Combining Molecular Docking and Graph Convolutional Network to Discover Potential Selective Limk1 Inhibitors

Weihe Zhong
Sun Yat-Sen University
Lu Zhao
Sun Yat-Sen University
Calvin Yu-Chian Chen (chenyuchian@mail.sysu.edu.cn)
Sun Yat-sen University https://orcid.org/0000-0001-9213-9832

Research Article

Keywords: LIM kinase 1 (Limk1), Molecular docking, Artificial intelligence (AI), Graph convolutional network (GCN), Molecular dynamics (MD) simulation

Posted Date: March 15th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-278141/v1

Combining molecular docking and graph convolutional network to discover potential selective Limk1 inhibitors

Weihe Zhong[a,b#], Lu Zhao[a,c#], Calvin Yu-Chian Chen[a,d,e]*

[a]*School of Intelligent Systems Engineering, Artificial Intelligence Medical Center, Sun Yat-sen University, Shenzhen, Guangdong 510275, China*

[b]*School of Biomedical Engineering, Sun Yat-sen University, Shenzhen, Guangdong 510275, China*

[c]*Department of Clinical Laboratory, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, Guangdong 510655, China*

[d]*Department of Medical Research, China Medical University Hospital, Taichung 40447, Taiwan*

[e]*Department of Bioinformatics and Medical Engineering, Asia University, Taichung 41354, Taiwan*

[#] *Equal contribution*

* *Corresponding Authors*

Calvin Yu-Chian Chen, Ph.D.

School of Intelligent Systems Engineering, Director of Artificial Intelligence Medical Center, Sun Yat-sen University, Shenzhen, Guangdong 510275, China.

TEL: 02039332153

E-mail: chenyuchian@mail.sysu.edu.cn

Abstract

The LIM kinase 1 (Limk1) has been demonstrated to be considered a therapeutic target and selective inhibitors of Limk1 rather Rho-associated kinase 2 (ROCK2) are considered of interest for the treatment of several indications such as Alzheimer's disease (AD), Parkinson's disease (PD) and cancer migration/invasion. Here, we utilized molecular docking to screen potential compounds of Limk1 from Traditional Chinese Medicine (TCM) database. Meanwhile, we performed a three-dimensional graph convolutional network (3DGCN), based on 3D molecular graph, to predict the inhibitory activity of Limk1 and ROCK2. Compared with the baseline models (RF, GCN and Weave), the 3DGCN achieved higher accuracy and the averaged RMSE values on test sets for Limk1 and ROCK2 were 0.721 and 0.852 respectively. In 3DGCN, above 80% of the test-set molecules from both two datasets were predicted within absolute error of 1.0 and the feature visualization suggested that it could automatically learn relevant structure features including 3D molecular information from a specific task for prediction. Furthermore, molecular dynamics (MD) simulations within 100 ns were employed to verify the stability of ligand-protein complexes and reveal the binding modes of the potential selective lead compounds of Limk1. Finally, integrating docking results, the predicted values by the 3DGCN and the MD analysis, we found that 7549, 2007_15649 and 3519 might be the potential and selective inhibitors for Limk1 receptor rather than ROCK2.

Keywords

LIM kinase 1 (Limk1); Molecular docking; Artificial intelligence (AI); Graph convolutional network (GCN); Molecular dynamics (MD) simulation

1. Introduction

The innovative drug research and development is a highly complex, time-consuming, risky and costly process, which aims to find chemical molecules able to bind (inhibit or activate) a drug target (usually a protein)[1, 2]. Currently, computer-aided drug design (CADD) is becoming increasingly important in the field of drug discovery as a means to reduce the time and cost of searching for novel lead compounds. Molecular docking is to place the molecules in the three-dimensional (3D) structure database one by one at the active site of the target protein. By continuously optimizing the position, conformation, dihedral angle of the rotatable bonds within the molecule, and the amino acid residue side chain of the receptor protein, the algorithms is trying to looking for the optimal conformation for the interaction between the small molecule and receptor target[3]. As a powerful tool and an essential technique in the CADD research field, docking is a theoretical simulation method that mainly studies the interactions between the drug molecule and receptor, predicts its binding mode and affinity. In the course of researches, docking can be used for virtual screening to filter the library of the compounds and reduce the cost of high-throughput screening testing or for lead optimization to find candidates that will be directly synthesized and optimized in the laboratory[4-6].

Over the past few years, artificial intelligence (AI), especially deep learning (DL)[7] technologies have achieved remarkable success in computer vision (CV)[8] and natural language processing (NLP)[9]. With the accumulation of a large amount of data concerning the biological effects of chemical compounds and several biochemical

databases are available for pharmaceutical research[10, 11], many DL methods[12-16] have been applied on molecular activity or property prediction and achieved excellent performance compared to traditional machine learning (ML) methods such as support vector machines (SVM)[17] and random forest (RF)[18]. However, most conventional DL approaches were based on hand-encoded features or predefined descriptors, instead of extracting structural information directly from raw inputs. Furthermore, the early DL architectures were not suitable for structured data such as molecules, and the structural information is neither considered nor fully utilized in the feature extraction process.

In recent years, graph neural networks (GNNs)[19] have been widely used and achieved impressive success in many research domains, such as social network[20], knowledge graphs[21] and recommendation[22], due to its capability of handling graph-structured data. For molecular graph structure, GNNs take the molecule graph (by treating atoms as nodes and bonds as edges) as input and extract feature automatically to accomplish the prediction task. Compared with sequence-based methods, such as SMILES[23], GNNs can extract the feature information from the molecular structure directly. To date, GNNs have been established to capture non-Euclidean structured data, which have attracted great attention in drug discovery and achieved state-of-the-art performance on molecular property (activity) prediction tasks. Duvenaud *et al.*[24] first proposed a graph convolutional network (GCN) that operates directly on molecular graphs. In this study, the final representation of a molecule was obtained by gathering the representation of all atoms and passing through a dense layer that can interpret the learned features. In addition to atom features, bond information is also critical for molecular properties and

can also be coded by graph convolution. Weave[25] and message passing neural network (MPNN)[26] model were developed to construct virtual edges through calculating a feature vector for each pair of atoms in a molecular graph and achieved the state-of art performance on most datasets in MoleculeNet[27]. Since the attention mechanism[28] has been widely utilized to focus on task-relevant features of the input information and improved the prediction accuracy in CV and NLP domain, several researchers[29-32] have applied the attention mechanism into the graph architecture for molecule property prediction and obtained excellent performance. Recently, Cho *et al*.[33] constructed a three-dimensional graph convolutional network (3DGCN), which receive 3D molecular graph with bond directions information, exhibited excellent performance on molecular properties and biochemical activities predictive tasks.

The LIM kinase (Limk) is a serine-threonine protein, including two isoforms, Limk1 and Limk2. These kinases are recognized by two amino-terminal LIM domains, an adjacent PDZ and a catalytic domain. Limk is located downstream of Rho-associated kinase (ROCK) and PAK, and regulates the polymerization of actin filaments in signaling pathway. Activated Limk can phosphorylate and inactivate confilin, leading to dynamic regulation of actin cytoskeleton (Figure 1)[34, 35]. The reported literatures suggest that many disorders are related to the regulatory mechanism of Limk1, and the Limk1 inhibitors are of potential drugs for the treatment of multiple indications including Alzheimer's disease (AD)[36], Parkinson's disease (PD)[37], and cancers and cancer cell migration/invasion[38-41]. ROCK is a challenging drug target, and inhibition of ROCK may lead to serious side effects, including a pronounced decrease in blood

pressure[42, 43], While Limk1 is downstream of the signal pathway and regulated by ROCK, especially ROCK2. Therefore, it is critical to discover potent and selective Limk1 inhibitors for the treatment of related diseases.

In this paper, we performed molecular docking technology to screen novel potential lead compounds of Limk1 protein from traditional Chinese medicines (TCM) database. Meanwhile, we introduced a 3DGCN to predict the inhibitory activity of LIMK1 and ROCK2, based on 3D molecular graph, and exhibits significantly higher performance compared with other GCN and machine-learning models. Furthermore, molecular dynamics (MD) simulations were employed to verify the stability of ligand-protein complexes and reveal the binding modes of the potential selective lead compounds of Limk1. The flow chart of our experimental design is shown in Figure 2.

2. Methods

2.1 Molecular Docking

The whole process could be divided into three steps: protein preparation, ligand database preparation and docking. The crystal structure of Limk1 protein was acquired from RCSB Protein Data Bank[44] (PDB ID: 3S95) with 1.65 Å resolution. We used the "Prepare Protein" module in Discovery Studio Client v17.2.0 (DS) software to prepare the Limk1 protein by removing the crystal water, supplementing the missing loops and protonation of the amino acids at PH 7.4. Then, the binding site was defined from a ligand selection in the prepared protein. Meanwhile, a total of 61,000 small molecules collected from TCM Database@Taiwan[45] (http://tcm.cmu.edu.tw) were filtered by Linpinski's Rule of Five and absorption, distribution, metabolism, excretion, and

toxicity (ADMET) protocol in DS. The screened 18,776 TCM compounds were then prepared for docking by energy minimization using the smart minimizer algorithm at Harvard Molecular Mechanics (CHARMm) force field. After that, we chose the "Dock Ligands (LigandFit)" protocol to dock the TCM compounds to Limk1 protein. During the docking procedure, the Monte-Carlo techniques were applied to generate ligand conformations and docked them into the active site using a shape-based initial docking. And the docked poses may optionally be minimized with CHARMm and evaluated with ten scoring functions including Jain, PMF, PMF04, LigScore1, LigScore2, PLP1, PLP2, Ludi1, Ludi2 and Ludi3. Also, the DREIDING force field with Gasteiger charges was used to calculate ligand-receptor interaction energies. Last, results of the molecular docking were ranked according to the "dock score" which was calculated mainly based on the minimum energy of the ligand docking conformation.

