JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

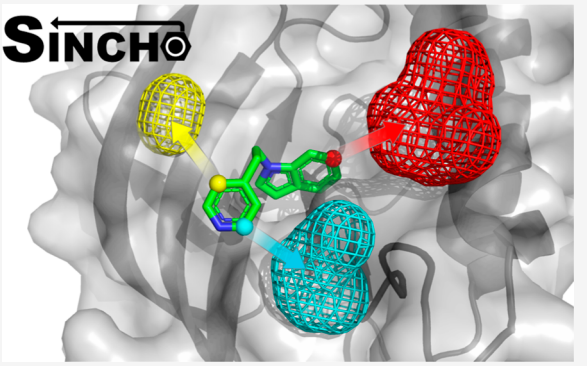
pubs.acs.org/jcim Article

Site Identification and Next Choice Protocol for Hit-to-Lead Optimization

Genki Kudo, Takumi Hirao, Ryunosuke Yoshino, Yasuteru Shigeta, and Takatsugu Hirokawa*

Cite This: J. Chem. Inf. Model. 2024, 64, 4475−4484 Read Online

ABSTRACT: Time efficiency and cost savings are major challenges in drug discovery and development. In this process, the hit-to-lead stage is expected to improve efficiency because it primarily exploits the trial-and-error approach of medicinal chemists. This study proposes a site identification and next choice (SINCHO) protocol to improve the hit-to-lead efficiency. This protocol selects an anchor atom and growth site pair, which is desirable for a hit-to-lead strategy starting from a 3D complex structure. We developed and fine-tuned the protocol using a training data set and assessed it using a test data set of the preceding hit-to-lead strategy. The protocol was tested for experimentally determined structures and molecular dynamics (MD) ensembles. The protocol had a high prediction accuracy for applying MD ensembles, owing to the consideration of protein flexibility. The SINCHO protocol enables medicinal chemists to visualize and modify functional groups in a hit-to-lead manner.



1. INTRODUCTION

Drug discovery processes up to the preclinical phase comprise three stages: target-to-hit, hit-to-lead, and lead optimization.[1] In the target-to-hit stage, hit compounds with affinity for the target are selected from the chemical library.[2] Each hit compound is added to the functional groups, and lead compounds with higher affinities than those of the hit compound are generated in the hit-to-lead stage.[3] In the third stage, the lead compound is optimized to maintain a favorable affinity and improve pharmacokinetics.[2] Recently, target-to-hit and lead optimization stages have become more efficient owing to computational tools such as docking simulation, molecular dynamics (MD) simulation, and artificial intelligence.[4−15]

In contrast, the hit-to-lead stage is expected to improve efficiency because it primarily exploits the trial-and-error approach of medicinal chemists.[16] In the early hit-to-lead stage, an appropriate selection of reaction points and a set of substitutes are required.[17−21] Medicinal chemists must make the right choices because their choices affect the affinity of hit derivatives and the efficiency of the hit-to-lead stage.[22,23] Even if structural information on the protein−hit compound complex is available, making decisions is difficult because numerous parameters must be considered, including physicochemical properties, target selectivity, and synthetic accessibility.[24] Therefore, a computational support tool for these decisions is required to improve the quality of lead compounds and the efficiency of the hit-to-lead stage.

In this study, we propose a support protocol for early decisions in the hit-to-lead stage, designated as the site identification and next choice (SINCHO) protocol. The SINCHO protocol suggests the anchor atom and growth site pairs as the lead strategy in the hit-to-lead stage based on the 3D protein−hit compound complex structure. Our protocol comprises three steps: site identification, anchor atom selection, and next choice. For site identification, the growth site, where the lead substructure can be placed around the hit compound, was searched using the Pocket to Concavity (P2C) tool, as proposed in our previous study.[35] For each detected growth site, the starting atom for modification with functional groups was specified based on the geometrical criteria between the heavy atom of the hit compound and the growth site for anchor atom selection. In the final step, desirable pairs between the anchor atom and growth site were selected based on the scoring function. To develop the SINCHO protocol, the geometrical criteria and coefficients of the scoring function were fine-tuned using 3D complex structures with preceding hit-to-lead cases. The fine-tuned protocol was tested using experimentally determined protein and hit compound complex structures and these MD ensembles. Furthermore, we discussed the validity of the protocol through an additional docking study.

Received: December 20, 2023
Revised: April 11, 2024
Accepted: April 11, 2024
Published: May 20, 2024

https://doi.org/10.1021/acs.jcim.3c02036
J. Chem. Inf. Model. 2024, 64, 4475−4484

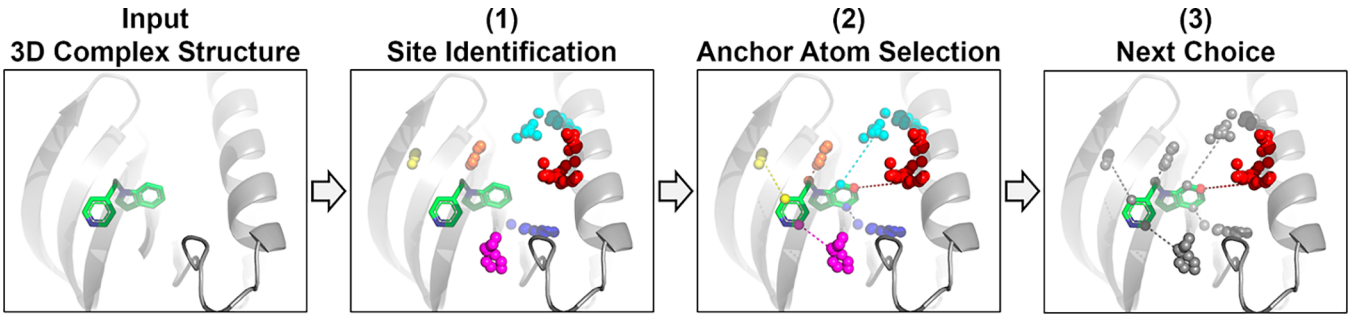


Figure 1. Workflow of the SINCHO protocol. The target protein and hit compound are shown as gray cartoon and green stick models, respectively. For site identification, the identified growth sites are shown as sphere cluster models. In the second step, the anchor atom candidates for each growth site are shown as ball models colored the same for each growth site.

2. METHODS

2.1. Training Data Set Preparation. To develop the SINCHO protocol, a training data set was prepared as follows. We first obtained preceding hit-to-lead cases from previous studies.[25−33] Among these cases, the training data set was selected based on the following criteria: (1) cases with known 3D complex structures between the target protein and lead compound and (2) cases with common structures between the hit and lead compounds. Consequently, 41 protein−lead compound complex structures were selected (Figure S1). Subsequently, starting from each lead compound 3D structure, a cognate hit compound was generated by removing the functional groups that were modified in the preceding hit-to-lead cases. The scaffold of each cognate hit compound was referenced from previous studies, and cases with scaffold differences between the hit and lead compounds were excluded from the training data set. Hydrogen atoms were then added to the cognate hit compound using OpenBabel 3.1.0.[34] The protonation state of each cognate hit compound was refined based on the bond order in the previous study. Prepared cognate hit compounds and original protein structures were merged, and the generated complex structure was designated as the "soaking structure". The correct anchor atoms and functional groups of the original 3D lead compounds were annotated for each soaked structure. Functional groups with fewer than three heavy atoms were excluded from the data set. Cognate hit and original lead compounds in the training data set are shown in the Supporting Information (Table S1). Consequently, 56 preceding pairs between the correct anchor atom and functional group in the 41 complex structures were prepared as the training data set. Note that cases with one functional group starting from two atoms were defined as one preceding pair, as shown in the example of 1QJ7 in Table S1.

2.2. SINCHO Protocol. The SINCHO protocol requires only a 3D complex structure of the target protein and hit compound as the input. Based on this structure, the SINCHO protocol proceeds in three steps: (1) site identification, (2) anchor atom selection, and (3) next choice (Figure 1).

