Master's Thesis of Artificial Intelligence

Deep Learning for the Prediction of Protein-Ligand Interactions: Applications to hERG and GPCRs

단백질-리간드 상호작용 예측을 위한 딥러닝: hERG와 GPCR에의 적용

February 2023

Graduate School of
Seoul National University
Interdisciplinary Program in Artificial Intelligence

Sumin Lee

Deep Learning for the Prediction of Protein-Ligand Interactions: Applications to hERG and GPCRs

Advisor Chaok Seok

Submitting a master's thesis of
Artificial Intelligence

February 2023

Graduate School of
Seoul National University
Interdisciplinary Program in Artificial Intelligence

Sumin Lee

Confirming the master's thesis written by
Sumin Lee
February 2023

Chair	백민경	(Seal)
Vice Chair	석차옥	(Seal)
Examiner	박한범	(Seal)

Abstract

The development of deep learning and the accumulation of data have made it possible to predict more accurate protein structures and molecular properties, and CADD research using deep learning has been actively conducted. In this research, we applied deep learning to ligand-protein interaction and protein structure prediction for the human ether-a-go-go related gene (hERG) and G protein-coupled receptors (GPCRs). First, hERG is a voltage-gated potassium ion channel expressed on the cardiomyocyte membrane. hERG affects repolarization and is related to drug cardiotoxicity. We trained hERG inhibitor prediction models using various molecular representations and machine learning/deep learning methods. The model's performance was measured by evaluating not only the accuracy but also the uncertainty calibration. As a result, it was observed that there was no significant difference between the pre-defined feature-based ML models and deep learning models in terms of both accuracy and the expected calibration error (ECE). Second, GPCRs are membrane proteins that transmit signals that regulate cell activity through intracellular G proteins by binding to various ligands and have been targeted by many drugs. The structure of GPCRs changes significantly depending on ligand binding. We tried to obtain a more accurate GPCR structure by using AlphaFold-multimer and AlphaFold-based multi-state modeling protocols. Docking and virtual screening were performed using various docking tools on the model structures. We measured the performance regarding pose prediction success rate and screening power. The receptor models were as accurate as cross-docking scenarios, and the docking tool considering receptor flexibility achieved the best performance. In both cases, the structures of the membrane protein receptor model were relatively accurate. Still, the ligand-receptor complex structure modeling and interaction prediction methods showed limitations. These results indicate the need for more accurate structural modeling approaches and uncertainty prediction to compensate for the lack of data.

Keyword : Computer-aided drug design, Deep learning, hERG, GPCR, ligand-receptor interaction prediction, virtual screening

Student Number: 2021-22785

Table of Contents

LIST OF FIGURES

LIST OF TABLES

1. Introduction

Drug discovery is a huge market. It has shown rapid growth driven by the increasing burden of chronic diseases and the medical unmet needs of many rare diseases [1]. The small molecule searching process of drug discovery can be roughly divided into four steps, gradually progressing from just a drug-like molecule library to efficient, safe, and free of patent-issue candidates. Hit discovery is the first step in which active molecules to the target protein are filtered through high-throughput screening. Hit-to-Lead is a process of selecting hits with low off-target effects on other proteins and higher affinity to the target. Then, leads are optimized for multiple objectives like high efficiency, improved absorption-distribution-metabolism-excretion (ADME) properties, low toxicity, and synthesizability. Finally, candidates that satisfy time- and resource-consuming clinical trial endpoints can be approved.

The whole process is searching for optimal molecules through the vast chemical space. The broader and more accurate search increases the probability of finding candidates, but this requires a lot of time and cost. Computer-aided drug design (CADD) is an approach to alleviate this burden. CADD can be applied through drug discovery processes, from hit discovery to lead optimization. For example, fast and coarse-grained binding affinity prediction models can be used for virtual screening tasks. Then more accurate fine-grained bioactivity and molecular property prediction models, such as QSAR, can provide a computational guide during the lead optimization process.

Recently, AlphaFold showed remarkable improvement in protein structure prediction with high atomic level accuracy, raising the possibility of a wide range of CADD applications. However, it has several limitations, such as ill performance for proteins without coevolutionary signals (e.g., antibody) and conformation biasing for multi-state proteins (e.g., GPCR). The GPCRs have different structures depending on the class of ligands. Thus predicting its conformation in the correct activation state is important to recover ligand-protein interaction.

Thanks to the advance of experimental techniques and computation resources, high throughput data has been accumulated, and more powerful prediction models have been developed with the help of machine learning and deep learning. However, the data set needs to be improved to cover the vast chemical space. In most cases, chemical property prediction tasks

suffer from overfitting and out-of-distribution issues. In addition, deep learning is usually regarded as a black box because it is hard to map hidden features to specific physicochemical properties. The uncertainty and lack of interpretability of the predictions limit the use of the model in real situations. Therefore, in order to supplement this limitation, a deep learning model should quantify the uncertainty about its prediction. In addition, the model will be more practical if it can provide which element was critical for the prediction. These factors can help later in the CADD lead optimization process.

In chapter 2, we trained the hERG blocker classification models. We used various molecular features and prediction models, including molecular fingerprint or descriptor-based machine learning models and 2D or 3D structure-based graph neural networks. In particular, we incorporated the Bayesian framework and attention pooling in GNN. We analyzed each molecular representation method's informativeness and discrimination power by visualizing feature distribution and classification power. The performance of models was evaluated in terms of accuracy and uncertainty calibration, which is important in real-world applications.

In chapter 3, we evaluated the accuracy of GPCR structures modeled by the various AlphaFold methods and how well ligand-protein interactions can be reproduced using these model structures. It is known that it is helpful to consider receptor flexibility when running docking simulations on different receptor structures rather than the binding conformation to the target ligand. Various docking methods of different degrees of flexibility were tested to determine the advantages and extent of considering flexibility for GPCR targets. Finally, we presented the best computational methodology guide for developing GPCR-targeted drugs.

2. Deep Learning Model for Cardiotoxicity Prediction

2.1. Research Background

The human ether-a-go-go-related gene (hERG) codes a voltage-gated potassium ion channel expressed on cardiomyocytes. This channel regulates cardiac repolarization and maintains regular heart activity by releasing potassium ions outward. Blocking of this channel delays the repolarization of the membrane potential, which is directly related to QT

interval prolongation [2]. QT interval prolongation can lead to irregular heartbeat, called torsade de pointes [2]. The determination of the hERG structure explained why this channel is a major cause of drug-induced cardiotoxicity and why the empirical approaches to alleviate hERG blocking worked [2,3]. hERG is a C4 symmetric homo-tetramer with four hydrophobic pockets and a negatively charged selectivity filter, providing more probability than other channel proteins to interact with small molecules [3]. The CryoEM map of hERG bound to astemizole, an antihistamine withdrawn from the market due to cardiotoxicity, showed electron density in the hydrophobic pockets and selectivity filter, supporting the hERG-ligand interaction hypothesis [2]. To prevent approval of cardiotoxic drugs such as astemizole, dofetilide, and cisapride [3–6], FDA released a guideline for cardiotoxicity evaluation at the preclinical stage [7].

The clinical evaluation of cardiotoxicity is determined by the QT interval prolongation measured by electrocardiogram (ECG) [7]. At the early drug discovery stage, *in vitro* approaches like patch-clamp and displacement assay are used to measure IC50 as an estimator of cardiotoxicity. However, using these assays in screening takes a lot of cost and time. Thus, Computational methods have been developed to predict channel inhibition in advance and to sample appropriate candidates. Conventionally, 3D-QSAR models using pharmacophore [8][9], fingerprint- and descriptor-based models [10,11] have been developed.

Ryu et al. developed a graph neural network (GNN) for hERG blocking prediction named DeepHIT [12]. Unlike conventional methods, the molecular graph is a more sophisticated representation in that deep learning automatically extracts meaningful features from raw structure data. Following DeepHIT, CardioTox net, an ensemble of all ligand-based methods, was developed [13]. However, fingerprints and molecular descriptors greatly impacted performance, while GNN had little. In addition, those models failed to show consistent performance across test sets in the recent benchmark [14]. This result indicates the publically available data is too limited to cover the whole chemical space. Recently, a GNN model using multi-task learning and Bayesian inference has been developed to overcome these limitations [14]. The authors increased accuracy by pre-training the model with a more extensive dataset but with low resolution. Applying Bayesian inference to GNN improved the explainability, presenting different attention weight patterns depending on whether a blocker or non-blocker.

The models that consider only the ligand structure recognizes ultimately the typical “substructural pattern” that appears in active ligands rather than

general ligand-receptor interactions. Therefore it is expected that these models have limits in generalization. A universal model can be created by considering ligand and receptor information together. However, in the case of hERG, only the receptor structure is revealed. There is a partial electron density of the ligand in the CryoEM map, but not enough to determine the exact binding mode. To solve this problem, researchers used docking poses on various receptor models, including CryoEM models, homology models, and MD conformations [15]. The researchers trained SVM using interaction fingerprints, defined as the existence of contact between ligand atoms and pre-assigned residue atoms.

In this chapter, we evaluated the hERG blocker prediction performance of machine learning (ML) and deep learning (DL) models using various molecular representations such as fingerprints, molecular descriptors, and molecular graphs. First, we checked if each molecular representation could express the characteristics of and the relationship between molecules. Then, ML/DL models were trained for each representation, and classification performance was evaluated for accuracy and uncertainty estimation. For 3D graph representation, we tested docking poses generated by three docking tools on four receptor models. We compared the results and discussed what kind of effort would be needed to improve prediction performance.

2.2. Related Works

2.2.1. Deep Learning Models for hERG Blocker Classification Using Diverse Molecular Representations

There are different ways to express molecules in computer-processable forms. SMILES (Simplified Molecular-Input Line-Entry System) represents molecules as strings by assigning single characters to atoms, bonds, and stereochemistry. [16]. Molecular fingerprint is a fixed-size bit array where substructures of a molecule are hashed and assigned to predefined bit locations [17]. Unlike the previous substructure-based expressions, molecular descriptors are a set of properties defined at the molecular level, such as molecular weight (MW), like a log of the partition coefficient (logP), and topological polar surface area (TPSA). However, they are not bijective functions that redundantly express or omit chemical features. Therefore, much hERG blocker prediction research has focused on selecting meaningful features like creating QSAR models or making ensemble models using different representations to improve prediction accuracy [8–11].

Deep learning has changed this feature selection problem to how to train models to extract features from high dimensional raw data for itself. A molecule can be represented as a graph in which atoms and bonds correspond to node and edge. DeepHit is the first model which applied graph neural network (GNN) to the hERG blocker prediction problem [12]. It ensembled two fingerprint- and molecular descriptor-based MLPs and one GNN achieving state-of-the-art performance at the time. In addition, it optimized hit compounds for Urotensin II receptor by decreasing hERG blocking while maintaining binding affinity to the target protein. CardioTox net is an extended version of DeepHIT, which added 1D-CNN and embedding to SMILES and fingerprint inputs [13]. Each model predicts the probability of hERG blocking in the range of 0-1, and CardioToxnet puts together all predictions to a 2-layer MLP to get the final single expected value. However, an ablation study discovered that their improved accuracy came from predefined features like molecular fingerprints and descriptors rather than GNN, showing the limitation of deep learning in the application of data-limited conditions.

2.2.2. Application of Bayesian Framework for Uncertainty Calibration of Deep Learning Model

SVM, RF, Gradient boost, and typical deep learning models use a training framework in which the optimal model parameter is set to the maximum-likelihood (ML) or maximum-a-posteriori (MAP) estimator given data and label. However, these models don't guarantee to match the predicted probability to the reliability or uncertainty. For example, bagging or random forest that average base models have fewer values away from 0 or 1 due to the variance of base models [18]. Ryu et al. applied variational inference to graph convolutional neural networks for chemical property prediction tasks [19]. They used Monte-Carlo dropout for sampling parameters and its output as approximated posterior distribution. According to their framework, training without MC dropout corresponds to ML estimation, training with MC dropout but inference without MC dropout to MAP estimation, and turning on MC dropout at both training and inference time to Bayesian. They showed an improvement in classification performance in terms of the accuracy and the expected calibration error, which is the weighted average of error of expected positive fraction and measured positive fraction in a certain range of predicted output values.

BayeshERG authors aimed to create a GCN-only model that can be used practically by explaining and predicting uncertainty rather than focusing on improving performance [14]. To make the most of the publically available

data, not only IC50 data but also inhibition percent data at 10 μM concentration were used (hERGCentral dataset) [20] through transfer learning. Transfer learning is a training strategy that first trains a model on a bigger general dataset and then trains it again on a smaller but specific dataset as a fine-tuning concept. For quantification of reliability and explainability, BayeshERG applied Monte Carlo dropout (MC dropout) as an approximate posterior distribution and used multi-head attention to analyze which nodes were influencing. It showed slightly higher accuracy than CardioTox net (about 80%) but 10% points higher than descriptor-based ML models on different test sets and decreased expected calibration error (ECE) than CardioTox net. GNN is a powerful model allowing direct use of a molecular graph, but it is prone to be overfitted. Given that BayeshERG's architecture is not so different from others, it seems that BayeshERG alleviated the overfitting issue by using various deep learning training techniques and finally built enough robust GNN model.