2.2 Three-Dimensional Graph Convolutional Network (3DGCN)

Graph-based deep learning algorithms, especially GNN and its variants that could handle graph-structured data, have attracted extensive attention in chemical science recently, and have shown remarkable performance in various applications, such as biological property and activity prediction, drug-target interaction (DTI) prediction, de novo molecular design, synthesis prediction and so on[46]. In the general model architecture of GNNs, a molecule is represented as a graph G, which consists of atom features and bond features. The graph-structured data G is converted into the node-embedding matrix through a neighborhood aggregation strategy. And then, the readout function is used to transform the node-embedding matrix into the graph-embedding

vector. In existing GNNs, the graph feature vector can be obtained by mean, sum and max operation for the last layer node states. Finally, the vector representations of the graph are feed into the FNN (fully connected network) to predict the target task. The process of the GNN model can be formulated as follows:

Aggregation:

$$H_i^{(k)} = \sigma \sum_{j \in \mathcal{N}_i} (\varphi(H_i^{(k-1)}, H_j^{(k-1)}, A_{ij})) \quad (1)$$

Readout:

$$\hat{y} = R(H_i^{(L)} \mid i \in G)$$

where $\mathcal{N}_i$ represents the neighbor atoms of atom i, A_{ij} denotes the edge between the atom i and the atom j, σ is an activation function, φ is an aggregation function and $H_i^{(k)}$ is the output of the kth aggregation layer. Given the final node-embedding representation $H_i^{(L)}$, we could obtain the feature vector $\hat{y}$ of the graph G from the readout phase and carry out the predictive task by the fully connected network.

Recently, various GNNs have been proposed with their own features' transformations based on graphs directly and achieved impressive success in drug-related tasks[27, 29-33]. Among them, GCNs have been the widely used architecture, shown powerful node and graph embedding capabilities and state-of-the-art performance on some molecular property/activity prediction tasks. Molecular compounds are 3D-shaped substance where the spatial distances and directions between atoms in the 3D space greatly impact molecular properties and activities. However, current mostly

GCNs operated on flat 2D molecular graph where only with nodes and edges and the structure information in the 3D space is neglected[46]. In this study, we applied a 3DGCN[33], which took a molecular graph with 3D bond features as the input, for molecular Limk1 and ROCK2 receptor activities prediction.

2.2.1 Graph Representation of Molecules

In 3DGCN, we represented a graph G as (X, A, R), where $X \in \mathbb{R}^{n\times m}$ is the node feature matrix, $A \in \mathbb{R}^{n\times n}$ is the adjacency matrix and $R \in \mathbb{R}^{n\times n\times 3}$ is the relative position matrix, n and m denote the number of atoms and the atom features respectively. The initial node feature matrix X represents the information of each atom features and the features include atomic type, number of heavy atom neighbors, hybridization, valence, formal charge, ring size, aromaticity, number of neighboring hydrogens, chirality, acid/base, and hydrogen bonding, which were described in Table 1. All features of a molecule are extracted through the RDKit package[47] and most of these features were encoded into a one-hot vector. The adjacency matrix represents the connection information between atoms and if there is a chemical bond between atom i and j, $A_{ij} = 1$; otherwise, $A_{ij} = 0$. Especially, the relative position matrix denotes the information of bond directions in 3D space which could be represented by the interatomic positions, rather than individual positions.

2.2.2 3DGCN architecture

Figure 3 provides a detailed schematic of the used model in this study. In 3DGCN, three modules are implemented including three graph convolution layers, a graph gather layer and two fully connected layers. Each layer has 128 units.

Graph Convolution Layer. After obtaining the 3D graph representation of molecules, the input matrices are further refined into the scalar features $S \in \mathbb{R}^m$ and vector features $V \in \mathbb{R}^{m\times 3}$ (initialized with zeros). In the first phase of the convolution layer, the linear combinations are performed to combined the scalar or vector features from adjacent atoms and generate intermediate features. In particular, the scalar-vector feature interconversions are conducted through dot product ($\odot$) or tensor product ($\otimes$) with the relative position vector $\vec{r}_{ij}$ (Figure 3a). The calculating process is formulated as follows (3)-(6):

$$H^k_{s\to s,i} = ReLu\Big[W_{s\to s}(S^k_i \parallel S^k_j) + b_{s\to s}\Big] \quad (3)$$

$$H^k_{s\to v,i} = tanh\Big[\big(W_{s\to v}\big(S^k_i \parallel S^k_j\big) + b_{s\to v}\big) \otimes \vec{r}_{ij}\Big] \quad (4)$$

$$H^k_{v\to s,i} = ReLu\Big[\big(W_{v\to s}\big(V^k_i \parallel V^k_j\big) + b_{v\to s}\big) \odot \vec{r}_{ij}\Big] \quad (5)$$

$$H^k_{v\to v,i} = tanh\Big[W_{v\to v}(V^k_i \parallel V^k_j) + b_{v\to v}\Big] \quad (6)$$

where S_i^k and V_i^k are the scalar and vector features of target node i on k layer, $\|$ represents concatenation, $W_{s\to s}$, $W_{s\to v}$, $W_{v\to s}$, $W_{v\to v}$ and $b_{s\to s}$, $b_{s\to v}$, $b_{v\to s}$, $b_{v\to v}$ are the learnable weight matrices and biases respectively. In the second phase, the four intermediate features ($H_{s\to s,i}^k$, $H_{s\to v,i}^k$, $H_{v\to s,i}^k$ and $H_{v\to v,i}^k$) are concatenated and followed by linear combination for the same feature form before convolution. Then, we implement the convolution operations with the normalized adjacency matrix proposed in Kipf's model[48] for aggregating the scalar and vector features of neighbor nodes and thus leading to generating the high-level features as shown in the following equation:

$$S_i^{k+1} = ReLu\left[\sum_{j\in\mathcal{N}_i} A_{ij}\left(W_s\left(H_{s\to s,i}^k \parallel H_{v\to s,i}^k\right)\right) + b_s\right] \quad (7)$$

$$V_i^{k+1} = tanh\left[\sum_{j\in\mathcal{N}_i} A_{ij}\left(W_v\left(H_{v\to v,i}^k \parallel H_{s\to v,i}^k\right)\right) + b_v\right] \quad (8)$$

where $\mathcal{N}_i$ is the set of neighbors for target node i, A_{ij} is the normalized adjacency matrix, and W_s, W_v, b_s and b_v are the trainable weigh matrices and biases.

Graph Gather Layer. After convolution, each atom obtains the optimal scalar and vector representative matrices, and the graph gather layer is needed to aggregate atomic information and obtain the descriptor of molecular graph for predicting specific task properties. In this study, in order to enhance the ability of the atom to integrate with surrounding information and be insensitive to the variance of molecular feature order, a readout operation is implemented to collect the features along the nodes with two

strategies for a comparison: the 3DGCN$_{sum}$ calculate the sum of all atomic features distributed on the nodes, leading to the generation of molecular scalar and vector features; the 3DGCN$_{max}$ choose the maximum value of the atomic features as a molecular feature.

Fully Connected Layer. The generated molecular features are fed into the fully connected neural network with ReLu activation for prediction. The scalar features are assigned to the two-layer stack of the fully connected neural network, while the vector features are assigned to the two-layer stack of the time-distributed fully connected neural network to suppress the separation between each axis during the linear combination. Meanwhile, in order to improve the generalization ability of the model and avoid overfitting, the dropout rate with 0.4 and L2 regularization with 0.0005 were applied. Finally, the outputs of molecular vector features are flattened, concatenated with scalar features and fed into a single-layer neural network, which predict the values of task activities from the datasets.

2.2.3 Cross-Validation (CV) and Model Evaluation

Cross-Validation is applied to assess the performance of the model and obtain a reliable and stable model. For the prediction performance evaluation, we use ten-fold stratified cross-validation. In detail, the datasets are randomly split into three sets in each fold: 80% of the dataset for train data, 10% for validation data, and 10% for test data. Cross-validation is repeated for 10 times, and each subsample is validated once. The average value of the 10 times test results is used to estimate the accuracy of the model, and then select a trial that achieve the best validation accuracy to perform our

regression prediction task. In order to compare with the baseline models in this paper, we apply the root mean square error (RMSE) to evaluate the performance of the model for regression tasks.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \quad (9$$

where $\hat{y}_i$ and y_i are the predicted and experimental molecular activity. RMSE represents the deviation between the predicted and experimental value and the smaller the value of RMSE, the better the performance of the predictive model.

2.2.4 Datasets of Limk1 and ROCK2 inhibitors

The SIMLES (simplified molecular input line entry specification) information of compounds with inhibitory effect on Limk1 and ROCK2 were acquired from ChEMBL database[10] (https://www.ebi.ac.uk/chembl/). We filtered the raw data by removing the duplicates and some data without 2D molecular structure or biological activity information. After filtering, a total of 578 Limk1 inhibitors and 1014 ROCK2 inhibitors remained and the activity values of inhibitors were normalized by Equation (10). The distribution histogram of Limk1 and ROCK2 inhibitors were shown in Figure 4. Particularly, the obtained datasets only contain the 2D structure information, not provide the 3D structures necessary for the 3D molecular graph representation in 3DGCN. Therefore, we generated a set of conformers for each molecule and optimized them with the Merck molecular force field (MMFF94) through RDKit package[49]. After optimization, the lowest-energy 3D conformer is selected and used as input of the model.

pIC_{50}

$= -\lg(IC_{50})$

2.2.5 Training Protocol

To implement the 3DGCN model, the open-source machine learning library Keras[50] 2.2.4 with TensorFlow[51] 1.12 was used as a backend. Model training used the Adam optimizer for gradient decent optimization and was controlled by learning rate decay and early stopping techniques[52]. The learning rate decay would monitor the validation loss and reduce the learning rate by multiplying a factor of 0.9, when the loss reached a plateau in 10 epochs. And the early stopping would terminate the training process to avoid overfitting and save training time, when the performance metric of the model has not improved in 15 epochs on validation set. The other training hyperparameters was set as: epochs = 150, batch size = 8, initial learning rate = 0.001, and minimum learning rate = 0.0005. All modeling experiments were carried out using a machine with an Intel(R) Core(TM) i7-9700K at 3.60GHz x 8 CPU, 15.6 GiB of RAM, and an NVIDIA GeForce RTX 2060 SUPER/PCIe/SSE2 graphics card.