2.2.1. Site Identification. In the first step of the SINCHO protocol, the growth sites where the lead functional groups can be placed were identified. P2C was used for the first step.[35] P2C accurately searches for deep concavities in the target protein based on the default pockets obtained from sphere-based pocket detection software. The ligand-bound (LB) mode in P2C provides growth sites around the hit compound. In our protocol, fpocket2[36] was used as the default pocket detection software for P2C. The LB mode was applied to the 3D complex structure, with the cutoff of the search for the growth site set to 10 Å from the hit compound. Site identification was conducted for each soaking structure in the training data set. The growth site that overlapped most with the functional group was annotated as the correct growth site.

2.2.2. Anchor Atom Selection. The anchor atom candidates, the starting atoms for modifying the functional groups, were selected for each growth site based on geometric criteria (Figure 2A). The criteria for the appropriate atom were defined as follows: (A) a heavy atom (HA) with a hydrogen atom (Hyd), (B) an angle between the center of mass (COM) of the growth site, HA, and Hyd of less than 90°, and (C) the first or second closest to the COM among the HAs satisfying A and B. These criteria were set based on the geometric conditions between the correct anchor atom and growth site in the training data set. Figure 2B shows the angle−distance plot between all HAs of the hit compound and the correct growth site in the training data set. Red/blue plots indicate the geometric conditions between the correct and incorrect anchor atom and the correct growth site. In fact, this plot shows that approximately 95% of the correct anchor atoms in the training data set had the second closest distance among the HAs with an angle less than 90°. This indicated that a HA with close and linear access to the growth site was appropriate as the anchor atom.

For each growth site detected by (1) in the training data set, two HAs were selected as anchor atom candidates. Thereafter, an anchor atom and a growth site were paired as one of the choices in the SINCHO protocol.

2.2.3. Next Choice. Among the candidate pairs between the anchor atom and growth site, the desirable pair for hit-to-lead was predicted based on the scoring function. We defined the scoring function as the "Extend Score (ES)." ES is given as follows

$$\text{Extend Score} = w_1 \times \text{Dist} + w_2 \times \log_{10}\left(\frac{1}{\text{DS}}\right) + w_3 \times \Delta\text{SA} \qquad (w_1 + w_2 + w_3 = 1) \tag{1}$$

The ES is the scoring function based on the descriptor of the complex structure for the emulation of a medicinal chemist's decision in the hit-to-lead process, balancing the ease of synthetic development and the prospect of increased activity. ES consists of three terms: Dist, DS, ΔSA, and their coefficients (w_1, w_2, and w_3). Dist represents the distance between the anchor atom and the COM of the growth site in angstrom units (Å). This term functions such that the closer this distance, the higher the evaluated growing availability. DS

https://doi.org/10.1021/acs.jcim.3c02036
J. Chem. Inf. Model. 2024, 64, 4475−4484

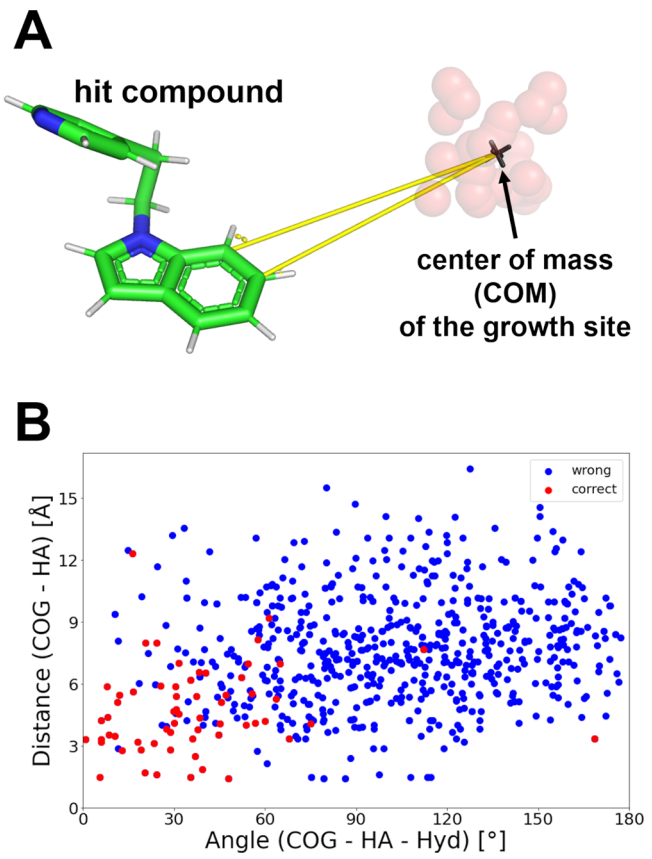


Figure 2. 3D representation and statistical analysis of the geometric conditions. (A) Example of geometric condition. The hit compound and growth site are shown as green stick and red sphere models, respectively. The black cross mark indicates the COM of the growth site. (B) Angle−distance scatter plot in the training data set. The *x*- and *y*-axes represent the angle formed by COM-HA-Hyd and the distance of COM-HA, respectively. The geometric conditions of the correct and wrong anchor atoms are shown as red and blue plots, respectively.

is the growth site druggability score referred to as fpocket2.[36] The druggability score also includes three terms: the hydrophobicity, volume, and penalty of the hydrophilic surface area. Because a high DS indicates high druggability and these values are very small (0−1), we scaled it by taking the logarithm of the reciprocal of the DS. This implied that a growth site with high druggability was suitable for growth. The difference in the synthetic accessibility (ΔSA) is the modifiability of the anchor atom in the hit compound based on the SA score. The SA score estimates the ease of synthesis of a compound based on its fragment contributions and complexity.[37] For the ΔSA calculation, 20 hypothetical lead compounds were generated by growing 20 functional groups from the anchor atom in the hit compound. 20 functional groups were selected based on the template moieties in the topless tree (Figure S2).[38] The SA scores of the hit compound (SA_{hit}) and the 20 hypothetical lead compounds were calculated, and those of the hypothetical lead compounds were averaged ($\overline{SA}_{hyp.lead}$). ΔSA is calculated as $\overline{SA}_{hyp.lead} - SA_{hit}$. The more difficult it was to grow from the anchor atom, the larger the value of $\overline{SA}_{hyp.lead}$. Hence, a high ΔSA indicates that the anchor atom is inappropriate as the starting atom for growing. Summarizing the above, it can be seen that the lower the ES value, the more appropriate the pairing between the anchor atom and the growth site.

2.3. Validation of the Protocol. The SINCHO protocol was applied to the test data set for two cases: experimentally determined complex structures and MD ensembles. The test data set consisted of 41 correct pairs of 30 protein−hit compound complex structures with known protein−lead complex structures. Compounds that were irrelevant to the binding site of the hit compound in the target protein were removed. The protonation states of the hit compounds were refined in the same manner as those in the training data set. For the application of MD ensembles, each complex structure in the test data set underwent MD simulation. After energy minimization of 2000 steps, 200 ps *NVT* equilibration at 300 K with the velocity-rescaling thermostat, and 800 ps *NPT* equilibration at 300 K and 1 bar with Berendsen barostat,[39−42] a production run was performed under the NPT conditions with a length of 10 ns using GROMACS2021.5.[43] The MD ensembles were obtained per 1 ns in the simulations. Only snapshots with a root-mean-square deviation (rmsd) of less than 2.5 Å were used to execute the SINCHO protocol. The detailed conditions of the simulations are provided in the Supporting Information (Appendix S1).

For each application of the SINCHO protocol, the anchor atom and growth site pairs were ranked based on ES. Subsequently, the ranking of the correct pair, which placed a known functional group, was annotated for each case in the test data set. For the application of MD ensembles, an annotation was performed for every snapshot of the MD ensemble. Note that growth sites with less than 15% volume overlap with the lead functional group were excluded as the correct growth site.