2.3. Methods

2.3.1. Dataset for Training and Evaluation

We collected bioactivity data for training from ChEMBL by searching for "hERG" (ChEMBL ID 240), consisting of IC50 measured by patch-clamp or displacement assay. To remove duplicated or inconsistent data, we applied the following steps; 1) molecules of IC50 lower than 10μM were regarded as "blockers" and the other as "non-blockers ." For inequality relation, data with "IC50 < a value larger than 10uM" or "IC50 > a value smaller than 10uM" were excluded. 2) SMILES was canonicalized; if the smiles had multiple fragments like salt, only the largest fragment was kept. 3) data with the same SMILES were collected, and the mean pIC50 was used as the reference value. If a ligand has an average pIC50 between 4 and 6 but a standard deviation larger than 1, it was excluded due to uncertainty. As an exception, cases where toxicity could be determined from the literature were included in the dataset. For example, cocaine, known as cardiotoxic, was included in the dataset, although its mean pIC was 5.7 and standard deviation was 1.18. 4) Finally, a ligand in test sets was removed from the training set. The training set size was 9463 (5342 blockers (55.4%), 4301 non-blockers (44.6%)) (**Table 1**).

We used three test sets by CardioToxNet without any modification for comparison with other models. The test set 1 of CardioToxNet was borrowed from DeepHIT (30 blockers, 14 non-blockers). Test set 2 and 3 were constructed by CardioToxNet authors by searching literature

consisting of 44 ligands (11 blockers and 30 non-blockers) and 839 ligands (53 blockers and 786 non-blockers), respectively [21–23] (**Table 1**).

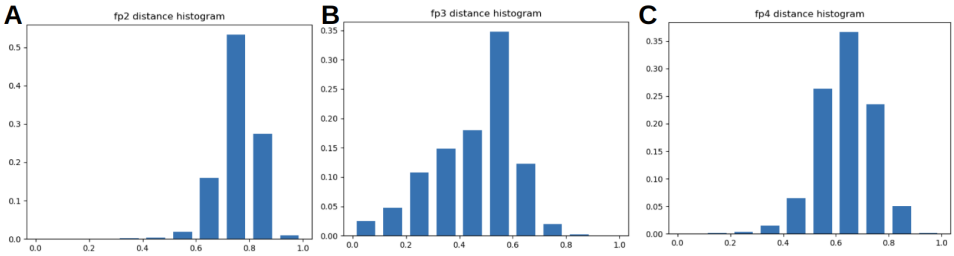
To search which molecular fingerprint well represents molecular similarity for train-valid splitting, we tested three types of fingerprints and two clustering methods; for fingerprints: 1) openbabel's fp2 (1024 bit, molecular fragments defined by atoms on the linear path up to 7 atoms)[24], 2) fp3 (SMARTS based functional group fingerprint, 55 bits)[24], 3) fp4 (SMART functional group defined by 'SMARTS_InteLigand.txt')[24]; for clustering, 1) hierarchical clustering based on Tanimoto similarity as distance using "scipy" python library, and 2) K-nearest neighbor (kNN) using the cartesian distance of fingerprints using "scikit-learn" python library. In addition, we tried scaffold splitting in which all ligands having the same ring and linker structure ignoring side chains, are categorized in the same cluster, but the distance between clusters is not defined. We used the "RDKit" python library to calculate the Bemis-Murko scaffold and assigned clusters to cross-validation sets, balancing the size of sets.

We observed that the distance distribution of ligands showed quite different trends depending on the fingerprints, and the correlation between fingerprints was very weak (**Figure 1**). In addition, when we tried hierarchical clustering using those fingerprints, all clustering results except FP2 were so biased that most data were collected in a single giant cluster (not shown). Because there is no external criterion to choose a better fingerprint, we used scaffold splitting, which is widely used in deep learning research. After clustering in consideration of molecular similarity, we divided the training set and validation set, making molecules to be included in the same cluster not included in the training set and validation set simultaneously. This data clustering is necessary because validation error is used to check whether overfitting occurs and adjust hyperparameters. Ideally, train-valid-test sets are expected to be sampled from the same distribution without replacement, and only in this case does evaluation on the valid set have a meaningful interpretation as a generalization error. In other words, to get reliable training and validation sets, they should have similar enough but different distributions. We concluded that it is reasonable to split data by cluster, not within a cluster, in the case of scaffold clustering, which collects almost the same molecules together.

Table 1. Dataset for cardiotoxicity prediction.

set type	train	test1	test2	test3
blocker	6565 (52.7%)	30 (68.1%)	11 (31.8%)	53 (6.3%)
non-blocker	5874 (47.2%)	14 (31.8%)	30 (68.1%)	786 (93.7%)
total	12439	44	44	839

Figure 1. **Distribution of ligands depending on molecular fingerprints**. A) distribution of distances of all pairs in the dataset by fp2, B) by fp3, C) by fp4. Distance is defined as 1 - Tanimoto coefficient. D) head-to-head comparison of distance by fp2 and fp3, E) fp2 and fp4.



2.3.2. Representations and Featurizations of Molecules

We used one fingerprint, two molecular descriptors and 2D molecular graph and 3D molecular graph from molecular docking. For molecular descriptors, features of zero variance, or which all molecules have same value, were removed. 3D conformation of ligands were obtained from docking pose using GalaxyDock3 [25].

1) ECFP of length 4 calculated by RDKit. 2048 bits

2) RDKit molecular descriptors. 196 features

3) Mordred descriptors. 1286 features

4) 2D molecular graph
   ● node_features (one node per one ligand heavy atom):
   - atom_type (40),
   - number of directly-bonded neighbors (0,1,2,3,4,5)
   - number of all hydrogens (0,1,2,3,4)
   - number of implicit hydrogens (0,1,2,3,4,5)
   - aromatic indicator (0 or 1)

   ● edge_features (if a direct bond between two atoms exists)
   - bond order (1,2,3)
   - aromatic indicator (0 or 1)
   - conjugate indicator (0 or 1)
   - ring indicator (0 or 1)

   all node features and edge features were calculated by RDKit and converted by one-hot-encoding or binary.

5) 3D molecular graph
   ● node_features
   - sybyl atom type (25)

   ● edge_features
   - bond_order (1,2,3)
   - amide indicator (0 or 1)
   - aromatic indicator (0 or 1)
   - no_bond (0 or 1)
   - distance (clipping at 5Å)

   Nodes include ligand atoms and protein atoms within 5Å to any

ligand atoms. This 3D molecular graph is basically a fully connected graph.

2.3.3. Molecular Docking for 3D Representation of Molecules

To get informative conformation of ligands for 3D representation, we tested different receptor models and three docking tools. Although hERG's apo and holo CryoEM structures were deposited to PDB, the structures were inaccurate with a local resolution of 3.5 Å, and the volume of the binding pockets was not enough to dock ligands, making clashes inevitable for some high-affinity ligands. For receptor, following four models were tested:

1) CryoEM apo (PDB ID: 5va1, deposit data 2017-05-03, resolution=3.7Å, tetramer modeled by applying C4 symmetry with single chain structure).
2) CryoEM with ligand density (PDB ID: 7cn1, deposit data 2021-01-20, resolution=3.7Å, tetramer modeled by clustering 2D image assuming C1 symmetry).
3) AlphaFold model of single chain and tetramer modeling using symmetry.
4) AlphaFold multimeter model of four identical chains at the same time.

For molecular docking, we used GalaxyDock2 (GD2), GalaxyDock3 (GD3), and GALigandDock (GALD). The degree of freedom (DOF) of each docking method is different. Ligand trans-rotational DOF, ring structure sampling, and receptor binding site backbone and side chain flexibility are considered.

2.3.4. Prediction Models for Each Molecular Representation

We trained three machine learning models and one deep learning model for molecular descriptors and fingerprint features; 1) support vector machine (SVM), 2) random forest (RF), 3) gradient boost (GB), and 4) simple multi-layer perceptron (MLP). The machine learning models were imported from the "scikit-learn" library and MLP was implemented by using the "pytorch" python library. We tested hyperparameters for each method; gamma (0.1, 1, 10, 100), input feature scaling (binary) and kennel types (radial basis function, sigmoid, linear) for SVM; number of estimators (50,100,500), minimum number of samples in a leaf node (1,5,10), and splitting criterions (gini, entropy) for RF; learning rate (0.01, 0.1, 1) and maximum depth (3, 20, 100) for GB. The other hyperparameters were set to

default. Additionally, to avoid overfitting, we performed principal component analysis (PCA) and used principal component vectors up to explaining variance 80%, 90%, 95% and the raw features.

For 2D and 3D molecular graphs, a graph neural network (GNN) architecture was used. The overall architecture is described in **Figure 2A**. Various message-passing and pooling functions were tested (**Figure 2B**). Graph convolution (GCN) is the basic message-passing function that makes messages by passing neighbor node features to a single perceptron layer and collects them to update the node features. Graph isomorphism (GIN) replaces single perceptron layers with MLP, and graph attention (GAT) generates key, query, value features for each node and updates the node features to the weighted average of values by attention weight, and the edge features to the weighted average of neighbor edge features.

2.4. Results and Discussion

2.4.1. Distribution and Clustering of Molecules by Representation and Featurization Methods

We visualized the distribution of blockers and non-blockers using different fingerprints and molecular descriptors by plotting the top 2 vectors of PCA to check if the difference between them is captured in the feature level (**Figure 3**). The distribution of ligands was significantly different according to the molecular fingerprint types. FP2, FP3, FP4, and Mordred descriptors showed similar distribution between blocker and non-blockers, while ecfp4 and rdkit descriptors showed separation between blockers and non-blockers. Especially unlike the other fingerprints, separation by ecfp would result from that higher dimension of ecfp could represent the shared and different substructures specific to each ligand type. As mentioned above, we just used scaffold splitting to train-valid-test set separation. RDKit descriptors showed highly concentrated distribution by PCA. This result means that data can be explained by a subset of discrete features, indicating active and inactive molecules have different molecular features. Although these differences, fingerprint types didn't affect prediction power (**Table 1**).

Figure 2. Architectures of 2D and 3D molecular graph GNN. A) GNN architecture, B) types of message passing functions and pooling methods.

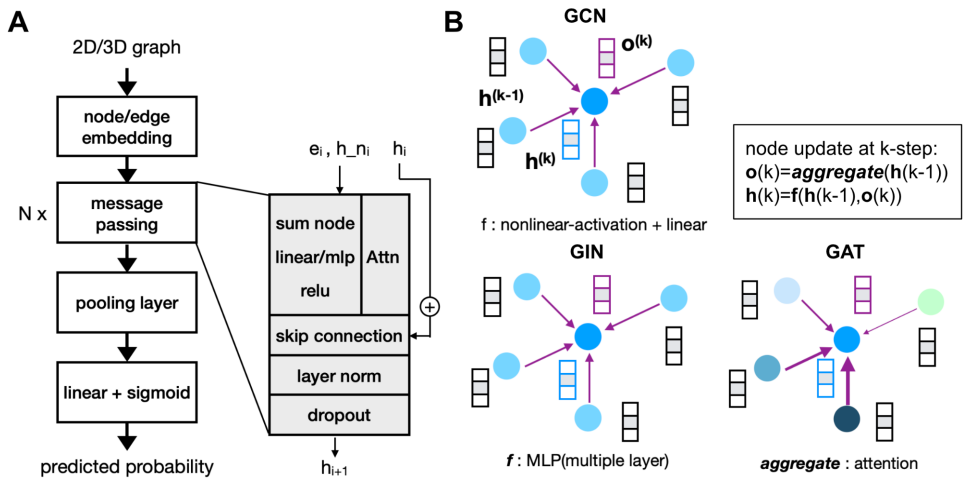
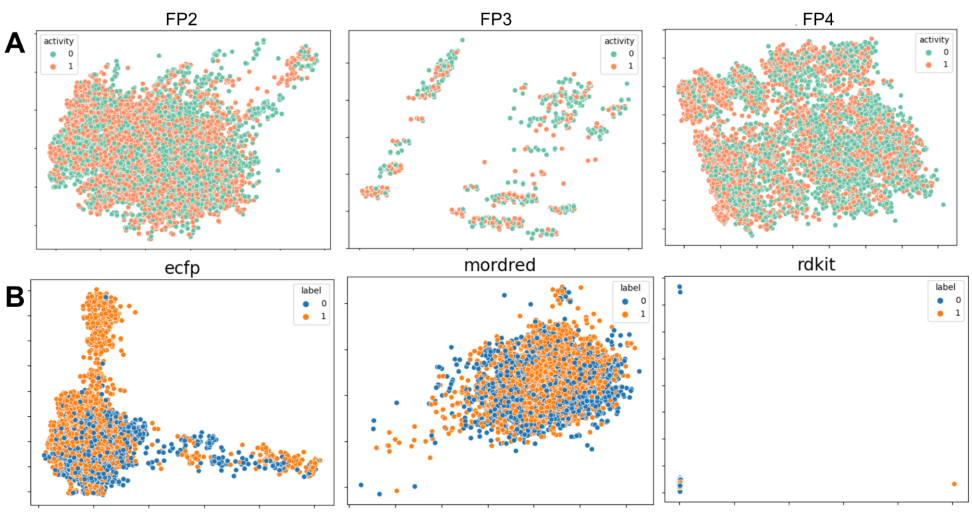


Figure 3. Distributions of molecules by representations. A) PCA plot of training set molecules by fingerprints. B) PCA plot of training set molecules by molecular descriptors for prediction.



