

Reducing hERG Toxicity Using hERG Classification Model and Fragment-growing Network

Yan Yang^{a,1}, Yanmin Zhang^{a,1}, Xingye Chen^a, Yi Hua^a, Guomeng Xing^a, Chenglong Deng^a, Li Liang^a, Tao Lu^{a,b,*}, Yadong Chen^{a,*}, Haichun Liu^{a,*}

^a Laboratory of Molecular Design and Drug Discovery, School of Science, China Pharmaceutical University, 639 Longmian Avenue, Nanjing 211198, China

^b State Key Laboratory of Natural Medicines, China Pharmaceutical University, 24 Tongjiiaxiang, Nanjing 210009, China

¹ The authors contributed equally to this work and should be considered as co-first authors.

Abstract

Drug-induced cardiotoxicity has become one of the major reasons leading to drug withdrawal in past decades, which is closely related to the blockade of human Ether-a-go-go-related gene (hERG) potassium channel. Developing reliable hERG predicting model and optimizing model can greatly reduce the risk faced in drug discovery. In this study, we constructed eight hERG classification models, the best of which shows desirable generalization ability on low-similarity clinical compounds, as well as advantages in perceiving activity gap caused by small structural changes. Furthermore, we developed a hERG optimizer based on fragment grow strategy and explored its usage in four cases. After reinforcement learning, our model successfully suggests same or similar compounds as chemists' optimization. Results suggest that our model can provide reasonable optimizing direction to reduce hERG toxicity when hERG risk is corresponding to lipophilicity, basicity, the number of rotatable bonds and pi-pi interactions. Overall, we demonstrate our model as a promising tool for medicinal chemists in hERG optimization attempts.

Introduction

Cardiotoxicity has become one of the major causes leading to drug failure, accounting for 18.8% of withdrawal between 1990 and 2010¹. The blockade of a cardiac potassium channel encoded by human Ether-a-go-go-related gene (hERG) is responsible for the drug-induced cardio disorder, which is associated with a prolongation of QT-interval in the electrocardiogram (ECG)²⁻³. Assessment of potential hERG-related cardiotoxicity has become a critical step in the drug discovery pipeline⁴. Many established experimental methods have been developed to measure hERG binding affinity in the past such as patch clamp assays, fluorescence-based assays and radioligand binding assays⁵, providing reliable statistics to medicinal chemists. However, it is time-consuming and expensive to test every candidate in the early stage. Therefore, it is of great use to build *in silico* model for the prediction of hERG risk.

Apart from hERG predicting, another important question is how to optimize compounds when they are found to be hERG channel blockers. It will greatly save the labor if the computer can suggest some structures with low hERG risk but high similarity to the original structures. Actually, this task is what we called property-oriented molecule generation. Previous works have proposed various approaches to optimize a certain property of a compound. For example, Marcus Olivecrona and co-workers introduced a SMILES-based model to generate active compounds against the dopamine receptor type 2 (DRD2) using Policy Gradient, a reinforcement learning method⁶. Same technique was implemented into the network of Reinforcement Learning for Structural Evolution (ReLeaSE) by Mariya Popova and co-workers⁷. Wengong Jin and co-workers⁸ employed a junction tree based network (JT-VAE) to automatic design molecules with specific chemical properties. Its performance was demonstrated to be better than other VAE-based methods. Jiaxuan You and co-workers⁹ adopted a Graph Convolutional Policy Network (GCPN) to perform goal-directed molecule generation, which uses graph information and Proximal Policy Optimization (PPO) algorithm. It can obtain 61% improvement compared with JT-VAE in the task of optimizing penalized logP. Łukasz Maziarka and co-workers proposed a CycleGAN-

based model that can generate optimized compounds with desired property and high structural similarity to the original ones¹⁰. It was reported to significantly outperform previous results in terms of optimizing penalized LogP of drug-like molecules. Instead of using network based on SMILES or molecular graph, Zhenpeng Zhou and co-workers¹¹ chose to directly modify the molecule by adding or removing atoms and bonds, which is called Molecule Deep Q-Networks (MolDQN), another reinforcement learning based method. MolDQN can generate better result than JT-VAE and comparable result to GCPN. Despite a wide range of options, few methods can specify the optimizing region. Sometimes chemists may desire keeping core fragments while growing new fragments on specific growing sites. However, approaches such as ReLeaSE, JT-VAE and CycleGAN are only suitable for *de novo* drug design. Users cannot define which fragment to be kept in the output structures. As for GCPN and MolDQN, they can start from a given fragment. But as they automatically define growing sites, sometimes they might reduce attention on the optimization in specific area that chemists care about.

Herein, we introduce a fragment-growing network that enable users to keep particular substructures and optimize a user-defined region. Reinforcement learning is applied to increase the network's ability to generate low-hERG-risk molecules. In order to make a reliable reward function, we first construct eight hERG classification models using eight machine learning methods. The best model shows desirable performance on different datasets and different level of tasks, especially good generalization capability on clinical external set and best competence in recognizing the activity difference in optimization cases. Then the fragment-growing network is built, which can keep user-defined substructures in the generated structures. The basic framework is a traditional sequence-based generator (Recurrent Neural Network, RNN)¹². However, the kept parts, which are imprecisely but understandably called "core fragments" in our study, are encoded by Gated Graph Neural Networks (GGNN)¹³ to facilitate fragment grow. The combination model of GGNN and RNN shows superior result to naive RNN in terms of validity and synthetic accessibility for the generated structures. We also demonstrate the application of the model in suggesting potential core fragments to chemists. The

grow model, assisted by reinforcement learning, is then tested on four cases extracted from the literature³. Each case represents a smart strategy to lower hERG binding affinity concluded by experienced chemists. The built hERG classification model works as a part of reward function for reinforcement learning. By comparing the generated molecules with chemists’ final optimized compounds, we demonstrate the potential of our model in providing chemists reasonable direction for hERG optimization.

Materials and Methods

Constructing hERG Classifying Model

Data Collection and Process. Data from ChEMBL (<https://www.ebi.ac.uk/chembl/>, ChEMBL Target ID: ChEMBL240), PubChem (<https://pubchem.ncbi.nlm.nih.gov/>, UniProt: Q12809) and literatures¹⁴⁻¹⁶ was merged. Compounds without IC₅₀ or Ki label were removed from the merged collection. All units were standardized to μM . Structures were neutralized first. The chiral marker ‘@’ and cis-trans marker ‘\’ ‘/’ in the SMILES of those structures were removed as we only used 2D descriptors in this work. The SMILES were subsequently normalized using Pipeline Pilot¹⁷. Then the data was sorted into two classes according to their activity, with IC₅₀/Ki equal or less than 10 μM labeled as blockers and IC₅₀/Ki equal or greater than 30 μM labeled as non-blockers according to Hongmao Sun’s work¹⁸. Compounds with conflicting activity labels were abandoned to ensure the optimal data consistency, while compounds with the same labels were deduplicated. After that, structures that had entered clinical phases were separated as external validation dataset for evaluation. The remaining 9,708 data was randomly divided into training set and test set (4:1) to constructing predicting model. The information of each dataset can be seen in **Table 1**.

Table 1. Information of Datasets used for hERG Classification Model

Training Set		Test Set		External Validation Set	
Blockers	Non-blockers	Blockers	Non-blockers	Blockers	Non-blockers
4885	2881	1254	688	140	107

Descriptors Calculation and Selection. It has been reported that some physicochemical properties of hERG channel blockers including lipophilicity, acidity or basicity, the pi-pi interaction of aromatic rings and the number of rotatable bonds are related to high hERG risk¹⁹. Strategies directing at improving these four properties have been successfully applied to mitigate hERG binding affinity³, indicating their importance in hERG model building. Therefore, related properties should be calculated and selected to build hERG toxicity model. The process was conducted as follows: Compounds were firstly added hydrogen and optimized in MMFF94 force field using RDKit package (<http://www.rdkit.org/>, version: 2018.09.1.0). Then 115 2D properties and 1024 ECFP4²⁰ bits calculated by RDKit, together with two properties calculated by Mordred²¹ (nAcid, nBase) for the training set were incorporated. Among these features, MolLogP is corresponding to lipophilicity; NumAromaticRings (NAroR) is related to the possibility of forming pi-pi interaction; the number of rotatable bonds is represented by NumRotatableBonds (NROT). nAcid and nBase might be related to the acidity or basicity, so we also put them into the property pond and further property selection. The incorporated 1,141 descriptors were reduced to 169 by low variance filter (threshold=0.1) and high correlation filter (threshold=0.9), and they were subsequently selected by Non-dominated Sorting Genetic Algorithm-II (NSGA-II)², which distinguished 61 out of 169 features to reach the best five-fold cross-validated Kappa of training set (**Scheme S1, Figure S1 in Supporting Information**). All above-mentioned special descriptors including MolLogP, nAcid, nBases, NAroR and NROT were selected by NSGA-II, demonstrating the algorithm can pick out relatively important features for constructing hERG models. The selected descriptors can be found in **Table S1**.

Model Construction. Eight widely used methods were adopted to build hERG classification model, including Support Vector Machine (SVM), Random Forest (RF), Extremely Randomized Trees (ExtraTree), Adaptive Boosting (Adaboost), Gradient Boosted Decision Trees (GBDT), Exterme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM) and Deep Neural Networks (DNN). Scikit-

learn (<https://scikit-learn.org/>, version: 0.23.1), XGBoost (<https://github.com/dmlc/xgboost>, version: 1.1.1), LightGBM (<https://github.com/microsoft/LightGBM>, version:2.3.1) and pyTorch (<https://pytorch.org/>, version:1.3.1) package were used to build models. Grid search was applied to explore the best combination of parameters for highest ten-fold cross validated accuracy, while class weight for non-blockers was fixed as the ratio of blockers to non-blockers uniformly. The detailed description of other parameters can be seen in **Supporting Information**.

