubstrate Recognition. *bioRxiv* **2024**,

(10) Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **2024**, *630*, 493–500.

(11) Wohlwend, J.; Corso, G.; Passaro, S.; Reveiz, M.; Leidal, K.; Swiderski, W.; Portnoi, T.; Chinn, I.; Silterra, J.; Jaakkola, T.; Barzilay, R. Boltz-1: Democratizing Biomolecular Interaction Modeling. *bioRxiv* **2024**,

(12) Team, T. S. M. I.; Che, X. YDS-Ternoplex: Surpassing AlphaFold 3-Type Models for Molecular Glue-Mediated Ternary Complex Prediction. *bioRxiv* **2024**,

(13) Wang, B.; Cao, S.; Zheng, N. Emerging strategies for prospective discovery of molecular glue degraders. *Current Opinion in Structural Biology* **2024**, *86*, 102811.

(14) Bose, S.; Kilinc, C.; Dickson, A. Markov State Models with Weighted Ensemble Simulation: How to Eliminate the Trajectory Merging Bias. *ChemRxiv* **2024**, This content is a preprint and has not been peer-reviewed.

(15) Qiu, Y.; Wiewiora, R. P.; Izaguirre, J. A.; Xu, H.; Sherman, W.; Tang, W.; Huang, X. Non-Markovian Dynamic Models Identify Non-Canonical KRAS-VHL Encounter Complex Conformations for Novel PROTAC Design. *JACS Au* **2024**, *4*, 3857–3868, Publisher: American Chemical Society.

(16) Dixon, T. et al. Predicting the structural basis of targeted protein degradation by integrating molecular dynamics simulations with structural mass spectrometry. *Nature Communications* **2022**, *13*, 5884.

(17) Mostofian, B.; Martin, H.-J.; Razavi, A.; Patel, S.; Allen, B.; Sherman, W.; Izaguirre, J. A. Targeted Protein Degradation: Advances, Challenges, and Prospects for Computational Methods. *Journal of Chemical Information and Modeling* **2023**, *63*, 5408–5432.

(18) Drummond, M. L. *Computational Drug Discovery*; John Wiley & Sons, Ltd, 2024; Chapter 22, pp 537–559.

(19) Lee, T.-S.; Allen, B. K.; Giese, T. J.; Guo, Z.; Li, P.; Lin, C.; McGee, T. D. J.; Pearlman, D. A.; Radak, B. K.; Tao, Y.; Tsai, H.-C.; Xu, H.; Sherman, W.; York, D. M. Alchemical Binding Free Energy Calculations in AMBER20: Advances and Best Practices for Drug Discovery. *Journal of Chemical Information and Modeling* **2020**, *60*, 5595–5623, Publisher: American Chemical Society.

(20) Li, P.; Li, Z.; Wang, Y.; Dou, H.; Radak, B. K.; Allen, B. K.; Sherman, W.; Xu, H. Precise Binding Free Energy Calculations for Multiple Molecules Using an Optimal Measurement Network of Pairwise Differences. *Journal of Chemical Theory and Computation* **2022**, *18*, 650–663, Publisher: American Chemical Society.

(21) Xu, H. Optimal Measurement Network of Pairwise Differences. *Journal of Chemical Information and Modeling* **2019**, *59*, 4720–4728, Publisher: American Chemical Society.

(22) Mey, A. S. J. S.; Allen, B. K.; Bruce McDonald, H. E.; Chodera, J. D.; Hahn, D. F.; Kuhn, M.; Michel, J.; Mobley, D. L.; Naden, L. N.; Prasad, S.; Rizzi, A.; Scheen, J.; Shirts, M. R.; Tresadern, G.; Xu, H. Best Practices for Alchemical Free Energy Calculations [Article v1.0]. *Living Journal of Computational Molecular Science* **2020**, *2*, 18378, Section: Articles.

(23) Angers, S.; Li, T.; Yi, X.; MacCoss, M. J.; Moon, R. T.; Zheng, N. Molecular architecture and assembly of the DDB1-CUL4A ubiquitin ligase machinery. *Nature* **2006**, *443*, 590–593.

(24) Chauvin, C.; Salhi, S.; Jean-Jean, O. Human eukaryotic release factor 3a depletion causes cell cycle arrest at G1 phase through inhibition of the mTOR pathway. *Molecular and Cellular Biology* **2007**, *27*, 5619–5629.

(25) Zhang, D.; Lin, P.; Lin, J. Molecular glues targeting GSPT1 in cancers: A potent therapy. *Bioorganic Chemistry* **2024**, *143*, 107000.

(26) Hu, Y.; Yan, Y.; Wang, J.; Hou, J.; Lin, Q. Molecular glue degrader for tumor treatment. *Frontiers in Oncology* **2024**, *14*, 1512666.

(27) Sellar, R. S. et al. Degradation of GSPT1 causes TP53-independent cell death in leukemia while sparing normal hematopoietic stem cells. *Journal of Clinical Investigation* **2022**, *132*, e153514.

(28) Wei, Y.; Xu, X.; Jiang, M.; Wang, Y.; Zhou, Y.; Wang, Z.; Zhang, Z.; Zhou, F.; Ding, K. Discovery of new Lenalidomide derivatives as potent and selective GSPT1 degraders. *European Journal of Medicinal Chemistry* **2023**, *258*, 115580.

(29) Matyskiela, M. E. et al. A novel cereblon modulator recruits GSPT1 to the CRL4CRBN ubiquitin ligase. *Nature* **2016**, *535*, 252–257.

(30) Zhang, S.; Nie, S.; Ma, G.; Shen, M.; Kong, L.; Zuo, Z.; Li, Y. Identification of novel GSPT1 degraders by virtual screening and bioassay. *European Journal of Medicinal Chemistry* **2024**, *273*, 116524.

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

(32) Noé, F.; Clementi, C. Kinetic Distance and Kinetic Maps from Molecular Dynamics Simulation. *Journal of Chemical Theory and Computation* **2015**, *11*, 5002–5011, PMID: 26574285.

(33) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; others Scikit-learn: Machine learning in Python. *Journal of machine Learning research* **2011**, *12*, 2825–2830.

(34) Molecular Operating Environment (MOE), 2024.0601. Chemical Computing Group ULC: 910-1010 Sherbrooke St. W., Montreal, QC H3A 2R7, 2024.

(35) Shrestha, U. R.; Smith, J. C.; Petridis, L. Full structural ensembles of intrinsically disordered proteins from unbiased molecular dynamics simulations. *Communications Biology* **2021**, *4*, 243.

(36) 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. *Journal of chemical theory and computation* **2015**, *11*, 3696–3713.

(37) Wang, L.; Friesner, R. A.; Berne, B. J. Replica Exchange with Solute Scaling: A More Efficient Version of Replica Exchange with Solute Tempering (REST2). *The Journal of Physical Chemistry B* **2011**, *115*, 9431–9438.