3. RESULTS

3.1. Parameter Tuning of the ES. Using the training data set, the coefficients of the ES were fine-tuned by execution under each condition with increments of 0.1, from 0 to 1, and their summation to 1. Figure 3 shows the success rate transitions for several conditions. The horizontal and vertical axes represent the number of pairs sorted by ES and the percentage of correct values in the training data set, respectively. The coefficient conditions with the best accuracy and those using only one term in the ES are plotted in this

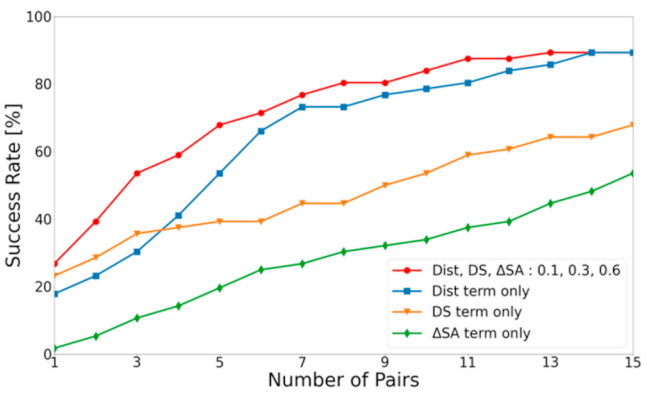


Figure 3. Success rate plot of the training data set with different coefficients. Success rate based on the ES with fine-tuned coefficient conditions, Dist term only, DS term only, and ΔSA term only are shown in red, blue, orange, and green, respectively.

https://doi.org/10.1021/acs.jcim.3c02036
J. Chem. Inf. Model. 2024, 64, 4475−4484

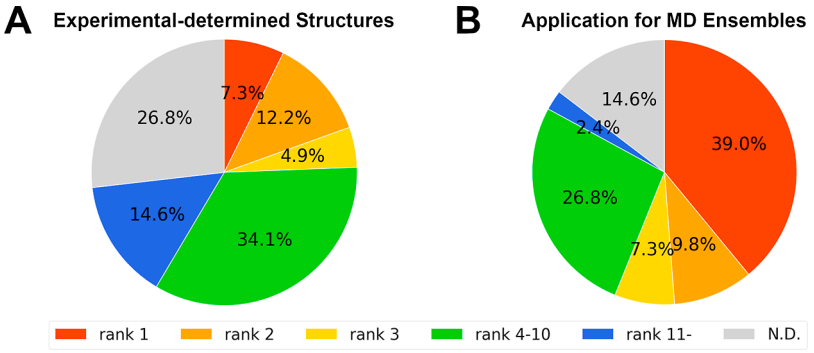


Figure 4. Application results of the SINCHO protocol for the test data set. (A) Pie chart of the application results for experimentally determined structures in the test data set. (B) Pie chart of the application results for MD ensembles in the data set. The rates of the preceding pairs ranked first, second, third, fourth to tenth, and eleventh onward are shown in red, orange, yellow, green, and blue, respectively. Rates where the preceding pair was not a candidate in the scoring are shown in gray.

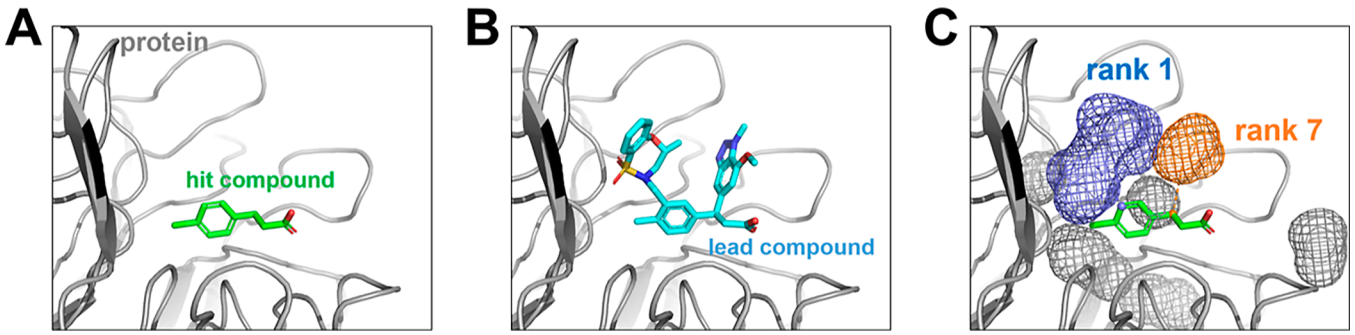


Figure 5. Case study of the SINCHO protocol. (A) 3D view of the complex structure between the target protein and hit compound. The target protein and hit compound are shown as gray cartoon and green stick models, respectively. (B) 3D view of the known lead compound. The lead compound is aligned with the protein backbone and is shown as a cyan line model. (C) Application results of the SINCHO protocol. Mesh models indicate the growth sites identified by the protocol. The preceding pairs between the anchor atom and growth site are shown as sphere and mesh models colored indigo and orange, respectively.

figure. For a low number of pairs, the plot with only the DS term has a higher success rate than those with only the Dist or ΔSA terms. This indicates that previous hit-to-lead studies tended to choose growth sites with high druggability. The plot with only the Dist term has a high success rate when considering the fourth pair onward. This result indicates that anchor atoms and growth sites close to each other tended to favor pairs in previous studies. In addition, the success rate transition of the Dist or DS term only suggests that the hit-to-lead strategy is not solely determined by either the Dist or DS term but by the balance of the two. The ΔSA term only resulted in a lower success rate than those of the Dist or DS terms only. This is because the lead strategy is designed to focus on the pocket environment and accessibility rather than the modifiability of the anchor atom. However, the steady linear increase in the success rate indicates that modifiability is partly related to the hit-to-lead strategy of medicinal chemists. The plot of the best coefficients ($w_1/w_2/w_3$ = 0.1:0.3:0.6) had a higher predictive accuracy than that of the other plots. In fact, the success rates for the top and fifth pairs with the best coefficients were 26.786 and 67.857%, respectively. The plot shows a high rising edge and an increased rate compared with only one term. This indicates that the scoring function, which combines the three terms, enables the prediction of a desirable pair by considering the balance of these characteristics. Figure S3 compares the success rate transitions under several coefficient conditions. The prediction results are similar. Among these results, the success rates of the best and top-three pairs in the selected conditions were superior to those of the others. Figure S4 shows the value distribution of each term and the tuned ES. It indicates that the tuned ES enables the distinction between the correct pairs and the others.

3.2. Application for the Test Data Set. Using the fine-tuned ES, we performed and assessed the SINCHO protocol on the test data set. First, the protocol was applied to experimentally determined complex structures in the test data set. Figure 4A presents a chart of the application results for the test data set. Detailed results for each preceding pair are shown in the Supporting Information (Table S2). Each preceding pair was annotated by ranking the ES among the candidates in the structure. N.D. means that the preceding pair was not a candidate for scoring, with misdetection of the P2C or failure of anchor atom selection. In the application for experimentally determined complex structures, 30 preceding pairs, 73.2% of the data set, were identified by the protocol as being among the candidates. Only 7.3% of the data set was correctly predicted at the top by the ES. Half of the scoring results ranked the preceding pairs from 4 to 10. These results indicate that applying the SINCHO protocol for experimentally determined complex structures works in terms of detecting preceding pairs but is insufficient for predicting the best hit-to-lead strategy.

Next, we applied the SINCHO protocol to the MD ensembles in the test data set. Each complex structure in the test data set was subjected to a 10 ns MD simulation, and snapshot intervals of 1 ns were obtained. The SINCHO

https://doi.org/10.1021/acs.jcim.3c02036
J. Chem. Inf. Model. 2024, 64, 4475−4484

protocol was applied in parallel for each snapshot with an rmsd of less than 2.5 Å from the initial pose of the hit compound, and the pairs were ranked in every snapshot. Figure 4B shows the results of applying the MD ensembles to the test data set. Note that for each preceding pair, the best ranking among the parallel executions of the protocol is counted in this chart. Compared with the application for experimentally determined structures, the rate of rank 1 increased from 7.3 to 39.0%. The cases with an improvement in the prediction rank include both cases ranked low by ES and cases in which the growth site was not detected when applying the experimentally determined structure. In addition, the rate of N.D. decreased from 26.8 to 14.6%. This implies that the parallel execution of the SINCHO protocol improved the detection of the preceding pairs and the prediction of desirable pairs.