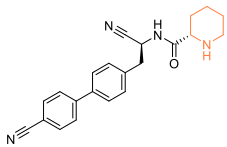
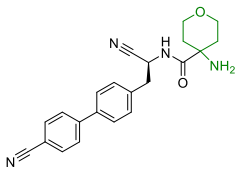
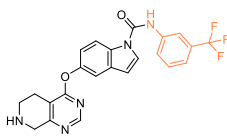
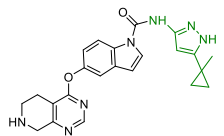
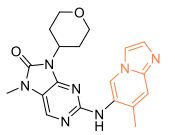
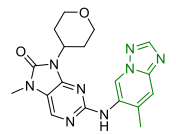
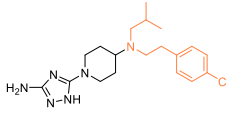
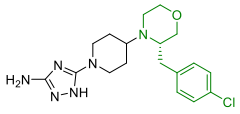
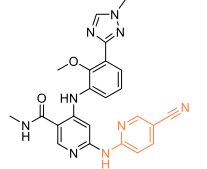
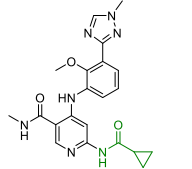
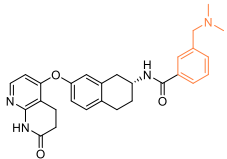
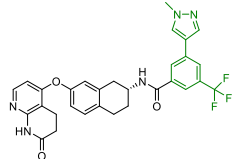
Model Evaluation. The obtained eight models were assessed using a variety of measures, which contained Area Under the ROC Curve (AUC), Accuracy (ACC), Balanced Accuracy (BAC), Matthews Correlation Coefficient (MCC), Cohen's Kappa (Kappa), F1_Score (F1), precision, recall (sensitivity), specificity (**Supporting Information**). The absolute values of all these metrics range from 0 to 1, with 1 representing the best and 0 representing the worst performance. 10-fold metrics of training set, test set and extra validation set were calculated to select top models used for further validation.

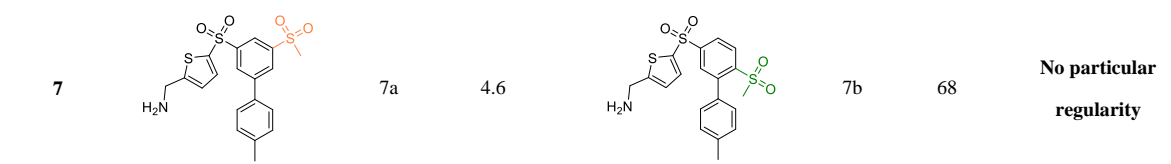
To investigate the selected models' practicability on correctly predicting new structures, we scaled the extra validation set according to their Tanimoto similarity /ECFP4 to training set. Six subsets were thus created, with similarity value <0.8, <0.7, <0.6, <0.5 and <0.4 respectively. It will be more convincing that if subsets with low similarity can also produce a good performance. We believe that structures with similarity below 0.4 are new enough compared with the training set. Balanced accuracy (BAC) for each subset was displayed for the chosen models. This step generated best four models with better generalization ability.

Validation on Optimization Cases. Generally speaking, it is more challenging for a model to recognize "activity cliffs", which means correctly classifying structures with huge gap of hERG binding affinity but subtle structural differences. Given that activity cliffs usually occur in chemists' compound optimization cases, herein, we tested our best models on seven cases provided by previous literature³. Each case is constituted

by two compounds with different IC₅₀ values, marked by **a** and **b** respectively. Compounds marked by **a** refer to the original structures with high hERG risk (<10μM), and compounds marked by **b** are optimized structures with much lower hERG binding affinity (>30μM). The original structures in case **1** will be named compound **1a**, while the optimized will be named compound **1b** and so on. The detailed information of these cases can be seen in **Table 2**. All the molecules in these cases do not appear in the training set. We compared the predicting accuracy of our four pre-selected models on these cases and finally picked the best one.

Table 2. Seven Optimization Cases Extracted from Literature

Case Index	hERG Channel Blockers	ID	hERG IC ₅₀ (μM)	hERG Improved	ID	hERG IC ₅₀ (μM)	Optimization Strategy
1		1a	1.4		1b	>33	Decreasing lipophilicity
2		2a	8.9		2b	>30	Decreasing lipophilicity
3		3a	8.1		3b	>198	Lowering basicity
4		4a	4.19		4b	39	Magnifying rigidity
5		5a	6.7		5b	>80	Altering pi-pi interactions
6		6a	6.3		6b	>100	No particular regularity



Fragment Grow Network

Model Overview. The main intention of our model is to optimize compound while keep substructures that chemists want to preserve because of activity demand or other reasons. Thus, we adopted a ‘fragment grow’ strategy that requires our model to learn how to generate target fragments on a given ‘core fragment’ instead of generating brand new molecules. In the training stage, we firstly broke a compound into two fragments. One fragment is called ‘core fragment’, as we assume that it is what chemists think is important and will not be changed. The other is taken as ‘source fragment’, which borrows the word ‘source’ from ‘source sequence’ in Natural Language Processing (NLP). As shown in **Figure 1**, The model mainly learns grammatical rules from the standardized SMILES of ‘source fragment’ using an encoder-decoder network. On the other hand, it also absorbs the graph information of ‘core fragment’ using GGNN to give the decoder an extra context. In other words, the model will ‘understand’ the scene, which means it can recognize which ‘core fragment’ it is supposed to grow on, which growing site it is supposed to start with and the surrounding information of the growing site. The two networks allow the model to reasonably grow target fragments based on the core fragments. Finally, target fragments are linked with core fragments on the growing site, generating the whole structures. Considering the existence of both GGNN and RNN network, we brief it as GGNN-RNN model.

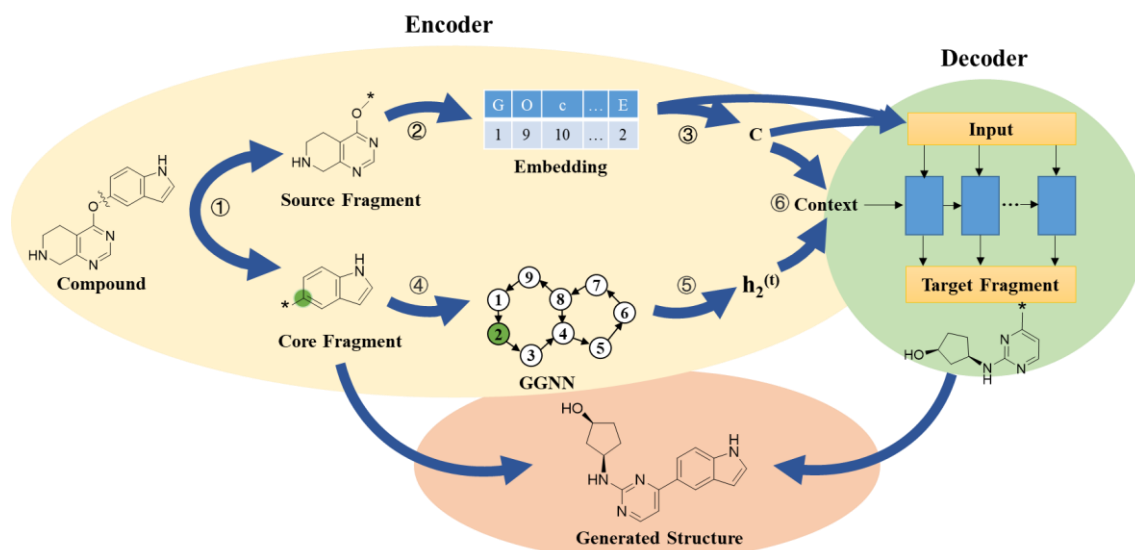


Figure 1. Brief Workflow of GGNN-RNN. The network adopts an encoder-decoder structure. The encoder involves six steps: ① The compound is broken into two pieces. ② The SMILES of source fragment is uniformed and processed using one hot method. ③ The processed information of the source fragment is encoded as a variable C . ④ The core fragment is represented as graph, where each node receives and send out message along edges that connecting nodes. ⑤ The message of the growing site (the node 2) at step t is extracted as $h_2^{(t)}$ ⑥ C and $h_2^{(t)}$ is combined as the total context of the decoder. The decoder produces target fragments using the information of the source target and the context. Finally, the core fragment is connected with the target fragment and generate the whole structure.

Data Preparation. Active kinase inhibitors with IC_{50}/K_i less than $10\mu M$ from ChEMBL database were collected, which belong to a total of 612 kinases target (**supporting information**). Inorganic structures and duplicates were removed. Structures containing phosphorus were also removed, as it is less commonly used in drugs. However, users can train their own model if needed. Molecular weight was controlled between 100 Da and 800 Da, after which 103, 292 cleaned kinase inhibitors were obtained. We broke these inhibitors by cutting on one single bond a time (bonds outside the ring but connecting with the ring), which generated many fragment pairs for each molecule. The larger fragments were used as ‘core fragment’, and the smaller were taken as ‘source

fragment', whose molecular weight was limited to no more than 350 Da. Both fragments carried a marker '*' to indicate the breaking site/growing site. To reduce the computation cost, only three pairs for each molecule were randomly selected and trained as our model input.

Source fragments were standardized using RDKit, which always puts the growing site '*' at the beginning of the SMILES. For example, SMILES 'O=C([*])[C@H]1NCCCC1' will be unified to '*C(=O)[C@@H]1CCCCN1' in RDKit. In this way, we can easily replace '*' with a 'G' to tell the generator a sequence of fragment is to begin and add an 'E' at the end of the standardized SMILES to inform the sequence's ending. The unified and modified sequence was then encoded using embedding method. The dictionary contains 39 words, covering the common symbols in the SMILES of organic molecules. Four numbers were used to represent special symbols, including 0 for 'pad', 1 for 'grow' (G), 2 for 'end' (E) and 39 for 'unused symbols'. Symbols containing two letters like 'Br' and 'Cl' were counted as single words. The overall encoded length for each fragment was set to 102. Fragments with longer length were abandoned. The dictionary can be seen as follows:

{0: 'pad', 1: 'G', 2: 'E', 3: '2', 4: 'F', 5: 'Cl', 6: 'N', 7: '[', 8: '6', 9: 'O',
10: 'c', 11: ']', 12: '#', 13: '=', 14: '3', 15: ')', 16: '4', 17: '-', 18: 'n', 19: 'o',
20: '5', 21: 'H', 22: '(', 23: 'C', 24: '1', 25: 'S', 26: 's', 27: 'Br', 28: '@', 29: '+',
30: '/', 31: '\\', 32: 'I', 33: 'P', 34: '7', 35: '8', 36: '9', 37: 'X', 38: 'Y', 39: 'unused'}

The graphs of core fragments utilized edges and nodes to indicate atoms and bonds separately. Each node was annotated by a total of 13 bits, standing for the presence of 'C', 'N', 'O', 'S', 'c', 'n', 'o', 's', 'H', 'F', 'I', 'Cl' and 'Br' atom respectively (as structures containing phosphorus were removed in our training data, we did not encode it in this model). The maximum number of nodes for each graph was confined to 60. There were four edge types, accounting for single, double, triple and aromatic bond respectively. Another important property for an edge is the direction, determined by the index of its begin node and end node. The edge type together with edge direction can help create the adjacency matrix which describes the communication between nodes and facilitates

the process of message traversing in the later stage. Examples and detailed illustration can be found in the original paper of GGNN¹³, which thoroughly explained how to represent the edge direction.