2.3 Molecular Dynamics Simulation (MDs)

The crystal structure obtained in the experiment could only represent the instantaneous conformation of the protein in complex changes and in order to effectively find the possible protein-ligand interaction mode in the large chemical space, most docking studies treated the receptor as a rigid substance, which often cause the docking poses do not match the protein-ligand interactions under physiological dynamic conditions. Molecular dynamics simulations could investigate the relationship

between dynamic experimental data established by computer simulation and static molecular structure and analyze the free energy of binding and the reaction process between the ligand and receptor. Therefore, we performed MD simulations with the protein−ligand complexes in 100 ns with Gromacs 2019 software. The whole process of MD simulation mainly includes four steps: the initial structure processing of the receptor-ligand complexes; the energy optimization of simulation system to remove inappropriate collisions in the structure; the balance of the system to prepare for long-term simulation; and the final long-term simulation to obtain the dynamic trajectory. Firstly, the ligands were submitted to SwissParam web server for processing to generate the topology file. The CHARMm27 force field was applied to the structure of ligands and protein, and the ligand-protein complexes were placed in the center of a periodic cubic box containing TIP3P explicit solvent water model and the distance between each atom of protein and the box boundary was not less than 1.2 nm. To simulate the physiological conditions, we also added appropriate amount of sodium ions to neutralize the system charge. Then, the energy minimization of the system was conducted with the steepest descent minimization algorithm in 5000 steps. After that, 10 ns canonical ensemble (NVT equilibration, Berendsen temperature coupling with constant particle count, volume and temperature) and 10 ns isothermal−isobaric ensemble (NPT equilibration, Parrinello-Rahman pressure coupling with constant particle count, pressure and temperature) were performed to maintain the simulation system in a stable environment (300 K, 1 bar). The coupling constants for temperature and pressure were set to 0.1 ps and 2 ps respectively. The Particle Mesh Ewald (PME)

algorithm with an interpolation order of 4 and a grid spacing of 1.6 Å was used to describe the long-range electrostatic interaction, and the Verlet scheme with a cutoff value of 10 Å was applied for calculating short-range electrostatics. All bond lengths were constrained by the linear constraint solver (LINCS) algorithm and the temperature was set at 310 K. Finally, after the thermodynamic properties were stable, the system was simulated for 100 ns with a time step of 2 ps and the coordinates of the system were saved every 10 ps. Through MD analysis, we could acquire the calculated results including the root-mean-square deviation (RMSD), the total energy, the radius of gyration (gyrate), mean square displacement (MSD), solvent accessible surface area (SASA), and root-mean-square fluctuation (RMSF).

3. Results and Discussion

3.1 Performance Evaluation of 3DGCN on Limk1 and ROCK2 Datasets

We conducted the performance evaluations of the 3DGCN with three popular baseline methods, which were shown in MoleculeNet[27], on Limk1 and ROCK2 datasets for regression tasks. Firstly, the Random Forests (RF)[18], one of the most famous machine learning methods, which is a classifier that uses many individual decision trees to train and predict samples. It could also be used for regression tasks. The second method was graph convolutional network (GCN)[48], which designed the circular fingerprints to encode molecular graphs directly and extracted effective molecular representations. And thirdly, the Weave model[25], which leveraged the node and edge attributes in the molecule to calculate the feature vector, was also used for comparison. We compared with these three models to show that 3DGCN provided an efficient way

to use molecular spatial topology information for binding affinity prediction.

Table 2 reported the mean absolute error (MAE) and root-mean-square error (RMSE) results of the models on Limk1 and ROCK2 datasets. The 10-fold cross-validation results showed that the 3DGCN$_{max}$ have achieved better performance than the GCN and Weave model on both two datasets. The averaged RMSE values of test sets for Limk1 and ROCK2 were 0.721 and 0.852 respectively, which achieved about 2.8% and 8.1% increases compared with 0.742 and 0.927 from the Weave and GCN model. The scatterplots of the predicted versus experimental values in the best worked model of the 10-fold cross-validation were shown in Figure 5 and indicated a linear relationship for Limk1 and ROCK2. It is worth noting that more than 50% of the test-set molecules from both two datasets were predicted within absolute error of 0.5 and above 80% were within absolute test error of 1.0 (Figure 6). In short, all this evidence showed that the 3DGCN has high prediction accuracy on molecular Limk1 and ROCk2 activity.

Since the 3DGCN achieved excellent performance for Limk1 and ROCK2 activity prediction, we further attempted to explored the interpretability issue of the model by visualizing the atomic contributions to the prediction result on a molecule. We collected the atomic features including scalar and vector features before aggregation process, and plotted a heat map for each molecule with color intensity according to the influence of individual atoms on molecule-level learned features. Example molecules were chosen randomly from the test sets of two datasets and was shown in Figure 7. From the heat maps, we could clearly observe that the highly attentive atoms matched with the structure-activity relationship reported in the literature. For example, the 2-

aminothiazole moiety of CHEMBL2070501, the pyrrolopyrimidine scaffold, dimethylamino and ester chain of CHEMBL3613609 played major impact on the IC_{50} values for Limk1[53, 54], and in CHEMBL2332093, the pyrazole group, the urea carbonyl, the hydroxyethyl group formed H-bonds interaction and the benzyl moiety formed hydrophobic interactions with ROCK2, all this were responsible for its high potency and selectivity[55]. These observations demonstrated that the 3DGCN had indeed successfully extracted relevant features including 3D molecular information by learning from a specific task, even though it had not been taught or given the information of functional groups.

3.2 Selection of TCM Candidates based on Docking Results and Bioactivity Predictions.

Compounds from the TCM database were used in high-throughput virtual screening against Limk1 and the results were ranked based on docking scores. The TCM compounds with the top 20 docking score were shown in Table 3, and then the best performance of the trained 3DGCN models were performed to predict their Limk1 and ROCK2 inhibitory activities. The results showed that several candidates including 7549, 2007_19424, 2007_15649 and 3519 have the obvious selective inhibitory effect on Lim1 rather than ROCK2. Considering the compound 7549 and 2007_19424 shared the same chemical structure, 7549, 2007_15649 and 3519 were finally selected as candidates for further analysis and their chemical scaffolds are displayed in Figure 8.

2D diagram of the chosen three candidates complexed with Limk1 in docking results were presented in Figure 9 and analysis of the docking poses revealed that all of the

TCM candidates formed H-bonding interactions with residues on Limk1 (Figure 10). The amide group of 7549 and the α, β-unsaturated carboxylate moiety of 2007_15649 both formed hydrogen bond interaction with GLY351 of Limk1 (2.3 Å and 2.4 Å respectively). Moreover, the two α, β-unsaturated carboxylate moiety of 2007_15649 also interacted with LYS368 and LYS426 through two hydrogen bonds (2.8 Å and 1.7 Å respectively), and the 1,3-dioxolo group of 3519 formed a 1.8 Å hydrogen bond with the LYS347 of Limk1. Further, the 1-methylenedecahydroazulene moiety of 7549 and the 1,3-Dihydroisobenzofuran moiety of 3519 generated hydrophobic interactions with its surrounding residues under the P-loop. Last but not least, Except for hydrogen bonds, each ligand engaged various potential interactions with key residues, including van der Waals, salt bridge, pi-interaction, and so on, which may improve the binding affinity with Limk1. Therefore, these three TCM candidates were used for further molecular dynamics simulation.

3.3 Molecular Dynamics Analysis.

The optimal conformation of the receptor-ligand complexes from the molecular docking were used as the initial structure for further MD simulations and all candidates' MD simulations were performed in 100 ns with Gromacs 2019 software to validate their stability by monitoring the structural variation over the simulations. We calculated the total energy of the receptor-ligand complexes in the 100_–ns procedure, analyzed the energy changes during the whole simulation process, and the results showed that the total energy of the system were stable and it had been maintained about -1 040 000 kJ/mol to -1 025 000 kJ/mol (Figure 11a).

The RMSD values (include complexes, protein and ligands) were calculated to evaluate the deviation of the structure from the original starting structure over the course of the simulations and also were an important basis for determining whether the system was stable. We plotted the RMSD changes of each complex with reference to the corresponding first frame in the simulation trajectory (shown in Figure 11b-d). The fluctuations in the complex RMSD plots of Limk1-candidates showed an upward trend at initial 20 ns, then tended to stabilize with a relatively flat curve and maintained around 0.25-0.5 nm. The protein RMSD profiles of Limk1 receptor resembled the RMSD change curve of the complex, suggesting an intrinsic stability of the protein structure in the complexes. Moreover, based on ligand RMSD, 2007_15649 had higher fluctuation rate than the other two ligands, which may explain high ligand gyration value. The RMSD analysis preliminary indicated that candidates 7549 and 3519 have higher binding stability to Limk1 receptor.