2.4.2. Binding Pose Prediction Results by Docking Tools

To generate a 3D representation of ligands, we tested two docking programs. Because no hERg-ligand complex structure has been resolved, it was hard to evaluate reliable docking poses by comparing experimental structures. Therefore we modeled an ideal binding pose that fits to a hydrophobic pocket and interacts with the selectivity filter electrostatically according to the literature (**Figure 4A**). Next, we aligned and compared the docking poses of the top 300 ligands with the highest binding affinity. **Figure 4D** shows the top4 high-affinity ligands' top1 conformation. While GD3 consistently docked ligands of the same scaffold to the well-overlapped conformation, GALD showed highly variable poses. In addition, GALD sometimes docked large molecules out of the binding pocket and rotated Y652 outward of the channel, a key interaction residue well observed in the CryoEM map. To maximize the sampling of reliable poses, we ran GALD with Y652 fixed or gradually decreased the weight of the inter-residue clash energy term. However, fixing Y652 hinders approaching hydrophobic pockets by limiting pore size, and the clash term weight change had little impact on docking poses.

We quantified the two key interactions of the ligands and compared the distribution of interactions of GD3 results and GALD results to check which method is better for explaining binding interactions. The hydrophobic interaction was quantified as the sum of logP of atoms in the spheres. The selectivity interaction was calculated as the sum of the partial charge of atoms in the cylinder. Despite binding pose difference, the distributions of interaction features were identical, and none distinguished blockers or non-blockers (**Figure 4E**). Thus, we used the top1 pose from GD3 results as 3D-GNN input structures.

2.4.3. Cardiotoxicity Prediction Performance

Among prediction models, GB ranked the highest accuracy (**Figure 5, Table 2**). All models except 3D-GNN suffered overfitting issues, which was severe for machine learning models. RF and GB showed the highest validation accuracy and comparable test accuracies, unlike SVM + ecfp4 showing large training and validation performance gap and lower test accuracy. Machine learning models using RDKit descriptors showed the best validation accuracy among other molecular descriptors. Considering that ecfp4 consistently ranked the highest training accuracy but lower validation accuracy, ecfp is vulnerable to overfitting due to its high dimensionality and redundancy.

Although 3D-GNN showed ideal training and validation gap, 3D-GNN's test accuracy was the worst, while 2D-GNN showed consistent accuracy on test sets. This result indicates that the training was done correctly but the model's generalization power is too limited. While the training and validation set were split by scaffold, the test sets were constructed considering molecular similarity (< 0.7 Tanimoto coefficient). The failure of 3D-GNN seems to come from the noise of docking poses. Fingerprints, descriptors, and the molecular graph would have common substructures even between molecules of small Tanimoto coefficient, but 3D docking poses of molecules with shared substructures can have different conformations due to the docking result's stochasticity and ambiguity. This randomness is maximized when just a single pose of top1 is used.

2.5. Conclusion

In this chapter, we checked the representation power of various methods and made machine learning and deep learning models to predict hERG blockers. As a result, we observed all fingerprints and descriptors have their specific representation patterns and unique distributions but the similar impact on classification performance. Unfortunately, as reported in the previous studies, GNNs didn't help to improve prediction performance compared to basic machine learning models using pre-defined molecular descriptors.

Especially, 3D-GNN showed ideal training results but failed at test sets. It is because of the noise and ambiguity in docking poses. Human insight should be applied to input generation and representation as an inductive bias to improve docking noise. For example, we can sample various binding poses of high-affinity ligands and cluster them into several binding modes to match known interactions. After selecting reference poses, we can align ligands to reference binding modes and merge the results to remove stochasticity from docking poses, thus emphasizing subtle differences between blockers and non-blockers.

Besides accuracy, we also applied bagging for ML models and 2D-GNN, but there was no improvement in uncertainty calibration (not shown). We checked the mathematical limit of the aleatoric and epistemic variance of bagging, and it matched the variance plot of random data. More rigorous statistical proof and foundation will be needed to apply Bayesian inference to ML and DL.

Figure4. Distribution of docking poses by docking tools. A) hERG receptor structure used fo docking (PDB ID: 7cn1, CryoEM structure) and the ideal ligand binding pose (modeled by GD3). it has four hydrophobic pockets and negatively charged area below the selectivity filter [26]. B) Binding site entrance structures of the models; 5va1 (blue), AlphaFold monomer (pink), AlphaFold multimer (light blue). C) GALigandDock docking results depending on the receptor models. D) Docking poses of Top 4 high affinity blockers by GalaxyDock3 and GALigandDock. Only top1 energy pose of each ligand was shown. E) Distribution of the sum of logP in the hydrophobic pocket and the sum of partial charges in the selectivity filter entrance. blockers (orange), non-blockers (blue).

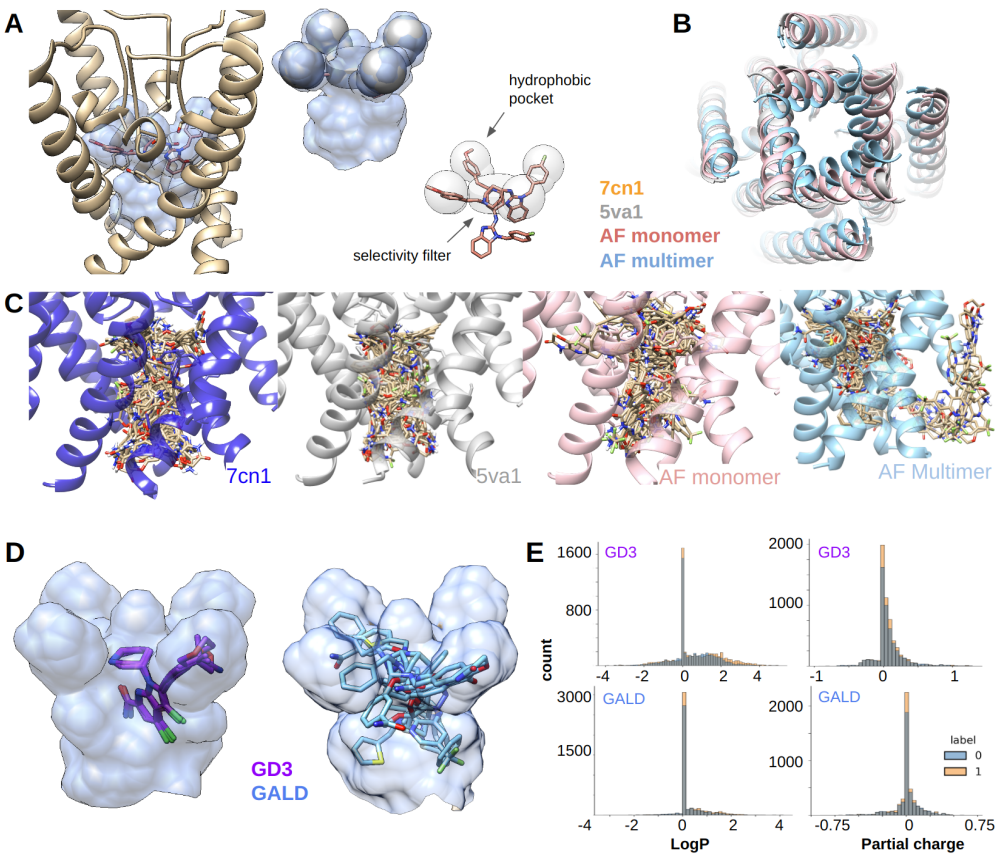


Figure 5. Classification performance of models.

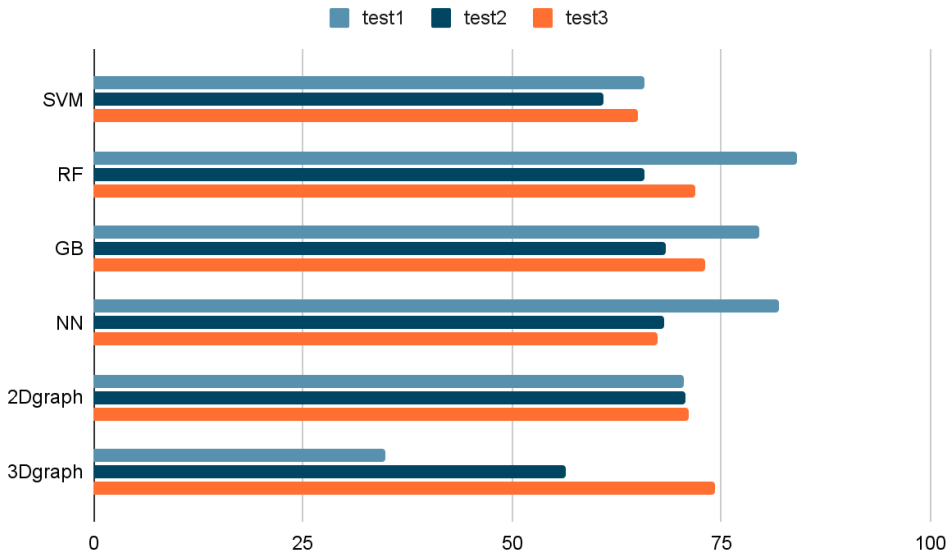


Table 2. Cardiotoxicity prediction accuracy. The test accuracy is of the highest validation accuracy model's prediction.

train acc (valid acc)	Mordred	RDKit	ecfp4	test1 acc test2 acc test3 acc
SVM	83.3 (62.3)	79.4 (67.0)	99.5 (70.4)	65.9 61.0 65.1
Random Forest	99.5 (75.8)	99.5 (76.5)	99.5 (71.7)	84.1 65.9 71.9
Gradient Boost	99.6 (72.7)	99.6 (77.5)	99.6 (73.6)	79.5 68.3 73.1
Neural Network	87.5, (71.0)	85.8 (72.0)	97.5 (69.7)	81.8 68.2 67.3
	best model		**train acc (valid acc)**	

2D GNN	message passing = GCN N message passing layers = 4 hidden dim = 128 pooling = max weight decay = 1e-5	94.7 (74.2)	70.6 70.8 71.1
3D GNN	message passing = GCN N message passing layers = 4 final_read_node = ligand hidden dim = 64 pooling = mean weight decay = 1e-6	80.6 (71.7)	34.9 56.4 74.3

3. Accuracy Evaluation of Protein Structure Predicted by Deep-Learning Model and Its Use for Ligand-Protein Interaction Prediction

3.1. Research Background

G-protein coupled receptors (GPCRs) are a large family of proteins with the same seven trans-membrane helix topology and the shared activation mechanism for many of them [27]. They transmit external signals into a cell, regulating cellular activities like metabolism and sensing. For this reason, one-third of FDA-proved drugs are designed to target GPCRs. However, the flexible structure of GPCR makes the determination of its structure very hard; therefore, only 103 structures out of 800 human GPCRs are available until January 2022 [28,29]. This lack of structural data limits the sequence identity of GPCR to homologs around 20~30% and also hiders the success of templated-based modeling (TBM), which works well with a close homolog. To overcome this limitation, multi-template and hybrid TBM protocols were proposed but showed limited performance. [30–33]).

In this context, the appearance of AlphaFold raises the question of whether AlphaFold can predict various structures of GPCRs, and whether the predicted model can be applied to drug discovery-related tasks like molecular docking and virtual screening. To answer these questions, He et al. evaluated AlphaFold models with experimental structures [34]; AlphaFold could predict the overall backbone structure of GPCRs, but the details have yet to be corrected. The orientation of the extracellular domain, the ligand-binding pocket, and the G protein binding interface conformation differed from the experimental structure.