Grow Model Pretraining. Model was pretrained before the inference of reinforcement learning, which is based on an encoder-decoder network, with two input (core fragment and source fragment). The encoder composes of two parts: layers for encoding the sequence information of source fragments and layers for encoding the graph information of core fragments. The former is made up by an embedding layer with an embedding size of 39 and a GRU layer with a hidden size of 128, which generate the embedding data and a context. The later is an implementation of the GGNN from YujiaLi and RichardZemel¹³. The structures of the decoder referred to that of MOSES on GitHub (<https://github.com/molecularsets/moses>), which is formed by three stacked GRU layers with the hidden size of 512 for each layer. Dropout rate was set to 0.2. The embedding data and context **C** was combined as the input of the decoder, while the overall context was used to initialize the hidden state of layers. Adam optimizer with different learning rate (0.0005, 0.001, 0.005, 0.01) was employed to train the network. Different batch size was explored (64, 128, 256).

Comparison with Medicinal Chemists' Optimization in Previous Cases. Chemists' optimization strategies includes decreasing lipophilicity, lowering basicity, magnifying rigidity and altering pi-pi interactions (**Table 2**). To assess whether the optimization direction by our model was in accordance with the initial intention of the chemists, we explored four representative cases (Case 1, 3~5 in **Table 2**) with our GGNN-RNN model, which covers all above-mentioned strategies. The model was further trained by reinforcement learning based on the fragment pair of original hERG-risky compound in each case, which implements previous-defined hERG-predicting model as reward. Then, molecules generated by the trained model were compared with the optimized compounds by chemists from the aspect of MolLogP, pKa, NROT and NAroR. In addition, we performed molecular docking²²⁻²³ using Schrodinger 2009 package²⁴ to

examine the docking conformation and docking score of optimized molecules in comparison with chemists' optimization.

Before the trained generator starts to grow, the user has to decide the fragment pair first. Considering chemists usually do not desire many variations in their designed structures, we aimed to select the best fragment pair that can be used to generate more non-blockers while keep the part of original structure (core fragment) as big as possible. Here we adopted the idea of NSGA-II to facilitate the selection. Compounds were broken into fragment pairs as above mentioned. The generator grew 1,000 structures for each fragment pair, which were then put into our hERG toxicity model. The fraction of predicted non-blockers for generated structures was calculated and defined as Object One for the non-dominated rank. The Object Two was defined as difference in molecular weight between core fragment and original compound. The top ranked fragment pairs were kept and selected to grow molecules.

We employed reinforcement learning to escalate the fraction of non-blockers in generated molecules. However, in the beginning of reinforcement learning, expert training on source fragments was first conducted to raise the similarity of target fragments to source fragments. Then, the Policy Gradient⁷ was applied to optimize the parameter of the network. The reward function for each generated molecule is defined as the non-blocker probability multiplied by the MACCS²⁵/Tanimoto similarity of the target fragment to the source fragment (we brief it as 'similarity'). However, it is influenced by many other factors including physiochemical properties such as the number of hydrogen acceptors and donors (HBA and HBD), the number of rotatable bonds (NROT) and the fraction of nitrogen and oxygen atoms in heavy atoms (FNO), the existence of pan-assay interference compounds alerts, toxic alerts and other self-defined undesired substructures (Alerts). The SMILES arbitrary target specification (SMARTS)²⁶ of Alerts can be seen in **supporting information**. Parameters for target fragments were set as follows: HBA and HBD should no more than 3; NROT should no more than 5; parameters for whole generated structures were set as: HBD should be no more than 5 according to rule of five; NROT should be no more than 8; FNO should be no more than 0.4 according to our previous work²⁷; structures should not contain any

alerts. Any violations of these rules will force the reward to be 0 regardless of hERG prediction results. Another manipulation is that if the similarity is higher than or equal to a certain threshold, then similarity value will be changed to 1 and otherwise it will be 0. This operation has two meanings: first, totally dissimilar fragments will be abandoned; second, once the similarity reach a certain threshold, the reward will only depend on the non-blocker probability, which guarantees the model to explore some new fragments while keep a certain degree of similarity. We explored different thresholds 0.1, 0.2, 0.3, 0.4, 0.5 and adopted best results for each case. The final thresholds are: 0.3 for Case 1, 0.5 for Case 3, 0.3 for Case 4 and 0.1 for Case 5. Learning rate for reinforcement learning was set as 0.0001. Optimization batch was set to 32, and the gamma factor was set to 0.97.

Result and discussion

Constructing hERG Classification Model

Classification Model Result. The result of eight hERG classification models can be seen in **Table 3**. In terms of the training set, ERF shows the highest ten-fold cross validated ACC (0.883), Kappa (0.745), MCC (0.747), second highest AUC (0.945) and largest BAC (0.867), indicating the best overall performance on training set. SVM takes the second place with a second largest ACC of 0.875, second largest Kappa and MCC of 0.732, third largest AUC of 0.937 and same BAC with ERF. The metrics for LightGBM are very close to SVM, which ranks the third place among all the methods. RF, XGB and DNN are equally competent in terms of ACC (around 0.87), Kappa (around 0.72) and MCC (around 0.72). Nevertheless, their AUC and BAC differ more, the AUC of which are 0.94, 0.935 and 0.927 respectively and the BAC of which is 0.851, 0.857 and 0.863 respectively. GBDT and AdaBoost exhibit weaker predicting performance on training set with Kappa less than 0.7 and 0.6 respectively.

Table 3. Result of Eight Classification Methods on Different Datasets

Performance on Training Set(cv=10)									
Method	AUC	ACC	Kappa	MCC	BAC	F1	Precision	Recall (Sensitivity)	Specificity
LightGBM	0.937	0.873	0.725	0.726	0.860	0.900	0.890	0.910	0.809
SVM	0.935	0.875	0.732	0.732	0.867	0.900	0.902	0.898	0.835
RF	0.940	0.870	0.716	0.719	0.851	0.900	0.874	0.928	0.773
ERF	0.945	0.883	0.745	0.747	0.867	0.909	0.889	0.931	0.802
XGB	0.935	0.869	0.719	0.719	0.857	0.897	0.890	0.904	0.811
GBDT	0.926	0.861	0.699	0.700	0.845	0.892	0.877	0.907	0.783
AdaBoost	0.781	0.798	0.565	0.565	0.781	0.840	0.835	0.846	0.716
DNN	0.927	0.868	0.720	0.721	0.863	0.894	0.904	0.885	0.841
Performance on Test Set									
Method	AUC	ACC	Kappa	MCC	BAC	F1	Precision	Recall (Sensitivity)	Specificity
LightGBM	0.942	0.881	0.735	0.736	0.861	0.909	0.892	0.927	0.795
SVM	0.936	0.869	0.713	0.713	0.857	0.898	0.899	0.897	0.817
RF	0.941	0.871	0.708	0.713	0.842	0.904	0.870	0.941	0.743
ERF	0.951	0.882	0.736	0.739	0.859	0.911	0.886	0.938	0.781
XGB	0.935	0.872	0.717	0.717	0.855	0.902	0.892	0.912	0.798
GBDT	0.925	0.859	0.687	0.688	0.838	0.893	0.877	0.909	0.767
AdaBoost	0.797	0.815	0.595	0.595	0.797	0.857	0.854	0.860	0.733
DNN	0.929	0.863	0.704	0.705	0.857	0.892	0.908	0.876	0.839
Performance on External Set									
Method	AUC	ACC	Kappa	MCC	BAC	F1	Precision	Recall (Sensitivity)	Specificity
LightGBM	0.889	0.838	0.675	0.678	0.842	0.851	0.891	0.814	0.869
SVM	0.861	0.761	0.526	0.535	0.769	0.770	0.846	0.707	0.832
RF	0.894	0.838	0.671	0.671	0.836	0.856	0.862	0.850	0.822
ERF	0.880	0.802	0.601	0.603	0.804	0.818	0.853	0.786	0.822
XGB	0.876	0.818	0.634	0.638	0.822	0.831	0.874	0.793	0.850
GBDT	0.874	0.794	0.585	0.589	0.797	0.809	0.850	0.771	0.822
AdaBoost	0.766	0.765	0.527	0.529	0.766	0.785	0.815	0.757	0.776
DNN	0.848	0.749	0.503	0.514	0.759	0.756	0.842	0.686	0.832

As for the test set, ERF, SVM, LightGBM, RF, XGB and DNN continue to show different degree of superiority over GBDT and AdaBoost in terms of the comprehensive indexes AUC, ACC, Kappa, MCC and BAC. Among the six superior methods, ERF demonstrates its best overall predicting competence on test set, with an AUC of 0.951,

ACC of 0.882, Kappa of 0.736, MCC of 0.739 and BAC of 0.859. LightGBM surpasses SVM with greater AUC (0.942), ACC (0.881), Kappa (0.735), MCC (0.736) and BAC (0.861), which is as robust as ERF. The other four methods (SVM, RF, XGB and DNN) are very close to each other, but there is a gap between them and ERF/LightGBM.

Different from the result of training set and test set where the gap between eight methods is relatively small, their predicting ability on clinical external set varies from method to method. As shown in **Table 3**, only four methods can reach an ACC of more than 0.8, which are LightGBM, RF, ERF and XGB. Among the four methods LightGBM and RF show the best predicting competence on external set with highest AUC (0.889 and 0.894 respectively), best ACC (both 0.838), highest Kappa (0.675 and 0.671 respectively), highest MCC (0.678 and 0.671 respectively) and largest BAC (0.842 and 0.836 respectively). SVM and DNN witnesses a significant drop on external set, even weaker than GBDT and AdaBoost. According to **Table 3**, we can conclude that RF, ERF, LightGBM and XGB exhibit favorable performance on all the three datasets.