4. DISCUSSION

4.1. Case Study. Figure 5 shows a case study of applying the SINCHO protocol to an experimentally determined complex structure. In this case, the hit compound was bound to the Keap1 Kelch domain in its experimentally determined structure [Figure 5A, Protein Data Bank (PDB) ID: 5FNQ[44]]. The lead compound was generated as described in a previous study,[44] and two functional groups were modified from different anchor atoms (Figure 5B). We assessed whether the protocol accurately predicted the preceding pairs. Site identification and anchor atom selection identified these pairs as candidates for the correct growth (Figure 5C). As a result of ranking by ES, the preceding pair between the meta-position atom and the growing functional group was predicted to be in rank 1, with a score of 1.028. However, the other pair between the alkyl atom and the growing functional group was not accurately predicted, with a rank of 7 and a score of 1.507. Table 1 presents the details of the contribution of each term to the ES for the preceding pairs. While the Dist term of rank 7 is comparable to that of rank 1, the DS term of rank 7 is inferior to that of rank 1. The poor druggability score of rank 7 indicated that the growth site was small and exposed to solvents. In addition, rank 7 exhibits a remarkably disadvantageous ΔSA term compared to rank 1. This disadvantage corresponds to the carbon atom in the alkyl groups having a lower reactivity than that in the benzyl groups. Hence, the SINCHO protocol works reliably, considering growth availability based on accessibility, druggability, and modifiability. Other results for the top 10 pairs are shown in the Supporting Information (Figure S5).

Table 1. Extend Score and Contributions of Each Term for the Preceding Pairs

	Dist term	DS term	ΔSA term	extend score
rank 1	0.453	0.386	0.189	1.028
rank 7	0.480	0.564	0.463	1.507

We assessed whether the application of MD ensembles has a higher accuracy than that of experimentally determined structures. For a detailed analysis, we present a case study applying an MD ensemble. Figure 6 shows a case study in which the prediction accuracy was improved by applying the MD ensemble. In this case, the SINCHO protocol was performed on the complex structure of the CBP bromodomain and the hit compound (PDB ID: 4YK0[45]). This hit compound contained a cognate lead compound optimized in a previous study.[46] We determined that the protocol could identify the anchor atom and growth site pair by tracing the optimized route (Figure 6A). However, this pair did not rank highly, ranking thirteenth among the candidates in the experimentally determined structure. The ES of the preceding pair was 2.232, with an inferior DS term compared to that of the other promising pairs. Specifically, the volume and hydrophobicity of the growth sites were inferior to those of the other sites. In contrast, the snapshot of the MD ensemble at 10 ns showed a growth site larger than that of the experimentally determined structure (Figure 6C). Accordingly, the DS and ES values improved. Among the snapshot candidates, the preceding pair ranked third. Moreover, the protein structure was compared to analyze conformational changes. Figure 6B,D shows the protein surfaces of the binding sites in each structure. The protein structure of the MD ensemble partially differed from the experimentally determined structure. In particular, the orientation of the Arg1173 side chain differed between the experimentally determined structure and the MD ensemble. For orientation in the MD ensemble, the side chain was largely exposed to the solvents and a potential pocket was formed. Consequently, the protocol identified the hidden growth site and predicted the priority of the preceding anchor atom and the growth site pair. The orientation of arginine in the MD

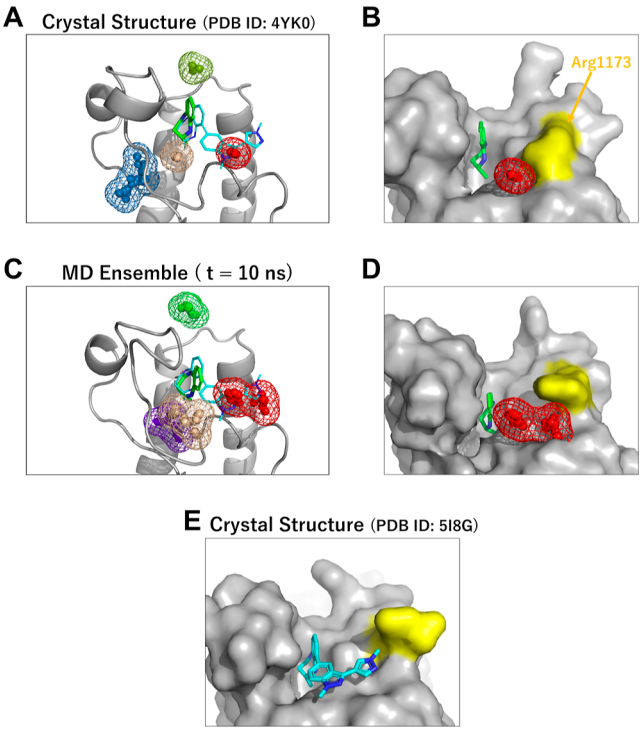


Figure 6. Comparison of the protein structure and the identified growth site. (A) Example of the application of the SINCHO protocol for the experimentally determined structures. (B) Protein surface representation of the experimentally determined structure and the preceding growth site. (C) Example of the application of the SINCHO protocol for MD ensembles. (D) Protein surface representation of the MD snapshot and the preceding growth site. (E) Protein surface representation of the experimentally determined structure with the lead compound. The protein structure and hit compound are shown as gray animation or surface and green stick models, respectively. The cognate lead compound aligned to the protein structure is shown as a cyan line model. Growth sites identified by the SINCHO protocol are shown as sphere and mesh models, and those placed on the lead substructure are colored in red.

https://doi.org/10.1021/acs.jcim.3c02036
J. Chem. Inf. Model. 2024, 64, 4475−4484

ensemble is similar to that in the experimentally determined structure of the cognate lead compound (PDB ID: 5I8G[46]), as shown in Figure 6D,E. This suggests that the application of a single complex structure may not adequately select the desirable pair because the conformation of the protein structure is inappropriate for lead compound binding. For accuracy improvement, parallel execution of the MD ensemble of the protocol is effective, owing to the consideration of protein flexibility.

Contrary to the abovementioned cases, there are cases that do not show improvements in prediction accuracy after applying the MD ensemble. We assessed SINCHO application for the MD ensemble starting from the complex structure between extracellular signal-regulated kinase 2 (ERK2) and the hit compound (PDB ID: 2OJG[47]). SINCHO application for neither the experimentally determined structure nor the MD ensemble could detect the growth site where the lead substructure was positioned (line 10 in Table S2). This misdetection derives from the local conformation of the P-loop in the protein structure. While the protein and hit compound complex have a closed P-loop conformation (Figure S6A), the protein and lead compound complex have open conformations, and the lead substructure is positioned in the space where tyrosine was located in the protein and hit compound complex structure (Figure S6B). These results indicate that lead strategy prediction considering these major conformational changes in the main chain atoms is challenging and beyond the scope of the current SINCHO protocol.

4.2. Effectiveness for MD Ensemble Application. Applying our protocol improves the accuracy of the growth strategy prediction. Even for a short simulation, an improvement in prediction accuracy is expected. Figure 7 shows the transitions in the best ranking for the MD simulation time. The *x*-axis represents the simulation time, and each bar shows the application results of the SINCHO protocol with all snapshots up to that time. The bars at 0 and 10 ns correspond to Figure 4A,B, respectively. Performance improvement in terms of protein flexibility was observed after 1 ns. In the results for the 1 ns time, the rate of N.D. fell by approximately 10%, and that of rank 1 increased by more than 7%. In the results for times of 4 ns onward, the rate of ranks 3 and above was over 50%. This indicates that protocol application considering protein flexibility is important for accurately selecting the desirable pair, even if the MD simulation is short. Meanwhile, the movement of the hit compound during the simulation should be considered. In fact, we often observed simulations with an unstable compound only in the 10 ns MD simulation for the experimentally determined structure. 12 of the 30 complexes in the test data set are shown as snapshots with significant movements of the hit compound from the initial coordinates. The SINCHO protocol cannot be applied to these snapshots because the desired pair can be misdirected. Complex structures with weak affinity for hit compounds will be susceptible to these significant movements. To deal with these unstable complexes, the implementation of positional constraints between the hit compounds and the binding site may be helpful in applying the SINCHO protocol to MD ensembles.