We also computed the radius of gyration for all of the three complexes over the simulations (Figure 12a and 12b). The radius of gyration could give a measure of the compactness of the structures, and could also give a measure of the atomic mass relative the molecular center mass. The smaller the value, the smaller the rotation change, indicating that the structures of the complexes are more closely bound during the simulations. It can be seen from the curve in the figures that protein gyrate and ligands gyrate were stable in general during the MD simulation process, and the gyrate of 2007_15649 was higher than other two ligands (include target protein system), which was consistent with the RMSD results. MSD revealed the movement of atoms from the

initial position to the final stage of MD simulations, which showed the movement trend of each ligand or protein. The low and stable MSD value of the ligand indicates the stability of the binding and the decrease of MSD value indicates that the ligand may be close to the binding pocket, while the high value of MSD and the increasing value of MSD indicate that the ligand tends to escape. As shown in Figure 12c and 12d, in the whole process of simulations, although the MSD values of Limk1 and three ligands both increased, the changes of all measured MSD values maintained in a low range. It was noteworthy that 2007_15649 obtained much higher ligand MSD value compared to other two ligands, which indicates the ligand 2007_15649 have a trend of escaping from the binding pocket. From the SASA calculated results, we can analyze the hydrophilicity and hydrophobicity of the simulation system. The solvent accessible area of Limk1 receptor decreased slowly during the period of 0-10 ns, and then remained to a relatively stable area (Figure 12e). Meanwhile, it could also be seen from Figure 12f, the SASA values of all ligands were very stable in the simulation process. This can further indicate the stability of the complex systems.

The RMSF of protein amino acids was used to analyze the fluctuation degree of single amino acid in protein during simulation. Some proteins may undergo obvious structural changes during the process of binding with ligands, even with "opening and closing" movements and through the calculation of RMSF, we can find the amino acids with drastic changes in protein structure. From Figure 13, we could observe the RMSF value of each protein residue for the three protein-ligand complexes and the obvious structural changes only occurred at the C-terminal (residue: 330-334), N-terminal

(residue: 663-667) and middle loop region (residue: 485-506) of the protein structure during the MD simulations, while the key binding pocket region showed good stability due to the low RMSF value here. Finally, the binding postures of three ligands and target protein in the initial and final conformations were displayed in Figure 14. The ligands rotated in the same pocket of Limk1 protein through binding the key residues and not detached from the target protein throughout the MD simulations.

4. Conclusions

Since the Limk1 is a promising drug target and few inhibitors with good Limk1/ROCK2 selectivity have been reported, discovering potential and selective Limk1 inhibitors with novel scaffolds is becoming an urgent need to develop new treatments for the related diseases. In this work, molecular docking technologies were utilized to identify novel potential lead compounds of Limk1 protein from TCM database and the screened compounds with top 20 docking score were further predicted the Limk1 and ROCK2 activities using the 3DGCN model. Meanwhile, we compared and discussed the performance of the 3DGCN and baseline models (RF, GCN and Weave), and the results showed that 3DGCN have achieved excellent accuracies on both Limk1 and ROCK2 datasets, owing to the property that can capture the atomic feature and 3D bond information from molecular graph directly. Through integrating docking results and the predicted values by the 3DGCN, we then found that 7549, 2007_15649 and 3519 could be the novel potential and selective inhibitors for Limk1 receptor rather than ROCK2. Furthermore, we also evaluated our results by MD simulations in 100 ns and the MD analysis displayed that all these ligand-receptor

complexes exhibited stable conditions during the entire simulation time. Finally, we provided a strategy for filtering large databases through combining molecular docking and 3DGCN model to discover selective Lmk1 inhibitors, and this also could be extended to other drug discovery domains.

Author contributions

Calvin Yu-Chian Chen designed research. Weihe Zhong worked together to completed the experiment and analyzed the data. Calvin Yu-Chian Chen contributed to analytic tools. Weihe Zhong, Lu Zhao and Calvin Yu-Chian Chen wrote the manuscript together.

Acknowledgements

This work was supported by Guangzhou Science and Technology fund (Grant No. 201803010072), Science, Technology & Innovation Commission of Shenzhen Municipality (JCYL 20170818165305521) and China Medical University Hospital (DMR-110-097). We also acknowledge the start-up funding from SYSU ‘‘Hundred Talent Program’’.

Declarations

Availability of data The datasets generated or analysed during the current study are available from the corresponding author on reasonable request.

Code availability The calculations have been carried out using Discovery Studio 2017 R2 Client and Gromacs 2019.

Ethics approval and consent to participate The manuscript is prepared in compliance with the Ethics in Publishing Policy as described in the Guide for Authors. The

manuscript is approved by all authors for publication.

Consent for publication All authors have agree with final submission and publication.

Conflicts of interest The author reports no conflicts of interest in this work.

References

1 Nosengo N, New tricks for old drugs, *Nature.* **2016**; 534: 314-316.
2 Kantarjian HM, Prat F, Steensma DP, Kurzrock R, Stewart DJ, Sekeres MA, Leveque J, Cancer Research in the United States: A Critical Review of Current Status and Proposal for Alternative Models, *Cancer.* **2018**; 124: 2881-2889.
3 Lionta E, Spyrou G, Vassilatis DK, Cournia Z, Structure-Based Virtual Screening for Drug Discovery: Principles, Applications and Recent Advances, *Curr Top Med Chem.* **2014**; 14: 1923-1938.
4 Chen YC, Beware of Docking! (vol 36, pg 78, 2015), *Trends Pharmacol Sci.* **2015**; 36: 617-617.
5 Kitchen DB, Decornez H, Furr JR, Bajorath J, Docking and scoring in virtual screening for drug discovery: methods and applications, *Nat Rev Drug Discov.* **2004**; 3: 935-949.
6 Pérez-Sianes J, Pérez-Sánchez H, Díaz F, Virtual screening meets deep learning, *Current computer-aided drug design.* **2019**; 15: 6-28.
7 LeCun Y, Bengio Y, Hinton G, Deep learning, *Nature.* **2015**; 521: 436-444.
8 Rawat W, Wang ZH, Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review, *Neural Comput.* **2017**; 29: 2352-2449.
9 Young T, Hazarika D, Poria S, Cambria E, Recent Trends in Deep Learning Based Natural Language Processing, *Ieee Comput Intell M.* **2018**; 13: 55-75.
10 Papadatos G, Gaulton A, Hersey A, Overington JP, Activity, assay and target data curation and quality in the ChEMBL database, *J Comput Aid Mol Des.* **2015**; 29: 885-896.
11 Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, Han LY, He JE, He SQ, Shoemaker BA, Wang JY, Yu B, Zhang J, Bryant SH, PubChem Substance and Compound databases, *Nucleic Acids Res.* **2016**; 44: D1202-D1213.
12 Ma JS, Sheridan RP, Liaw A, Dahl GE, Svetnik V, Deep Neural Nets as a Method for Quantitative Structure-Activity Relationships, *J Chem Inf Model.* **2015**; 55: 263-274.
13 Mayr A, Klambauer G, Unterthiner T, Hochreiter S, DeepTox: Toxicity Prediction using Deep Learning, *Front Env Sci-Switz.* **2016**; 3.
14 Ramsundar B, Liu BW, Wu ZQ, Verras A, Tudor M, Sheridan RP, Pande V, Is Multitask Deep Learning Practical for Pharma?, *J Chem Inf Model.* **2017**; 57: 2068-2076.
15 Lavecchia A, Deep learning in drug discovery: opportunities, challenges and future prospects, *Drug Discov Today.* **2019**; 24: 2017-2032.
16 Feinberg EN, Sur D, Wu ZQ, Husic BE, Mai HH, Li Y, Sun SS, Yang JY, Ramsundar

B, Pande VS, PotentialNet for Molecular Property Prediction, *ACS central science.* **2018**; 4: 1520-1530.
17 Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B, Support vector machines, *IEEE Intelligent Systems and their applications.* **1998**; 13: 18-28.
18 Ho TK, Random decision forests, Proceedings of 3rd international conference on document analysis and recognition, IEEE, **1995**, pp. 278-282.
19 Scarselli F, Gori M, Tsoi AC, Hagenbuchner M, Monfardini G, The Graph Neural Network Model, *Ieee T Neural Networ.* **2009**; 20: 61-80.
20 Huang WB, Zhang T, Rong Y, Huang JZ, Adaptive Sampling Towards Fast Graph Representation Learning, *Advances in Neural Information Processing Systems 31 (Nips 2018).* **2018**; 31.
21 Hamilton WL, Bajaj P, Zitnik M, Jurafsky D, Leskovec J, Embedding Logical Queries on Knowledge Graphs, *Advances in Neural Information Processing Systems 31 (Nips 2018).* **2018**; 31.
22 Monti F, Bronstein MM, Bresson X, Geometric Matrix Completion with Recurrent Multi-Graph Neural Networks, *Advances in Neural Information Processing Systems 30 (Nips 2017).* **2017**; 30.
23 Weininger D, SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules, *Journal of chemical information and computer sciences.* **1988**; 28: 31-36.
24 Duvenaudt D, Maclaurin D, Aguilera-Iparraguirre J, Gomez-Bombarelli R, Hirzel T, Aspuru-Guzik A, Adams RP, Convolutional Networks on Graphs for Learning Molecular Fingerprints, *Adv Neur In.* **2015**; 28.
25 Kearnes S, McCloskey K, Berndl M, Pande V, Riley P, Molecular graph convolutions: moving beyond fingerprints, *J Comput Aid Mol Des.* **2016**; 30: 595-608.
26 Gilmer J, Schoenholz SS, Riley PF, Vinyals O, Dahl GE, Neural message passing for quantum chemistry, International Conference on Machine Learning, PMLR, **2017**, pp. 1263-1272.
27 Wu ZQ, Ramsundar B, Feinberg EN, Gomes J, Geniesse C, Pappu AS, Leswing K, Pande V, MoleculeNet: a benchmark for molecular machine learning, *Chem Sci.* **2018**; 9: 513-530.
28 Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I, Attention is all you need, *arXiv preprint arXiv:1706.03762.* **2017**.
29 Withnall M, Lindelöf E, Engkvist O, Chen H, Building attention and edge message passing neural networks for bioactivity and physical–chemical property prediction, *J Cheminformatics.* **2020**; 12: 1-18.
30 Tang B, Kramer ST, Fang M, Qiu Y, Wu Z, Xu D, A self-attention based message passing neural network for predicting molecular lipophilicity and aqueous solubility, *J Cheminformatics.* **2020**; 12: 1-9.
31 Li P, Li Y, Hsieh C-Y, Zhang S, Liu X, Liu H, Song S, Yao X, TrimNet: learning molecular representation from triplet messages for biomedicine, *Briefings in Bioinformatics.* **2020**.
32 Xiong Z, Wang D, Liu X, Zhong F, Wan X, Li X, Li Z, Luo X, Chen K, Jiang H, Pushing the boundaries of molecular representation for drug discovery with the graph