Table 4. BAC on Different External Subsets Scaled by Similarity

Similarity ^a	Count	RF	ERF	LightGBM	XGB
<0.8	217	0.828	0.801	0.833	0.815
<0.7	163	0.817	0.787	0.816	0.807
<0.6	111	0.784	0.757	0.775	0.766
<0.5	90	0.767	0.746	0.746	0.734
<0.4	59	0.779	0.729	0.729	0.711

^a Compounds' similarity to the training set.

Table 4 illustrates the four models' BAC on different external subsets scaled by similarity to the training set. LightGBM shows highest BAC of 0.833 when the similarity of external set is less than 0.8, revealing its good generalization ability. RF is better than the other three when similarity threshold is set to 0.4~0.7. Especially when similarity is less than 0.4, the BAC of RF remains 0.779, compared with 0.729 of ERF, LightGBM and 0.711 of XGB. Overall, the four methods do not differ much in this experiment and all of them can reach a BAC of more than 0.7 even when the compounds

are very different from training sets.

Validation on Optimization Cases. RF, ERF, LightGBM and XGB is further evaluated on seven optimization cases to investigate their sensitivity to subtle structural difference, which can be seen in **Table 5**. The before-optimization compounds (hERG channel blockers) are labeled as 1, and after-optimization compounds (hERG channel non-blockers) are labeled as 0. Only when the model correctly predicts both the before- and after- structures can we name it a successful case. LightGBM successfully classifies five cases in total, dominating the first place, which is followed by 4 cases of XGB, 3 cases of ERF and only 2 cases of RF. The results demonstrate LightGBM may be the most suitable model to recognize the hERG binding affinity variations when chemists conduct small structural changes on lead compounds.

Table 5. Predicting Result of Literature Compounds

Case	Compound ID	True Label	RF Pred Label	ERF Pred Label	LightGBM Pred Label	XGB Pred Label
1	1a	1	1	1	1	1
	1b	0	0	0	0	0
2	2a	1	1	1	1	1
	2b	0	1	1	0	0
3	3a	1	0	0	0	0
	3b	0	0	0	0	0
4	4a	1	1	1	1	1
	4b	0	1	1	0	1
5	5a	1	0	0	1	1
	5b	0	0	0	0	0
6	6a	1	1	1	1	1
	6b	0	0	0	0	0
7	7a	1	1	1	1	1
	7b	0	1	0	1	1
Correct Cases^a			2	3	5	4

^a The number of correctly predicted cases. The before-optimization compounds are labeled as 1 (blockers), and after-optimization compounds are labeled as 0 (non-blockers). Only when the model correctly predicts both the before- and after- structures can we name it a successful case.

Compared with Other Works. Our work is compared with the previous studies in **Table 6**. From the table we can see, there is no unified standard on setting the thresholds for computational chemists. Thresholds include both single threshold such as 1 μM , 10 μM , 50% efficacy, 50% inhibition at 10 μM , as well as double threshold such as ' $\leq 10 \mu\text{M}$, $> 10 \mu\text{M}$ ', ' $\leq 10 \mu\text{M} \geq 80 \mu\text{M}$ ', ' $< 10 \mu\text{M} > 30 \mu\text{M}$ '. Different thresholds may impose different influence on model performance. If the IC_{50} gap between the defined 'blocker' and 'non-blocker' is large, the model performance will be improved according to Chuipu Cai et al.²⁸. However, this improvement might suffer from losing some data. So we should take both model performance and dataset size into account. According to the table, from the perspective of model performance our LightGBM model generates a good ACC (0.881), high AUC (0.942), good BAC (0.861), good Kappa (0.735) and MCC (0.736) on test set compared with other models. Only three works shows obvious advantages over us with high ACC (> 0.9) and at least one other better metric ($\text{MCC} > 0.8$ or $\text{AUC} > 0.96$), which are studies from Chuipu Cai et al.²⁸, Li-li LIU et al.²⁹ and Yiwei Wang et al.³⁰. However, Chuipu Cai and co-workers chose a double threshold with large gap ' $\leq 10 \mu\text{M} \geq 80 \mu\text{M}$ ', with a dataset of only 4,447 data points, in comparison with our 9,708 data points. Similarly, although Li-li LIU et al.²⁹ and Yiwei Wang et al.³⁰ employed almost the same thresholds as us and obtained a quite good result, their data size is only 2,644, which is only about one third of ours.

Table 6. Comparison of Our hERG Classification Model with Previous work.

Year	Author	Method	Descriptors	Threshold	Data Size	Performance
2015	Rodolpho C. Braga et al. ³¹	SVM	Fingerprints and CDK descriptors	1 μM , 10 μM	5,984	$\text{ACC}_{\text{train}}(\text{CV}=5)=0.88$ $\text{AUC}_{\text{train}}(\text{CV}=5)=0.88$ $\text{MCC}_{\text{train}}(\text{CV}=5)=0.56$
2016	Remigijus Didziapetris et al. ³²	GBM	Physicochemical and topological parameters	$\leq 10 \mu\text{M} > 10 \mu\text{M}$	6,124	$\text{ACC}_{\text{test}}=0.72-0.75$ $\text{AUC}_{\text{test}}=0.80-0.83$ $\text{MCC}_{\text{test}}=0.42-0.50$
2017	Hongmao Sun et al. ³³	SVM	Atom-type-based molecular descriptor	$> 50\%$ efficacy	3,024	$\text{AUC}_{\text{train}}=0.93$
2018	Vishal B. Siramshetty et al. ³⁴	NN, SVM, RF	Fingerprints	$\leq 1 \mu\text{M} \geq 10 \mu\text{M}$	3,223	$\text{BAC}_{\text{train}}(\text{CV}=10)=0.87$ $\text{AUC}_{\text{train}}(\text{CV}=10)=0.94$

2019	Keiji Ogura et al. ²	SVM	Fingerprint, MOE, Pipeline Pilot descriptors	10 μ M, 50% inhibition at 10 μ M	291,202	BAC _{test} =0.839 Kappa _{test} =0.749
2019	Hyang-Mi Lee et al. ³⁵	Neural Network	Fingerprints and physicochemical descriptors	<10 μ M \geq 10 μ M	2,130	ACC _{train} (CV=10)=0.910 AUC _{train} (CV=10)=0.764 MCC _{train} (CV=10)=0.368
2019	ChuiPu Cai et al. ²⁸	DNN,GCNN,NB, SVM,RF	Mol2vec,MOE(2D)	\leq 10 μ M \geq 80 μ M	4,447	ACC _{test} =0.925 AUC _{test} =0.967
2020	Hyunho Kim et al. ³⁶	Self-attention network	Fingerprints	\leq 10 μ M, >10 μ M	11,860	AUC _{test} =0.89
2020	Jae Yong Ryu et al. ³⁷	DNN,GCN	Moredred Descriptors,Graph	<10 μ M \geq 10 μ M	14,440	ACC _{test} =0.812 MCC _{test} =0.641
2010	Munikumar R. Doddareddy et al. ¹⁶	SVM,LDA	Fingerprints	<10 μ M >30 μ M	2,644	ACC _{train} (CV=5)=0.93 MCC _{train} (CV=5)=0.70
2014	Li-li LIU et al. ²⁹	NBC	Fingerprints and 4 molecular properties	<10 μ M >30 μ M	2,644	ACC _{test} =0.91
2020	Yiwei Wang et al. ³⁰	Capsule Networks	Fingerprints, 13 molecular properties	<10 μ M >30 μ M	2,644	ACC _{test} =0.918 AUC _{test} =0.940 MCC _{test} =0.835 ACC _{test} =0.881 AUC _{test} =0.942
-	Our Model	SVM,RF,Adaboost,DNN	Fingerprints,descriptors	\leq10μM \geq30 μM	9,708	BAC_{test}=0.861 MCC_{test}=0.736 Kappa_{test}=0.735

Fragment Grow Network

Grow Model Construction. Different GGNN-RNN models parameterized by different batch size and learning rate were constructed. 1000 molecules were generated on core fragment '[*]C1=CC=C2C(C=CN2)=C1' and source fragment '*Oc1ncnc2c1CCNC2' for each model. The fraction of valid structures is shown in **Figure 2**. As it suggests, model performance is very sensitive to the change of learning rate, varying largely from 0.001 to 0.96. The best value is 0.001, at which all models can reach equal to or more than 0.9 validity, regardless of the batch size. In comparison, variation in batch size seems to matter less, but it still generates a difference of 0.33 at learning rate 0.0005. Batch size 128 and learning rate 0.001 were chosen to build our model, the validity of which is 0.96.

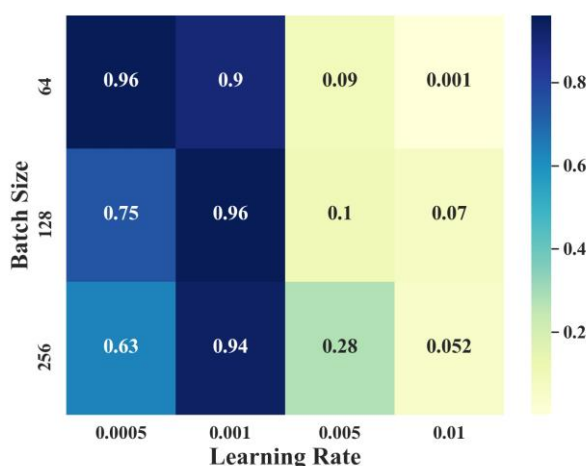


Figure 2. Fraction of Valid Structures Generated by Different GGNN-RNN Models

In order to verify the addition of GGNN can facilitate the molecule-growing process, we compare our GGNN-RNN model with the raw RNN model, which replaces the GGNN encoder with a single GRU layer to encode the context information of the core fragment. 1000 molecules were generated based on core fragment [*]C1=CC=C2C(C=CN2)=C1 and source fragment *Oc1ncnc2c1CCNC2. The proportion of valid SMILES produced by raw GNN-RNN is much higher than the raw RNN model, which is only 0.792. Moreover, it has lower average Synthetic Accessibility Score (SA Score)³⁸ (a score between 1 (easy to make) and 10 (very difficult to make)), which is 2.316, suggesting the structures it generates may be more easier to synthesize compared with raw RNN model (2.745). The difference of SA Score is significant at 95% confidence level. The reason for these results may be that the node level GGNN can extract the information of a particular growing site on core fragments, instead of broad information that expresses the whole fragments. Therefore, the network can gather more surrounding information of the starting point before growing, which may make the link between core fragments and target fragments to be more authentic (or more valid). Thus the fraction of valid structures and the synthetic accessibility will increase.