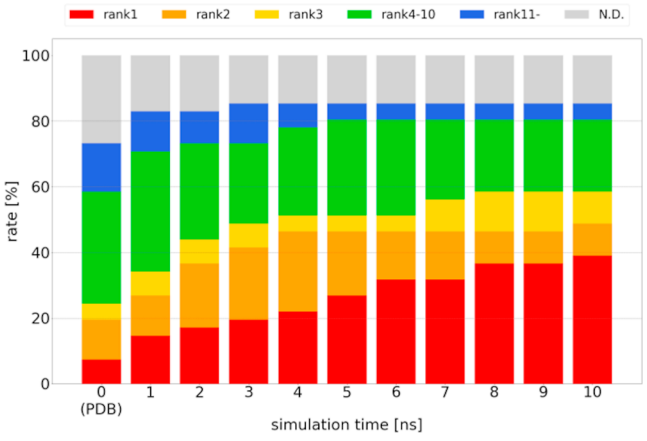


Figure 7. Ranking transitions per simulation time. Each bar represents the application results using snapshots. The rates of the preceding pair ranked first, second, third, fourth to tenth, and eleventh onward are colored red, orange, yellow, green, and blue, respectively. Rates at which the preceding pair was not a candidate for scoring are colored in gray.

For more detailed analysis, we focused on the amino acid distribution around the lead substructure, with respect to the differences between cases in which application for the MD ensemble was effective and those in which it was not. For the lead substructures of the test data set, the protein environment around 5 Å of the lead substructure in the crystal structure was extracted. Then, the amino acid residues included in the environment were counted for distribution analysis. The distributions of the cases with and without prediction accuracy improvement in the MD ensemble are shown in Figure 8. Note that cases that have already been highly ranked by the application of the crystal structure are not included in the analysis. The distributions indicate that the pocket environment in cases with improvement after MD ensemble application tended to have high distributions of Val, Ile, Met, Lys, and Glu. It is suggested that these cases may have a flexible pocket environment that contains linear-like amino acid residues. As these flexible pockets show a high fluctuation in the MD simulation, the SINCHO application of MD ensembles enabled the formation of a pre-existing pocket shape acceptable for lead substructure binding. In contrast, for cases that did not show improvement after MD ensemble application, the distributions of Gly, Ala, Ser, Phe, Tyr, and Trp were high values. These residues were divided into two types: tiny side chain residues and bulky side chain residues. In general, these two types of residues show less fluctuation in MD simulations. As the pocket environment that contains these rigid residues is less affected by fluctuation, the prediction accuracy improvement offered by the application of the MD ensemble may be less effective. These insights are helpful for understanding the SINCHO results of applied MD ensembles.

4.3. Interpretation of Failure Case. As mentioned above, the SINCHO protocol can identify and predict desirable anchor atom and growth site pairs for hit-to-lead. The pairs suggested by the protocol include several preceding pairs. However, the top-ranked pairs are positioned at anchor atoms and growth sites different from the known lead substructure. These cases are regarded as failure cases of the protocol at first glance. To validate these pairs, an additional docking study was conducted. Here, we present a case study of the complex structure of bromodomains and inhibitors (Figure 9). Among the results of the SINCHO protocol for the experimentally determined structure of the hit compound (PDB ID: 4A9H[48]), the pair that ranked second corresponds to the position of the

https://doi.org/10.1021/acs.jcim.3c02036
J. Chem. Inf. Model. 2024, 64, 4475−4484

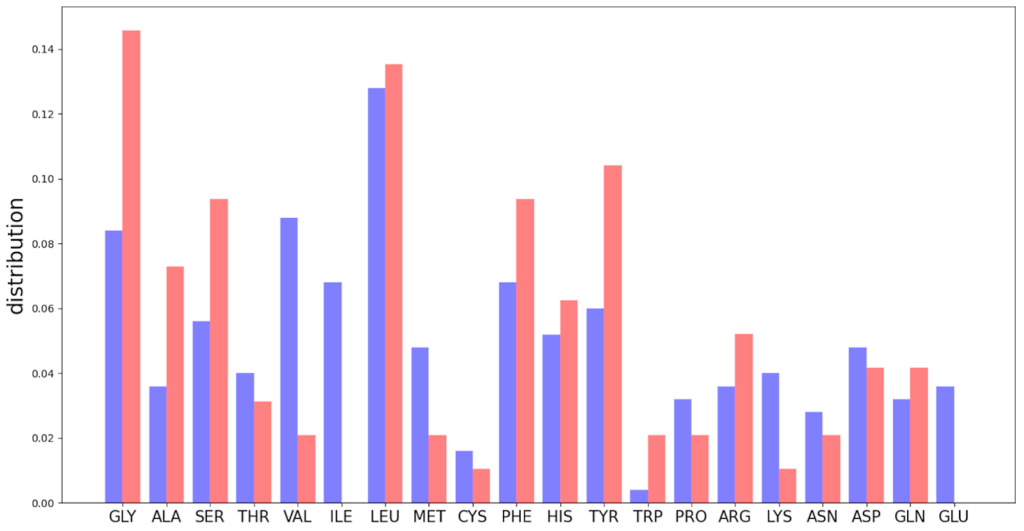


Figure 8. Amino acid distributions around the lead substructures. The distributions of the cases showing prediction accuracy improvement after the MD ensemble and those without improvement are shown as blue and red bars, respectively.

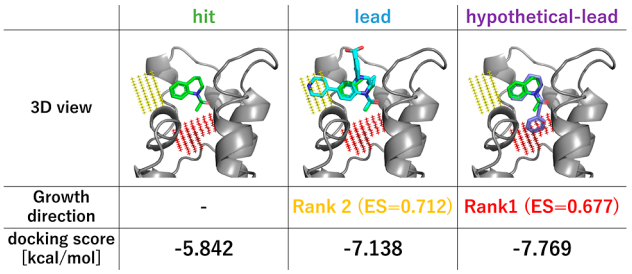


Figure 9. Docking validation for the known and hypothetical compounds. The protein structure and hit/lead/hypothetical lead compounds are shown as gray cartoon and green/cyan/purple stick models, respectively. Red and yellow grids indicate growth sites ranked 1 and 2, respectively, by the SINCHO protocol application for the experimentally determined structure of the hit compound.

lead substructure in the experimentally determined structure of the lead compound (PDB ID: 6FFG[49]). While the preceding pair was predicted as a high priority, the SINCHO protocol suggested another pair as a rank 1. To validate the rank 1 pair, 20 hypothetical lead compounds with growing lead substructures, termed the pair, were generated and used for docking simulation. Note that the lead substructures are the same as the functional groups for the ΔSA calculation (Figure S2). Docking simulations were performed using the Glide XP docking program (Schrödinger LLC, New York, NY, USA). The protein structure of 4A9H was prepared for docking simulations using the Protein Preparation Wizard Script in Maestro. For comparison of known compounds, the hit and lead compounds in each experimentally determined structure were calculated using docking scores in place of the Glide XP docking program. The docking scores of the hit and lead compounds were −5.842 and −7.138 kcal/mol, respectively. Notably, the hypothetical lead compound with a phenyl group has a docking score of −7.769 kcal/mol. These results indicate that this hypothetical lead compound can be a promising candidate for improving the affinity of the hit compound and the pair ranked 1 by ES can be a desirable anchor atom and growth site pair. Therefore, it is suggested that the SINCHO protocol not only traces the known hit-to-lead strategy but also can suggest a new strategy for hit-to-lead optimization. In addition, it is also suggested that we need to consider several high-ranked pairs as a promising lead strategy.