In addition to the local inaccuracy of models, AF has another limitation in that it returns structures with bias in only one of the activation states. Because GPCRs show very dynamic structure changes including a binding pocket depending on their activation states, modeling GPCR in a specific activation state would help improve reproducing receptor-ligand interactions. Lim and Feig showed that AlphaFold could model the structures in the intended state by constraining template databases determined by a specific activation state and removing MSA from input features [29]. They also applied this activation-annotated AlphaFold modeling to get receptor structures for the binding pose prediction task. They showed improved performance on active-state targets, as expected given that molecular docking heavily depends on binding site accuracy.

Although structures of GPCR can be predicted accurately, the docking pose prediction is still a hard task because GPCR has a flexible binding site that changes its conformation variously upon ligand binding. For example, an intruding loop or side chain to the binding pocket can hinder even the sampling of the near-native pose. For this reason, the need to consider receptor flexibility during docking has been stressed [29,35]. There are docking methods considering a variety of flexibility from ligand torsion angle to side chain to backbone flexibility. It is expected that the full flexibility of ligand and receptor molecules during docking can solve this problem. However, the computation cost would be high, therefore the efficiency of flexibility consideration should be validated for bulky data experiments like virtual screening.

In this chapter, we evaluated receptor modeling methods and small molecule docking methods on the benchmark set covering all classes of GPCRs for a general evaluation. We investigated the accuracy of models by classical TBM, AlphaFold multimer, and AlphaFold with template biasing. Then, we compared the performance of docking methods in binding pose prediction and virtual screening using the best receptor models. We hope this research can provide a useful guide for GPCR modeling and molecular docking.

3.2. Related Works

3.2.1. multi-chain modeling using AlphaFold Multimer

AlphaFold multimer (AF-multimer) is a modified version of AlphaFold for simultaneous modeling of multiple chains [Protein complex prediction with AlphaFold-Multimer, bioRxiv, 2021]. AF-multimer has the same architecture to AF, which consists of evoformer and structure module, but training scheme is slightly different. As its name implies, protein complexes dataset was used as training set with interface-centric cropping to enhance the orientation accuracy. The clash term between chains was added to the loss function and changed summing to averaging for the local clash loss. As as results, It outperformed classical docking program showing DockQ score of 0.63.

We used AF-multimer for modeling GPCRs in active states. We modeled GPCRs with a Gα subunit or whole G protein subunits expecting G proteins can provide the information of active state by forming a specific interacting environment.

3.2.2. Multi-state modeling of GPCRs using AlphaFold

As mentioned earlier, AlphaFold tends to predict GPCRs in a single state. To overcome this problem, Lim and Feig tried different modifications to AlphaFold to seek the best method for activation modeling [29]. For benchmark, they modeled 68 GPCRs deposited after AlphaFold training. All GPCRs were modeled in either active or initiative states, not in both, and inactive GPCRs showed more accurate predictions. To make multi-state models, they constrained templates to the predefined GPCR sets according to the activation states. However, this biased template selection made little changes to the original model because MSAs were sufficiently deep. They removed MSA and only used templates like “template-based modeling”. This approach lowered accuracy slightly for inactive states but significantly increased the accuracy of active states, including key residues’ conformation changes.

3.3. Methods

3.3.1. Dataset for Benchmark

We selected GPCR targets to cover diverse GPCR classes. This set consists of 37 inactive-state and 24 active-state small molecule-receptor complex structures with resolutions lower than 3.5Å, comprising a total of 51 (**Table 2**). GPCRs of classes A, B1, C, and F are included. 18 of them are Cryo-EM structures, and the others are X-crystallography structures.

For virtual screening, we selected 10 GPCRs with a sufficient number of known ligands considering protein family diversity (**Table 3**). The overview is described in **Figure 6**. AA2AR, ADRB1, ADRB2, CXCR4, and DRD3, all in class A, are GPCR targets included in the DUD-E benchmark set, which is generally used to evaluate virtual screening performance in the academy. We used 3D conformation files of ligands and decoys in mol2 format downloaded from the website [36]. DUD-E set provides different protonation states of the same molecule. If there are multiple protonated forms, we randomly selected one state for ligands of the same ChEMBL ID.

Protonation assignment is a crucial factor affecting the binding affinity prediction. For general comparison with other results, we used protonation states given by DUD-E dataset as mol2 format except for CXCR4 ligands, for which DUD-E method failed to reproduce important key proton in ligand-receptor interaction [37] (**Figure S1**). We tested two other protonation methods. We used all ligands from DUD-E set and randomly

sampled 1600 decoys for each target, and then applied openbabel or chimera for protonation prediction at pH=7. We ran docking using Vina and GALD on the experimental structures. In **Figure S2**, while Vina showed no difference according to the protonation methods, GALD results on CXCR4 were dependent on protonation methods. Although openbabel showed the highest performance, it attached hydrogen to pyrimidine nitrogen, which is not preferred at physiological pH, making an artificial hydrogen bond with the receptor (**Figure S1**). Therefore, we used chimera for protonation prediction of CXCR4 ligands.

AGTR1, CFCR1, GRM2, OPRD, and S1PR1 were selected additionally to cover other classes of GPCRs (class A, B1, and C). We collected ligands from GPCRdb [38] and curated as the following process; 1) remove duplicates by smiles and ChEMBL IDs, 2) filter ligands of atom numbers smaller than 70, 3) cluster by scaffolds and pick the highest affinity ligand, 4) filter ligands with an affinity higher than 10nM. We used decoys generated by DUD-E server [36]. DUD-E predicts ligand's protonation states in pH 6~8 and then selects a set of molecules from ZINC database [39] with identical molecular properties but different scaffolds to ligands. This approach assumes that different scaffolds would have a lower probability of binding. There were some cases in which the same molecule had different smiles and different ChEMBL IDs. These cases resulted in duplicated DUD-E protonated ligands. Therefore removing duplicates by DUD-E result smiles was added at the end. We sampled decoys randomly at a ratio of 1:40 to ligands.

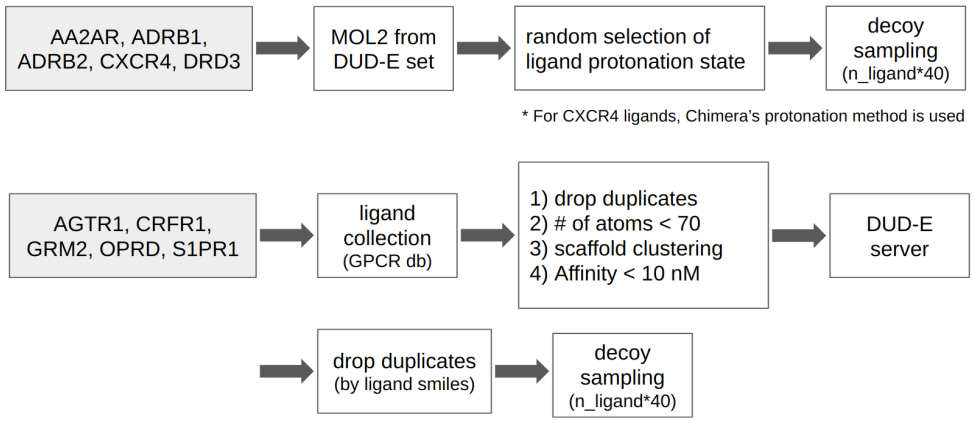
Table 3. List of benchmark targets.

PDBID	Receptor	Activation state	PDBID	Receptor	Activation state
5NM4	AA2AR	I	5WF5	AA2AR	A
4N6H	OPRD	I	7BU7	ADRB1	A
5WQC	OX2R	I	4LDE	ADRB2	A
6ZFZ	ACM1	I	5XRA	CNR1	A
6HLP	NK1R	I	6BQG	5HT2C	A
6PS2	ADRB2	I	6B73	OPRK	A
7WC8	5HT2A	I	6PT3	OPRD	A
4JKV	SMO	I	6X1A	GLP1R	A
7BVQ	ADRB1	I	7M3G	CASR	A
7F8Y	CCKAR	I	7TD4	S1PR1	A
7EPE	GRM2	I	7CMV	DRD3	A
5U09	CNR1	I	7NA8	GHSR	A
6BQH	5HT2C	I	7TD0	LPAR1	A
5ZBQ	NPY1R	I	7VKT	LT4R1	A
4MBS	CCR5	I	7C7Q	GABR2	A
4ZUD	AGTR1	I	7L1V	OX2R	A
6ME2	MTR1A	I	7VGY	MTR1A	A
3V2Y	S1PR1	I	7LD3	AA1R	A
3PBL	DRD3	I	7MTS	GRM2	A
4Z36	LPAR1	I	6XBK	SMO	A
4DJH	OPRK	I	6OIJ	ACM1	A
7F83	GHSR	I	7DFL	HRH1	A
4K5Y	CRFR1	I	7D7M	PE2R4	A
3RZE	HRH1	I	6WHA	5HT2A	A
6TPK	OXYR	I			
5YWY	PE2R4	I			
7C7S	GABR2	I			

Table 4. List of GPCRs for virtual screening

GPCR	biological ligand	# of unique ligands (ChEMBL)	# decoys
AGTR1	Angiotensin (peptide hormone)	186	7440
CRFR1	Corticotropin-releasing factor (peptide hormone)	101	4040
GRM2	metabotropic glutamate (neurotransmitter)	49	1960
OPRD	endorphin (peptide)	378	15120
S1PR1	Lysophospholipid	380	15200
AA2AR	Adenosine (neurotransmitter)	482	19280
ADRB1	Adrenaline (neurotransmitter)	247	9880
ADRB2	Adrenaline (neurotransmitter)	231	9240
CXCR4	C-X-C chemokine (protein)	40	1600
DRD3	dopamine	480	19200

Figure 6. Overview of virtual screening dataset construction



3.3.2. Receptor Modeling Protocols

We used four types of AlphaFold and TBM.
For active-state GPCRs, the following four types of AlphaFold were used:

1. AlphaFold (GPCR sequence only, with MSA and templates)
2. AlphaFold with template biasing (templates only)
3. AlphaFold multimer with G protein alpha subunit sequence
4. AlphaFold multimer with all G proteins sequences

For inactive-state GPCRs, the following two types of AlphaFold and TBM were used:

1. AlphaFold
2. AlphaFold with template biasing
3. TBM models provided by Bender et. al. [40]

We used TBM only for inactive conformations because RosettaCM returned models in inactive states. Templates for Alphafold protocols except template basing was searched on PDB70 (version May 2020) using HHsearch in HH-suite 3 [41], while the maximum template date was limited to the day before the release date of the query PDB. Sequence for MSA building was searched using jackhmmer 3.3.2 [42] against UniRef90 (version 2022) [43], BFD (version Mar 2019) [44]and Mgnify (version Dec 2018) [45]. We tried different parameters for recycle number and AMBER energy relaxation and used default parameters. Template biasing models were downloaded from [29] if available. Otherwise, GPCR was modeled using run scripts and the state-annotated database provided by the authors (borrowed in Aug 2020, github.com/huhlim/alphafold-multistate). Meilerlab's TBM was constructed by RosettaCM with multiple templates of sequence identity lower than 40% assuming limited conditions [30]. It should be noted that AlphaFold has no sequence identity criteria for templates except the deposit date.

3.3.3. Small-molecule Docking Protocols

We used five docking tools covering different levels of ligand-receptor flexibility. We also tried DL-based docking tool (EquiBind [arXiv.2202.05146]) but the result was excluded because its training set would contain the benchmark sample and the reported accuracy was not reproduced in our own running.

1. AutoDock Vina [46]
   Degree of freedom: ligand translation, rotation, and torsion angles

We also used the flexible side chain option. However, flexible side chains should be assigned manually. Therefore, We defined flexible residues which satisfy the following criteria; 1) any of side-chain atoms within 3Å to ligand atoms, 2) χ angle is different from that of experimental structure more than 40°. 49 of 51 GPCRs have flexible side chains by these criteria ranging from 1 to 21. We also tried docking with up to two flexible residues following [47]. The consideration of full residues showed slightly better results, and only its result was reported.

2. GalaxyDock3 [25]
   Degree of freedom: DOFs considered in Vina, ligand ring flexibility by sampling ring conformations from a crystal ring structure library

3. CSAlign-Dock [48]
   Degree of freedom: DOFs in GalaxyDock3, a shape score measuring the similarity of the query ligand to a reference ligand is added to the GalaxyDock2 energy score. GalaxySite [49] was used to search three reference molecules of the highest Tanimoto coefficients.

4. Rosetta GALigandDock [50]
   Degree of freedom: DOFs in GalaxyDock3, a set of side chains and backbones selected automatically during the docking.
   Unlike other tools, the docking process is repeated 15 times following the guidelines for receptor-flexible docking for increased convergence.

5. Galaxy7TM [51]
   Degree of freedom: all residues' backbones and side chains by sampling an ensemble of perturbed receptors and side chain repacking followed by relaxation.