Choosing Suitable Fragment Pair. Before reinforcement learning starts, it is necessary to decide which fragment pair to feed into the model, as cutting on different bonds will generate different fragment pairs. In order to search for the most suitable fragment with less hERG risk as well as reserve the original structure as much as possible, fragments are put into the pretrained GGNN-RNN model and the fraction of non-blockers in the output is recorded as a reference for selection. The idea of NSGA-II was employed to find the best solution. As we can see from **Figure 3**, this method successfully offers suitable core fragments which are in agreement with chemists' final decisions in the explored four cases (source fragments are the remaining parts of the structures and thus are not displayed). In Case 1, four core fragments are suggested by NSGA-II sorting. Model that grows on core fragment **1-1** can generate 94% non-blockers in valid SMILES, but 178.66 Da of the molecule has to be cut down. By contrast, fragment **1-3** only abandons 84.08 Da molecular weight but provides 65.2% chance to obtain a non-blocker even when reinforcement learning is not performed, which is also promising to be a core fragment. Chemists in this case actually chose the fragment **1-3** as core fragment for hERG optimization, demonstrating the practicability of our technique. Fragment **3-2** in Case 3 and **5-2** in Case 5 can give very high fraction of non-blockers, which is 92.4% and 85.7 % respectively. Both of them were taken as core fragments by chemists. In Case 4, no fragments are exactly the same as the choice of chemists. However, all of them reserve the triazole and piperidine parts, which may give chemists some ideas to choose the core fragment. It should be noted that circumstances vary from case to case, as drugs usually directing at a special target through special functional groups. The method cannot tell which fragment is the best one using unified standards but offer chemists with several options. Although it needs some human involvement in this step, it greatly reduces the human labor by limiting the scope of selection.

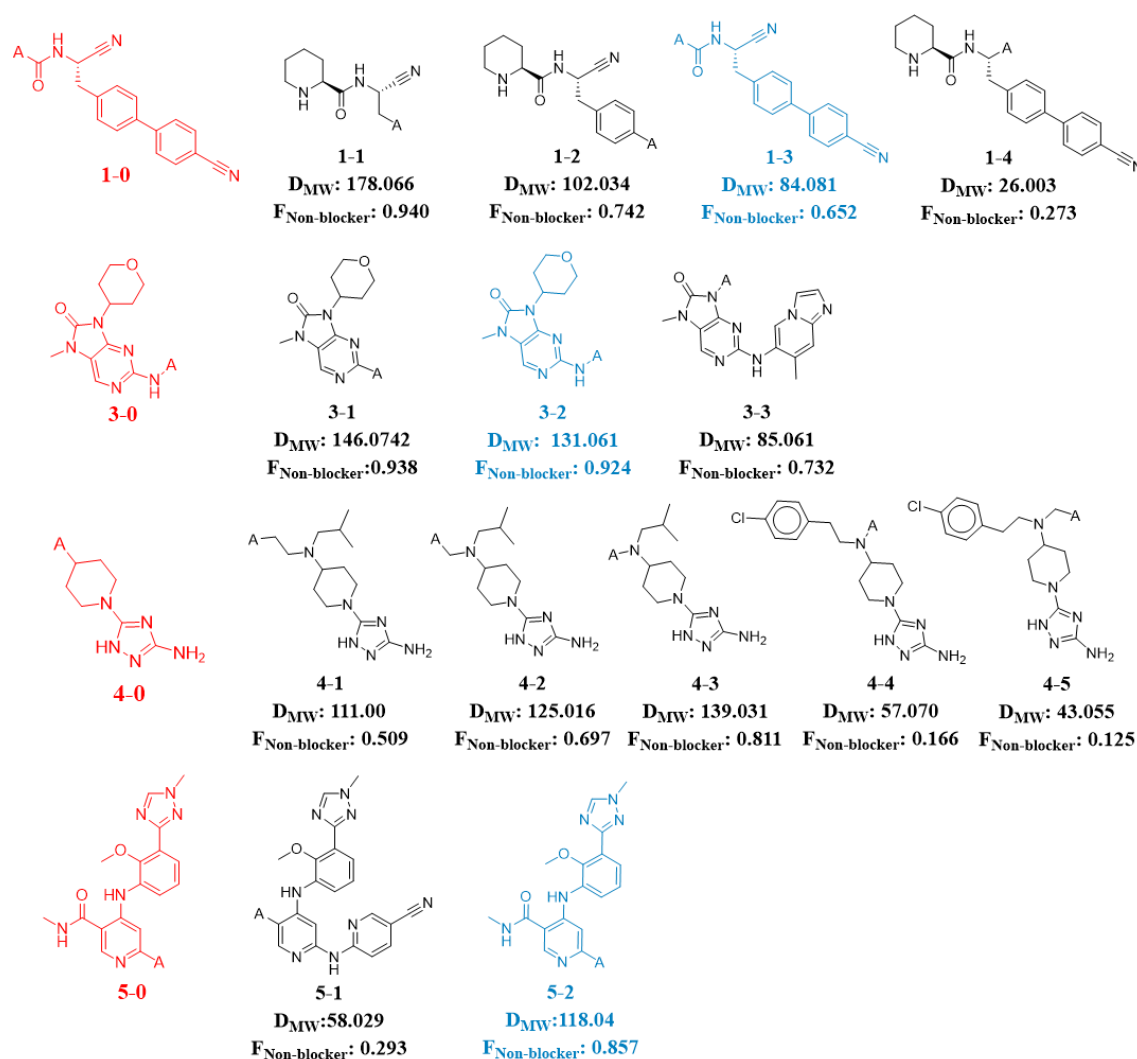


Figure 3. Comparison of Core Fragments Chosen by Chemists with Our Models' Suggestion. Core fragments selected by chemists are marked by red. Core fragments suggested by NSGA-II are marked by black. Blue-colored core fragments are recommended core fragments which are the same as chemists' choice. Core fragments are displayed with their difference in molecular weight (MW) to the original compound (D_{MW}) and fraction of non-blockers (F_{Non-blocker}) they generated after being fed into the pretrained GGNN-RNN model.

Comparison with Medicinal Chemists' Optimization. Assuming that chemists have chosen red parts as core fragments and the corresponding rest as source fragments in **Figure 3**, the next step is to feed those fragments into the model. We performed about

80 epochs for each case to improve the hERG property. 1000 molecules were generated after the model was trained. Invalid molecules, duplicates and predicted hERG channel blockers were removed. Here we will explain the result case by case.

Decreasing lipophilicity. Case 1 is a representative of the strategy of decreasing lipophilicity to reduce hERG risk³⁹. 80 epochs of reinforcement learning were conducted on core fragment **1-3** and corresponding source fragment. 65 predicted non-blockers (1-3-1~1-3-65) were obtained after removing duplicates. Their SMILES, hERG-channel-blocker probability and docking score to the original target Cathepsin C³⁹ can be seen in **Table S2**. According to the result of One-Sample T-Test using SPSS, we found the mean cLogP (2.151) of the 65 non-duplicated generated molecules is significantly lower than the original structure **1a** (2.706) at 95% confidence level, demonstrating our optimization direction is consistent with chemists' experience. After checking the structures one by one, we find that structure **1-3-27** is most similar to chemists' optimization, both of which have a free primary amino and a heterocyclic to lower the lipophilicity of the whole molecule (**Figure 4A**). Furthermore, **1-3-27** can be well accommodated in the binding site of its original target Cathepsin C protein and keep the key interactions as **1b** according to the covalent docking result (**Figure 4B**), which suggests our method can perform reasonable hERG optimization by lowering cLogP while reserving the binding mode.

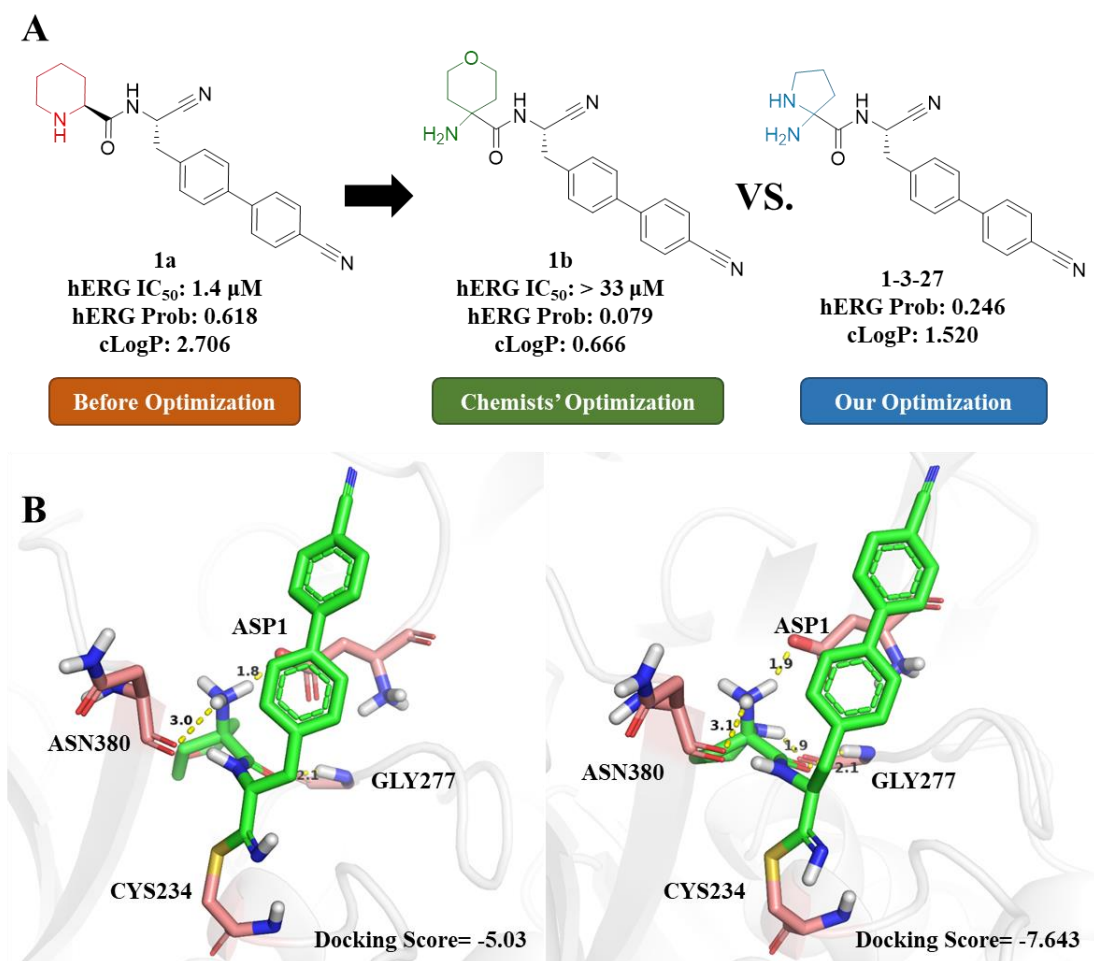


Figure 4. Conformation Comparison of Chemists' and Our Model's Optimization in Case 1. **A**, structures, cLogP and predicted hERG channel blocking probability (hERG Prob; higher score means higher blocking possibility) of original compound and optimized results. **B**, covalent docking conformations of compound **1b** (left) and structure **1-3-27** (right) to Cathepsin C (PDB code: 4CDD). For ligands, carbon atoms are colored in green, nitrogen atoms in blue, oxygen atoms in red, sulfur atoms in yellow. Yellow dash line stands for hydrogen bond.

