Structure−Kinetic Relationship for Drug Design Revealed by a PLS Model with Retrosynthesis-Based Pre-Trained Molecular Representation and Molecular Dynamics Simulation

Feng Zhou, Shiqiu Yin, Yi Xiao, Zaiyun Lin, Weiqiang Fu, and Yingsheng J. Zhang*

ACCESS | Metrics & More | Article Recommendations | SI Supporting Information

ABSTRACT: Drug design based on kinetic properties is growing in application. Here, we applied retrosynthesis-based pre-trained molecular representation (RPM) in machine learning (ML) to train 501 inhibitors of 55 proteins and successfully predicted the dissociation rate constant (k_{off}) values of 38 inhibitors from an independent dataset for the N-terminal domain of heat shock protein 90α (N-HSP90). Our RPM molecular representation outperforms other pre-trained molecular representations such as GEM, MPG, and general molecular descriptors from RDKit. Furthermore, we optimized the accelerated molecular dynamics to calculate the relative retention time (RT) for the 128 inhibitors of N-HSP90 and obtained the protein−ligand interaction fingerprints

(IFPs) on their dissociation pathways and their influencing weights on the k_{off} value. We observed a high correlation among the simulated, predicted, and experimental $-\log(k_{off})$ values. Combining ML, molecular dynamics (MD) simulation, and IFPs derived from accelerated MD helps design a drug for specific kinetic properties and selectivity profiles to the target of interest. To further validate our k_{off} predictive ML model, we tested our model on two new N-HSP90 inhibitors, which have experimental k_{off} values and are not in our ML training dataset. The predicted k_{off} values are consistent with experimental data, and the mechanism of their kinetic properties can be explained by IFPs, which shed light on the nature of their selectivity against N-HSP90 protein. We believe that the ML model described here is transferable to predict k_{off} of other proteins and will enhance the kinetics-based drug design endeavor.
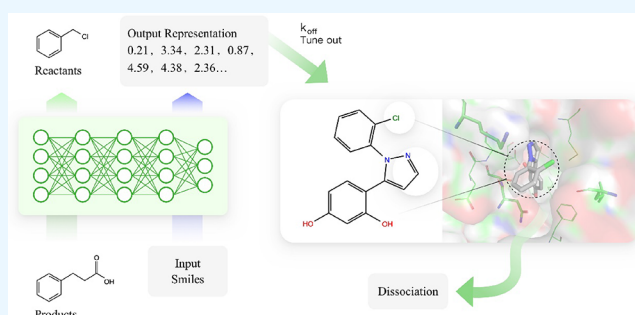
1. INTRODUCTION

Protein−drug interactions can be described by both thermodynamic properties (i.e., equilibrium constant K_i, K_D, and half-maximal inhibitory concentration IC50) and kinetic properties (i.e., association k_{on} and dissociation rate constant, k_{off}). Ligands with the same affinity may have different association and dissociation rates, so kinetic parameters often provide more useful information for drug design. Compounds having a short retention time in active sites may lead to low drug occupancy at targets and poor efficacy. In contrast, for compounds with a longer retention time or slower dissociation rate, their doses can be reduced to achieve better selectivity and fewer side effects. Studies show that kinetic properties are strongly correlated with the pharmacological activity of drugs in cells and in vivo.[1−4] Heitman et al. studied the relationship among intracellular efficacy, affinity, and residence time of 10 A_{2A} adenosine receptor agonists, finding that there is a stronger correlation between intracellular efficacy and residence time ($R^2 = 0.90$) than equilibrium inhibition constant K_i ($R^2 = 0.13$) of these agonists.[2] One strategy to increase selectivity of drugs is to optimize the kinetic properties to obtain drugs with longer retention time.[5,6] Applying kinetics-based drug design

techniques, Khanna et al. optimized the kinetic property of EZH2 inhibitors and achieved a 10 times increase of retention time only by changing $-OCH_3$ to $-SCH_3$.[7] Zhou et al. designed and synthesized 44 novel derivatives and discovered an ACE inhibitor as a clinical candidate with much lower k_{off} than donepezil.[8] Among them, compound **12** demonstrated a much better efficacy and a lower effective dose than that of donepezil. Using ensemble molecular docking and relative binding free-energy calculation, Miller et al. found a novel compound TDI-11861, which shows higher Sac (ADCY10) binding affinity and longer residence time than its precursor (3181 versus 25 s).[9] Combining induced fit docking and MM/GBSA calculation cubic by cubic, Bai and Jiang constructed a sophisticated energy landscape of ligand dissociation process and found a potential lead compound, which gives better
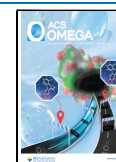
binding kinetic and thermodynamic properties with either TcAChE or mAChE.[10] Because the wet lab experiments to determine the kinetic properties of protein−drug interactions are expensive and time-consuming, new in silico methods such as molecular dynamics (MD) simulation and ML are imperative.

Free MD without any bias is the most straightforward way to calculate kinetic properties like k_{on} and k_{off}. For example, using the supercomputer ANTON, Shaw et al. first reported the binding process of a molecule directly from outside of the pocket by unbiased MD.[11] With 400 μs MD, Pantsar et al. found the important role of water and protein conformational change on the kinetic properties of two p38α MARK inhibitors with short and long residence times but nearly identical activities.[12] Buch et al. ran 495 short free MD and used MSM to construct the complete binding process of the trypsin/benzamidine complex, in which 187 trajectories were consistent with the binding mode found in the crystal.[13] In general, the dissociation time of molecules from their binding pockets is often on the time scale of seconds to hours. Most computing tasks cannot afford such expensive MD simulations. Since 2000, many enhanced sampling methods have been developed to calculate k_{off}. Capelli et al. used the infrequent metadynamics (InMetaD) to calculate the k_{off} of the M2 receptor/iperoxo complex, and the result is only an order of magnitude away from the experimental value.[14] InMetaD has also been used to calculate the k_{off} of trypsin-benzamidine,[15] kinase,[16−18] biotin−streptavidin,[19] and FKBP.[20] Umbrella sampling (US) is useful for studying the dissociation pathway and calculating the free energy of different metastable states. For example, US was used to study the dissociation process and binding free energy of various protein complexes such as benzamidine-trypsin,[21] acetylcholinesterase,[22] cathepsin K, type I dehydrogenase, HSP90, and factor Xa.[23] For more complex systems in which the dissociation pathway is straight, the steered MD (SMD) was used together with the US to find the optimal dissociation path and to calculate the potential of mean force (PMF). For example, US+SMD was used to calculate the binding free energy between maltose-binding protein and maltose.[24] You et al. have applied US+MM/GBSA (molecular mechanics with generalized Born and surface area solvation) to study the mechanism of dissociation of two small molecules from ATP allosteric channels of p38 MAP kinase.[25] PMF can also be used to estimate the retention time ($1/k_{off}$) by calculating the mean free passage times (MFPTs) from the initial unbinding barrier.[26] SMD was also used to calculate the relative residence times of p38α kinase and FAK.[27,28] Potterton et al. predicted the relative residence time of 17 small molecules on the A$_{2A}$ adenosine receptor from changes in water−ligand interaction energy.[29] Mollica et al. used scaled MD (sMD) to calculate the relative residence time of the same series of ligands on proteins such as HSP90, glucose regulation (Grp78), and the A$_{2A}$ GPCR.[30−32] Zhou et al. combined sMD and InMetaD to calculate the residence time and dissociation mechanism of ASEM.[33]

In 2018, using random accelerated MD (RAMD), Kokh et al. performed large-scale k_{off} calculation to calculate the relative residence time of 70 diverse drug-like ligands of N-HSP90 protein and discussed the effect of different substituents on the residence time of small molecules.[34] The number of inhibitors of N-HSP90 was extended to 94 in a following paper.[35] They found that there is a strong correlation between the calculated retention time and the experimental value. By combining RAMD and experimental results, Berger et al. found that PF-562271 is more selective for focal adhesion kinase (FAK) kinase than proline-rich tyrosine kinase 2.[36] RAMD was also used to calculate the kinetic behaviors of a series of inhibitors of the T4 lysozyme, and the results showed that the ligands with a longer residence time have more intermediate metastable states along the dissociation pathways.[37] Recently, RAMD was successfully applied to studying the relative residence times, dissociation mechanisms, and the allosteric effects for the two important membrane-embedded drug targets: β2-adrenergic receptor and muscarinic acetylcholine receptor M2.[38] They found that the dissociation mechanisms observed in the relatively cheap RAMD simulations are consistent with the much more computationally expensive free MD and MetaD simulations. They also uncovered the relationship between the residence time and the allosteric modulation and associated changes in the ligand dissociation pathways. Although there are many successful cases of MD in the calculation of retention time and k_{off}, due to the intrinsic drawback of the force field and insufficient sampling, the calculated retention time and k_{off} often deviate from the ground true experimental values especially when the ligands have diverse scaffolds.[34,35,39] Additionally, most enhanced sampling methods can only produce relative retention times.