For AlphaFold-derived protocols, model number 1 was used for docking. All docking tools used true binding site information, and default parameters were used.

Ligand extracted from experimental PDB structure was converted to SMILES by OpenBabel [24] and then 3D structure by CORINA to get random initial conformation. For GALigandDock, ligands were protonated at pH 7, and partial charges were assigned by MMFF94 forcefield [24,52] using OpenBabel. For the other docking tools, Chimera [53] was used for protonation and Gasteiger partial charge assignment [54].

3.3.4. Virtual Screening protocols

We performed a virtual screening on 10 GPCR targets using GALigandDock which showed the best docking pose prediction accuracy. AutoDock Vina and Δ_{vina} RF_{20} [55] were also tested for comparison as a baseline and a state-of-the-art model, respectively. Virtual screening mode of GALigandDock was used following the guideline (https://www.rosettacommons.org/docs/latest/scripting_documentation/RosettaScripts/Movers/GALigandDock) The predicted affinity and ΔG of the top1 pose was used for Vina and GALigandDock, respectively. For Δ_{vina} RF_{20} , the top1 pose of Vina results was used to calculate input features using modified delta vina provided by the authors and MSMS in MGLTools (version 1.5.7). The predicted affinity and scores were used to measure the ROC and enrichment factors (EFs).

3.4. Results and Discussion

3.4.1. Protein Model Accuracy Evaluation

We evaluated the model structures in two aspects: 1) the overall quality of GPCR structures and 2) binding pocket shape accuracy, which is closely related to docking performance. We used TM-scores and backbone root-mean-squared-distance (RMSD) of binding site residues. As expected, AlphaFold models were better than TBM in global and local accuracy for inactive conformations (**Figure 7A, B**). However, this is not a fair comparison because the sequence identity of TBM is limited to 40%, otherwise unlimited. To complement the observation, we also analyzed AlphaFold models with template sequence restraint in **Figure S3**. Although the overall structure and binding site accuracy decreased, it still showed better results than TBM.

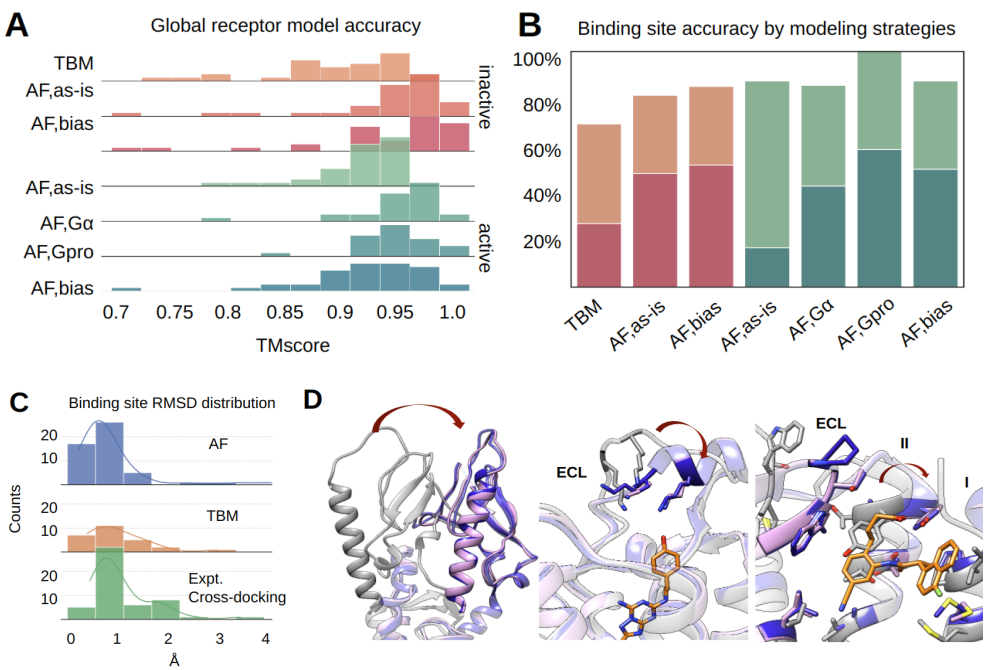
For active state GPCRs, AlphaFold multimer results modeled together with the G α subunit (“AF,Gα”) showed the highest TM-score (**Figure 7A**). However, the binding site accuracy was the highest when using the AlphaFold with entire G protein subunits (“AF,Gpro”), followed by AlphaFold with template biasing (“AF,bias”) and “AF,Gα” (**Figure 7B**). Considering that binding site accuracy is highly correlated with the docking performance, “AF,Gpro” is the best choice as receptor modeling. However, when adding 14 peptide binding complexes, AlphaFold multimer models and “AF,bias” showed similar accuracy (**Figure S4**). Although the superiority between AlphaFold multimers and “AF,bias” is hard to figure out due to limited data size, it should be noted that both AlphaFold multimers and “AF,bias”

modeling outperformed naive AlphaFold (“AF,as-is”). For example, **Figure 7D** shows models by “AF,bias” (blue) having correct backbone structures in the binding site, while “AF,as-is” failed to reconstruct the key interaction in the extracellular loop and blocked the binding pocket with TM helix.

To weigh up how difficult docking tasks are when using AlphaFold models, we compared the binding site RMSD distribution of three structures (**Figure 7C**): AlphaFold with template biasing, TBM, and the same GPCRs bound with other ligands, which corresponds to the cross-docking scenario. AlphaFold models showed accuracy comparable to that of when using receptors bound to other ligands. Docking task difficulty is determined by receptor accuracy. The easiest case is when the bound form of the receptor to the same ligand is used. In other words, RSMD is 0. Because the receptor's structure bound to other ligands has an almost accurate backbone structure compared to model structures, the cross-docking scenario has more chance to get the right binding poses than model docking.

Interestingly, the improved TM-score of AlphaFold multimer models for the active state came from correcting the orientation of the large extracellular domain (**Figure 7D**). AlphaFold and AlphaFold multimer share the key architectures and training strategy, except template searching method, input cropping for training, and some minor adjustment to the loss function. This difference can affect the domain orientation problem: template selection and model parameter due to different training schemes. To figure out which component is dominant, we compared the selected templates of AlphaFold and AlphaFold multimer to the experimental structure. Surprisingly, the accuracy of the templates’ domain orientation is the same, indicating that the template structure has little influence on the domain orientation and that the improved performance might stem from the training strategy. To distinguish the effect of the model parameter from the effect of simultaneous modeling with the Gα subunit, GPCR modeling without Gα subunit using AlphaFold multimer should be compared to “AF,Gα” models.

Figure 7. **GPCR model quality by different modeling methods**. “TBM”, template-based modeling; “AF,as-is”, plain AlphaFold without any modification, “AF,bias”, AlphaFold with a biased template set matching the functional state [29]; “AF,Gα”, AlphaFold-Multimer modeling of receptor + Gα subunit; “AF,Gpro”, AlphaFold-Multimer modeling of receptor + whole G protein subunits. A) Global receptor model accuracy measured by TM-score. B) Binding site accuracy measured by the fraction of models with backbone RMSD < 0.5 Å and < 1.0 Å shown in dark and light bars, respectively. Backbone RMSD refers to RMSD between the backbone atoms of the model and the experimental structure. C) Distribution of binding site accuracy for AlphaFold models (top), TBMs (middle), and experimental structures bound to other ligand molecules (“cross”, bottom). y-axis; the number of models belonging to the corresponding RMSD bin D) Examples of receptor models showing large differences with the following color scheme: experimental structure (pink), best model (blue), worst model (grey) (left, GLR, 6wpw) the extracellular orientation was corrected when modeled with the G alpha subunit together. (middle, AA2AR, 5nm4) AF-bias correctly modeled the extracellular loop alpha helix fragment, which contains lysine interacting with the ligand, while AF as-is failed with the unfolded loop. (right, PE2R4, 5ywy) The TM helix I and II modeled by TBM invade ligand binding pocket, which makes sampling the correct pose impossible, while AF-bias showed a structure similar to the native.



3.4.2. Docking Pose Prediction Evaluation

Based on the receptor modeling results, we selected four types of receptor models for docking: 1) experimental structure, 2) TBM, 3) “AF,as-is”, and 4) “AF,bias”. Docking on experimental structures provides the upper limit of performance for the targets. We used “AF,bias” as the representative for advanced AlphaFold modeling methods because it can be applied to both inactive and active states and showed comparable binding site accuracy to other modeling methods (**Figure 7**). TBM was used as the best classical method before deep learning.

The receptor modeling methods affect the performance of all docking methods (**Figure 8A**). The success rates of docking on the experimental structures were ranging from 70 to 80%. However, docking on TBM showed the worst performance limiting the success rate to lower than 20% except for CSAlign which is relatively less sensitive to receptor accuracy than molecular docking methods. Through all docking methods, “AF,bias” were slightly better than “AF,as-is” although the difference was not statistically significant. **Figure 8E** shows the example of the receptor modeling effects on ligand-receptor interaction recovery and binding pocket blocking. Thus, we focused on the results of “AF,bias” for further analysis.

The degree of flexibility considered in the docking process also affected the pose prediction performance (**Figure 8B**). Comparing the results of rigid or flexible receptor docking protocols of the same docking tool showed significant improvements of 10% points for Vina and 25% points for GALigandDock. Ligand ring flexibility also helps, as shown in GD3 results (**Figure 8B**). This improvement came from the enabled sampling of near-native conformations by widening the searching space by considering ligand and receptor flexibility (**Figure 8C**). Docking on model structures often fails to recover residue conformations for the key interaction and blocks binding sites even by minor difference, as shown in **Figure 8E, F**. Thus, consideration of side chain rotamer and backbone flexibility can make up for this difference when backbone structures are quite accurate (**Figure 8D**). However, when receptor models are not similar enough (binding site RMSD larger than 0.5Å), flexible docking didn’t help, and the success rates dramatically decreased (**Figure 8D**). Unfortunately, this result is unsatisfying. This range of difference (up to RMSD 1Å) should be solved in the flexible docking process rather than the receptor modeling stage, considering that it is observed in the cross-docking scenario as in **Figure 7C**. GalaxyDock7, which considers full flexibility, showed worse performance than other methods (**Figure S5**).

CSAlign was relatively less affected by receptor accuracy and reached comparable performance to the best method GALigandDock in success rate and sampling power (**Figure 8**). It is an expected result because it uses as guidance the conformation of the most similar ligand from similar complexes selected by receptor sequence identity or binding site conformation to supplement the receptor inaccuracy. For this reason, CSAlign could sample near-native pose at least for one target on receptor models of binding site RMSD larger than 1Å. However, this approach highly depends on the database, therefore its application is limited if there is no complex information close enough.

In brief, multi-state modeling using AlphaFold with template-biasing helped to improve the global and local accuracy of receptor models, and GALigandDock which considers receptor flexibility showed the best performance in pose prediction. As a result, we suggest using this AlphaFold modeling protocol and the flexible docking tool for the best result (47% of success rate in this benchmark). In a real situation, the accuracy of the receptor model is unknown because the answer structure is not available. It was confirmed that there is a correlation between the average of predicted lddt (plddt) by Alphafold and the success rate, and reliable results are expected to be obtained at plddt 0.95 or higher (**Figure 9**).

3.4.3. Virtual Screening Performance Evaluation

Based on pose prediction results and the docking method's generality, we used GALigandDock for virtual screening. Although we have shown that "AF,bias" was the best modeling method, we used "AF,as-is" for simplicity, assuming that small molecules are usually inhibitors.

All virtual screening methods showed a trend that the AlphaFold model was better than TBM in terms of the average AUC (**Figure 10A,B,C**). GALigandDock was better than the other methods for most GPCRs in respect of both AUC and enrichment factor (EF). Especially, GALigandDock outperformed other methods for AGTR1 (**Figure 10F**). GALigandDock's AUC was 0.79, while others were around 0.58, slightly better than random classification. In addition, EF of GALigandDock at various cutoffs was more than 5 times higher (**Table 4**). However, the variance of performance was large because performance depending on receptor type was quite different (**Figure 10D,E**).