Lowering basicity. Lowering basicity is the optimization strategy employed by chemists in Case 3⁴⁰. Thus, we utilize pKa to analyze the result. After 80 runs of training on core fragment **3-2** and corresponding source fragment as well as molecule clean, 80 potential non-blockers were attained. Their SMILES, pKa, hERG-channel blocker probability and docking score to the original target DNA-PK are recorded in **Table S3**.

The average pKa (14.097) of the generated molecules is significantly lower than the original structure **3a** (15.178) at 95% confidence level, according to the One-Sample T-Test result, revealing that same optimization direction can be found in our optimization and chemists' optimizations. More importantly, the model successfully reproduces the optimized structure **3b** given by chemists (**Figure 5, 3-2-76**), which demonstrates that our model is an useful to tool for medicinal chemists to reduce hERG risk.

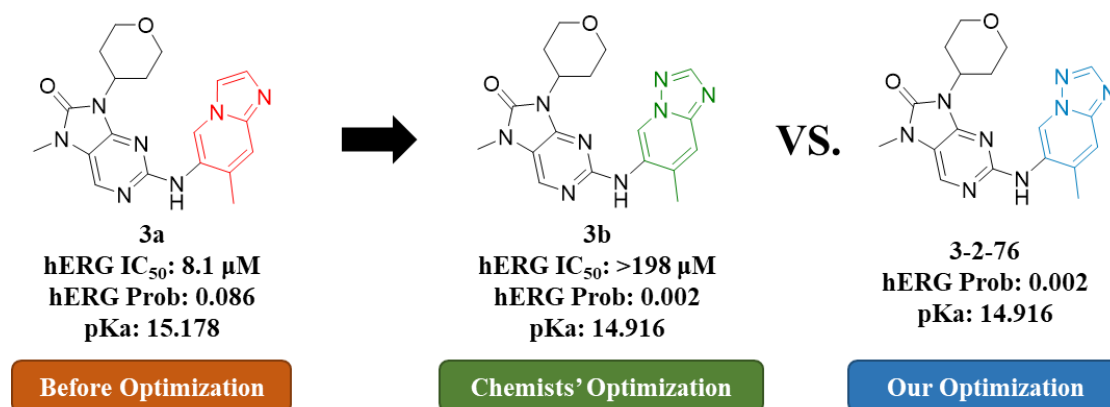


Figure 5. Structures, pKa and predicted hERG channel blocking probability (hERG Prob; higher score means higher blocking possibility) of original compound and optimized results.

Magnifying rigidity. Flexible linkers allow compounds to adopt better conformations in hERG channel^{3, 41}. Therefore, rigidification has been proposed a reliable approach to reduce hERG risk. We conducted reinforcement learning on core **4-0** and its corresponding source fragment, which is actually the fragment pair derived from chemists' choice (as in this case, we do not generate exactly the same fragment pair). 56 non-redundant non-blockers were obtained, the SMILES, NROT, hERG-channel blocker probability and docking score to the original target acidic mammalian chitinase (AMCase)⁴² can be found in **Table S4**. Their average NROT is 3.411, significantly smaller than **4a** (7) at 95% confidence level. Most structures produced by our model abandon the chlorobenzene ring and directly introduce a saturated heterocycle to the growing sites, as **4-0-9** in **Figure 6A**. The loss of the aromatic ring may help attenuate the hERG channel binding affinity, because it may be an important motif that occupies the highly favorable hydrophobic pocket⁴¹ in the hERG channel. In addition, it is also

next to PHE656, which is a frequently binded residue of many hERG channel blockers⁴³ (**Figure S2**). Thus, the hERG binding affinity of **4-0-9** is expected to drop, as it abandons the aromatic ring. However, the binding affinity to the original target AMCase suffers from a drop. Take **4-0-9** as an example (**Figure 6B**), the loss of chlorobenzene results in the loss of pi-pi interactions with TYR267 and larger docking score (-7.741 of **4-0-9** versus -9.273 of **4b**). Nevertheless, our model does recommend structures with greater rigidity and suggest a morpholine, which was also adopted by chemists. Overall, the case shows that hERG optimization may still need some human involvement, but our model can act as a good adviser and bring chemists some reasonable motifs to consider.

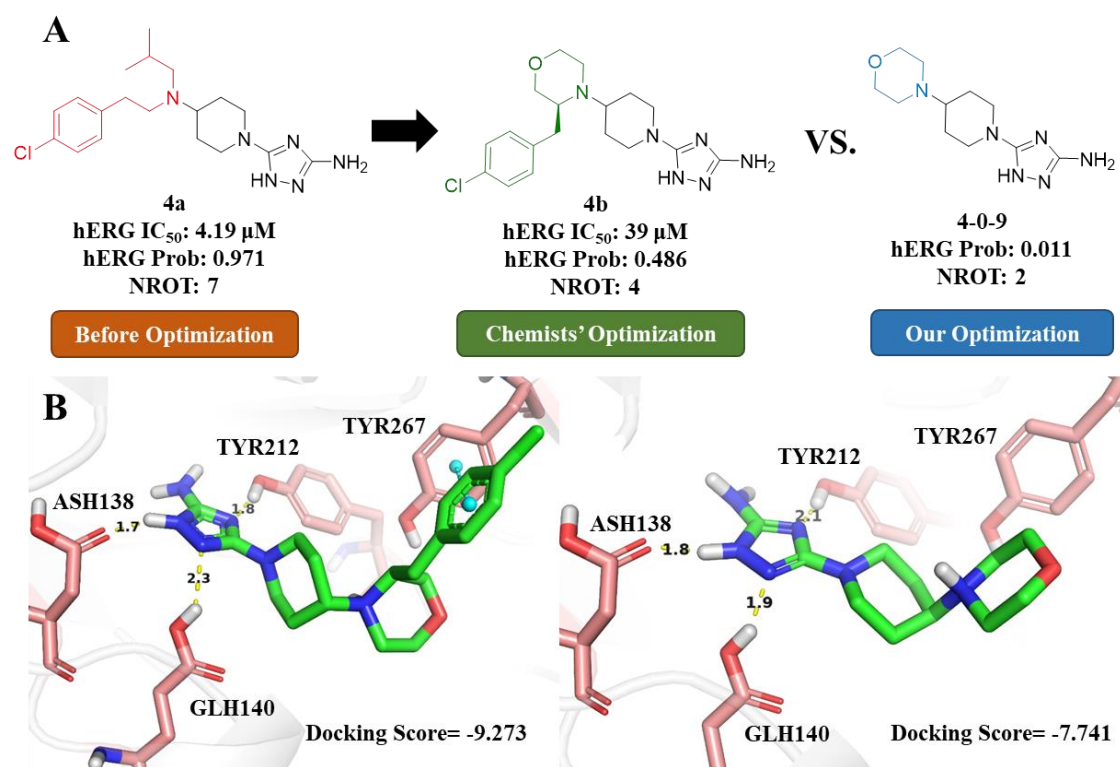


Figure 6. Conformation Comparison of Chemists' and Our Model's Optimization. **A**, structures, NROT and predicted hERG channel blocking probability (hERG Prob; higher score means higher blocking possibility) of original compound and optimized results. **B**, docking conformations of compound **4b** (left) and structure **4-0-9** (right) to AMCase (PDB code: 3RM4). For ligands, carbon atoms are colored in green, nitrogen atoms in blue, oxygen atoms in red, sulfur atoms in yellow. Yellow dash line stands for hydrogen bond. Blue dash line stands for pi-pi interactions.

Altering *pi-pi* interactions. Case 5 illustrates the strategy of decreasing aromaticity to reducing hERG binding affinity⁴⁴. Reinforcement learning were carried out on core fragment **5-2** and its corresponding source fragment. Detailed information for 38 non-duplicated potential non-blockers can be seen in **Table S5**. Most structures abandon the aromatic pyridine rings and introduce non-aromatic rings after reinforcement learning, which agrees with the chemists' optimization strategy. The most similar structure to **5b** is **5-2-10**, which replaces the pyridine with a cyclohexane (**Figure 7A**). Their docking conformation is also similar, despite that the docking score of **5-2-10** is larger than **5b**, which might be caused by the weaker hydrogen bond between 'O' on VAL690 and 'NH' on ligand (**Figure 7B**). In this case, our model also recommends reasonable optimization direction for chemists.