In recent years, through the development of artificial intelligence (AI), more accurate predictions of k_{off} were seen by combining molecular simulation and ML. In 2016, a multi-target machine learning (MTML) with interaction fingerprints was used to predict the k_{off} of HIV-1 protease.[40] Based on electrostatic interaction terms and conformational dynamics, a classification model was used to train and classify 39 inhibitors of HIV-1 protease into four classes, predicting effects with 74% accuracy. This paper argues that electrostatic interactions contribute more to the k_{off} value than van der Waals interactions. Similarly, Zhang et al. used a partial least squares (PLS) regression model to learn interaction features from position-restrained MD to predict the k_{off} of p38 MAPK class II inhibitors and obtained good results.[41] A 3D QSAR (quantitative structure−activity relationship) model was also applied to predict k_{off}, and Qu et al. used VolSurf descriptors as features to predict the k_{off} of HIV-1 protease inhibitors using PLS regression models.[42] The Wade group predicted k_{off} with high accuracy using the comparative binding energy (COMBINE) method for N-HSP90 and HIV-1 protease inhibitors.[43] The same group also used interaction fingerprints (IFPs) from a large number of RAMD trajectories to predict the k_{off} of N-HSP90.[35] They found that information along the dissociation pathway is more important than the bound state. Similarly, Huang et al. predicted the k_{off} of 37 inhibitors of HIV-1 using IFPs derived from SMD.[44] Based on graph methods like contact principal component and pairwise mean Euclidean distance analysis, Bray et al. developed an ML model for clustering the unbinding trajectories and path reaction coordinate detection.[45] A recent literature reported the prediction of k_{off} for a large array of proteins and ligands using the random forest method with RF-score scoring functions[46] and protein secondary structures as features.[47] Although the correlation coefficient r_P reached 0.78, the prediction power of this model became very poor for proteins that had not been seen in the training set and the model was not tested on another dataset. Compared to the traditional molecular simulation methods, the ML method to predict k_{off} still faces many challenges such as proper feature representa-
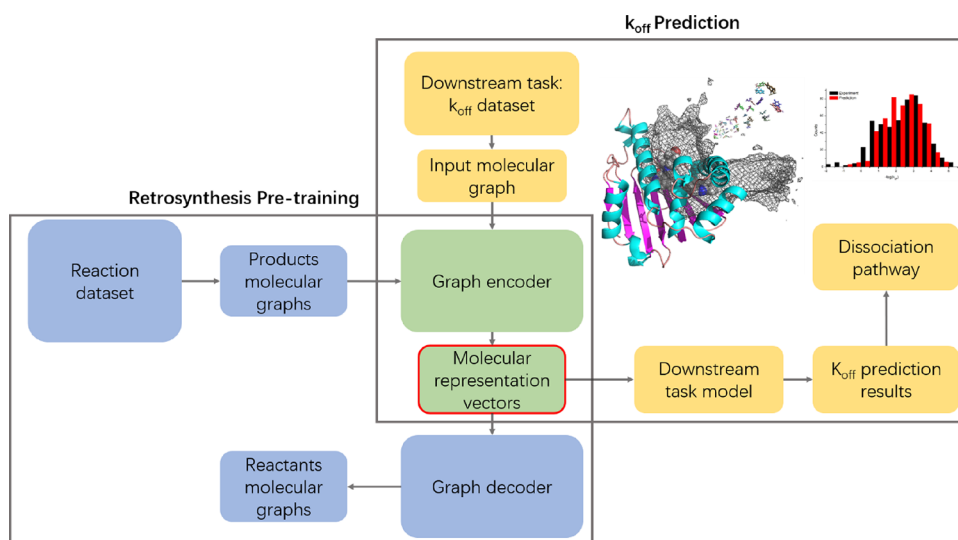
Figure 1. Flowchart of using retrosynthesis-based pre-trained molecular representation (RPM) to predict k_{off}.

tion and model selection as well as methods to increase its interpretability to guide drug design.

To enhance ML predictability and increase model interpretability, we trained a new molecular representation named RPM to predict the kinetic k_{off} and compared its performance to other popular pre-trained molecular representations, i.e., GEM (geometry-enhanced molecular representation),[48] MPG (molecular pre-training graph-based deep learning framework),[49] and a commonly used molecular predictor from RDKit.[50] In our previous work, we trained a molecular pre-training model (G2GT)[51] based on retrosynthesis reaction data. By cutting off the connections between encoder and decoder stages of the model, we used the output of the encoder as a representation (RPM) of the molecule to predict the k_{off} property. The model was successfully tested in an independent dataset and showed a high correlation with experimental data. The mechanisms of the kinetic properties can be explained using IFPs. As a case study, we applied our ML/MD protocol to two N-HSP90 inhibitors, which were not in our model training dataset, and the predicted k_{off} values showed a consistent trend with experimental values. By combining ML and MD simulation, coupled with the molecular representation features and IFP analysis, we unveil the features that have the most impact on the kinetic properties of the protein target and the interpretable details on how the molecules achieve their long retention time.

2. RESULTS AND DISCUSSION

2.1. Molecular Representation RPM Generation. G2GT[51] is a transformer model that uses a large amount of retrosynthetic reaction data as input to train. The model combines the encoder and decoder architecture of Transformer with a molecular graph. The encoder structure of G2GT extracts the structural information, the interaction relationship between molecules and output vectors, which the decoder uses to predict possible synthons. In this stage, we obtain the intrinsic relationship between input and output structures such as functional groups and intermolecular relationships. By cutting off the connections between encoder and decoder stages, we use the latent vector output as a representation of the molecule (RPM). The RPM was used with the PLS regression method to predict the k_{off} of 501

ligands of 55 proteins. This new molecular representation takes consideration of molecular reactivity, function of chemical groups, and intermolecular relationships, thus showing superior performance in property prediction such as k_{off} (complete process of k_{off} prediction in Figure 1).

2.2. k_{off} Prediction Using RPM. A common method to measure a pre-trained model's performance is to test and evaluate it on downstream tasks.[52] In order to show that our model can learn a rich and generalized representation, especially for downstream k_{off} tasks, we compare our RPM with two mainstream pre-trained models: MPG and GEM, which are currently recognized as two state-of-the-art pre-trained models for molecular property prediction. MPG is a pre-trained molecular representation based on molecular graphs.[49] It combines the classical Neural Message Passing (MPNN) for the Quantum Chemistry framework with the powerful transformer block to learn molecular representations. GEM is also a GNN-based molecular pre-training representation.[48] In order to solve the problem of traditional GNN that three-dimensional molecular information will partially lose in graph representation,[53] GEM introduces molecular graphs and bond angle graphs in which bond-angle information is calculated and shared to jointly model molecular representation. Our RPM and the two aforementioned models are all self-supervised pre-trained molecular representations. We also added the traditional RDKit molecular descriptors and then use a unified predictor (i.e., PLS) to predict the k_{off} value. The training data is from the dataset of 501 inhibitors (dataset A) collected by Amangeldiuly et al.[47] We adopt the leave-one-out strategy to train and predict these samples (Table 1). The results show that our RPM molecular representation gives the

Table 1. Comparing Different Molecular Representation Performances in 501 Inhibitors $-\text{Log}(k_{off})$ Prediction Using the Partial Least Squares (PLS) Regression Method

	r_P	MSE
RPM	0.76	0.74
RDKit	0.58	1.19
GEM	0.47	1.34
MPG	0.65	1.01

best prediction for the k_{off} task, which refers to a higher Pearson correlation coefficient ($r_P = 0.76$), compared with GEM ($r_P = 0.47$) and MPG ($r_P = 0.65$), and the lowest MSE (0.74).

2.3. Test of RPM on the New Dataset. To demonstrate the generalizability of our pre-trained molecular representations, we tested 38 ligands on an independent dataset collected by Liu et al.[54] (dataset B). The 38 ligands are all from the same target protein, N-HSP90. HSP90 is a chaperone protein that assists in protein folding, blocking its N-terminal domain by small-molecule inhibitors leading to degradation of the client proteins and diminished tumor growth. Brough et al. reported the new HSP90 inhibitor (VER-52296/NVP-AUY922), which inhibits tumor growth by ~50%.[55] Specifically, a significant reduction of tumor growth in HCT116 colon cancer xenografts was observed with VER-52296/NVP-AUY922.[56] In comparison with other proteins, N-HSP90 consists of more kinetic data and there are lots of literatures on k_{off} simulation and prediction for N-HSP90.[34,35,43,57−62] Figure 2 shows that our
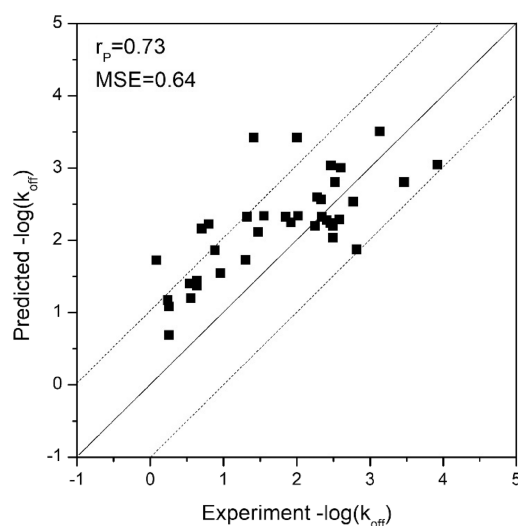

Figure 2. $-\log(k_{off})$ prediction on the new 38 inhibitors of N-HSP90 from the independent database (dataset B).[54] The Pearson correlation coefficient (r_P) is 0.73, and the mean square error (MSE) is 0.64.

prediction value has a similar high correlation and lower MSE than the leave-one-out validation. This suggests that our pre-trained molecular representation has superior quality for cross-dataset representations and exhibits enormous potential for other molecular property prediction task.