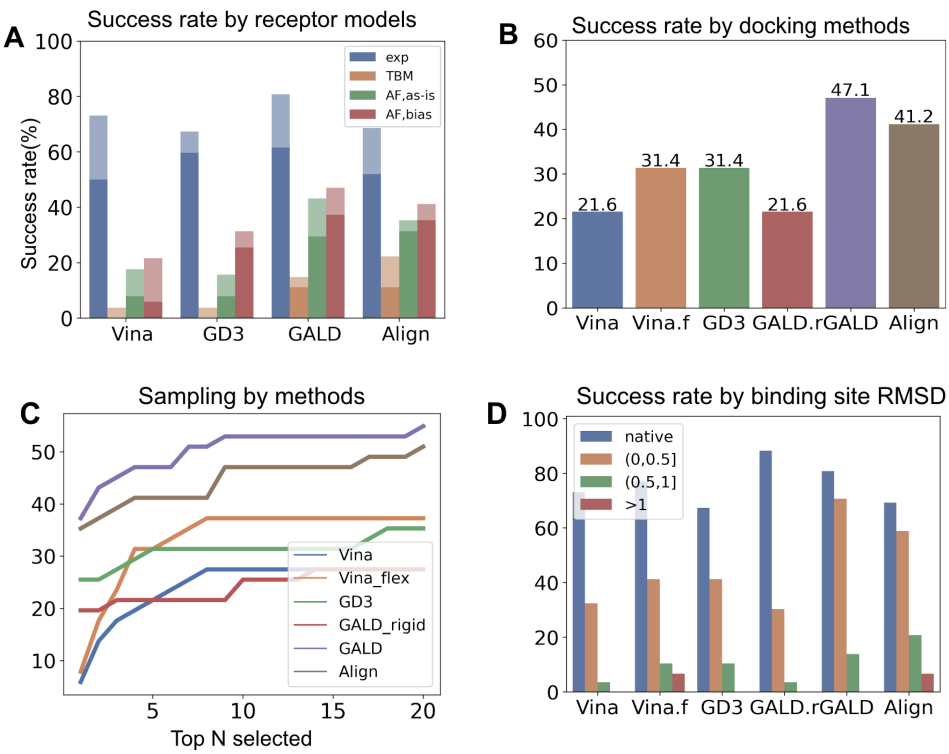
Notably, Δ_{vina} RF_{20} failed to improve classification performance, unlike the CASF16 results, even though it showed minor improvement in EFs (**Figure 10A,B,C**). This result indicates that rescoring approach is not enough to

improve performance.

The classification performance of the other methods was also different for each GPCR target, as shown in **Figure 10F**. ROC curves of AA2AR, ADRB1, ADRB2, DRD3, OPRD, and S1PR1 had a typical shape with a preferred feature of a high true positive rate (TPR) at a low false positive rate (FPR), which is related to EF. On the other hand, EF values of CXCR4 and GRM2 were unstable for all methods (**Table 4**), which seems like because of small ligand numbers (40 and 49, respectively). Interestingly, Vina and Δ_{vina} RF_{20} predicted the opposite on the TBM and AlphaFold model of CRFR1 (**Figure 10F**). TBM and AlphaFold modeled the TM6 helix inside the binding site, overlapping the true binding pose (**Figure S6**). These wrong TM6 helix structures made most ligands docked outside of the binding site and some decoys docked in the binding site with low energy, resulting in the inverse classification phenomenon.

To summarize, using the AF model and flexible docking increased virtual screening performance as the docking pose prediction task. On the other hand, Vina and Δ_{vina} RF_{20} showed no consistency on the receptor types for EF. Comparing the results of GALigandDock without the flexible-receptor option would help to figure out the effect of flexible docking clearly.

Figure 8. Pose prediction performance by receptor and docking methods. “exp”, experimental structure; “TBM”, template-based modeling; “AF,as-is”, plain AlphaFold without any modification, “AF,bias”, AlphaFold with template biasing; A) Top1 (dark bars) and top5 (light bars) success rate with success criterion of ligand RMSD < 2.5A. B) Top5 success rate by docking methods. C) Conformation sampling ability of docking tools. D) Dependence of docking performance on receptor binding site accuracy. E) Example of binding site accuracy importance. Inaccurate binding site backbone structure failed to reconstruct interactions by positioning key side chains away from the binding site. AA2AR (PDB 5wf5). wrong side chain in dark grey. F) Examples of flexible docking importance. Rotating side chains made enough volume in the binding pocket, allowing sampling of near-native conformations. ACM1 (PDB 6zfz) on the left, OPRD(PDB 6pt3) on the right.



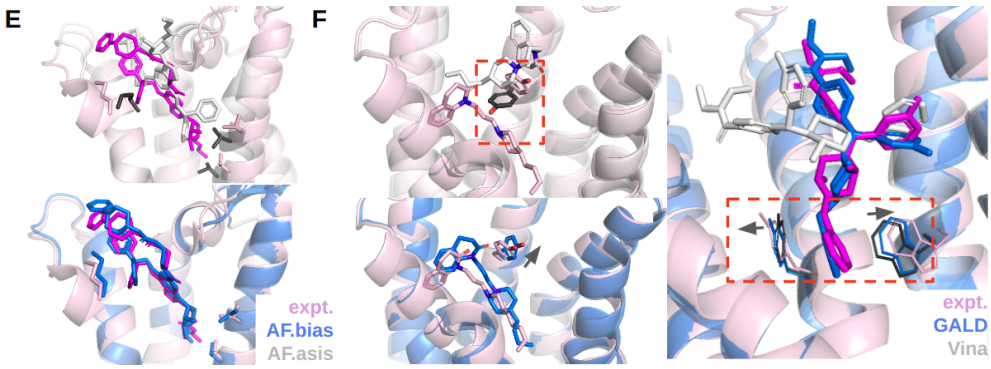


Figure 9. Pose prediction success rate depending on the average of binding site plddt. the success rate of binding pose prediction using "AF,bias" receptor models. The number on the bar indicates the number of receptor models belonging to the plddt bin.

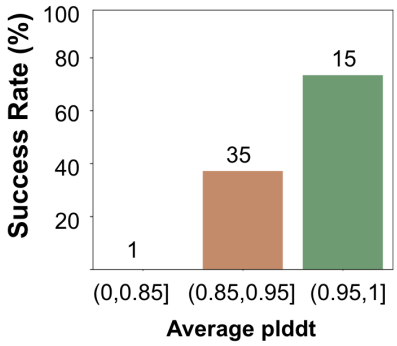


Figure 10. Virtual screening performance by receptor and binding affinity prediction methods. A) mean AUC of ROC curve, B-C) Enrichment factor at 0.5, 1%. D) AUC of GALD by GPCR targets, E) enrichment factor at 1% of GALD by GPCR targets F) ROC curves by virtual screening methods of each target on the AlphaFold model.

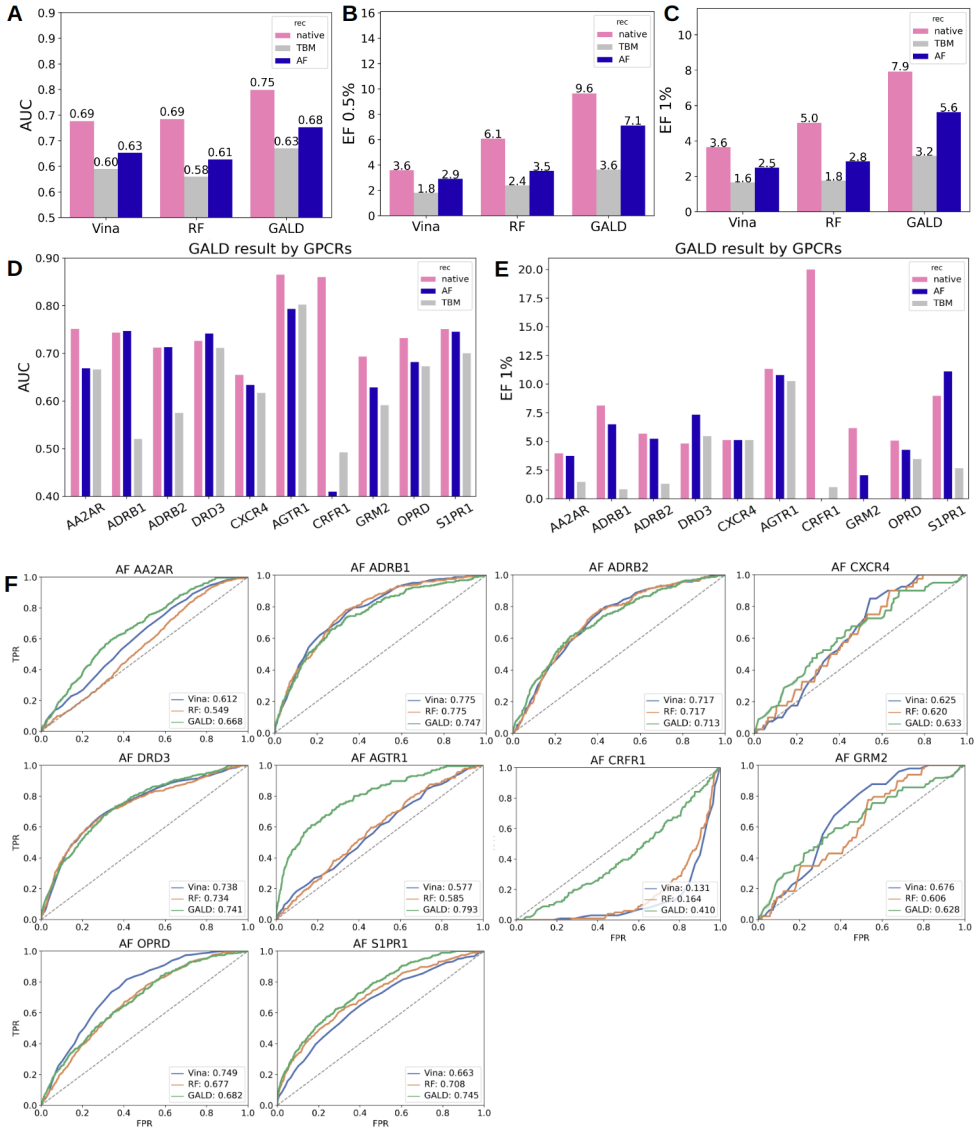


Table 5. Virtual screening results of each target (EFs)

Vina

cut-off	AA2AR	ADRB1	ADRB2	DRD3	CXCR4	AGTR1	CRFR1	GRM2	OPRD	S1PR1
0.1%	0	4.1	0	9.3	13.6	1.95	0	0	3.64	2.73
0.5%	0.98	3.83	1.74	6.15	1.71	2.88	0	2.73	4.61	3.02
1%	1.32	3.25	1.74	4.61	0.85	3.06	0	4.1	4.26	2.73
5%	1.73	2.89	2.31	3.37	0.83	2.51	0.4	2.87	3.55	1.88
10%	1.49	2.66	2.34	3.17	0.67	2.26	0.33	2.32	3.23	2.07

Δ_{vina} RF_{20}

cut-off	AA2AR	ADRB1	ADRB2	DRD3	CXCR4	AGTR1	CRFR1	GRM2	OPRD	S1PR1
0.1%	0	10.9	3.04	10.8	0	3.9	0	6.83	7.29	7.29
0.5%	0.42	6.01	3.2	5.31	1.71	2.88	0	8.2	6.74	5.5
1%	0.76	4.87	3.05	3.98	1.71	2.34	0	6.15	4.97	4.23
5%	1.04	3.57	2.48	3.52	1	2.26	0.73	3.01	3.11	2.7
10%	1.13	2.98	2.45	3.15	0.92	2.19	0.66	2.46	2.72	2.17

GALigandDock

cut-off	AA2AR	ADRB1	ADRB2	DRD3	CXCR4	AGTR1	CRFR1	GRM2	OPRD	S1PR1
0.1%	5.75	6.83	4.55	12.9	27.3	19.5	6.83	6.83	1.82	20.0
0.5%	3.07	5.47	4.07	7.82	6.83	13.6	10.9	2.73	4.08	9.23
1%	3.05	5.14	4.07	5.87	5.12	10.7	7	2.73	4.26	7.58
5%	2.28	3.24	2.54	3.5	3.17	7.1	3.9	2.46	2.98	3.93
10%	2.35	2.71	2.47	3.08	2.42	4.73	2.64	2.25	2.53	3.08

3.5. Conclusion

In this paper, we compared protein receptor model quality using deep learning and examined how much performance can be expected in the current situation by using a docking method considering various degrees of flexibility. Based on the results, we expect to get the best results when using AF with template biasing and GALigandDock.

The deep learning method showed significantly improved performance compared to the classical modeling method, TBM. The binding site accuracy was also improved, showing modeling results comparable to cross-docking scenarios. Also, the accuracy of the active state conformation was improved when using Alphafold multimer or AF with template biasing than naive Alphafold. The difference in the binding site backbone of AF models is comparable to a difference that can be observed in the experimental structures of the same protein.

When using the AlphaFold model, GALigandDock showed the best performance in the pose prediction task. This is a result of considering receptor flexibility and can be confirmed more clearly by comparing the flexible and rigid modes of Vina and GALigandDock. However, the performance of 7TM considering full flexibility was not good. By checking whether a near-native receptor structure was sampled in the ensemble generation step and then checking the sampling and scoring performance of ligand conformation, it would be possible to confirm which part is the bottleneck in full flexible docking.