attention mechanism, *J Med Chem.* **2019**; 63: 8749-8760.
33 Cho H, Choi IS, Enhanced Deep‐Learning Prediction of Molecular Properties via Augmentation of Bond Topology, *ChemMedChem.* **2019**; 14: 1604-1609.
34 Charles MD, Brookfield JL, Ekwuru TC, Stockley M, Dunn J, Riddick M, Hammonds T, Trivier E, Greenland G, Wong AC, Cheasty A, Boyd S, Crighton D, Olson MF, Discovery, Development, and SAR of Aminothiazoles as LIMK Inhibitors with Cellular Anti-Invasive Properties, *J Med Chem.* **2015**; 58: 8309-8313.
35 Yin Y, Zheng K, Eid N, Howard S, Jeong JH, Yi F, Guo J, Park CM, Bibian M, Wu WL, Hernandez P, Park H, Wu YT, Luo JL, LoGrasso PV, Feng YB, Bis-aryl Urea Derivatives as Potent and Selective LIM Kinase (Limk) Inhibitors, *J Med Chem.* **2015**; 58: 1846-1861.
36 Henderson BW, Greathouse KM, Ramdas R, Walker CK, Rao TC, Bach SV, Curtis KA, Day JJ, Mattheyses AL, Herskowitz JH, Pharmacologic inhibition of LIMK1 provides dendritic spine resilience against beta-amyloid, *Sci Signal.* **2019**; 12.
37 Saal KA, Koch JC, Tatenhorst L, Szego EM, Ribas VT, Michel U, Bahr M, Tonges L, Lingor P, AAV.shRNA-mediated downregulation of ROCK2 attenuates degeneration of dopaminergic neurons in toxin-induced models of Parkinson's disease in vitro and in vivo, *Neurobiol Dis.* **2015**; 73: 150-162.
38 Xia H, Sun SJ, Wang B, Wang T, Liang CY, Li G, Huang CB, Qi DL, Chu XY, miR-143 Inhibits NSCLC Cell Growth and Metastasis by Targeting Limk1, *Int J Mol Sci.* **2014**; 15: 11973-11983.
39 Chen PX, Zeng MJ, Zhao Y, Fang XL, Upregulation of Limk1 caused by microRNA-138 loss aggravates the metastasis of ovarian cancer by activation of Limk1/cofilin signaling, *Oncol Rep.* **2014**; 32: 2070-2076.
40 Wang J, He ZW, Xu J, Chen P, Jiang JX, Long noncoding RNA LINC00941 promotes pancreatic cancer progression by competitively binding miR-335-5p to regulate ROCK1-mediated LIMK1/Cofilin-1 signaling, *Cell death & disease.* **2021**; 12.
41 Li ZF, Yao YD, Zhao YY, Liu Y, Liu ZH, Hu P, Zhu ZR, Effects of PAK4/LIMK1/Cofilin-1 signaling pathway on proliferation, invasion, and migration of human osteosarcoma cells, *J Clin Lab Anal.* **2020**; 34.
42 Kast R, Schirok H, Figueroa-Perez S, Mittendorf J, Gnoth MJ, Apeler H, Lenz J, Franz JK, Knorr A, Hutter J, Lobell M, Zimmermann K, Munter K, Augstein KH, Ehmke H, Stasch JP, Cardiovascular effects of a novel potent and highly selective azaindole-based inhibitor of Rho-kinase, *Brit J Pharmacol.* **2007**; 152: 1070-1080.
43 Boland S, Bourin A, Alen J, Geraets J, Schroeders P, Castermans K, Kindt N, Boumans N, Panitti L, Fransen S, Vanormelingen J, Stassen JM, Leysen D, Defert O, Design, Synthesis, and Biological Evaluation of Novel, Highly Active Soft ROCK Inhibitors, *J Med Chem.* **2015**; 58: 4309-4324.
44 Burley SK, Berman HM, Christie C, Duarte JM, Feng Z, Westbrook J, Young J, Zardecki C, RCSB Protein Data Bank: Sustaining a living digital data resource that enables breakthroughs in scientific research and biomedical education, *Protein Science.* **2018**; 27: 316-330.
45 Chen CYC, TCM Database@Taiwan: The World's Largest Traditional Chinese Medicine Database for Drug Screening In Silico, *Plos One.* **2011**; 6.

46 Sun M, Zhao S, Gilvary C, Elemento O, Zhou J, Wang F, Graph convolutional networks for computational drug development and discovery, *Briefings in bioinformatics.* **2020**; 21: 919-935.
47 Landrum G, RDKit: Open-source cheminformatics from machine learning to chemical registration, *Abstr Pap Am Chem S.* **2019**; 258.
48 Kipf TN, Welling M, Semi-supervised classification with graph convolutional networks, *arXiv preprint arXiv:1609.02907.* **2016**.
49 Tosco P, Stiefl N, Landrum G, Bringing the MMFF force field to the RDKit: implementation and validation, *J Cheminformatics.* **2014**; 6.
50 Chollet F. Keras. https://keras.io. (Accessed Feb 12, 2021).
51 Abadi MA, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G. S.; Davis, A.; Dean, J.; Zheng, X. . TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. https://www.tensorflow.org/ (Accessed Feb 12, 2021).
52 Yao Y, Rosasco L, Caponnetto A, On early stopping in gradient descent learning, *Constr Approx.* **2007**; 26: 289-315.
53 He L, Seitz SP, Trainor GL, Tortolani D, Vaccaro W, Poss M, Tarby CM, Tokarski JS, Penhallow B, Hung CY, Attar R, Lin TA, Modulation of cofilin phosphorylation by inhibition of the Lim family kinases, *Bioorg Med Chem Lett.* **2012**; 22: 5995-5998.
54 Boland S, Bourin A, Alen J, Geraets J, Schroeders P, Castermans K, Kindt N, Boumans N, Panitti L, Vanormelingen J, Fransen S, Van de Velde S, Defert O, Design, synthesis and biological characterization of selective LIMK inhibitors, *Bioorg Med Chem Lett.* **2015**; 25: 4005-4010.
55 Yin Y, Lin L, Ruiz C, Khan S, Cameron MD, Grant W, Pocas J, Eid N, Park H, Schroter T, Lograsso PV, Feng Y, Synthesis and biological evaluation of urea derivatives as highly potent and selective rho kinase inhibitors, *J Med Chem.* **2013**; 56: 3568-3581.

Table 1. Description of Atom Features

feature	description	size
atom type	[C, O, N, S, Cl, F, Br, P, I, Si, B, Na, Sn, Se, other] (one-hot)	15
degree	the number of heavy atom neighbors [0, 1, 2, 3, 4, 5, 6] (one-hot)	7
hybridization	[sp, sp^2, sp^3, sp^3d, sp^3d^2] (one-hot)	5
valence	number of implicit valences [0, 1, 2, 3, 4, 5, 6] (one-hot)	7
formal charge	integer electronic charge [-3, -2, -1, 0, 1, 2, 3] (one-hot)	7
atom in ring of size	[3, 4, 5, 6, 7, 8] (binary)	6
aromatic	whether the atom is part of an aromatic systems [0, 1] (binary)	1
explicit hydrogen	the number of neighboring hydrogens [0, 1, 2, 3, 4] (one-hot)	5
chirality	[R, S, or nonchiral] (one-hot)	3
acid/base	Whether this atom is acidic or basic (binary)	2
hydrogen bonding	Whether this atom is a hydrogen bond donor or acceptor (binary)	2
		60

Table 2. Ten-fold cross-validation performances of various models for the Limk1 and ROCK2 datasets.