4.4. Robustness Assessment Using Unseen Data. For validation using unseen data, we applied SINCHO to isocitrate dehydrogenase 1 (IDH1). As regulators against IDH1, allosteric inhibitors have been developed. Levell et al. identified the hit compound (IDH125) that selectively inhibits the

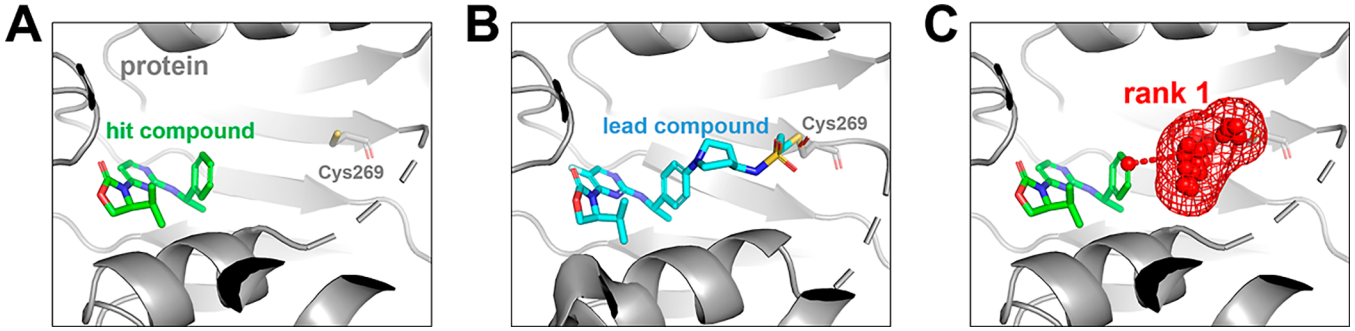


Figure 10. IDH1R132H−compound complex structures and SINCHO results. (A) 3D view of the complex structure between the target protein and the hit compound. The target protein and hit compound are shown as gray cartoon and green stick models, respectively. (B) 3D view of the complex structure between the target protein and known lead compound. The lead compound covalently binds to Cys269. (C) Application results of the SINCHO protocol. Predicted anchor atom and growth site (rank 1) are shown as sphere and mesh models.

https://doi.org/10.1021/acs.jcim.3c02036
J. Chem. Inf. Model. 2024, 64, 4475−4484

IDH1^{R132H} mutant with IC50 = 0.22 μM (Figure 10A).[50] Based on the crystal complex structure, Liang et al. designed the irreversible lead compound (IHMT-IDH1-053) with IC50 = 0.0047 μM (Figure 10B).[51] Using the crystal structure between IDH1R132H and the hit compound (PDB ID: 5SVF[52]), we applied the SINCHO protocol and assessed whether the lead strategy could be predicted. As a result, the SINCHO protocol successfully predicted the preceding lead strategy as ES rank 1 (Figure 10C). Notably, the SINCHO protocol can be applied to suggest the lead strategy of irreversible compounds without any information regarding the covalently bondable amino acid residues. This demonstrates that our protocol functions not only for proteins in the data set but also for other protein−compound complexes. In future work, it is expected that the SINCHO protocol will be used in practical hit-to-lead processes, allowing further experimental validation of our protocol.

5. CONCLUSIONS

In this study, we developed the SINCHO protocol, which is a method for predicting desirable anchor atoms and growth site pairs for hit-to-lead optimization. This protocol is based only on the 3D complex structure and consists of three steps: (1) site identification, (2) anchor atom selection, and (3) next choice. These steps were developed and fine-tuned using a training data set and assessed using a test data set. In applying MD ensembles to the test data set, approximately 40% of the preceding pairs were predicted as rank 1. A case study of the applications showed that the parallel execution of the protocol for the MD ensemble has high accuracy, owing to the consideration of protein flexibility. In addition, the SINCHO protocol may suggest not only a known hit-to-lead strategy but also a new one. The SINCHO protocol helps medicinal chemists visualize and modify functional groups in a hit-to-lead manner. Moreover, the protocol will be available to improve the efficiency of de novo molecular design by using it as a filter for molecular generation.

■ ASSOCIATED CONTENT

Data Availability Statement

3D structures of the complex in the training and test data sets were downloaded from the Protein Data Bank (PDB). The P2C was used for site identification using the SINCHO protocol. Schrödinger suite 2023-1 was used for protein preparation and docking simulations. OpenBabel 3.1.0 was used for protonating the hit compounds. AmberTools21 and Gaussian 16 Rev B.01 were used for the MD simulations. GROMACS2021.5 was used as the MD engine. PyMOL was used for visualization. The SINCHO protocol is provided on GitHub (https://github.com/genki-kudo/SINCHO-H2L.git). The application result for the test data set have been deposited in Zenodo (DOI: 10.5281/zenodo.10408661).

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.3c02036.

Training data set preparation scheme, functional groups for synthetic accessibility score calculation of hypothetical lead compounds, success rate plots of the training data set with different coefficient conditions, value ranges of the Extend Score, SINCHO results for the case study, comparison of protein conformation, training data set, application results for the test data set, and details of the molecular dynamics simulation for the test data set (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Takatsugu Hirokawa − *Division of Biomedical Science, Faculty of Medicine and Transborder Medical Research Center, University of Tsukuba, Tsukuba, Ibaraki 305-8575, Japan;* orcid.org/0000-0002-3180-5050; Email: t-hirokawa@md.tsukuba.ac.jp

Authors

Genki Kudo − *Physics Department, Graduate School of Pure and Applied Sciences, University of Tsukuba, Tsukuba, Ibaraki 305-8571, Japan;* orcid.org/0000-0002-4727-7002

Takumi Hirao − *Doctoral Program in Medical Sciences, Graduate School of Comprehensive Human Sciences, University of Tsukuba, Tsukuba, Ibaraki 305-8575, Japan; Division of Biomedical Science, Faculty of Medicine, University of Tsukuba, Tsukuba, Ibaraki 305-8575, Japan*

Ryunosuke Yoshino − *Division of Biomedical Science, Faculty of Medicine and Transborder Medical Research Center, University of Tsukuba, Tsukuba, Ibaraki 305-8575, Japan*

Yasuteru Shigeta − *Center for Computational Sciences, University of Tsukuba, Tsukuba, Ibaraki 305-8577, Japan;* orcid.org/0000-0002-3219-6007

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.3c02036

Author Contributions

G.K., T.H., R.Y., Y.S., and T.H. conceived the study. G.K. and T.H. implemented the SINCHO protocol. G.K. performed the MD simulations. The manuscript was written with contributions from all authors. All the authors approved the final version of the manuscript.

Funding

This research was supported by the Research Support Project for Life Science and Drug Discovery [Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS)] (grant no. JP23ama121029j0003) of the Japan Agency for Medical Research and Development (AMED) and JST SPRING (grant no. JPMJSP2124).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Research Institute of Systems Planning, Inc.for providing the synthetic accessibility score calculation module. The authors thank Kenichiro Imai and Chie Motono at the AIST for their constructive discussions and feedback. The MD simulations used Cygnus computational resources provided by the Multidisciplinary Cooperative Research Program at the Center for Computational Sciences (Project Code: CADD) at the University of Tsukuba. G.K. acknowledges the support from a Grant-in-Aid for JSPS Fellows under grant no. 24KJ0483.

■ ABBREVIATIONS

COM, center of mass; Dist, distance; DS, druggability score; ERK2, extracellular signal-regulated kinase 2; HA, heavy atom; Hyd, hydrogen atom; IDH1, isocitrate dehydrogenase 1; MD,

https://doi.org/10.1021/acs.jcim.3c02036
J. Chem. Inf. Model. 2024, 64, 4475−4484

molecular dynamics; LB, ligand-bound; P2C, pocket to concavity; PDB, protein data bank; rmsd, root-mean-square deviation; SA, synthetic accessibility; SINCHO, site identification and next choice

REFERENCES

(1) Paul, S. M.; Mytelka, D. S.; Dunwiddie, C. T.; Persinger, C. C.; Munos, B. H.; Lindborg, S. R.; Schacht, A. L. How to Improve RD Productivity: The Pharmaceutical Industry's Grand Challenge. *Nat. Rev. Drug Discovery* **2010**, *9*, 203−214.

(2) Hughes, J. P.; Rees, S. S.; Kalindjian, S. B.; Philpott, K. L. Principles of Early Drug Discovery. *Br. J. Pharmacol.* **2011**, *162*, 1239−1249.

(3) Poulain, R.; Horvath, D.; Bonnet, B.; Eckhoff, C.; Chapelain, B.; Bodinier, M. C.; Déprez, B. From Hit to Lead. Analyzing Structure-Profile Relationships. *J. Med. Chem.* **2001**, *44* (21), 3391−3401.

(4) Vemula, D.; Jayasurya, P.; Sushmitha, V.; Kumar, Y. N.; Bhandari, V. CADD, AI and ML in Drug Discovery: A Comprehensive Review. *Eur. J. Pharm. Sci.* **2023**, *181*, 106324.

(5) Bassani, D.; Moro, S. Past, Present, and Future Perspectives on Computer-Aided Drug Design Methodologies. *Molecules* **2023**, *28*, 3906.

(6) Gurung, A. B.; Ali, M. A.; Lee, J.; Farah, M. A.; Al-Anazi, K. M. An Updated Review of Computer-Aided Drug Design and Its Application to COVID-19. *BioMed Res. Int.* **2021**, *2021*, 1−18.