2.4. Accelerated MD Unveils the Important Residues for k_{off} of HSP90 and Guides the Kinetic-Based Drug Design. The k_{off} prediction by pre-trained molecular representations and the PLS model is based only on the molecular features and lack of the protein pocket information. For example, the same molecule may have different binding affinities and k_{off} to different targets. To increase the selectivity of the target, one needs to consider both the features of molecules and the pocket information and protein−ligand interactions. RAMD is an efficient method to rapidly obtain the relative retention time, the dissociation pathways, and the IFPs. We performed 90 independent RAMD simulations starting from six different initial structures for each complex to avoid stochastic error. The initial structures of the complex are the cocrystal structures either from the Protein Data Bank (PDB)[63] or from docking poses by Amangeldiuly et al.[47] Short,

unbiased MD simulations were independently performed six times, 10 ns each run, for each ligand to obtain the equilibrium structures of the bound state, followed by 90 RAMDs to calculate the retention time. The detailed MD method is in the Materials and Methods. The calculated $-\log(k_{off})$ was plotted against experimental values (Figure 3).

The Pearson correlation coefficient between the calculated $-\log(k_{off})$ from RAMD and the experimental values is not high ($r_P = 0.534$) for the 100 inhibitors from dataset A. The outliers, which were also observed by Kokh et al.,[34,35] can be attributed to the structures of the ligands, the binding modes, and the force field. However, the correlation between the calculated and the measured $\log(k_{off})$ is significantly improved when grouping the 100 inhibitors by scaffolds. For example, the calculated Pearson correlation coefficient for scaffold I, which contains the hydroxy-indazole structure, is 0.76. Kokh et al. also found that the retention time for the 10 compounds of the amino-quinazoline and amino-pyrrolopyrimidine was systematically underestimated.[34] Consistent with the experiment, we also found that ligands bound to the helical conformation display slow dissociation rates in comparison with those bound to the loop conformation (Figure 3c).[64] The predicted and simulated $-\log(k_{off})$ values have similar populations as the experiment values and the largest population in the range from 2.0 to 3.0 for helical conformation. We also performed RAMD simulation on the inhibitors of N-HSP90 protein from dataset B. The initial structures of the complexes are the cocrystal structures from PDB[63] or from docking poses by Liu et al.[54] The inhibitors that do not share a common scaffold with others were excluded, and the remaining 28 new inhibitors were used for RAMD simulation. Both the predicted $-\log(k_{off})$ values and simulated ones have a high correlation with the experimental values ($r_P = 0.73$) (figure in the Supporting Information).

As mentioned above, the ML model is good at predicting the k_{off} but the factors behind the ML model that contribute to the prediction can only be understood by analyzing the IFPs. Therefore, we used the IFPs extracted from RAMD as features to study their relationship to the kinetic properties of N-HSP90 inhibitors. The IFPs that are generated from the 90 RAMD trajectories using the python script from Kokh et al.[39] consist of hydrogen bond donor (HD), hydrogen bond acceptor (HA), hydrophobic (HY), salt bridges (IN), $\pi−\pi$ stacking (AR), and halogen−π interaction (HL).

From the PLS coefficient (Figure 4), one can analyze the most important IFPs, which have a strong correlation with the kinetic properties of ligands. These IFPs involve residues like ASN51, SER52, ASP54, ASP93, GLY97, ASP102, and LEU107, most of which are polar and surrounding the ATP binding pocket, TYR139 and THR184 inside the hydrophobic sub-pocket, and PHE138 at the entrance of the hydrophobic sub-pocket. The results are in line with the findings by Kokh et al.[35] A distal residue ILE110 at the exit of the hydrophobic sub-pocket does not interact with ligands in the bound state but shows the importance during the dissociation process. Figure 5 outlines these important residues and their positions on N-HSP90.

Figure 6 shows the schematic visualization of the RAMD dissociation trajectories of the four compounds with the slowest (Figure 6a,b) and fastest (Figure 6c,d) dissociation rates among the 28 simulated inhibitors from dataset B. There are more interactions of 5j6m_ligand than 6ei5_ligand along their dissociation pathways. Aforementioned important resi-
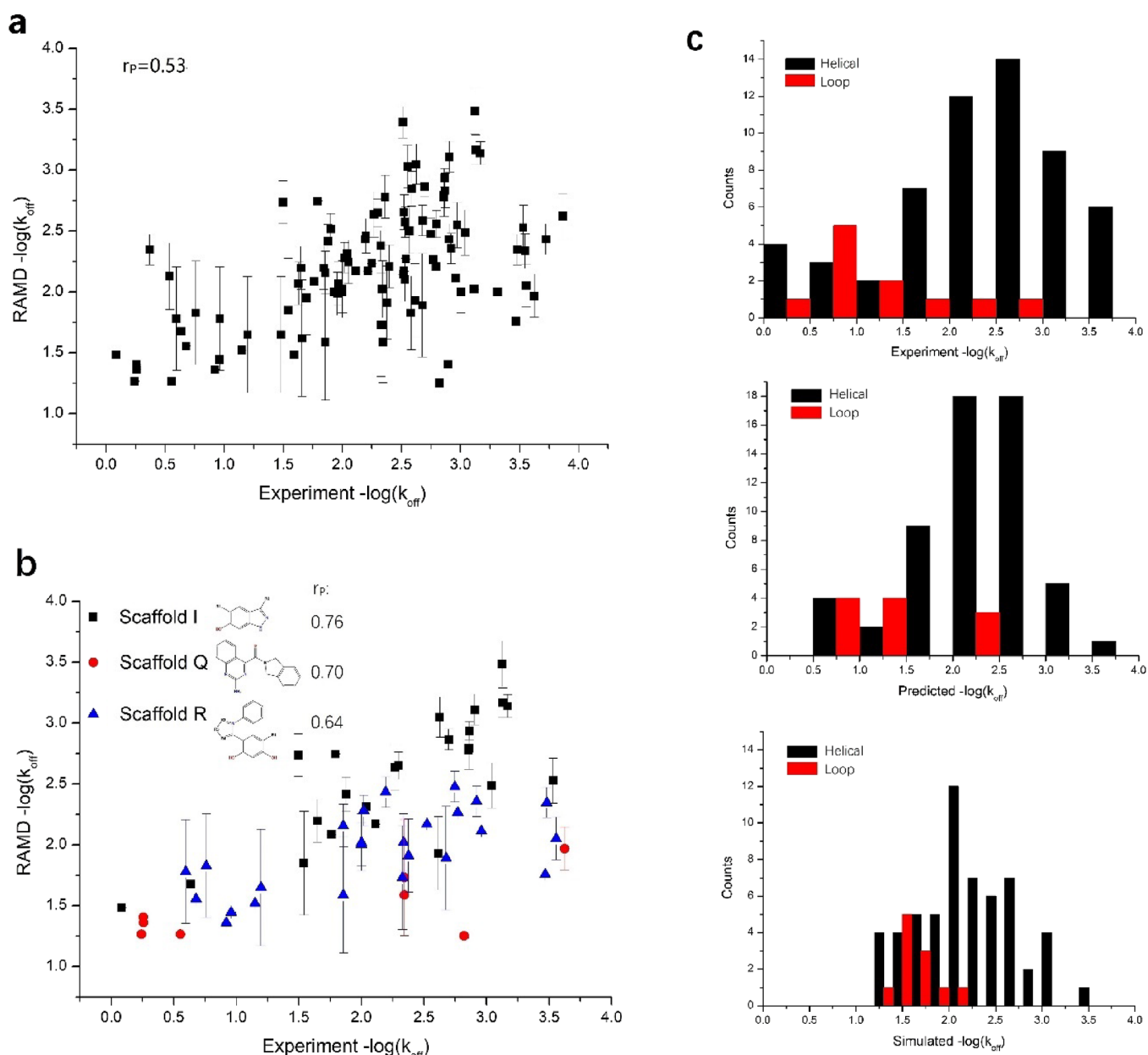
Figure 3. Simulation results for 100 inhibitors of N-HSP90 from dataset A. Scaled residence times plotted vs experiment $-\log(k_{off})$ values on a logarithmic scale (a) for the complete 100 sets of compounds and (b) grouped by the different scaffolds. Scaffold I = hydroxy-indazole, scaffold Q = amino-quinazoline, scaffold R = resorcinol. (c) The $-\log(k_{off})$ population for helical and loop conformation. The correlation coefficient for each set is labeled. k_{off} is scaled according to the linear fitting $-1.2 \times \log(k_{off}) + 3$ of all compounds.

dues, ASN51, ASP93, GLY97, PHE138, and THR184, are all existent and retained during the dissociation process for 5j6m_ligand. However, for 6ei5_ligand, these interactions are either absent or transient during the dissociation process. Furthermore, the dissociation pathways show that for 6ei5_ligand, the transition flow between each intermediate is small while the transition to the unbound state has a larger flow (thicker gray arrow in Figure 6), which suggests a faster dissociation. On the other hand, for 5j6m_ligand, there is a large transition flow between clusters 1 to 6 before it transits into cluster 7, which is the last intermediate before fully unbinding. That means it costs more time for 5j6m_ligand to transit between different intermediates and the transition barrier to the unbound state is higher.