Dataset	Model	RMSE		MAE	
		Validation	Test	Validation	Test
Limk1	RF	0.997 ± 0.115	0.902 ± 0.078	0.734 ± 0.086	0.671 ± 0.076
	GCN	0.988 ± 0.083	0.930 ± 0.067	0.777 ± 0.074	0.732 ± 0.076
	Weave	0.815 ± 0.060	0.742 ± 0.065	0.649 ± 0.053	0.595 ± 0.061
	3DGCN$_{sum}$	0.709 ± 0.060	0.773 ± 0.076	0.576 ± 0.047	0.628 ± 0.050
	3DGCN$_{max}$	**0.631 ± 0.039**	**0.721 ± 0.056**	**0.504 ± 0.032**	**0.578 ± 0.045**
ROCK2	RF	1.102 ± 0.076	1.065 ± 0.068	0.850 ± 0.055	0.817 ± 0.059
	GCN	0.873 ± 0.030	0.927 ± 0.038	0.705 ± 0.034	0.762 ± 0.036
	Weave	0.966 ± 0.037	1.013 ± 0.029	0.794 ± 0.043	0.832 ± 0.026
	3DGCN$_{sum}$	0.805 ± 0.054	**0.839 ± 0.087**	0.659 ± 0.041	**0.685 ± 0.074**
	3DGCN$_{max}$	**0.802 ± 0.078**	0.852 ± 0.083	**0.653 ± 0.078**	0.686 ± 0.072

Table 3. Docking Score and Predicted Activity Value for the Top 20 TCM Candidates

Compound	Docking score	Predicted activity	
		Limk1	ROCK2
8797	144.503	5.621	6.896
2007_22325	141.452	5.493	6.595
2007_15317	137.164	6.027	7.357
2007_22057	127.319	6.611	6.900
210	126.113	5.546	6.229
2007_12264	118.568	6.165	5.960
7549	**115.768**	**6.704**	**5.870**
2007_19424	**115.768**	**6.704**	**5.870**
2007_8260	115.645	5.370	6.325
2007_8902	115.428	6.530	6.788
3274	114.745	5.525	6.244
2007_15649	**113.547**	**7.096**	**5.612**
2007_14881	113.544	6.623	6.825
2007_9349	113.476	5.215	5.723
8672	113.325	5.557	6.981
3519	**113.248**	**6.847**	**6.395**
8909	112.468	5.632	6.357
2007_9018	110.930	6.553	6.332
radicamine A	110.470	6.635	6.803
(3R)-3-hydroxy-12-[(1R,4R,5S)-4-hydroxy-5-methyl-piperidin-1-yl]-dodecanoic acid	110.419	5.377	6.016

Figures

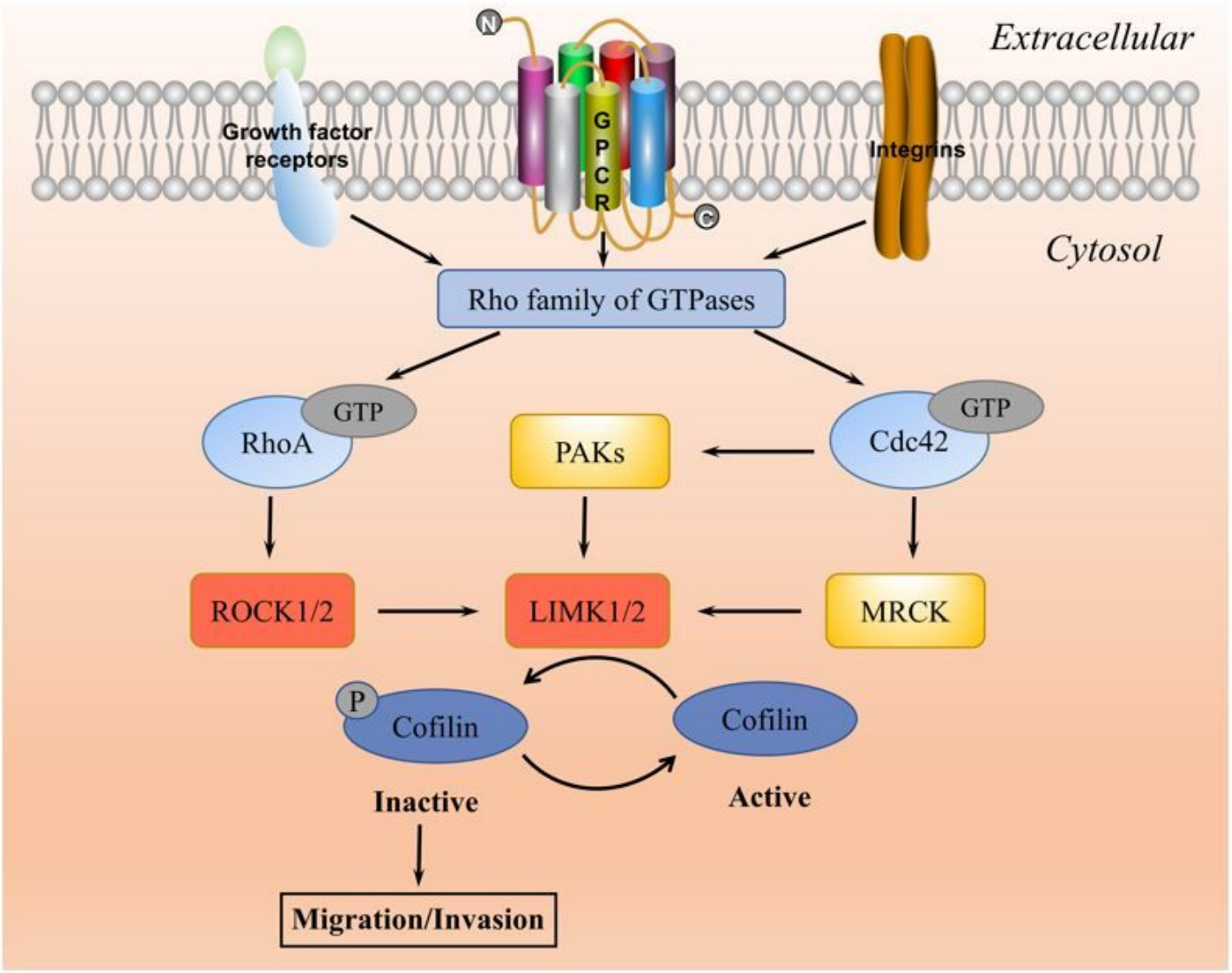


Figure 1

LIMKs are key regulators of the actin cytoskeleton through their modulation of cofilin function.

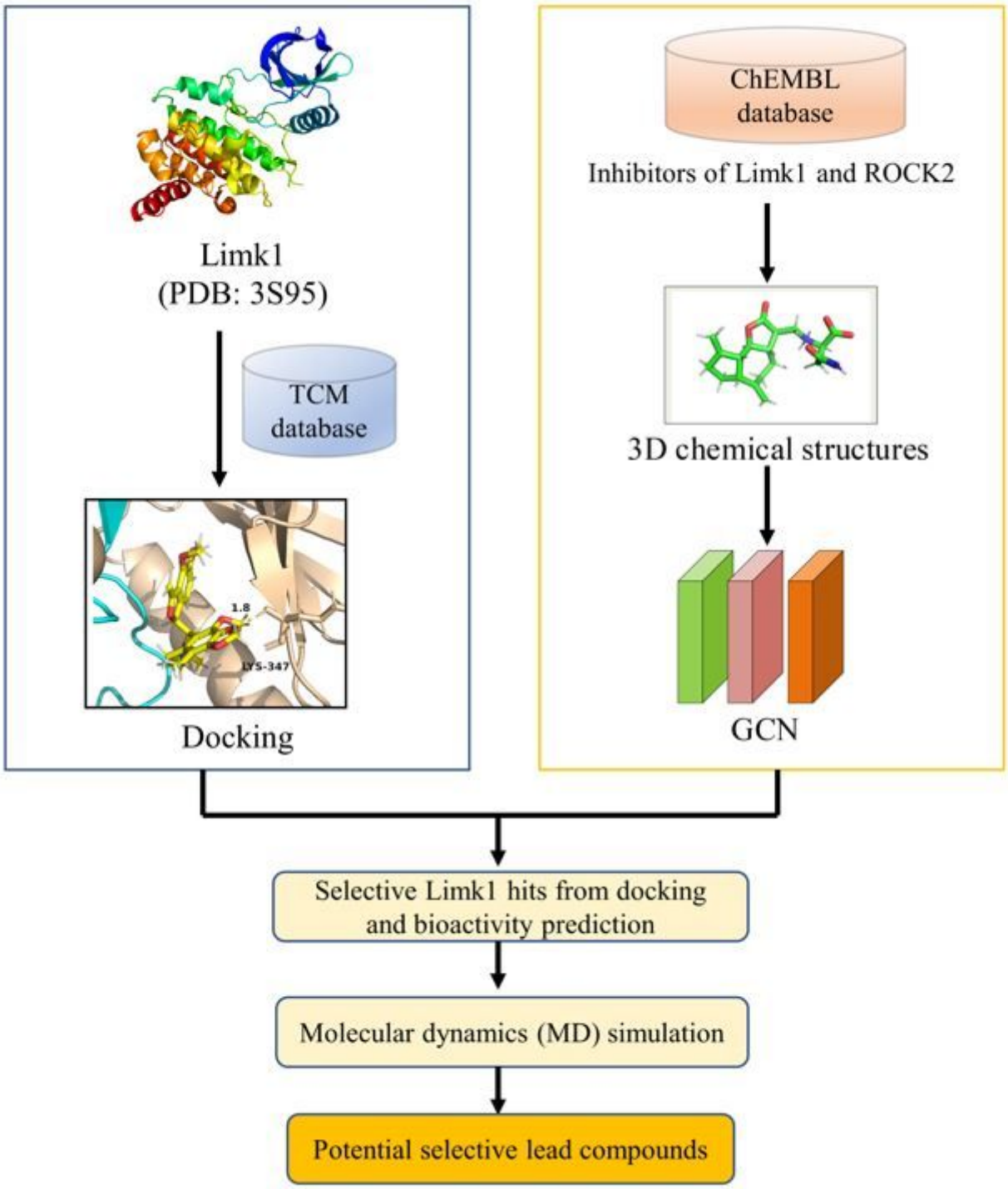


Figure 2

The flow chart of experimental design.