(7) Ou-Yang, S. S.; Lu, J. Y.; Kong, X. Q.; Liang, Z. J.; Luo, C.; Jiang, H. Computational Drug Discovery. *Acta Pharmacol. Sin.* **2012**, *33*, 1131−1140.

(8) Patel, L.; Shukla, T.; Huang, X.; Ussery, D. W.; Wang, S. Machine Learning Methods in Drug Discovery. *Molecules* **2020**, *25* (22), 5277.

(9) Shaker, B.; Ahmad, S.; Lee, J.; Jung, C.; Na, D. In Silico Methods and Tools for Drug Discovery. *Comput. Biol. Med.* **2021**, *137*, 104851.

(10) Sakamoto, K.; Asano, S.; Ago, Y.; Hirokawa, T. AlphaFold Version 2.0 Elucidates the Binding Mechanism between VIPR2 and KS-133, and Reveals an S-S Bond (Cys25-Cys192) Formation of Functional Significance for VIPR2. *Biochem. Biophys. Res. Commun.* **2022**, *636*, 10−16.

(11) Asamitsu, K.; Hirokawa, T.; Okamoto, T. Identification of a Novel CDK9 Inhibitor Targeting the Intramolecular Hidden Cavity of CDK9 Induced by Tat Binding. *PLoS One* **2022**, *17*, No. e0277024.

(12) Yoshino, R.; Yasuo, N.; Hagiwara, Y.; Ishida, T.; Inaoka, D. K.; Amano, Y.; Tateishi, Y.; Ohno, K.; Namatame, I.; Niimi, T.; Orita, M.; Kita, K.; Akiyama, Y.; Sekijima, M. Discovery of a Hidden Trypanosoma Cruzi Spermidine Synthase Binding Site and Inhibitors through In Silico, In Vitro, and X-Ray Crystallography. *ACS Omega* **2023**, *8* (29), 25850−25860.

(13) Yoshino, R.; Yasuo, N.; Sekijima, M. Identification of Key Interactions between SARS-CoV-2 Main Protease and Inhibitor Drug Candidates. *Sci. Rep.* **2020**, *10* (1), 12493.

(14) Cáceres, E. L.; Tudor, M.; Cheng, A. C. Deep Learning Approaches in Predicting ADMET Properties. *Future Med. Chem.* **2020**, *12*, 1995−1999.

(15) Fralish, Z.; Chen, A.; Skaluba, P.; Reker, D. DeepDelta: Predicting ADMET Improvements of Molecular Derivatives with Deep Learning. *J. Cheminf.* **2023**, *15* (1), 101.

(16) Hoffer, L.; Muller, C.; Roche, P.; Morelli, X. Chemistry-Driven Hit-to-Lead Optimization Guided by Structure-Based Approaches. *Mol. Inf.* **2018**, *37*, 1800059.

(17) Poyraz, O. ··; Jeankumar, V. U.; Saxena, S.; Schnell, R.; Haraldsson, M.; Yogeeswari, P.; Sriram, D.; Schneider, G. Structure-Guided Design of Novel Thiazolidine Inhibitors of O -Acetyl Serine Sulfhydrylase from Mycobacterium Tuberculosis. *J. Med. Chem.* **2013**, *56* (16), 6457−6466.

(18) Mologni, L.; Dalla Via, M.; Chilin, A.; Palumbo, M.; Marzaro, G. Discovery of WtRET and V804MRET Inhibitors: From Hit to Lead. *ChemMedChem* **2017**, *12* (16), 1390−1398.

(19) Halby, L.; Menon, Y.; Rilova, E.; Pechalrieu, D.; Masson, V.; Faux, C.; Bouhlel, M. A.; David-Cordonnier, M. H.; Novosad, N.; Aussagues, Y.; Samson, A.; Lacroix, L.; Ausseil, F.; Fleury, L.; Guianvarc'H, D.; Ferroud, C.; Arimondo, P. B. Rational Design of Bisubstrate-Type Analogues as Inhibitors of DNA Methyltransferases in Cancer Cells. *J. Med. Chem.* **2017**, *60* (11), 4665−4679.

(20) Newton, R.; Waszkowycz, B.; Seewooruthun, C.; Burschowsky, D.; Richards, M.; Hitchin, S.; Begum, H.; Watson, A.; French, E.; Hamilton, N.; Jones, S.; Lin, L. Y.; Waddell, I.; Echalier, A.; Bayliss, R.; Jordan, A. M.; Ogilvie, D. Discovery and Optimization of Wt-RET/KDR-Selective Inhibitors of RETV804M Kinase. *ACS Med. Chem. Lett.* **2020**, *11* (4), 497−505.

(21) Klug, D. M.; Tschiegg, L.; Diaz, R.; Rojas-Barros, D.; Perez-Moreno, G.; Ceballos, G.; García-Hernández, R.; Martinez-Martinez, M. S.; Manzano, P.; Ruiz, L. M.; Caffrey, C. R.; Gamarro, F.; Pacanowska, D. G.; Ferrins, L.; Navarro, M.; Pollastri, M. P. Hit-to-Lead Optimization of Benzoxazepinoindazoles As Human African Trypanosomiasis Therapeutics. *J. Med. Chem.* **2020**, *63* (5), 2527−2546.

(22) Schnecke, V.; Boström, J. Computational Chemistry-Driven Decision Making in Lead Generation. *Drug Discovery Today* **2006**, *11*, 43−50.

(23) Holenz, J.; Stoy, P. Advances in Lead Generation. *Bioorg. Med. Chem. Lett.* **2019**, *29*, 517−524.

(24) Bleicher, K. H.; Böhm, H. J.; Müller, K.; Alanine, A. I. Hit and Lead Generation: Beyond High-Throughput Screening. *Nat. Rev. Drug Discovery* **2003**, *2*, 369−378.

(25) Johnson, C. N.; Erlanson, D. A.; Murray, C. W.; Rees, D. C. Fragment-to-Lead Medicinal Chemistry Publications in 2015. *J. Med. Chem.* **2017**, *60*, 89−99.

(26) Johnson, C. N.; Erlanson, D. A.; Jahnke, W.; Mortenson, P. N.; Rees, D. C. Fragment-to-Lead Medicinal Chemistry Publications in 2016. *J. Med. Chem.* **2018**, *61*, 1774−1784.

(27) Mortenson, P. N.; Erlanson, D. A.; De Esch, I. J. P.; Jahnke, W.; Johnson, C. N. Fragment-to-Lead Medicinal Chemistry Publications in 2017. *J. Med. Chem.* **2019**, *62* (8), 3857−3872.

(28) Erlanson, D. A.; De Esch, I. J. P.; Jahnke, W.; Johnson, C. N.; Mortenson, P. N. Fragment-to-Lead Medicinal Chemistry Publications in 2018. *J. Med. Chem.* **2020**, *63*, 4430−4444.

(29) Jahnke, W.; Erlanson, D. A.; De Esch, I. J. P.; Johnson, C. N.; Mortenson, P. N.; Ochi, Y.; Urushima, T. Fragment-to-Lead Medicinal Chemistry Publications in 2019. *J. Med. Chem.* **2020**, *63* (24), 15494−15507.

(30) De Esch, I. J. P.; Erlanson, D. A.; Jahnke, W.; Johnson, C. N.; Walsh, L. Fragment-to-Lead Medicinal Chemistry Publications in 2020. *J. Med. Chem.* **2022**, *65*, 84−99.

(31) Walsh, L.; Erlanson, D. A.; de Esch, I. J. P.; Jahnke, W.; Woodhead, A.; Wren, E. Fragment-to-Lead Medicinal Chemistry Publications in 2021. *J. Med. Chem.* **2023**, *66*, 1137−1156.

(32) Ichihara, O.; Shimada, Y.; Yoshidome, D. The Importance of Hydration Thermodynamics in Fragment-to-Lead Optimization. *ChemMedChem* **2014**, *9* (12), 2708−2717.

(33) Orita, M.; Ohno, K.; Warizaya, M.; Amano, Y.; Niimi, T. Lead Generation and Examples: Opinion Regarding How to Follow up Hits. *Methods Enzymol.* **2011**, *493*, 383−419.