Although considerable improvement has been made in the receptor modeling part with the help of deep learning, the success rate of the best practice docking method was 47.1%, which is only half of the performance when docked to the experimental structure, indicating that there is still a lot of room for improvement. When the binding site backbone RMSD was less than 0.5, the performance was comparable to that of docking to the experimental structure, but the success rate decreased dramatically when the receptor structure difference was larger. This shows that the current flexible docking does not fully consider flexibility. In addition, GALD showed better performance in virtual screening, but it took about 40 times more time than Vina. This limits the practical application of flexible docking to a huge drug library. In this context, alignment can help with this limitation. We observed that CSAlignment can guide docking quite accurately regardless of receptor difference if there is sufficiently similar complex data. Considering both accuracy and time-saving aspects, a hybrid docking

method that can utilize a wider dimension of flexibility and utilize appropriate reference ligand-receptor interaction information be an alternative approach.

4. Conclusion: limitations of using predicted structure information for deep learning

In this research, we made the cardiotoxicity prediction model using hERG-ligand docking poses and evaluated GPCR receptor modelings and docking performance from the perspective of drug discovery.

Unlike the expectation, docking pose information didn't improve the hERG blocker prediction. This result indicates that inaccurate and noisy structure data hinders the extraction of meaningful patterns and the generalization of interactions. To solve this problem, removing artificial noise from predicted docking poses should precede before constructing structure-based models. For example, building reference structures according to literature and aligning ligands to the reference structures could help. This process puts ligands in standard binding modes before inference, thus making deep learning models focus on differences in the interactions between the ligand and experimentally proven key residues. This constraint is called inductive bias, which is useful when applying DL to complex problems with insufficient data.

Like hERG blocker prediction, virtual screening on GPCR depends on the quality of predicted docking poses. AF showed sufficiently good receptor model quality comparable to cross-docking scenarios. However, insufficient flexibility of docking methods limited the accuracy of docking pose prediction and virtual screening. Although it is hard to evaluate the binding pose quality of screening library ligands without experimental structures, the more consistent docking method (GALD) also showed high virtual screening power. To determine the impact of input binding pose, we can collect docking pose data and apply them to different re-scoring models.

In conclusion, these results showed the need for more accurate and consistent docking methods for cross-docking tasks and applying physicochemical knowledge to deep learning models for generalization and reliability due to data-insufficient cases.

Supplementary Information

Figure S1. Protonation results by DUD-E, openbabel, and chimera. protonation state from DUD-E lacks protonated nitrogen which makes a salt bridge with Asp97 of CXCR4 [37]. Openbabel added hydrogen to both nitrogens in pyrimidine which is unfavorable.

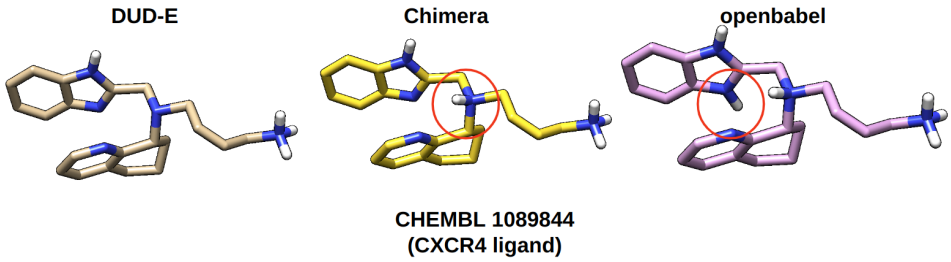


Figure S2. Receiver Operating Characteristic (ROC) curves by ligand protonation methods. row) docking tools; Vina and GALD, columns) GPCRs; ADRB1, ADRB2, CXCR4, and DRD3, hue) protonation methods; DUD-E server (pink), openbabel (grey), and chimera (blue). AUC of ROC curve is in the legend.

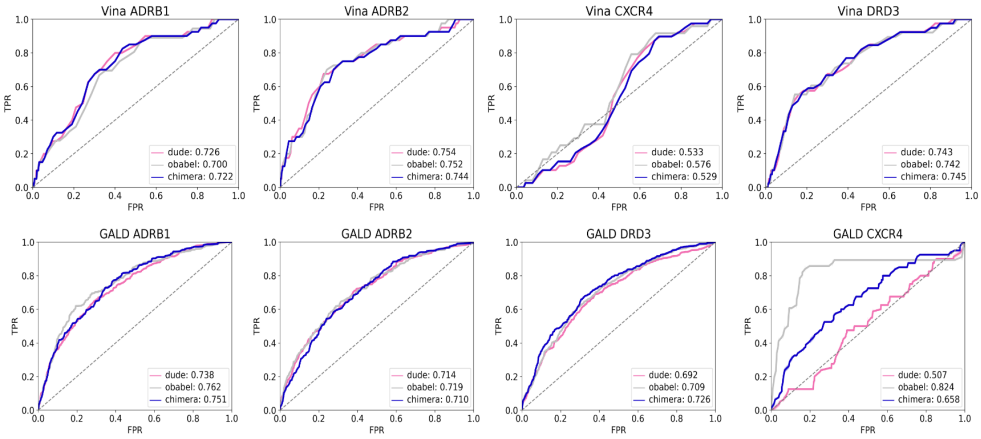


Figure S3. Receptor model quality with template sequence identity constraint. A) TMsocre distribution, B) fraction of binding site backbone RMSD within < 0.5Å (dark) and < 1Å (light). C) distribution of binding site backbone RMSD.

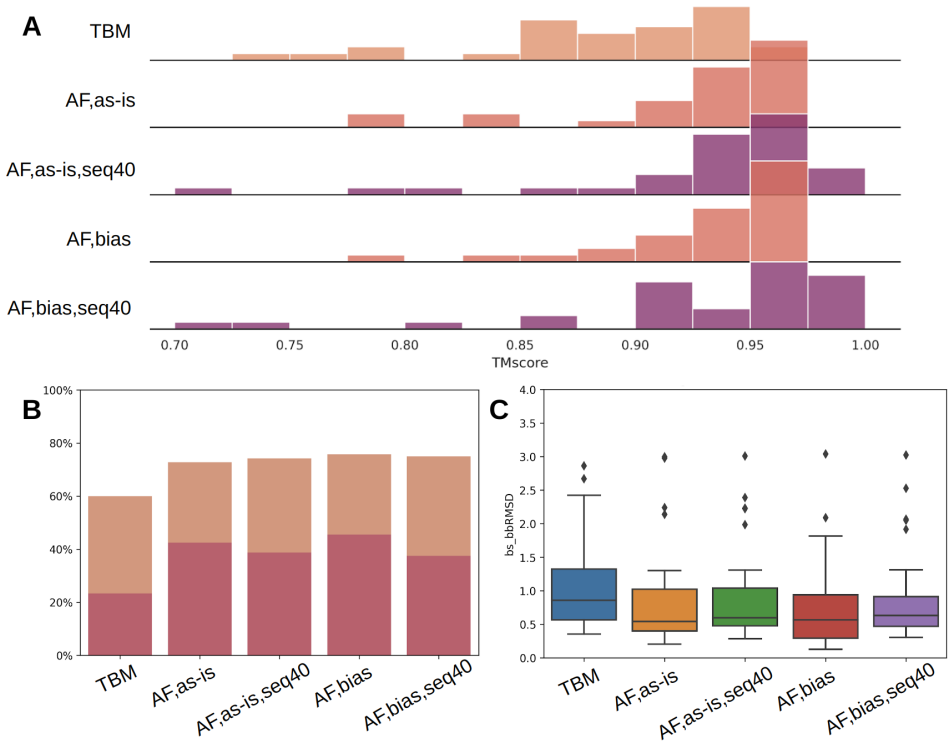


Figure S4. Receptor model quality including peptide binding complex. local accuracy of receptor models including 19 peptide-bound GPCRs (14 actives, 5 inactives). y-axis is the fraction of binding site backbone RMSD within < 0.5Å (dark) and < 1Å (light).

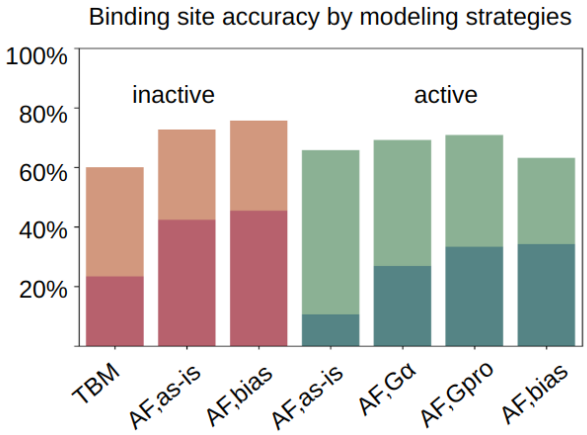


Figure S5. Pose prediction performance of Galaxy7TM. ligand RMSD of Top5 poses of Galaxy7TM is compared to GALD_rigid and GALD_flex (the best model)

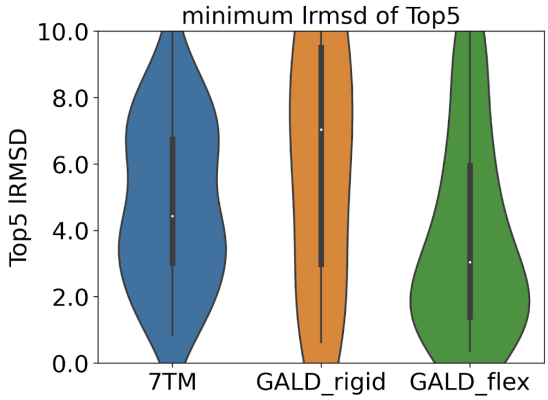
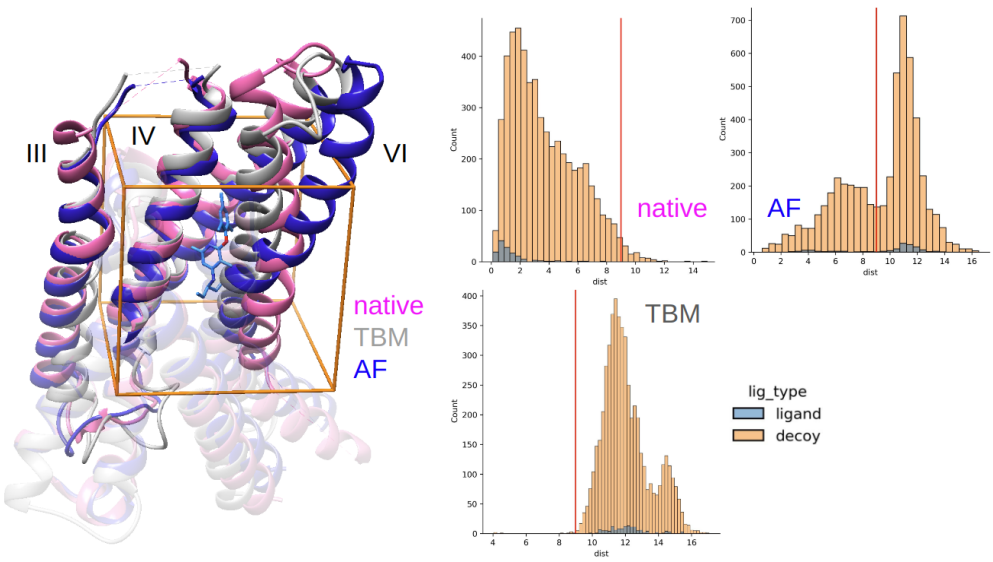


Figure S6. TBM, AlphaFold, and experimental structure of CRFR1 and distance-from-binding-site distribution of ligands and decoys. left) receptor models and binding site box of Vina (orange) and native ligand pose (light blue) . right) Histogram of distance between center of binding box (cetner of mass of native ligand atoms) and center of mass of ligand atoms. ligand (blue), decoy (orange).



Bibliography

1. Hughes JP, Rees S, Kalindjian SB, Philpott KL. Principles of early drug discovery. British Journal of Pharmacology. 2011. pp. 1239–1249. doi:10.1111/j.1476-5381.2010.01127.x

2. Redfern WS, Carlsson L, Davis AS, Lynch WG, MacKenzie I, Palethorpe S, et al. Relationships between preclinical cardiac electrophysiology, clinical QT interval prolongation and torsade de pointes for a broad range of drugs: evidence for a provisional safety margin in drug development. Cardiovasc Res. 2003;58: 32–45.

3. Wang W, MacKinnon R. Cryo-EM Structure of the Open Human Ether-à-go-go-Related K Channel hERG. Cell. 2017;169: 422–430.e10.