2.5. Validation on Two New N-HSP90 Inhibitors—A Case Study of the Structure–Kinetic Relationship for Drug Design. To conduct a case study, we searched the PDBbind database (http://www.pdbbind.org.cn/) and found two N-HSP90 inhibitors (PDB ID: 6fcj[65] and 6hhr[66]), which have experimental k_{off} data and have not been included in our training datasets. These two inhibitors will be ideal to test our ML/MD protocol performance. 6fcj has a pyrazole scaffold and 6hhr has a triazolone one. The experimental $-\log(k_{off})$ values for 6fcj and 6hhr are 1.47 and 0.95, respectively. This suggests that 6fcj has a retention time four times longer than that of 6hhr. Our model's predicted $-\log(k_{off})$ values for 6fcj and 6hhr are 2.12 and 1.55, respectively, which also show a four times difference in retention time. These results suggest that our model shows promise in predicting $-\log(k_{off})$ values.
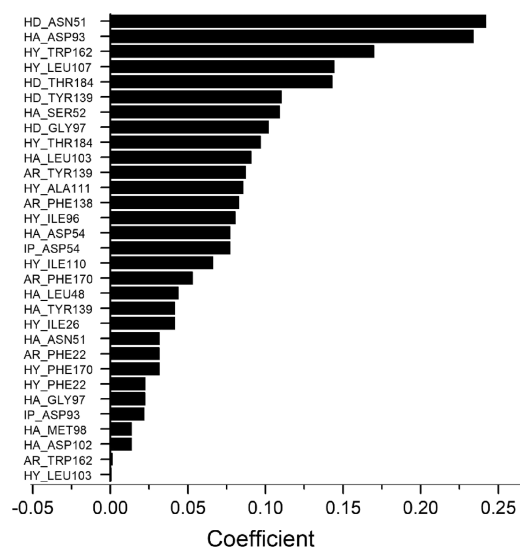
Figure 4. Coefficients of the PLS model in the leave-one-out cross validation built on RAMD IFPs (only the top 30 are shown).

Furthermore, we plan to expand our model to include other scaffolds and explore different scenarios. Additionally, we intend to verify our model's predictions through prospective experimental validation in the future.

To decipher the mechanism behind the improved kinetic properties, we analyzed the trajectories and plot the IFPs along the dissociation path (Figure 7). The detailed RMSD with respect to the bound state of the 90 trajectories and the dissociation pathway in 2D graph representation are in the Supporting Informations. 6hhr_ligand and 6fcj_ligand are both inhibitors for the loop conformation of N-HSP90. There are two structural differences between the two molecules. First, 6hhr_ligand has a triazolone and 6fcj_ligand a pyrazole scaffold. Second, 6hhr_ligand has fluorine and sulfur as the substitution group and 6fcj_ligand has chlorine as a substitution group. Comparing the IFPs between the two

ligands, we found that 6fcj has extra IFPs like HD_ASN51, HA_SER52, HD_LYS58, AR_LYS58, and HY_LEU107, while 6hhr has extra interactions with HA_GLY97, and HY_THR109. As shown in Figure 4, HD_ASN51, HA_SER52, and HY_LEU107 are all important residues. The fact that HA_GLY97 is not as important as HD_GLY97 means that a longer retention time is expected when GLY97 is a hydrogen bond donor not acceptor. Furthermore, both LEU107 and THR109 are at the entrance of the pocket, but LEU107 has more impact on the kinetic properties than THR109, likely due to forming hydrophobic interactions with chlorine.

3. CONCLUSIONS

The RPM-based PLS ML model was successfully applied to predict the k_{off} of 138 inhibitors of N-HSP90 protein. The relative retention time is calculated using RAMD simulation. Both methods can produce the $-\log(k_{off})$, which has a high correlation to the corresponding experimental data. The molecular features from the pre-trained model and the IFPs from RAMD together give more information for the drug design. Using our ML k_{off} prediction approach, we studied two new N-HSP90 inhibitors [6fcj_ligand and 6hhr_ligand]. The predicted $-\log(k_{off})$s show consistent trends with experiments, and the mechanism behind is explained by IFPs. The resulting molecular mechanisms were elucidated in both molecular features and IFPs along the dissociation pathways. Our proposed protocol offers a feasible way for the kinetics-driven drug design approach and increases the success of finding molecules with proper kinetic profiles to the interested targets.

4. MATERIALS AND METHODS

4.1. Retrosynthesis-Based Pre-Trained Molecular Representations (RPM). We present a novel deep learning model using numerous authentic reaction data for molecular representation learning. Based on our previous work for retrosynthesis prediction,[51] we cut off the connections between the encoder and decoder stages and output the vectors as the
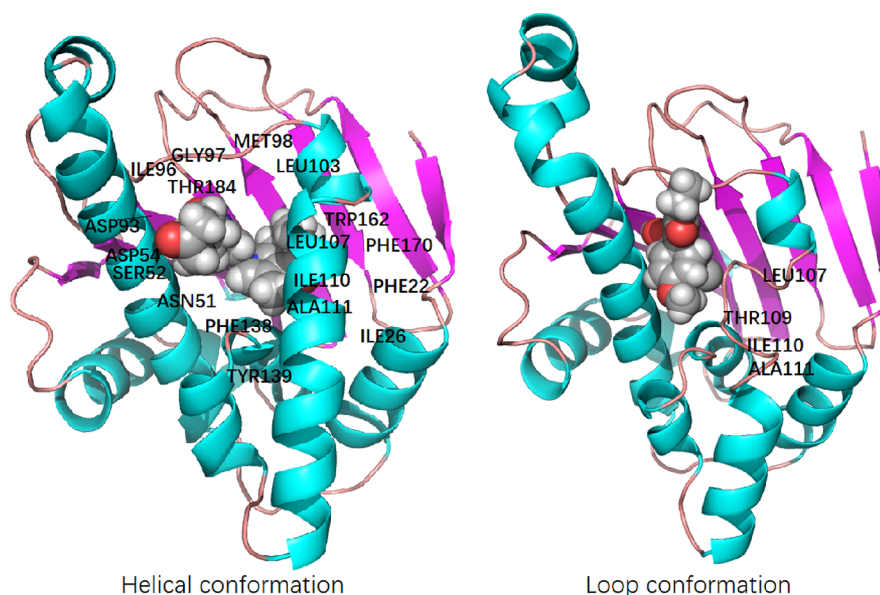