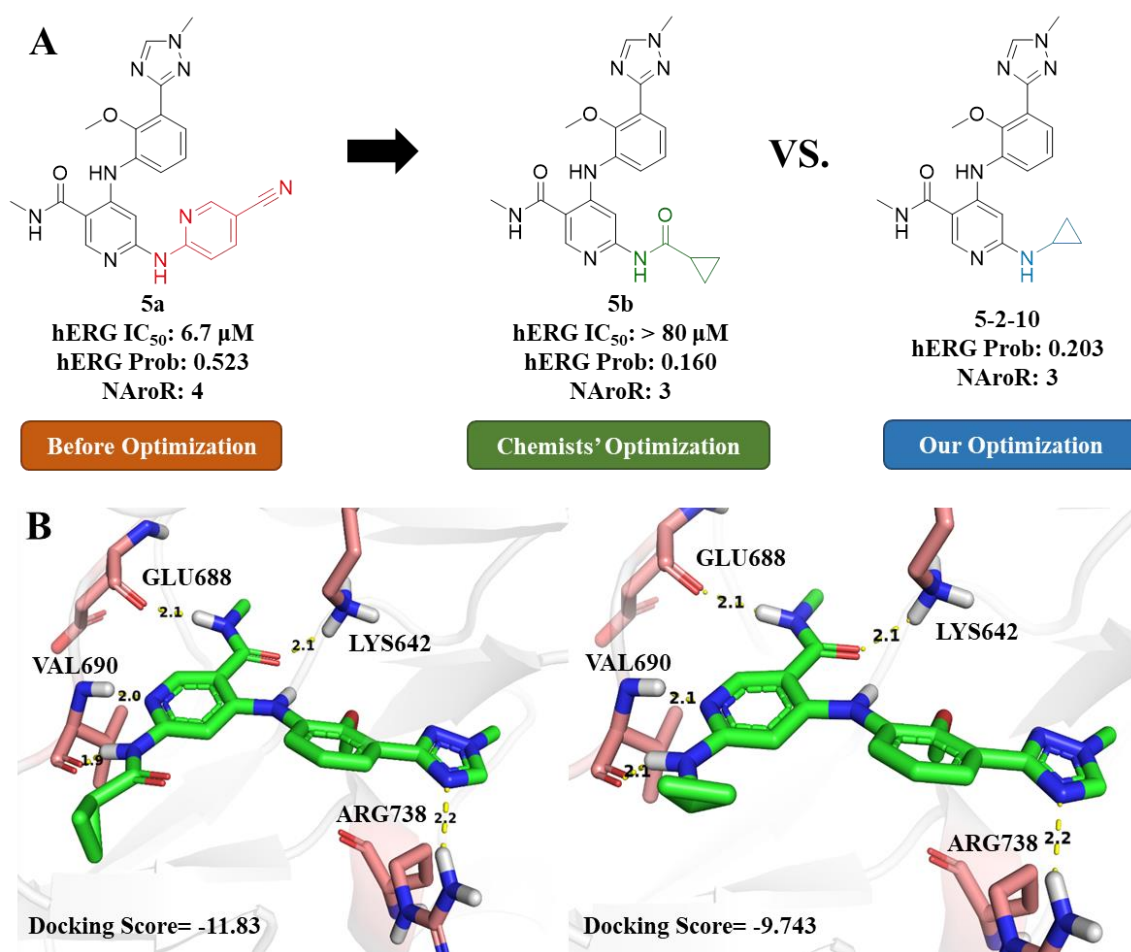


Figure 7. Conformation Comparison of Chemists' and Our Model's Optimization. **A**,

structures, NAroR and predicted hERG channel blocking probability (hERG Prob; higher score means higher blocking possibility) of original compound and optimized results. **B**, docking conformations of compound **5b** (left) and structure **5-2-10** (right) to TYK2 (PDB code: 6NZP). For ligands, carbon atoms are colored in green, nitrogen atoms in blue, oxygen atoms in red, sulfur atoms in yellow. Yellow dash line stands for hydrogen bond.

Conclusion

Assessing and reducing hERG-related cardiotoxicity is a critical step in the early stage of drug discovery. In this paper, we first built eight hERG classification models based on 9,708 molecules. The selected LightGBM model shows with good generalization ability on clinical external sets, which can reach a BAC of 0.729 even when the similarity of compounds to the training set is less than 0.4. Moreover, it correctly predicts five out of seven optimization cases provide by a previous study, demonstrating its ability to identify ‘activity cliffs’ caused by small structural changes. Then we propose a fragment-growing framework termed as GGNN-RNN to reduce hERG binding affinity while keep particular fragments. Our model can grow more valid structures with better synthetic accessibility than naive RNN. Comparison with chemists’ optimization in previous cases shows that our model can provide reliable optimizing directions including lowering lipophilicity, decreasing basicity and altering pi-pi interactions after being trained by reinforcement learning, which successfully reproduces or generates similar compounds as chemists’ optimization. The hERG classification model and the GGNN-RNN model can work as a useful tool for chemists to evaluate and improve hERG property in the drug discovery pipeline.

SUPPORTING INFORMATION

Supporting Methods

Scheme S1 Feature Selection Method

Scheme S2 Parameters for Eight Classification Methods

Scheme S3 Equations and Explanations for Classification Metrics

Scheme S4 SMARTS of Alerts Used in This Study

Supporting Tables

Table S1 Selected Descriptors by NSGA-II

Table S2-5 Optimized Structures by Our Model in Five Cases

Supporting Figures

Figure S1 The Variation of Dominance Ratio in 150 Generations

Figure S2 Induced Fit Docking of **4b** and **4-0-9** to hERG Homology Model

Supporting Excels

Excel S1 Datasets for hERG Classification Model

Excel S2 ChEMBL Kinase Targets

ABBREVIATIONS USED

hERG	Human Ether-a-go-go-related Gene
GGNN	Gated Graph Neural Networks
RNN	Recurrent Neural Network
NAroR	NumAromaticRings
NROT	NumRotatableBonds
NSGA-II	Non-dominated Sorting Genetic Algorithm-II
SVM	Support Vector Machine
RF	Random Forest
ExtraTree	Extremely Randomized Trees
Adaboost	Adaptive Boosting
GBDT	Gradient Boosted Decision Trees
XGBoost	Extrem Gradient Boosting
LightGBM	Light Gradient Boosting Machine

DNN	Deep Neural Networks
AUC	Area Under the ROC Curve
ACC	Accuracy
BAC	Balanced Accuracy
MCC	Matthews Correlation Coefficient
Kappa	Cohen's Kappa
F1	F1_Score
NLP	Natural Language Processing
HBA	The number of hydrogen acceptors
HBD	The number of hydrogen donors
NROT	The number of rotatable bonds
FNO	The fraction of nitrogen and oxygen atoms in heavy atoms
MW	Molecular weight
Alerts	Pan-assay interference compounds alerts, toxic alerts and other self-defined undesired substructures
SAR	Structure-activity relationship

CORRESPONDING AUTHORS

For Haichun Liu: Tel.: +86-25-86185163. Fax: +86-25-86185182. E-mail:

haichunliu@cpu.edu.cn

For Yadong Chen: Tel.: +86-25-86185163. Fax: +86-25-86185182. E-mail:

ydchen@cpu.edu.cn

For Lu Tao: Tel.: +86-25-86185163. Fax: +86-25-86185182. E-mail:

lutao@cpu.edu.cn

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. Yan Yang and Yanmin Zhang contributed equally to this work and should be considered as co-first authors.

Conflicts of interest

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was financially supported by National Natural Science Foundation of China (No. 81803370, No. 82073704), Natural Science Foundation of Jiangsu Province (No. BK20180559), State Key Laboratory Innovation Research and Cultivation Fund (No. SKLNMZZCX201812), and “Double World-classes” Construction Program of China Pharmaceutical University (No. CPU2018GF02).

Reference

1. Craveiro, N. S.; Lopes, B. S.; Tomás, L.; Almeida, S. F., Drug Withdrawal Due to Safety: A Review of the Data Supporting Withdrawal Decision. *Curr. Drug Saf.* **2020**, *15*, 4-12.
2. Ogura, K.; Sato, T.; Yuki, H.; Honma, T., Support Vector Machine model for hERG inhibitory activities based on the integrated hERG database using descriptor selection by NSGA-II. *Sci. Rep.* **2019**, *9*, 12220.
3. Garrido, A.; Lepailleur, A.; Mignani, S. M.; Dallemagne, P.; Rochais, C., hERG toxicity assessment: Useful guidelines for drug design. *Eur. J. Med. Chem.* **2020**, *195*, 112290.
4. Wang, S. Q.; Sun, H. Y.; Liu, H.; Li, D.; Li, Y. Y.; Hou, T. J., ADMET Evaluation in Drug Discovery. 16. Predicting hERG Blockers by Combining Multiple Pharmacophores and Machine Learning Approaches. *Mol. Pharm.* **2016**, *13*, 2855-2866.
5. Villoutreix, B. O.; Taboureau, O., Computational investigations of hERG channel blockers: New insights and current predictive models. *Adv. Drug Deliv. Rev.* **2015**, *86*, 72-82.
6. Olivecrona, M.; Blaschke, T.; Engkvist, O.; Chen, H., Molecular de-novo design through deep reinforcement learning. *J. Cheminf.* **2017**, *9*, 48.
7. Popova, M.; Isayev, O.; Tropsha, A., Deep Reinforcement Learning for de-novo Drug Design. *Sci. Adv.* **2018**, *4*, eaap7885
8. Jin, W.; Barzilay, R.; Jaakkola, T., Junction Tree Variational Autoencoder for Molecular Graph Generation. **2018**, arXiv:1802.04364.
9. You, J.; Liu, B.; Ying, R.; Pande, V.; Leskovec, J. In *Graph Convolutional Policy Network for Goal-Directed Molecular Graph Generation*, NIPS, 2018.
10. Maziarka, Ł.; Pocha, A.; Kaczmarczyk, J.; Rataj, K.; Danel, T.; Warchoń, M., Mol-CycleGAN: a generative model for molecular optimization. *J. Cheminf.* **2020**, *12*.
11. Zhou, Z.; Kearnes, S.; Li, L.; Zare, R. N.; Riley, P., Optimization of Molecules via Deep Reinforcement Learning. *Sci. Rep.* **2019**, *9*, 10752.
12. Gupta, A.; Muller, A. T.; Huisman, B. J. H.; Fuchs, J. A.; Schneider, P.; Schneider, G., Generative Recurrent Networks for De Novo Drug Design. *Mol Inform* **2018**, *37*.
13. Li, Y.; Tarlow, D.; Brockschmidt, M.; Zemel, R., Gated Graph Sequence Neural Networks. **2015**, arXiv:1511.05493.
14. Li, Q.; Jørgensen, F. S.; Oprea, T.; Brunak, S.; Taboureau, O., hERG Classification Model Based on a Combination of Support Vector Machine Method and GRIND Descriptors. *Mol. Pharm.* **2008**, *5*, 117-127.
15. Broccatelli, F.; Mannhold, R.; Moriconi, A.; Giuli, S.; Carosati, E., QSAR Modeling and Data Mining Link Torsades de Pointes Risk to the Interplay of Extent of Metabolism, Active Transport, and hERG Liability. *Mol. Pharm.* **2012**, *9*, 2290-2301.
16. Doddareddy, M. R.; Klaasse, E. C.; Shagufta; Ijzerman, A. P.; Bender, A., Prospective validation of a comprehensive in silico hERG model and its applications to commercial compound and drug databases. *ChemMedChem* **2010**, *5*, 716-29.
17. Inc. AS (2008) Pilot Pipeline version 7.5. Accelrys Software Inc: San Diego.
18. Sun, H., An accurate and interpretable bayesian classification model for prediction of HERG liability. *ChemMedChem* **2006**, *1*, 315-22.
19. Kalyaanamoorthy, S.; Barakat, K. H., Development of Safe Drugs: The hERG Challenge. *Med. Res. Rev.* **2018**, *38*, 525-555.