(34) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open Chemical Toolbox. *J. Cheminf.* **2011**, *3* (1), 33.

(35) Kudo, G.; Hirao, T.; Yoshino, R.; Shigeta, Y.; Hirokawa, T. Pocket to Concavity: A Tool for the Refinement of Protein-Ligand Binding Site Shape from Alpha Spheres. *Bioinformatics* **2023**, *39* (4), btad212.

(36) Le Guilloux, V.; Schmidtke, P.; Tuffery, P. Fpocket: An Open Source Platform for Ligand Pocket Detection. *BMC Bioinf.* **2009**, *10*, 168.

(37) Ertl, P.; Schuffenhauer, A. Estimation of Synthetic Accessibility Score of Drug-like Molecules Based on Molecular Complexity and Fragment Contributions. *J. Cheminf.* **2009**, *1* (1), 8.

https://doi.org/10.1021/acs.jcim.3c02036
J. Chem. Inf. Model. 2024, 64, 4475−4484

(38) Topliss, J. G. Utilization of Operational Schemes for Analog Synthesis in Drug Design. *J. Med. Chem.* **1972**, *15* (10), 1006−1011.
(39) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126* (1), 014101.
(40) Bussi, G.; Parrinello, M. Stochastic Thermostats: Comparison of Local and Global Schemes. *Comput. Phys. Commun.* **2008**, *179* (1−3), 26−29.
(41) Bussi, G.; Zykova-Timan, T.; Parrinello, M. Isothermal-Isobaric Molecular Dynamics Using Stochastic Velocity Rescaling. *J. Chem. Phys.* **2009**, *130* (7), 074101.
(42) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81* (8), 3684−3690.
(43) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26* (16), 1701−1718.
(44) Davies, T. G.; Wixted, W. E.; Coyle, J. E.; Griffiths-Jones, C.; Hearn, K.; McMenamin, R.; Norton, D.; Rich, S. J.; Richardson, C.; Saxty, G.; Willems, H. M. G.; Woolford, A. J. A.; Cottom, J. E.; Kou, J. P.; Yonchuk, J. G.; Feldser, H. G.; Sanchez, Y.; Foley, J. P.; Bolognese, B. J.; Logan, G.; Podolin, P. L.; Yan, H.; Callahan, J. F.; Heightman, T. D.; Kerns, J. K. Monoacidic Inhibitors of the Kelch-like ECH-Associated Protein 1: Nuclear Factor Erythroid 2-Related Factor 2 (KEAP1:NRF2) Protein-Protein Interaction with High Cell Potency Identified by Fragment-Based Discovery. *J. Med. Chem.* **2016**, *59* (8), 3991−4006.
(45) Ghosh, S.; Taylor, A.; Chin, M.; Huang, H. R.; Conery, A. R.; Mertz, J. A.; Salmeron, A.; Dakle, P. J.; Mele, D.; Cote, A.; Jayaram, H.; Setser, J. W.; Poy, F.; Hatzivassiliou, G.; DeAlmeida-Nagata, D.; Sandy, P.; Hatton, C.; Romero, F. A.; Chiang, E.; Reimer, T.; Crawford, T.; Pardo, E.; Watson, V. G.; Tsui, V.; Cochran, A. G.; Zawadzke, L.; Harmange, J. C.; Audia, J. E.; Bryant, B. M.; Cummings, R. T.; Magnuson, S. R.; Grogan, J. L.; Bellon, S. F.; Albrecht, B. K.; Sims, R. J.; Lora, J. M. Regulatory T Cell Modulation by CBP/EP300 Bromodomain Inhibition. *J. Biol. Chem.* **2016**, *291* (25), 13014−13027.
(46) Taylor, A. M.; Côté, A.; Hewitt, M. C.; Pastor, R.; Leblanc, Y.; Nasveschuk, C. G.; Romero, F. A.; Crawford, T. D.; Cantone, N.; Jayaram, H.; Setser, J.; Murray, J.; Beresini, M. H.; De Leon Boenig, G.; Chen, Z.; Conery, A. R.; Cummings, R. T.; Dakin, L. A.; Flynn, E. M.; Huang, O. W.; Kaufman, S.; Keller, P. J.; Kiefer, J. R.; Lai, T.; Li, Y.; Liao, J.; Liu, W.; Lu, H.; Pardo, E.; Tsui, V.; Wang, J.; Wang, Y.; Xu, Z.; Yan, F.; Yu, D.; Zawadzke, L.; Zhu, X.; Zhu, X.; Sims, R. J.; Cochran, A. G.; Bellon, S.; Audia, J. E.; Magnuson, S.; Albrecht, B. K. Fragment-Based Discovery of a Selective and Cell-Active Benzodiazepinone CBP/EP300 Bromodomain Inhibitor (CPI-637). *ACS Med. Chem. Lett.* **2016**, *7* (5), 531−536.
(47) Aronov, A. M.; Baker, C.; Bemis, G. W.; Cao, J.; Chen, G.; Ford, P. J.; Germann, U. A.; Green, J.; Hale, M. R.; Jacobs, M.; Janetka, J. W.; Maltais, F.; Martinez-Botella, G.; Namchuk, M. N.; Straub, J.; Tang, Q.; Xie, X. Flipped out: Structure-Guided Design of Selective Pyrazolylpyrrole ERK Inhibitors. *J. Med. Chem.* **2007**, *50* (6), 1280−1287.
(48) Chung, C. W.; Dean, A. W.; Woolven, J. M.; Bamborough, P. Fragment-Based Discovery of Bromodomain Inhibitors Part 1: Inhibitor Binding Modes and Implications for Lead Discovery. *J. Med. Chem.* **2012**, *55* (2), 576−586.
(49) Law, R. P.; Atkinson, S. J.; Bamborough, P.; Chung, C. W.; Demont, E. H.; Gordon, L. J.; Lindon, M.; Prinjha, R. K.; Watson, A. J. B.; Hirst, D. J. Discovery of Tetrahydroquinoxalines as Bromodomain and Extra-Terminal Domain (BET) Inhibitors with Selectivity for the Second Bromodomain. *J. Med. Chem.* **2018**, *61* (10), 4317−4334.
(50) Levell, J. R.; Caferro, T.; Chenail, G.; Dix, I.; Dooley, J.; Firestone, B.; Fortin, P. D.; Giraldes, J.; Gould, T.; Growney, J. D.; Jones, M. D.; Kulathila, R.; Lin, F.; Liu, G.; Mueller, A.; Van der Plas, S.; Slocum, K.; Smith, T.; Terranova, R.; Touré, B. B.; Tyagi, V.; Wagner, T.; Xie, X.; Xu, M.; Yang, F. S.; Zhou, L. X.; Pagliarini, R.; Cho, Y. S. Optimization of 3-Pyrimidin-4-Yl-Oxazolidin-2-Ones as Allosteric and Mutant Specific Inhibitors of IDH1. *ACS Med. Chem. Lett.* **2017**, *8* (2), 151−156.
(51) Liang, Q.; Wang, B.; Zou, F.; Guo, G.; Wang, W.; Wang, W.; Liu, Q.; Shen, L.; Hu, C.; Wang, W.; Wang, A.; Huang, T.; He, Y.; Xia, R.; Ge, J.; Liu, J.; Liu, Q. Structure-Based Discovery of IHMT-IDH1−053 as a Potent Irreversible IDH1 Mutant Selective Inhibitor. *Eur. J. Med. Chem.* **2023**, *256*, 115411.
(52) Xie, X.; Baird, D.; Bowen, K.; Capka, V.; Chen, J.; Chenail, G.; Cho, Y. S.; Dooley, J.; Farsidjani, A.; Fortin, P.; Kohls, D.; Kulathila, R.; Lin, F.; McKay, D.; Rodrigues, L.; Sage, D.; Touré, B. B.; van der Plas, S.; Wright, K.; Xu, M.; Yin, H.; Levell, J.; Pagliarini, R. A. Allosteric Mutant IDH1 Inhibitors Reveal Mechanisms for IDH1 Mutant and Isoform Selectivity. *Structure* **2017**, *25* (3), 506−513.

https://doi.org/10.1021/acs.jcim.3c02036
J. Chem. Inf. Model. 2024, 64, 4475−4484