Figure 5. The helical and loop conformation of N-HSP90; the important residues are labeled. The protein is shown in cartoon representation with the helix in cyan, the strand in magenta, and the loop in orange. The inhibitor inside the pocket is shown in sphere.
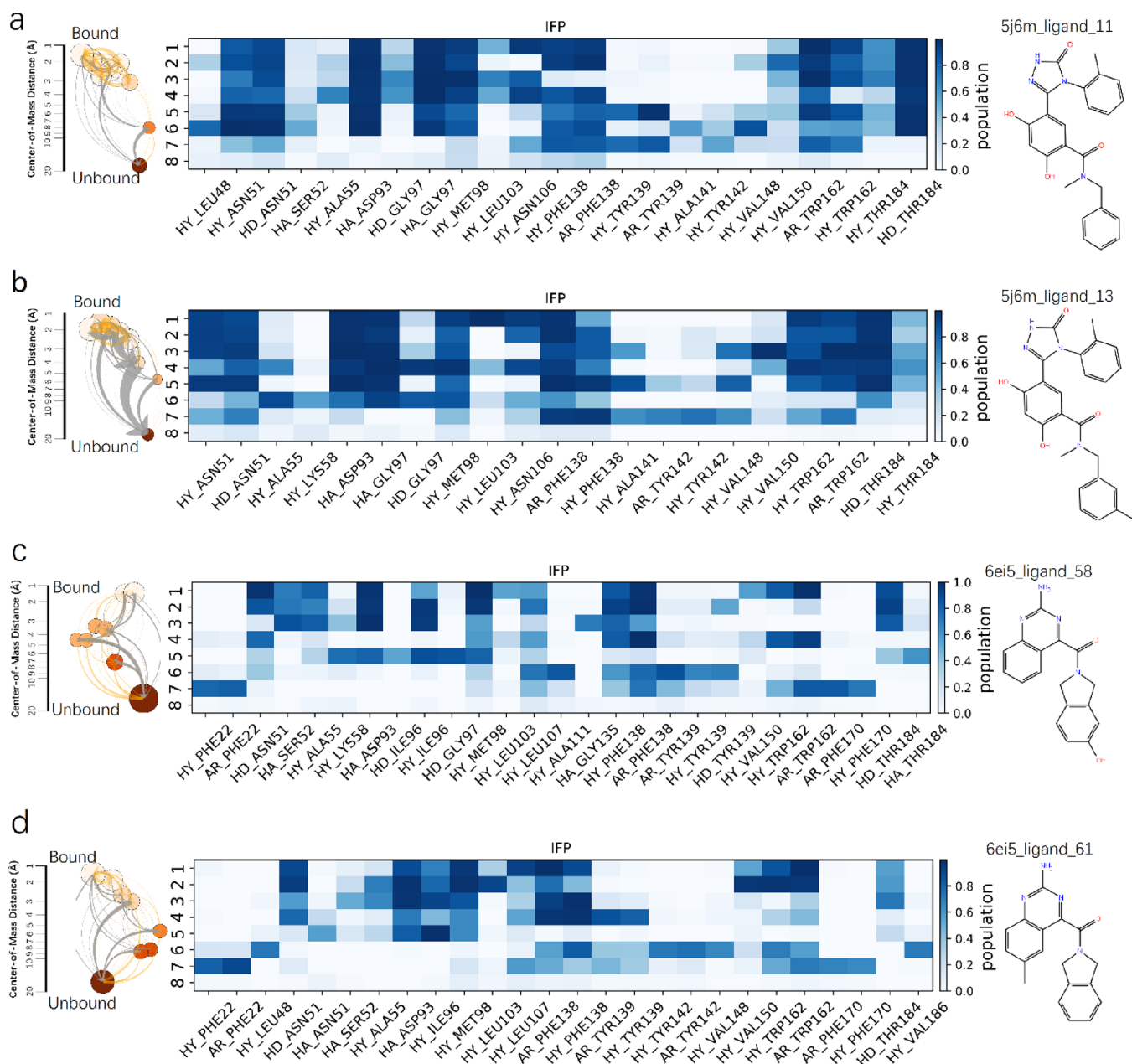
Figure 6. Schematic visualization of the RAMD dissociation trajectories of (a, b) 5j6m_ligand (PDB ID: 5j6m) and (c, d) 6ei5_ligand (PDB ID: 6ei5) of N-HSP90. From left to right: 2D graph representation of the dissociation pathways; IFP composition of each cluster along the dissociation pathways (cluster 1 is the bound state, and cluster 8 is the unbound state). Chemical structures of the compound. Each cluster is shown by a node with the size indicating the cluster population and the color indicating the position from the pocket. The width of the light-orange arrows is proportional to the number of corresponding transitions between two nodes, and the gray arrows indicate the total flow between two nodes. The thicker the gray arrow is, the more chance of transfer from one node to another the larger flow is.

representation of the input molecule. Because our model utilizes the data of the paired reactant product of chemical reactions, rather than the traditional single-molecule properties, we believe our model extracts more information on molecular functional groups and reactivities, especially certain interactions between molecules that current existing models do not take advantage of during training. Therefore, as a representation that takes consideration of molecular reactivity and intermolecular relationships, our model shows superior performance in property prediction such as k_{off} (Figure 1).

4.2. PLS Regression. As an extension of linear regression, multiple linear regression (MLR), and principal component regression (PCR), PLS regression is used to investigate multicollinearity between two groups of variables. For big-p, little-n problems, using PLS to build a model has advantages over MLR and other methods.[67] Moreover, PLS has been successfully applied in k_{off} prediction for HSP90, HIV-1, p38-MARK, and adenosine receptor with a small kinetics dataset.[41,43,58,68] In this paper, we use PLS regression and RPM molecular representation to predict k_{off}.

4.3. RAMD Simulations and IFPs. The relative retention times for the inhibitors were calculated from RAMD, developed by Lüdemann et al.[69] Short, unbiased MD simulations were independently performed six times, 10 ns each run, for each ligand to obtain the equilibrium ensemble of the bound state. A random force was then applied to the ligand
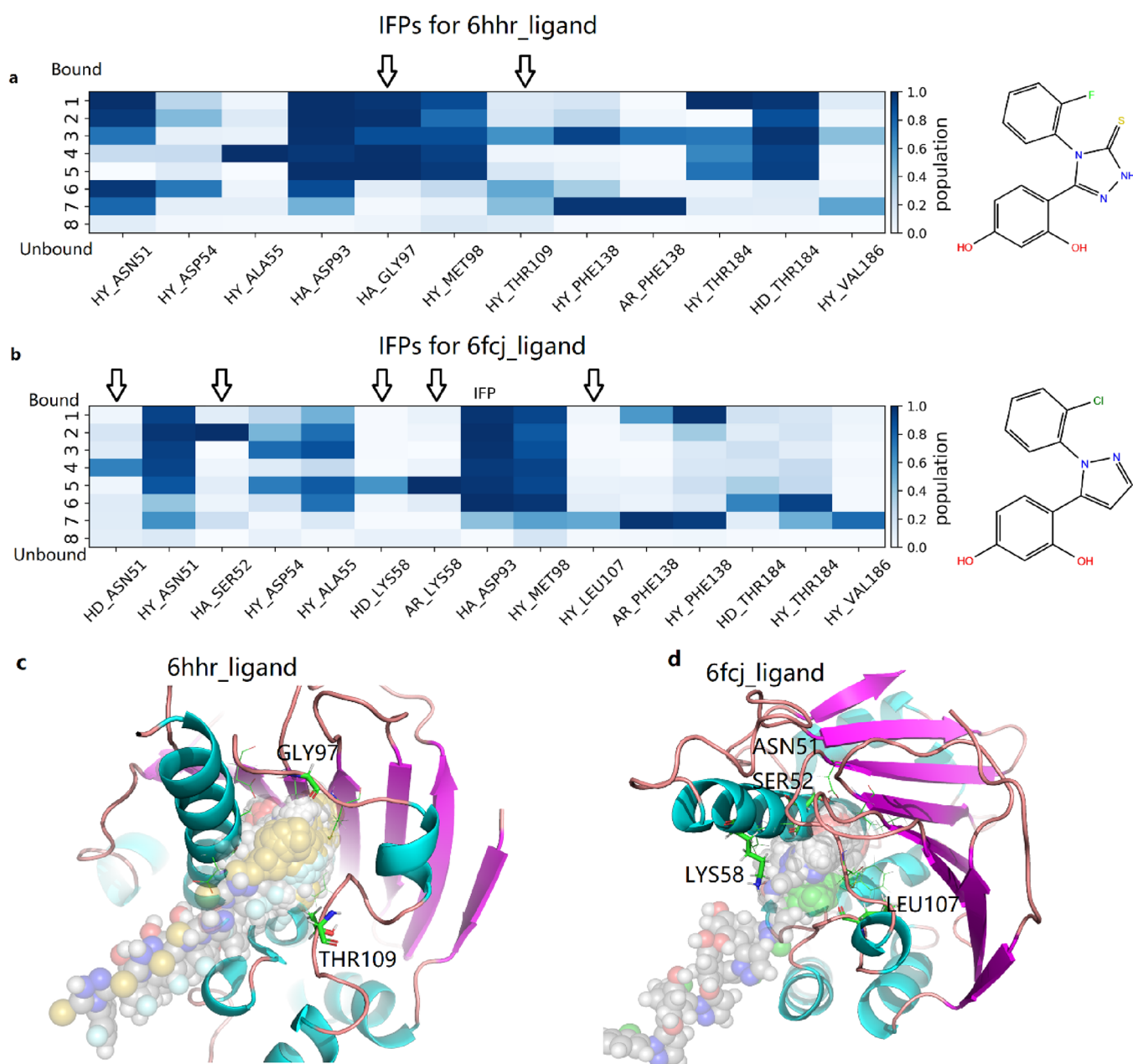
Figure 7. Schematic visualization of the RAMD dissociation trajectories of 6fcj_ligand (PDB ID: 6fcj) and 6hhr_ligand (PDB ID: 6hhr). (a, b) IFP composition of each cluster along the dissociation pathways and the chemical structures of the compound. The important residues, which make the difference between 6fcj_ligand and 6hhr_ligand, are indicated by an arrow. (c, d) The pocket and the important residues along the dissociation pathway. The ligand is show in sphere for each frame along the dissociation pathway.

to accelerate the dissociation rate of the ligands. The force was set to be 20 kcal/mol/Å, and the calculated retention time can be controlled to be at a nanosecond scale. The retention time was calculated as the average value for the 90 independent RAMDs. In all MD simulations, the amber99sb-star-ildn force field[70] and TIP3P model[71] were used for protein and water. The ligand force field was parametrized with Gaff2[72] using ACPYE[73] with AM1-BCC charges.[74,75] The overall temperature of the system was kept constant, pairing independently for each group at 300 K with a Nosé−Hoover thermostat.[76,77] The pressure was coupled to a Parrinello−Rahman barostat[78] at 1 atm separately in every dimension. The temperature and pressure time constants of the coupling were 1 and 5 ps, respectively. The integration of the equations of motion was

performed by using a leapfrog algorithm with a time step of 2 fs. Periodic boundary conditions were implemented in all systems. A cutoff of 1.2 nm was implemented for the Lennard-Jones and the direct space part of the Ewald sum for Coulombic interactions. The Fourier space part of the Ewald splitting was computed by using the particle mesh Ewald method,[79] with a grid length of 0.12 nm on the side and a cubic spline interpolation. The MD simulation package used is Gromacs 2019.[80,81]

■ ASSOCIATED CONTENT

Data Availability Statement

All data are included in the Supporting Information. The training and test datasets are downloaded from PDB and the

supporting information of two literatures by Amangeldiuly et al.[47] and Liu et al.[54] The GEM, MPG, RDKit, and RPM features and the simulated retention time values in csv format can be downloaded from the Supporting Information.