20. Rogers, D.; Hahn, M., Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742-54.
21. Moriawaki, H.; Tian, Y.-S.; Kawashita, N.; Takagi, T., Mordred: a molecular descriptor calculator. *J. Cheminf.* **2018**, *10*, 4.
22. Ding, K.; Kong, X.; Wang, J.; Lu, L.; Zhou, W.; Zhan, T.; Zhang, C.; Zhuang, S., Side chains of parabens modulate antiandrogenic activity: in vitro and molecular docking studies. *Environ. Sci. Technol.* **2017**, *51*, 6452-6460.
23. Zhang, Y.; Chen, Y.; Zhang, D.; Wang, L.; Lu, T.; Jiao, Y., Discovery of Novel Potent VEGFR-2 Inhibitors Exerting Significant Antiproliferative Activity against Cancer Cell Lines. *J. Med.Chem.* **2018**, *61*, 140-157.
24. Schrodinger, L., Schrodinger software suite. *New York: Schrödinger, LLC* **2011**, 670.
25. Durant, J. L.; Leland, B. A.; Henry, D. R.; Nourse, J. G., Reoptimization of MDL Keys for Use in Drug Discovery. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1273-1280.
26. Daylight Chemical Information Systems Inc., 27401 Los Altos, Suite 360, Mission Viejo, CA 92691.
27. Yang, Y.; Zhang, Y.; Hua, Y.; Chen, X.; Fan, Y.; Wang, Y.; Liang, L.; Deng, C.; Lu, T.; Chen, Y.; Liu, H., In Silico Design and Analysis of a Kinase-Focused Combinatorial Library Considering Diversity and Quality. *J. Chem. Inf. Model* **2020**, *60*, 92-107.
28. Cai, C. P.; Guo, P. F.; Zhou, Y. D.; Zhou, J. W.; Wang, Q.; Zhang, F. X.; Fang, J. S.; Cheng, F. X., Deep Learning-Based Prediction of Drug-Induced Cardiotoxicity. *J. Chem. Inf. Model.* **2019**, *59*, 1073-1084.
29. Liu, L. L.; Lu, J.; Lu, Y.; Zheng, M. Y.; Luo, X. M.; Zhu, W. L.; Jiang, H. L.; Chen, K. X., Novel Bayesian classification models for predicting compounds blocking hERG potassium channels. *Acta Pharmacol. Sin.* **2014**, *35*, 1093-102.
30. Wang, Y. W.; Huang, L.; Jiang, S. W.; Wang, Y. F.; Zou, J.; Fu, H. G.; Yang, S. Y., Capsule Networks Showed Excellent Performance in the Classification of hERG Blockers/Nonblockers. *Front. Pharmacol.* **2020**, *10*.
31. Braga, R. C.; Alves, V. M.; Silva, M. F.; Muratov, E.; Fourches, D.; Liao, L. M.; Tropsha, A.; Andrade, C. H., Pred-hERG: A Novel web-Accessible Computational Tool for Predicting Cardiac Toxicity. *Mol. Inform.* **2015**, *34*, 698-701.
32. Didziapetris, R.; Lanevskij, K., Compilation and physicochemical classification analysis of a diverse hERG inhibition database. *J. Comput. Aided Mol. Des.* **2016**, *30*, 1175-1188.
33. Sun, H.; Huang, R.; Xia, M.; Shahane, S.; Southall, N.; Wang, Y., Prediction of hERG Liability - Using SVM Classification, Bootstrapping and Jackknifing. *Mol. Inform.* **2017**, *36*.
34. Siramshetty, V. B.; Chen, Q. F.; Devarakonda, P.; Preissner, R., The Catch-22 of Predicting hERG Blockade Using Publicly Accessible Bioactivity Data. *J. Chem. Inf. Model.* **2018**, *58*, 1224-1233.
35. Lee, H. M.; Yu, M. S.; Kazmi, S. R.; Oh, S. Y.; Rhee, K. H.; Bae, M. A.; Lee, B. H.; Shin, D. S.; Oh, K. S.; Ceong, H.; Lee, D.; Na, D., Computational determination of hERG-related cardiotoxicity of drug candidates. *BMC Bioinf.* **2019**, *20*, 250.
36. Kim, H.; Nam, H., hERG-Att: Self-attention-based deep neural network for predicting hERG blockers. *Comput. Biol. Chem.* **2020**, *87*, 107286.
37. Ryu, J. Y.; Lee, M. Y.; Lee, J. H.; Lee, B. H.; Oh, K. S., DeepHIT: a deep learning framework for prediction of hERG-induced cardiotoxicity. *Bioinformatics* **2020**, *36*, 3049-3055.
38. Ertl, P.; Schuffenhauer, A., Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *J. Cheminf.* **2009**, *1*, 8.
39. Furber, M.; Tiden, A.-K.; Gardiner, P.; Mete, A.; Ford, R.; Millichip, I.; Stein, L.; Mather, A.;

Kinchin, E.; Luckhurst, C.; Barber, S.; Cage, P.; Sanganee, H.; Austin, R.; Chohan, K.; Beri, R.; Thong, B.; Wallace, A.; Oreffo, V.; Hutchinson, R.; Harper, S.; Debreczeni, J.; Breed, J.; Wissler, L.; Edman, K., Cathepsin C Inhibitors: Property Optimization and Identification of a Clinical Candidate. *J. Med. Chem.* **2014**, *57*, 2357-2367.

40. Goldberg, F. W.; Finlay, M. R. V.; Ting, A.; Beattie, D.; Lamont, G.; Fallan, C.; Wrigley, G.; Schimpl, M.; Howard, M. R.; Williamson, B.; Vazquez-Chantada, M.; Barratt, D.; Davies, B.; Cadogan, E.; Montoy, A. R.; Dean, E., The Discovery of 7-Methyl-2-[(7-methyl[1,2,4]triazolo[1,5-a]pyridin-6-yl)amino]-9-(tetrahydro-2H-pyran-4-yl)-7,9-dihydro-8H-purin-8-one (AZD7648), a Potent and Selective DNA-Dependent Protein Kinase (DNA-PK) Inhibitor. *J. Med. Chem.* **2020**, *63*, 3461-3471.

41. Cho, Y. S.; Whitehead, L.; Li, J.; Chen, C. H. T.; Jiang, L.; Vögtle, M.; Francotte, E.; Richert, P.; Wagner, T.; Traebert, M.; Lu, Q.; Cao, X.; Dumotier, B.; Fejzo, J.; Rajan, S.; Wang, P.; Yan-Neale, Y.; Shao, W.; Atadja, P.; Shultz, M., Conformational Refinement of Hydroxamate-Based Histone Deacetylase Inhibitors and Exploration of 3-Piperidin-3-ylindole Analogues of Dacinostat (LAQ824). *J. Med. Chem.* **2010**, *53*, 2952-2963.

42. Mazur, M.; Olczak, J.; Olejniczak, S.; Koralewski, R.; Czestkowski, W.; Jedrzejczak, A.; Golab, J.; Dzwonek, K.; Dymek, B.; Sklepkiewicz, P. L.; Zagozdzon, A.; Noonan, T.; Mahboubi, K.; Conway, B.; Sheeler, R.; Beckett, P.; Hungerford, W. M.; Podjarny, A.; Mitschler, A.; Cousido-Siah, A.; Fadel, F.; Golebiowski, A., Targeting Acidic Mammalian chitinase Is Effective in Animal Model of Asthma. *J. Med. Chem.* **2018**, *61*, 695-710.

43. Mitcheson, J. S.; Chen, J.; Lin, M.; Culberson, C.; Sanguinetti, M. C., A structural basis for drug-induced long QT syndrome. *PNAS* **2000**, *97*, 12329.

44. Wroblewski, S. T.; Moslin, R.; Lin, S.; Zhang, Y.; Spergel, S.; Kempson, J.; Tokarski, J. S.; Strnad, J.; Zupa-Fernandez, A.; Cheng, L.; Shuster, D.; Gillooly, K.; Yang, X.; Heimrich, E.; McIntyre, K. W.; Chaudhry, C.; Khan, J.; Ruzanov, M.; Tredup, J.; Mulligan, D.; Xie, D.; Sun, H.; Huang, C.; D'Arienzo, C.; Aranibar, N.; Chiney, M.; Chimalakonda, A.; Pitts, W. J.; Lombardo, L.; Carter, P. H.; Burke, J. R.; Weinstein, D. S., Highly Selective Inhibition of Tyrosine Kinase 2 (TYK2) for the Treatment of Autoimmune Diseases: Discovery of the Allosteric Inhibitor BMS-986165. *J. Med. Chem.* **2019**, *62*, 8973-8995.

Table of Contents

Reducing hERG Toxicity Using hERG Classification Model and Fragment-growing Network

Yan Yang^{a,1}, Yanmin Zhang^{a,1}, Xingye Chen^a, Yi Hua^a, Guomeng Xing^a, Chenglong Deng^a, Li Liang^a, Tao Lu^{a,b,*}, Yadong Chen^{a,*}, Haichun Liu^{a,*}

^a Laboratory of Molecular Design and Drug Discovery, School of Science, China Pharmaceutical University, 639 Longmian Avenue, Nanjing 211198, China

^b State Key Laboratory of Natural Medicines, China Pharmaceutical University, 24 Tongjiaxiang, Nanjing 210009, China

¹ The authors contributed equally to this work and should be considered as co-first authors.