4. Rampe D, Roy M-L, Dennis A, Brown AM. A mechanism for the proarrhythmic effects of cisapride (Propulsid): high affinity blockade of the human cardiac potassium channel HERG. FEBS Letters. 1997. pp. 28–32. doi:10.1016/s0014-5793(97)01249-0

5. Annual Review of Pharmacology and Toxicology. doi:10.1146/pharmtox.711

6. Zhou Z, Vorperian VR, Gong Q, Zhang S, January CT. Block of HERG potassium channels by the antihistamine astemizole and its metabolites desmethylastemizole and norastemizole. J Cardiovasc Electrophysiol. 1999;10: 836–843.

7. Darpo B, Nebout T, Sager PT. Clinical Evaluation of QT/QTc Prolongation and Proarrhythmic Potential for Nonantiarrhythmic Drugs: The International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use E14 Guideline. The Journal of Clinical Pharmacology. 2006. pp. 498–507. doi:10.1177/0091270006286436

8. Chemi G, Gemma S, Campiani G, Brogi S, Butini S, Brindisi M. Computational Tool for Fast in silico Evaluation of hERG K Channel Affinity. Frontiers in Chemistry. 2017. doi:10.3389/fchem.2017.00007

9. Kratz JM, Schuster D, Edtbauer M, Saxena P, Mair CE, Kirchebner J, et al. Experimentally validated HERG pharmacophore models as cardiotoxicity prediction tools. J Chem Inf Model. 2014;54: 2887–2901.

10. Braga RC, Alves VM, Silva MFB, Muratov E, Fourches D, Lião LM, et al. Pred-hERG: A Novel web-Accessible Computational Tool for Predicting Cardiac Toxicity. Mol Inform. 2015;34: 698–701.

11. Lee H-M, Yu M-S, Kazmi SR, Oh SY, Rhee K-H, Bae M-A, et al. Computational determination of hERG-related cardiotoxicity of drug

candidates. BMC Bioinformatics. 2019;20: 250.

12. Ryu JY, Lee MY, Lee JH, Lee BH, Oh K-S. DeepHIT: a deep learning framework for prediction of hERG-induced cardiotoxicity. Bioinformatics. 2020;36: 3049–3055.

13. Karim A, Lee M, Balle T, Sattar A. CardioTox net: a robust predictor for hERG channel blockade based on deep learning meta-feature ensembles. Journal of Cheminformatics. 2021. doi:10.1186/s13321-021-00541-z

14. Kim H, Park M, Lee I, Nam H. BayeshERG: a robust, reliable and interpretable deep learning model for predicting hERG channel blockers. Brief Bioinform. 2022;23. doi:10.1093/bib/bbac211

15. Creanza TM, Delre P, Ancona N, Lentini G, Saviano M, Mangiatordi GF. Structure-Based Prediction of hERG-Related Cardiotoxicity: A Benchmark Study. J Chem Inf Model. 2021;61: 4758–4770.

16. Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. Journal of Chemical Information and Modeling. 1988. pp. 31–36. doi:10.1021/ci00057a005

17. Rogers D, Hahn M. Extended-Connectivity Fingerprints. Journal of Chemical Information and Modeling. 2010. pp. 742–754. doi:10.1021/ci100050t

18. Niculescu-Mizil A, Caruana R. Predicting good probabilities with supervised learning. Proceedings of the 22nd international conference on Machine learning - ICML ’05. 2005. doi:10.1145/1102351.1102430

19. Ryu S, Kwon Y, Kim WY. A Bayesian graph convolutional network for reliable prediction of molecular properties with uncertainty quantification. Chem Sci. 2019;10: 8438–8446.

20. Du F, Yu H, Zou B, Babcock J, Long S, Li M. hERGCentral: a large database to store, retrieve, and analyze compound-human Ether-à-go-go related gene channel interactions to facilitate cardiotoxicity assessment in drug development. Assay Drug Dev Technol. 2011;9: 580–588.

21. Siramshetty VB, Chen Q, Devarakonda P, Preissner R. The Catch-22 of Predicting hERG Blockade Using Publicly Accessible Bioactivity Data. J Chem Inf Model. 2018;58: 1224–1233.

22. Konda LSK, Keerthi Praba S, Kristam R. hERG liability classification models using machine learning techniques. Computational Toxicology. 2019. p. 100089. doi:10.1016/j.comtox.2019.100089

23. Critical Assessment of Artificial Intelligence Methods for Prediction of hERG Channel Inhibition in the Big Data Era.

doi:10.1021/acs.jcim.0c00884.s001

24. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An open chemical toolbox. J Cheminform. 2011;3: 33.

25. Yang J, Baek M, Seok C. GalaxyDock3: Protein-ligand docking that considers the full ligand conformational flexibility. J Comput Chem. 2019;40: 2739–2748.

26. Asai T, Adachi N, Moriya T, Oki H, Maru T, Kawasaki M, et al. Cryo-EM Structure of K-Bound hERG Channel Complexed with the Blocker Astemizole. Structure. 2021;29: 203–212.e4.

27. Trzaskowski B, Latek D, Yuan S, Ghoshdastider U, Debinski A, Filipek S. Action of molecular switches in GPCRs--theoretical and experimental studies. Curr Med Chem. 2012;19: 1090–1109.

28. Congreve M, de Graaf C, Swain NA, Tate CG. Impact of GPCR Structures on Drug Discovery. Cell. 2020;181: 81–91.

29. Heo L, Feig M. Multi-state modeling of G-protein coupled receptors at experimental accuracy. Proteins. 2022;90: 1873–1885.

30. Bender BJ, Marlow B, Meiler J. Improving homology modeling from low-sequence identity templates in Rosetta: A case study in GPCRs. PLoS Comput Biol. 2020;16: e1007597.

31. G Protein-Coupled Receptors - Part A. Elsevier; 2022.

32. Zhang J, Yang J, Jang R, Zhang Y. GPCR-I-TASSER: A Hybrid Approach to G Protein-Coupled Receptor Structure Modeling and the Application to the Human Genome. Structure. 2015. pp. 1538–1549. doi:10.1016/j.str.2015.06.007

33. Miszta P, Pasznik P, Jakowiecki J, Sztyler A, Latek D, Filipek S. GPCRM: a homology modeling web service with triple membrane-fitted quality assessment of GPCR models. Nucleic Acids Res. 2018;46: W387–W395.

34. He X-H, You C-Z, Jiang H-L, Jiang Y, Eric Xu H, Cheng X. AlphaFold2 versus experimental structures: evaluation on G protein-coupled receptors. Acta Pharmacologica Sinica. 2022. doi:10.1038/s41401-022-00938-y

35. Lee GR, Seok C. Galaxy7TM: flexible GPCR-ligand docking by structure refinement. Nucleic Acids Res. 2016;44: W502–6.

36. Mysinger MM, Carchia M, Irwin JJ, Shoichet BK. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. J Med Chem. 2012;55: 6582–6594.

37. Wu B, Chien EYT, Mol CD, Fenalti G, Liu W, Katritch V, et al. Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. Science. 2010;330: 1066–1071.

38. Pándy-Szekeres G, Munk C, Tsonkov TM, Mordalski S, Harpsøe K, Hauser AS, et al. GPCRdb in 2018: adding GPCR structure models and ligands. Nucleic Acids Res. 2018;46: D440–D446.

39. Sterling T, Irwin JJ. ZINC 15 – Ligand Discovery for Everyone. Journal of Chemical Information and Modeling. 2015. pp. 2324–2337. doi:10.1021/acs.jcim.5b00559

40. Bender BJ, Marlow B, Meiler J. RosettaGPCR: Multiple Template Homology Modeling of GPCRs with Rosetta. doi:10.1101/2019.12.13.875237

41. Steinegger M, Meier M, Mirdita M, Vöhringer H, Haunsberger SJ, Söding J. HH-suite3 for fast remote homology detection and deep protein annotation. BMC Bioinformatics. 2019;20: 473.

42. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web server: 2018 update. Nucleic Acids Research. 2018. pp. W200–W204. doi:10.1093/nar/gky448

43. Suzek BE, Wang Y, Huang H, McGarvey PB, Wu CH, UniProt Consortium. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. Bioinformatics. 2015;31: 926–932.

44. Steinegger M, Mirdita M, Söding J. Protein-level assembly increases protein sequence recovery from metagenomic samples manyfold. Nat Methods. 2019;16: 603–606.

45. Forster SC, Kumar N, Anonye BO, Almeida A, Viciani E, Stares MD, et al. A human gut bacterial genome and culture collection for improved metagenomic analyses. Nat Biotechnol. 2019;37: 186–192.

46. Eberhardt J, Santos-Martins D, Tillack A, Forli S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. doi:10.26434/chemrxiv.14774223

47. Heo L, Feig M. Multi-state modeling of G-protein coupled receptors at experimental accuracy. Proteins. 2022. doi:10.1002/prot.26382

48. Kwon S, Seok C. CSAlign and CSAlign-Dock: Structure alignment of ligands considering full flexibility and application to protein–ligand docking. Computational and Structural Biotechnology Journal. 2023. pp. 1–10. doi:10.1016/j.csbj.2022.11.047

49. Heo L, Shin W-H, Lee MS, Seok C. GalaxySite: ligand-binding-site prediction by using molecular docking. Nucleic Acids Res. 2014;42:

W210–4.

50. Park H, Zhou G, Baek M, Baker D, DiMaio F. Force Field Optimization Guided by Small Molecule Crystal Lattice Data Enables Consistent Sub-Angstrom Protein-Ligand Docking. J Chem Theory Comput. 2021;17: 2000–2010.

51. Lee GR, Seok C. Galaxy7TM: flexible GPCR-ligand docking by structure refinement. Nucleic Acids Res. 2016;44: W502–6.

52. Halgren TA. Force Fields: MMFF94. Encyclopedia of Computational Chemistry. 2002. doi:10.1002/0470845015.cma012m

53. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem. 2004;25: 1605–1612.

54. Gasteiger J, Marsili M. Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. Tetrahedron. 1980. pp. 3219–3228. doi:10.1016/0040-4020(80)80168-2

55. Wang C, Zhang Y. Improving scoring-docking-screening powers of protein-ligand scoring functions using random forest. J Comput Chem. 2017;38: 169–177.

국문초록

딥러닝의 발달과 데이터의 축적으로 보다 정확한 단백질 구조와 분자 특성을 예측할 수 있게 되었고, 딥러닝을 이용한 CADD 연구가 활발히 진행되고 있다. 이 논문에서는 딥러닝을 활용하여 두 가지 단백질에 대한 리간드-단백질 상호작용 예측 연구를 수행하였다. 첫번째로 hERG(human ether-a-go-go related gene) 이온 통로의 억제제 분류 모델을 학습시키고 독성 예측을 진행하였다. hERG는 심근 세포막에 발현되는 전압 개폐 칼륨 이온 채널로 재분극을 조절하는데 중요한 역할을 하며 심장 독성과 연관되어 의약품 개발시 hERG와의 상호작용을 고려해야할 필요성이 있다. 분자지문, 분자설명자, 분자그래프 등 컴퓨터가 이해할 수 있는 형태로 분자를 변환하여 나타내는 다양한 방법을 이용해 분류 모델을 만든 후 성능을 평가하였다. 그 결과 사전에 정의한 특성 기반 머신러닝 모델과 딥러닝 모델 간에 정확도와 예상 보정 오류(ECE)를 비교하였을 때 유의미한 차이가 없는 것을 관찰하였다. 두번째로 G-protein 결합 수용체(GPCR)에 대한 복합체 구조 예측과 가상스크리닝을 진행하였다. GPCR은 다양한 리간드와 결합하여 세포 내 G protein을 통해 세포 활동을 조절하는 신호를 전달하는 막 단백질로, 다양한 생리학적/병리학적 기전과 관련되어 있어 의약품 개발시 표적단백질로 꼽힌다. GPCR은 리간드 결합에 따라 활성과정에서 구조가 크게 바뀌는 것으로 알려져있는데, AlphaFold-multimer와 AlphaFold 기반의 다중 상태 모델링을 사용하여 보다 정확한 GPCR 구조를 얻고자 하였다. 이렇게 얻은 모델 구조에 다양한 도킹 도구를 이용하여 도킹과 가상 스크리닝을 수행하여 실제 문제에 적용가능한지 확인하고자 하였다. 수용체 모델은 cross-docking을 수행할 때와 유사한 정확도를 보였으며, 리간드 유연성 및 수용체 유연성도 고려한 도킹 방법이 가장 높은 성능을 나타내었다. hERG와 GPCR 두 사례 모두 막단백질 수용체 모델 구조가 비교적 정확한 반면 리간드-수용체 복합체 구조 모델링 및 상호작용 예측 방법의 한계를 보였다. 이러한 결과는 보다 정확한 구조 모델링 접근법과 데이터 부족을 보완하기 위한 불확실성 예측의 필요성을 보여준다.

주요어: 컴퓨터 기반 의약품 개발, 딥러닝, hERG, GPCR, 리간드-수용체 상호작용 예측, 가상 스크리닝

학번 : 2021-22785