ⓈⒾ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.3c02294.

Comparison of the performance of different molecular features and fingerprint for k_{off} prediction of the 100 inhibitors of N-HSP90 by PLS regression. Predicted and simulated $-\log(k_{off})$ vs experiment values of the ligands of N-HSP90 from datasets A and B. The chemical structures of 6fcj_ligand and 6hhr_ligand, their RMSD with respect to the bound state, and the dissociation pathway in 2D graph representation. The ligand's optimized structure coordinates with AM1-BCC charges in mol2 format (PDF)

The RPM, GEM, MPG, RDKit features, and the simulated retention time values in csv format (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

Yingsheng J. Zhang − *Beijing StoneWise Technology Co Ltd., Beijing 100080, China;* ⓞ orcid.org/0000-0003-2520-3923; Phone: (508)397-7337; Email: zhangyingsheng@stonewise.cn

Authors

Feng Zhou − *Beijing StoneWise Technology Co Ltd., Beijing 100080, China;* ⓞ orcid.org/0000-0001-9040-1796

Shiqiu Yin − *Beijing StoneWise Technology Co Ltd., Beijing 100080, China;* ⓞ orcid.org/0000-0001-8125-0367

Yi Xiao − *Beijing StoneWise Technology Co Ltd., Beijing 100080, China*

Zaiyun Lin − *Beijing StoneWise Technology Co Ltd., Beijing 100080, China*

Weiqiang Fu − *Beijing StoneWise Technology Co Ltd., Beijing 100080, China*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acsomega.3c02294

Author Contributions

The manuscript was written through the contributions of all authors. All authors have approved the final version of the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ ABBREVIATIONS

AI, artificial intelligence; ML, machine learning; MTML, multi-target machine learning; GEM, geometry-enhanced molecular representation; MPG, molecular pre-training graph-based deep learning framework; RPM, retrosynthesis-based pre-trained molecular representation; PLS, partial least squares; MD, molecular dynamics; InMetaD, infrequent metadynamics; US, umbrella sampling; SMD, steered molecular dynamics; PMF, potential of mean force; sMD, scaled molecular dynamics; RAMD, random accelerated molecular dynamics; MM/GBSA, molecular mechanics with generalized Born and surface area solvation; QSAR, quantitative structure−activity relationship; QSKR, quantitative structure-kinetics relationship; COMBINE, comparative binding energy; IFPs, protein−ligand interaction fingerprints; RT, retention time; N-HSP90, N-terminal domain of heat shock protein 90α; kon, association rate constant; koff, dissociation rate constant; K_i, inhibition constant; K_D, dissociation constant; IC50, the half maximal inhibitory concentration

■ REFERENCES

(1) Copeland, R. A.; Pompliano, D. L.; Meek, T. D. Drug-Target Residence Time and Its Implications for Lead Optimization. *Nat. Rev. Drug Discovery* **2006**, *5*, 730−739.

(2) Guo, D.; Mulder-Krieger, T.; IJzerman, A. P.; Heitman, L. H. Functional Efficacy of Adenosine A2A Receptor Agonists Is Positively Correlated to Their Receptor Residence Time. *Br. J. Pharmacol.* **2012**, *166*, 1846−1859.

(3) Schuetz, D. A.; de Witte, W. E. A.; Wong, Y. C.; Knasmueller, B.; Richter, L.; Kokh, D. B.; Sadiq, S. K.; Bosma, R.; Nederpelt, I.; Heitman, L. H.; Segala, E.; Amaral, M.; Guo, D.; Andres, D.; Georgi, V.; Stoddart, L. A.; Hill, S.; Cooke, R. M.; De Graaf, C.; Leurs, R.; Frech, M.; Wade, R. C.; de Lange, E. C. M.; IJzerman, A. P.; Müller-Fahrnow, A.; Ecker, G. F. Kinetics for Drug Discovery: An Industry-Driven Effort to Target Drug Residence Time. *Drug Discovery Today* **2017**, *22*, 896−911.

(4) Folmer, R. H. A. Drug Target Residence Time: A Misleading Concept. *Drug Discovery Today* **2018**, *23*, 12−16.

(5) Yang, T.; Cuesta, A.; Wan, X.; Craven, G. B.; Hirakawa, B.; Khamphavong, P.; May, J. R.; Kath, J. C.; Lapek, J. D., Jr.; Niessen, S.; Burlingame, A. L.; Carelli, J. D.; Taunton, J. Reversible Lysine-Targeted Probes Reveal Residence Time-Based Kinase Selectivity. *Nat. Chem. Biol.* **2022**, *18*, 934−941.

(6) Ferguson, F. M. Tuned Out. *Nat. Chem. Biol.* **2022**, *18*, 917−918.

(7) Khanna, A.; Côté, A.; Arora, S.; Moine, L.; Gehling, V. S.; Brenneman, J.; Cantone, N.; Stuckey, J. I.; Apte, S.; Ramakrishnan, A.; Bruderek, K.; Bradley, W. D.; Audia, J. E.; Cummings, R. T.; Sims, R. J., III; Trojer, P.; Levell, J. R. Design, Synthesis, and Pharmacological Evaluation of Second Generation EZH2 Inhibitors with Long Residence Time. *ACS Med. Chem. Lett.* **2020**, *11*, 1205−1212.

(8) Zhou, Y.; Fu, Y.; Yin, W.; Li, J.; Wang, W.; Bai, F.; Xu, S.; Gong, Q.; Peng, T.; Hong, Y.; Zhang, D.; Zhang, D.; Liu, Q.; Xu, Y.; Xu, H. E.; Zhang, H.; Jiang, H.; Liu, H. Kinetics-Driven Drug Design Strategy for Next-Generation Acetylcholinesterase Inhibitors to Clinical Candidate. *J. Med. Chem.* **2021**, *64*, 1844−1855.

(9) Miller, M.; Rossetti, T.; Ferreira, J.; Ghanem, L.; Balbach, M.; Kaur, N.; Levin, L. R.; Buck, J.; Kehr, M.; Coquille, S.; Van Den Heuvel, J.; Steegborn, C.; Fushimi, M.; Finkin-groner, E.; Myers, R. W.; Kargman, S.; Liverton, N. J.; Huggins, D. J.; Meinke, P. T. Design, Synthesis, and Pharmacological Evaluation of Second- Generation Soluble Adenylyl Cyclase (SAC, ADCY10) Inhibitors with Slow Dissociation Rates. *J. Med. Chem.* **2022**, *65*, 15208−15226.

(10) Bai, F.; Jiang, H. Computationally Elucidating the Binding Kinetics for Different AChE Inhibitors to Access the Rationale for Improving the Drug Efficacy. *J. Phys. Chem. B* **2022**, 7797.

(11) Shan, Y.; Kim, E. T.; Eastwood, M. P.; Dror, R. O.; Seeliger, M. A.; Shaw, D. E. How Does a Drug Molecule Find Its Target Binding Site? *J. Am. Chem. Soc.* **2011**, *133*, 9181−9183.

(12) Pantsar, T.; Kaiser, P. D.; Kudolo, M.; Forster, M.; Rothbauer, U.; Laufer, S. A. Decisive Role of Water and Protein Dynamics in Residence Time of P38α MAP Kinase Inhibitors. *Nat. Commun.* **2022**, *13*, 569.

(13) Buch, I.; Giorgino, T.; De Fabritiis, G. Complete Reconstruction of an Enzyme-Inhibitor Binding Process by Molecular Dynamics Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 10184−10189.

(14) Capelli, R.; Lyu, W.; Bolnykh, V.; Meloni, S.; Olsen, J. M. H.; Rothlisberger, U.; Parrinello, M.; Carloni, P. Accuracy of Molecular Simulation-Based Predictions of KoffValues: A Metadynamics Study. *J. Phys. Chem. Lett.* **2020**, *11*, 6373−6381.