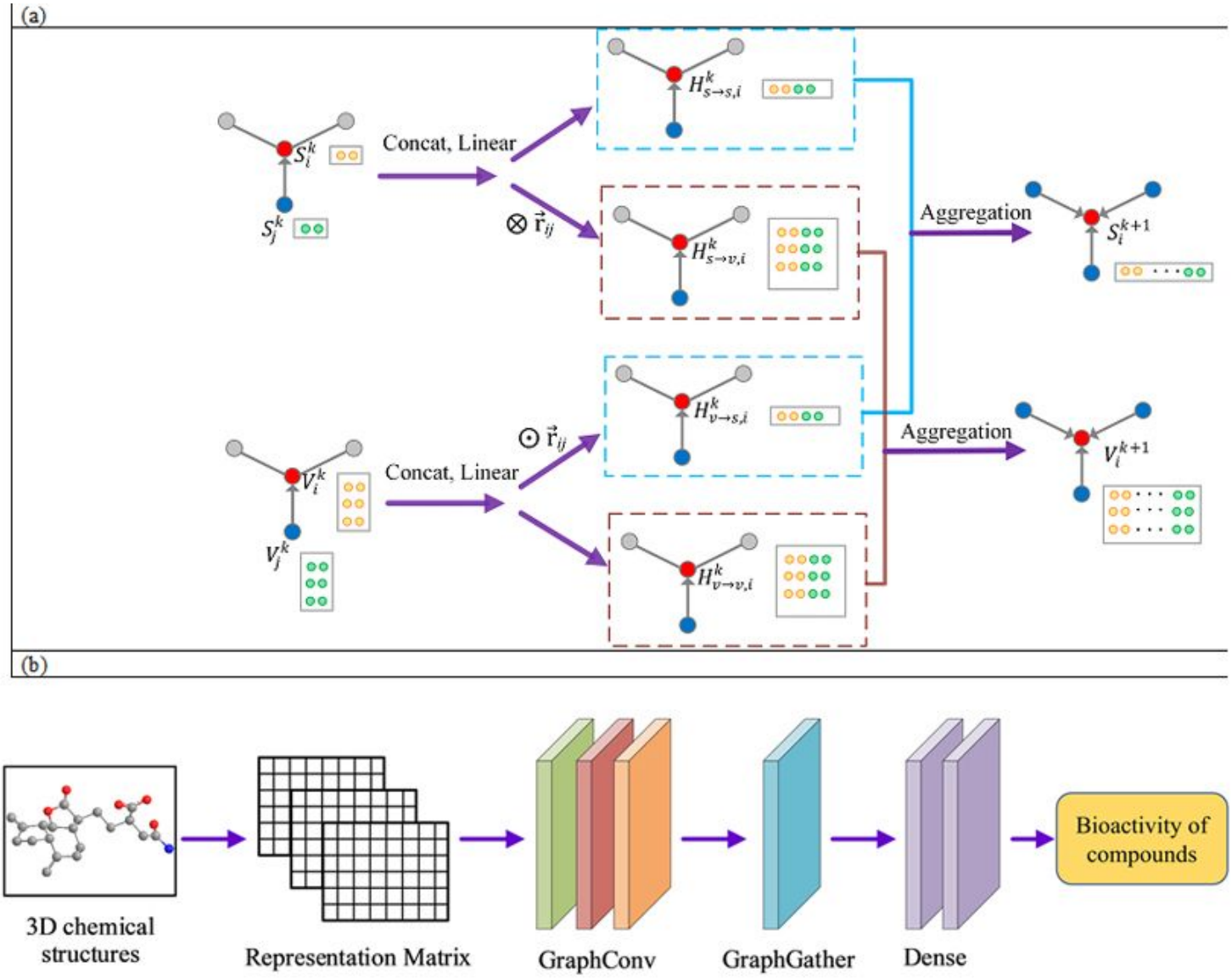


Figure 3

(a) illustration of handling 3D molecular scalar and vector features, (b) the overall architectures of 3DGCN

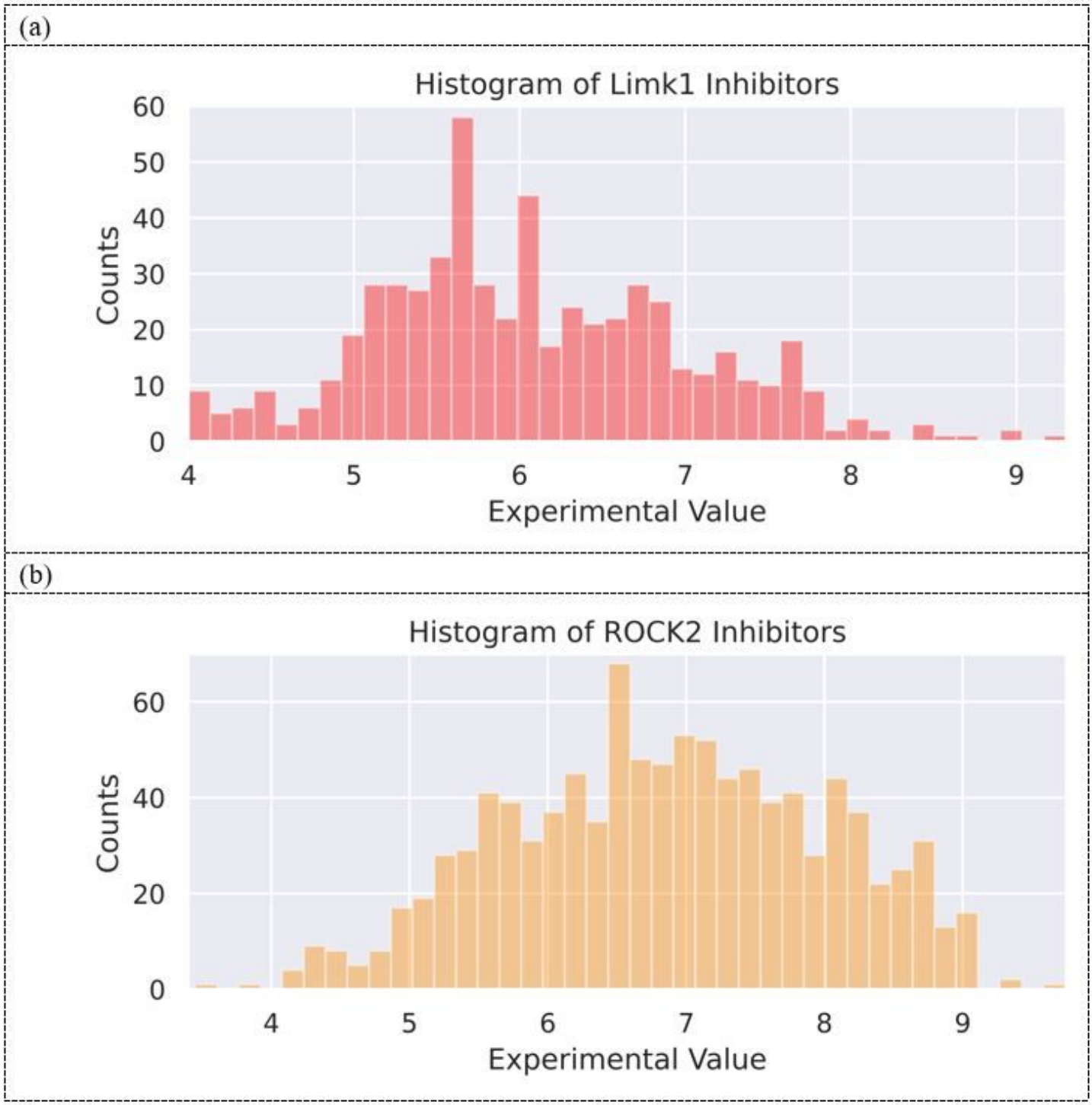
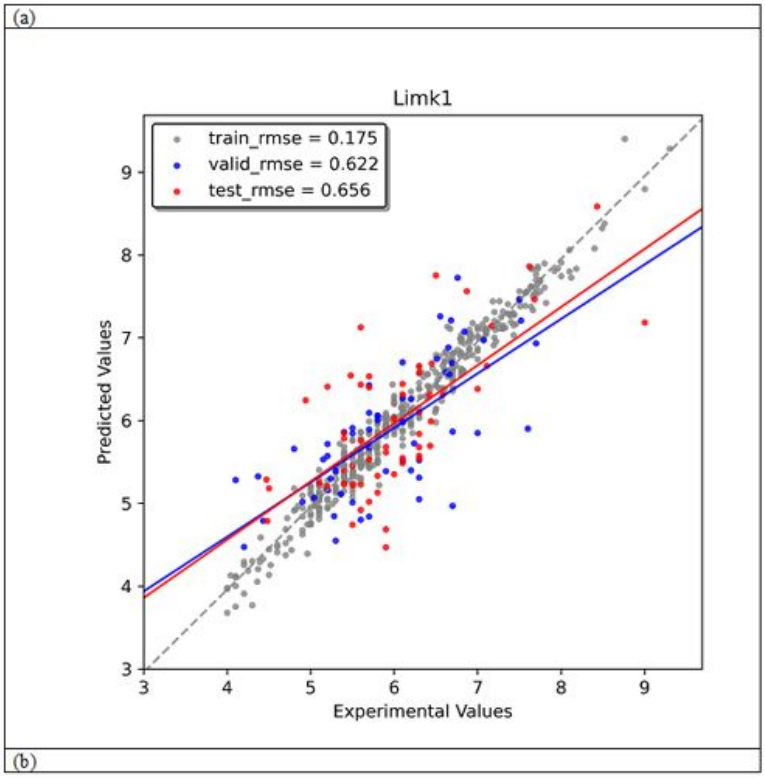


Figure 4

Data set used for building prediction models (a) Limk1, (b) ROCK2.