(15) Tiwary, P.; Limongelli, V.; Salvalaglio, M.; Parrinello, M. Kinetics of Protein-Ligand Unbinding: Predicting Pathways, Rates, and Rate-Limiting Steps. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, E386−E391.

(16) Callegari, D.; Lodola, A.; Pala, D.; Rivara, S.; Mor, M.; Rizzi, A.; Capelli, A. M. Metadynamics Simulations Distinguish Short- and Long-Residence-Time Inhibitors of Cyclin-Dependent Kinase 8. *J. Chem. Inf. Model.* **2017**, *57*, 159−169.

(17) Tiwary, P.; Mondal, J.; Berne, B. J. How and When Does an Anticancer Drug Leave Its Binding Site? *Sci. Adv.* **2017**, *3*, No. e1700014.

(18) Casasnovas, R.; Limongelli, V.; Tiwary, P.; Carloni, P.; Parrinello, M. Unbinding Kinetics of a P38 MAP Kinase Type II Inhibitor from Metadynamics Simulations. *J. Am. Chem. Soc.* **2017**, *139*, 4780−4788.

(19) Tiwary, P. Molecular Determinants and Bottlenecks in the Dissociation Dynamics of Biotin-Streptavidin. *J. Phys. Chem. B* **2017**, *121*, 10841−10849.

(20) Pramanik, D.; Smith, Z.; Kells, A.; Tiwary, P. Can One Trust Kinetic and Thermodynamic Observables from Biased Metadynamics Simulations?: Detailed Quantitative Benchmarks on Millimolar Drug Fragment Dissociation. *J. Phys. Chem. B* **2019**, *123*, 3672−3678.

(21) Doudou, S.; Burton, N. A.; Henchman, R. H. Standard Free Energy of Binding from a One-Dimensional Potential of Mean Force. *J. Chem. Theory Comput.* **2009**, *5*, 909−918.

(22) Lan, N. T.; Vu, K. B.; Ngoc, M. K. D.; Tran, P.-T.; Hiep, D. M.; Tung, N. T.; Ngo, S. T. Prediction of AChE-Ligand Affinity Using the Umbrella Sampling Simulation. *J. Mol. Graphics Modell.* **2019**, *93*, No. 107441.

(23) Ngo, S. T.; Vu, K. B.; Bui, L. M.; Vu, V. V. Effective Estimation of Ligand-Binding Affinity Using Biased Sampling Method. *ACS Omega* **2019**, *4*, 3887−3893.

(24) Mascarenhas, N. M.; Kästner, J. How Maltose Influences Structural Changes to Bind to Maltose-Binding Protein: Results from Umbrella Sampling Simulation. *Proteins: Struct., Funct., Bioinf.* **2013**, *81*, 185−198.

(25) You, W.; Tang, Z.; Chang, C. E. A. Potential Mean Force from Umbrella Sampling Simulations: What Can We Learn and What Is Missed? *J. Chem. Theory Comput.* **2019**, *15*, 2433−2443.

(26) Tang, Z.; Chen, S. H.; Chang, C. E. A. Transient States and Barriers from Molecular Simulations and the Milestoning Theory: Kinetics in Ligand-Protein Recognition and Compound Design. *J. Chem. Theory Comput.* **2020**, *16*, 1882−1895.

(27) Braka, A.; Garnier, N.; Bonnet, P.; Aci-Sèche, S. Residence Time Prediction of Type 1 and 2 Kinase Inhibitors from Unbinding Simulations. *J. Chem. Inf. Model.* **2020**, *60*, 342−348.

(28) Spiriti, J.; Wong, C. F. Qualitative Prediction of Ligand Dissociation Kinetics from Focal Adhesion Kinase Using Steered Molecular Dynamics. *Life* **2021**, *11*, 74.

(29) Potterton, A.; Husseini, F. S.; Southey, M. W. Y.; Bodkin, M. J.; Heifetz, A.; Coveney, P. V.; Townsend-Nicholson, A. Ensemble-Based Steered Molecular Dynamics Predicts Relative Residence Time of A 2A Receptor Binders. *J. Chem. Theory Comput.* **2019**, *15*, 3316−3330.

(30) Mollica, L.; Decherchi, S.; Zia, S. R.; Gaspari, R.; Cavalli, A.; Rocchia, W. Kinetics of Protein-Ligand Unbinding via Smoothed Potential Molecular Dynamics Simulations. *Sci. Rep.* **2015**, *5*, 11539.

(31) Bernetti, M.; Rosini, E.; Mollica, L.; Masetti, M.; Pollegioni, L.; Recanatini, M.; Cavalli, A. Binding Residence Time through Scaled Molecular Dynamics: A Prospective Application to HDAAO Inhibitors. *J. Chem. Inf. Model.* **2018**, *58*, 2255−2265.

(32) Mollica, L.; Theret, I.; Antoine, M.; Perron-Sierra, F.; Charton, Y.; Fourquez, J. M.; Wierzbicki, M.; Boutin, J. A.; Ferry, G.; Decherchi, S.; Bottegoni, G.; Ducrot, P.; Cavalli, A. Molecular Dynamics Simulations and Kinetic Measurements to Estimate and Predict Protein-Ligand Residence Times. *J. Med. Chem.* **2016**, *59*, 7167−7176.

(33) Zhou, Y.; Zou, R.; Kuang, G.; Långström, B.; Halldin, C.; Ågren, H.; Tu, Y. Enhanced Sampling Simulations of Ligand Unbinding Kinetics Controlled by Protein Conformational Changes. *J. Chem. Inf. Model.* **2019**, *59*, 3910−3918.

(34) Kokh, D. B.; Amaral, M.; Bomke, J.; Grädler, U.; Musil, D.; Buchstaller, H. P.; Dreyer, M. K.; Frech, M.; Lowinski, M.; Vallee, F.; Bianciotto, M.; Rak, A.; Wade, R. C. Estimation of Drug-Target Residence Times by τ-Random Acceleration Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2018**, *14*, 3859−3869.

(35) Kokh, D. B.; Kaufmann, T.; Kister, B.; Wade, R. C. Machine Learning Analysis of TRAMD Trajectories to Decipher Molecular Determinants of Drug-Target Residence Times. *Front. Mol. Biosci.* **2019**, *6*, 36.

(36) Berger, B. T.; Amaral, M.; Kokh, D. B.; Nunes-Alves, A.; Musil, D.; Heinrich, T.; Schröder, M.; Neil, R.; Wang, J.; Navratilova, I.; Bomke, J.; Elkins, J. M.; Müller, S.; Frech, M.; Wade, R. C.; Knapp, S. Structure-Kinetic Relationship Reveals the Mechanism of Selectivity of FAK Inhibitors over PYK2. *Cell Chem. Biol.* **2021**, *28*, 686−698.e7.

(37) Nunes-Alves, A.; Kokh, D. B.; Wade, R. C. Ligand Unbinding Mechanisms and Kinetics for T4 Lysozyme Mutants from τRAMD Simulations. *Curr. Res. Struct. Biol.* **2021**, *3*, 106−111.

(38) Kokh, D. B.; Wade, R. C. G Protein-Coupled Receptor-Ligand Dissociation Rates and Mechanisms from TrAMD Simulations. *J. Chem. Theory Comput.* **2021**, *17*, 6610−6623.

(39) Kokh, D. B.; Doser, B.; Richter, S.; Ormersbach, F.; Cheng, X.; Wade, R. C. A Workflow for Exploring Ligand Dissociation from a Macromolecule: Efficient Random Acceleration Molecular Dynamics Simulation and Interaction Fingerprint Analysis of Ligand Trajectories. *J. Chem. Phys.* **2020**, *153*, No. 125102.

(40) Chiu, S. H.; Xie, L. Toward High-Throughput Predictive Modeling of Protein Binding/Unbinding Kinetics. *J. Chem. Inf. Model.* **2016**, *56*, 1164−1174.

(41) Zhang, D.; Huang, S.; Mei, H.; Kevin, M.; Shi, T.; Chen, L. Protein−Ligand Interaction Fingerprints for Accurate Prediction of Dissociation Rates of P38 MAPK Type II Inhibitors. *Integr. Biol.* **2019**, *11*, 53−60.

(42) Qu, S.; Huang, S.; Pan, X.; Yang, L.; Mei, H. Constructing Interconsistent, Reasonable, and Predictive Models for Both the Kinetic and Thermodynamic Properties of HIV-1 Protease Inhibitors. *J. Chem. Inf. Model.* **2016**, *56*, 2061−2068.

(43) Ganotra, G. K.; Wade, R. C. Prediction of Drug-Target Binding Kinetics by Comparative Binding Energy Analysis. *ACS Med. Chem. Lett.* **2018**, *9*, 1134−1139.

(44) Huang, S.; Zhang, D.; Mei, H.; Kevin, M. Y.; Qu, S.; Pan, X.; Lu, L. SMD-Based Interaction-Energy Fingerprints Can Predict Accurately the Dissociation Rate Constants of HIV-1 Protease Inhibitors. *J. Chem. Inf. Model.* **2019**, *59*, 159−169.

(45) Bray, S.; Tänzel, V.; Wolf, S. Ligand Unbinding Pathway and Mechanism Analysis Assisted by Machine Learning and Graph Methods. *J. Chem. Inf. Model.* **2022**, 4591.

(46) Ballester, P. J.; Mitchell, J. B. O. A Machine Learning Approach to Predicting Protein-Ligand Binding Affinity with Applications to Molecular Docking. *Bioinformatics* **2010**, *26*, 1169−1175.

(47) Amangeldiuly, N.; Karlov, D.; Fedorov, M. V. Baseline Model for Predicting Protein-Ligand Unbinding Kinetics through Machine Learning. *J. Chem. Inf. Model.* **2020**, *60*, 5946−5956.

(48) Fang, X.; Liu, L.; Lei, J.; He, D.; Zhang, S.; Zhou, J.; Wang, F.; Wu, H.; Wang, H. Geometry-Enhanced Molecular Representation Learning for Property Prediction. *Nat. Mach. Intell.* **2022**, *4*, 127−134.

(49) Li, P.; Wang, J.; Qiao, Y.; Chen, H.; Yu, Y.; Yao, X.; Gao, P.; Xie, G.; Song, S. An Effective Self-Supervised Framework for Learning Expressive Molecular Global Representations to Drug Discovery. *Briefings Bioinf.* **2021**, *22*, bbab109.

(50) Landrum, G. *RDKit Documentation*; (2019.09.1). Https://Www.Rdkit.Org/ 2019.

(51) Lin, Z.; Yin, S.; Shi, L.; Zhou, W.; Zhang, Y. G2GT: Retrosynthesis Prediction with Graph to Graph Attention Neural Network and Self-Training. *J. Chem. Inf. Model.* **2023**, 1894.

(52) Hendrycks, D.; Lee, K.; Mazeika, M. Using Pre-Training Can Improve Model Robustness and Uncertainty. In *International Conference on Machine Learning*; Proceedings of Machine Learning Research, 2019, 97, 2712.

(53) David, L.; Thakkar, A.; Mercado, R.; Engkvist, O. Molecular Representations in AI-Driven Drug Discovery: A Review and Practical Guide. *Aust. J. Chem.* **2020**, 12, 1−22.

(54) Liu, H.; Su, M.; Lin, H. X.; Wang, R.; Li, Y. Public Data Set of Protein-Ligand Dissociation Kinetic Constants for Quantitative Structure-Kinetics Relationship Studies. *ACS Omega* **2022**, 7, 18985−18996.

(55) Brough, P. A.; Aherne, W.; Barril, X.; Borgognoni, J.; Boxall, K.; Cansfield, J. E.; Cheung, K.-M. J.; Collins, I.; Davies, N. G. M.; Drysdale, M. J.; Dymock, B.; Eccles, S. A.; Finch, H.; Fink, A.; Hayes, A.; Howes, R.; Hubbard, R. E.; James, K.; Jordan, A. M.; Lockie, A.; Martins, V.; Massey, A.; Matthews, T. P.; McDonald, E.; Northfield, C. J.; Pearl, L. H.; Prodromou, C.; Ray, S.; Raynaud, F. I.; Roughley, S. D.; Sharp, S. Y.; Surgenor, A.; Walmsley, D. L.; Webb, P.; Wood, M.; Workman, P.; Wright, L. 4,5-Diarylisoxazole HSP90 Chaperone Inhibitors: Potential Therapeutic Agents for the Treatment of Cancer. *J. Med. Chem.* **2008**, 51, 196−218.

(56) Drysdale, M. J.; Dymock, B. W.; Finch, H.; Webb, P.; McDonald, E.; James, K. E.; Cheung, K.-M.; Mathews, T. P. Preparation of Isoxazoles as Inhibitors of Heat Shock Proteins. PCT International Application, WO04072051 2004.

(57) Bianciotto, M.; Gkeka, P.; Kokh, D. B.; Wade, R. C.; Minoux, H. Contact Map Fingerprints of Protein-Ligand Unbinding Trajectories Reveal Mechanisms Determining Residence Times Computed from Scaled Molecular Dynamics. *J. Chem. Theory Comput.* **2021**, 17, 6522−6535.

(58) Huang, S.; Chen, L.; Mei, H.; Zhang, D.; Shi, T.; Kuang, Z.; Heng, Y.; Xu, L.; Pan, X. In Silico Prediction of the Dissociation Rate Constants of Small Chemical Ligands by 3D-Grid-Based Volsurf Method. *Int. J. Mol. Sci.* **2020**, 21, 2456.

(59) Babbitt, G. A.; Fokoue, E. P.; Evans, J. R.; Diller, K. I.; Adams, L. E. DROIDS 3.0—Detecting Genetic and Drug Class Variant Impact on Conserved Protein Binding Dynamics. *Biophys. J.* **2020**, 118, 541−551.

(60) Nunes-Alves, A.; Kokh, D. B.; Wade, R. C. Recent Progress in Molecular Simulation Methods for Drug Binding Kinetics. *Curr. Opin. Struct. Biol.* **2020**, 64, 126−133.

(61) Baumann, H. M.; Gapsys, V.; De Groot, B. L.; Mobley, D. L. Challenges Encountered Applying Equilibrium and Nonequilibrium Binding Free Energy Calculations. *J. Phys. Chem. B* **2021**, 125, 4241−4261.

(62) D'Annessa, I.; Raniolo, S.; Limongelli, V.; Di Marino, D.; Colombo, G. Ligand Binding, Unbinding, and Allosteric Effects: Deciphering Small-Molecule Modulation of HSP90. *J. Chem. Theory Comput.* **2019**, 15, 6368−6381.

(63) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, 28, 235−242.

(64) Amaral, M.; Kokh, D. B.; Bomke, J.; Wegener, A.; Buchstaller, H. P.; Eggenweiler, H. M.; Matias, P.; Sirrenberg, C.; Wade, R. C.; Frech, M. Protein Conformational Flexibility Modulates Kinetics and Thermodynamics of Drug Binding. *Nat. Commun.* **2017**, 8, 2276.

(65) Wolf, S.; Amaral, M.; Lowinski, M.; Vallée, F.; Musil, D.; Güldenhaupt, J.; Dreyer, M. K.; Bomke, J.; Frech, M.; Schlitter, J.; Gerwert, K. Estimation of Protein-Ligand Unbinding Kinetics Using Non-Equilibrium Targeted Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2019**, 59, 5135−5147.

(66) Schuetz, D. A.; Bernetti, M.; Bertazzo, M.; Musil, D.; Eggenweiler, H. M.; Recanatini, M.; Masetti, M.; Ecker, G. F.; Cavalli, A. Predicting Residence Time and Drug Unbinding Pathway through Scaled Molecular Dynamics. *J. Chem. Inf. Model.* **2019**, 59, 535−549.

(67) Göktaş, A.; Akkuş, Ö. Comparison of Partial Least Squares with Other Prediction Methods via Generated Data. *J. Stat. Comput. Simul.* **2020**, 90, 3009−3024.

(68) Nunes-Alves, A.; Ormersbach, F.; Wade, R. C. Prediction of the Drug−Target Binding Kinetics for Flexible Proteins by Comparative Binding Energy Analysis. *J. Chem. Inf. Model.* **2021**, 61, 3708−3721.

(69) Lüdemann, S. K.; Lounnas, V.; Wade, R. C. How Do Substrates Enter and Products Exit the Buried Active Site of Cytochrome P450cam? 1. Random Expulsion Molecular Dynamics Investigation of Ligand Access Channels and Mechanisms. *J. Mol. Biol.* **2000**, 303, 797−811.

(70) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber Ff99SB Protein Force Field. *Proteins: Struct., Funct., Bioinf.* **2010**, 78, 1950−1958.

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

(72) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graphics Modell.* **2006**, 25, 247−260.

(73) Sousa da Silva, A. W.; Vranken, W. F. ACPYPE - AnteChamber PYthon Parser InterfacE. *BMC Res. Notes* **2012**, 5, 367.

(74) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. *J. Comput. Chem.* **2000**, 21, 132−146.

(75) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *J. Comput. Chem.* **2002**, 23, 1623−1641.

(76) Nosé, S. A Unified Formulation of the Constant Temperature Molecular Dynamics Methods. *J. Chem. Phys.* **1984**, 81, 511−519.

(77) Hoover, W. G. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A* **1985**, 31, 1695−1697.

(78) Parrinello, M.; Rahman, A. POLYMORPHIC TRANSITIONS IN SINGLE-CRYSTALS - A NEW MOLECULAR-DYNAMICS METHOD. *J. Appl. Phys.* **1981**, 52, 7182−7190.

(79) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, 98, 10089−10092.

(80) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, 4, 435−447.

(81) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. Gromacs: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, 1-2, 19−25.