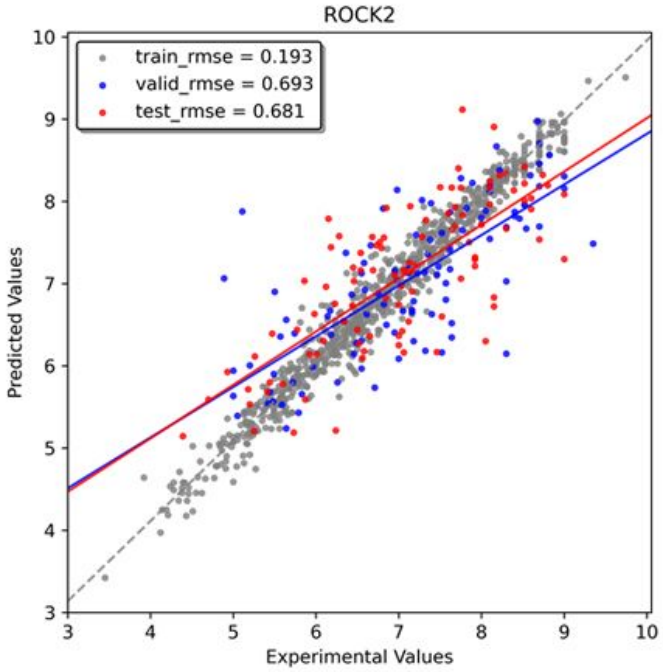


Figure 5

Prediction results of 3DGCNmax model.

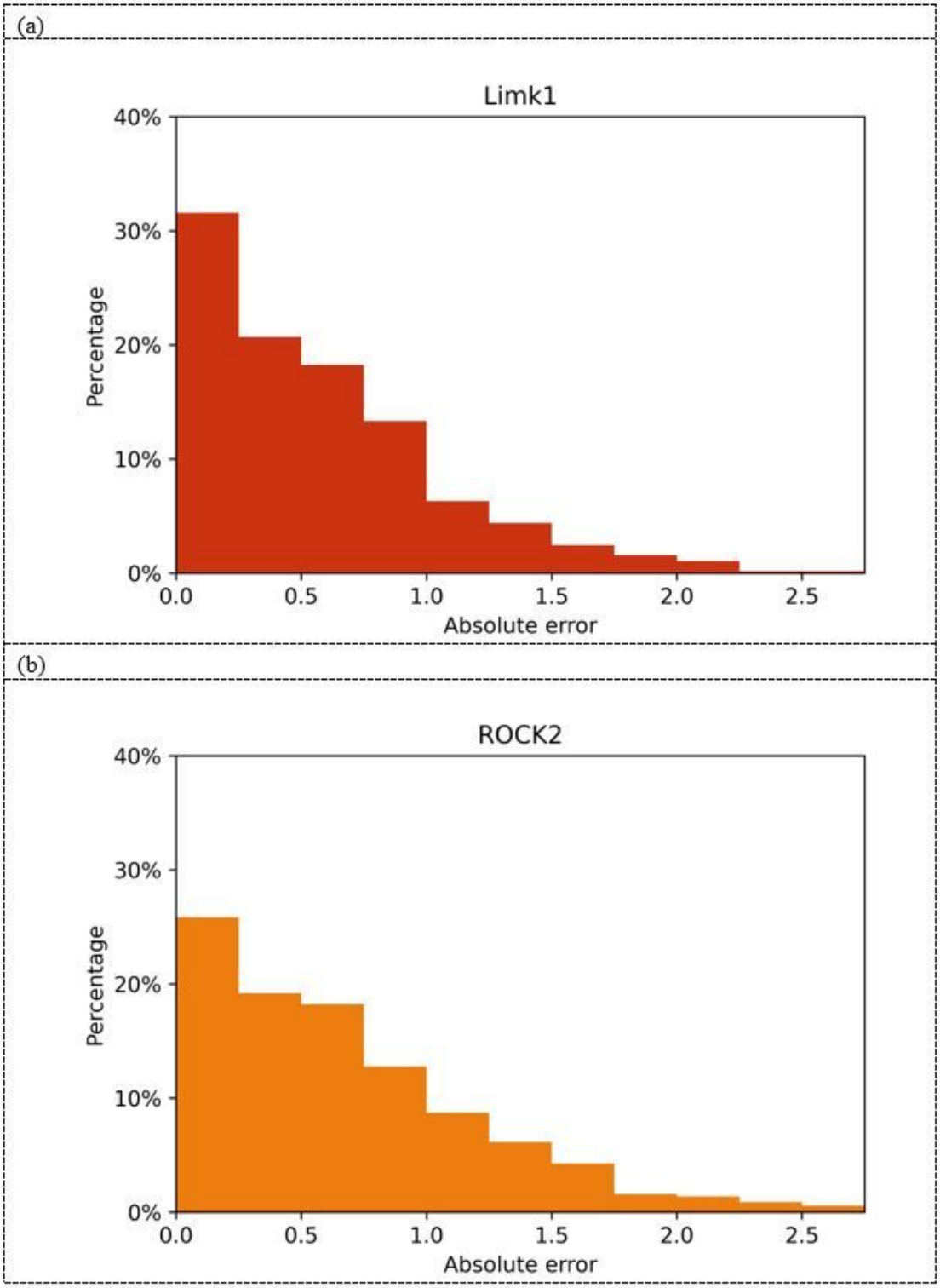


Figure 6

Histograms of absolute errors from the test sets for the Limk1 (a) and ROCK2 (b) datasets of 3DGCNmax.

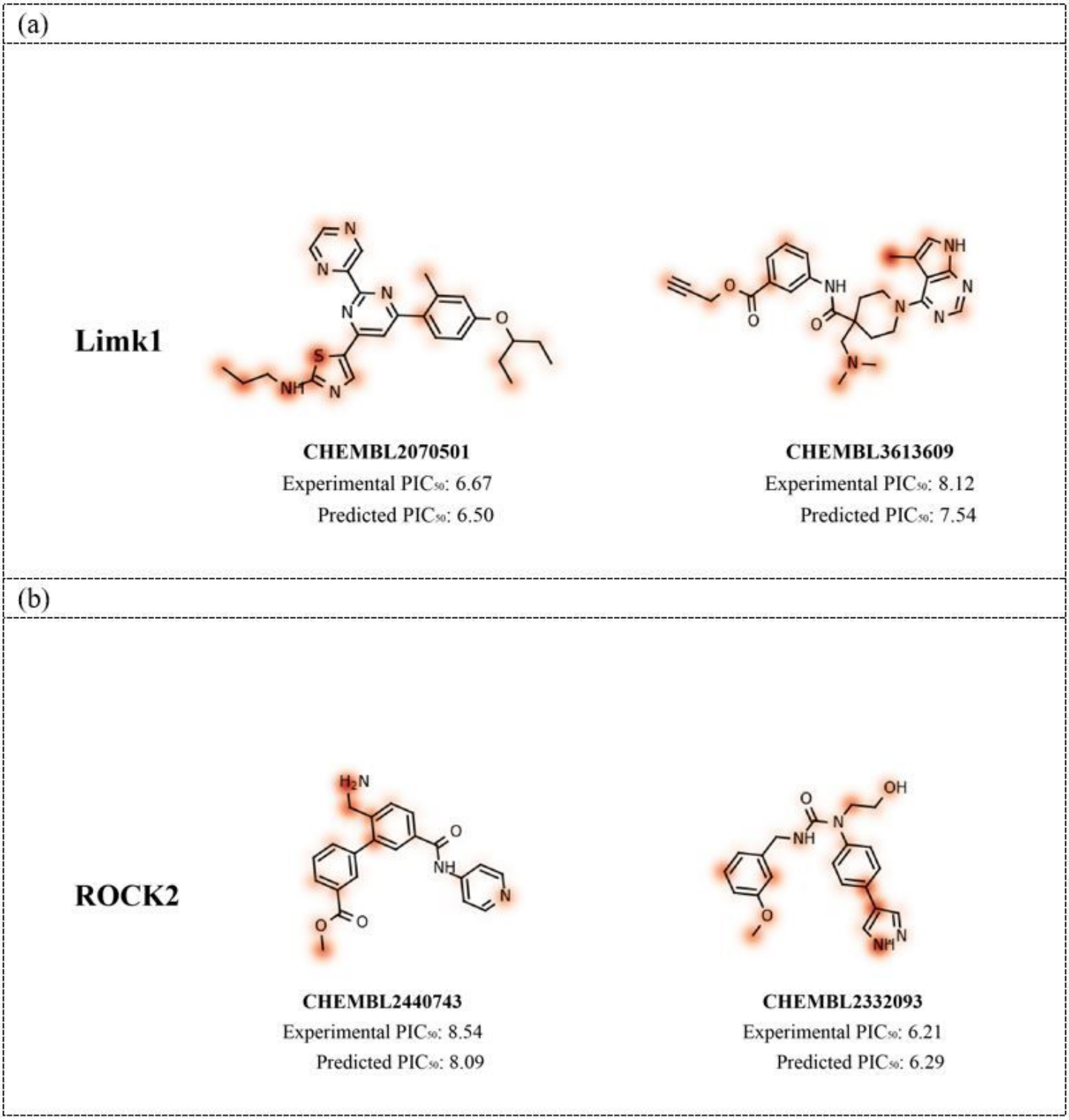


Figure 7

Visualization of atomic influence on the molecular activity prediction after training with (a) Limk1 and (b) ROCK2 dataset.

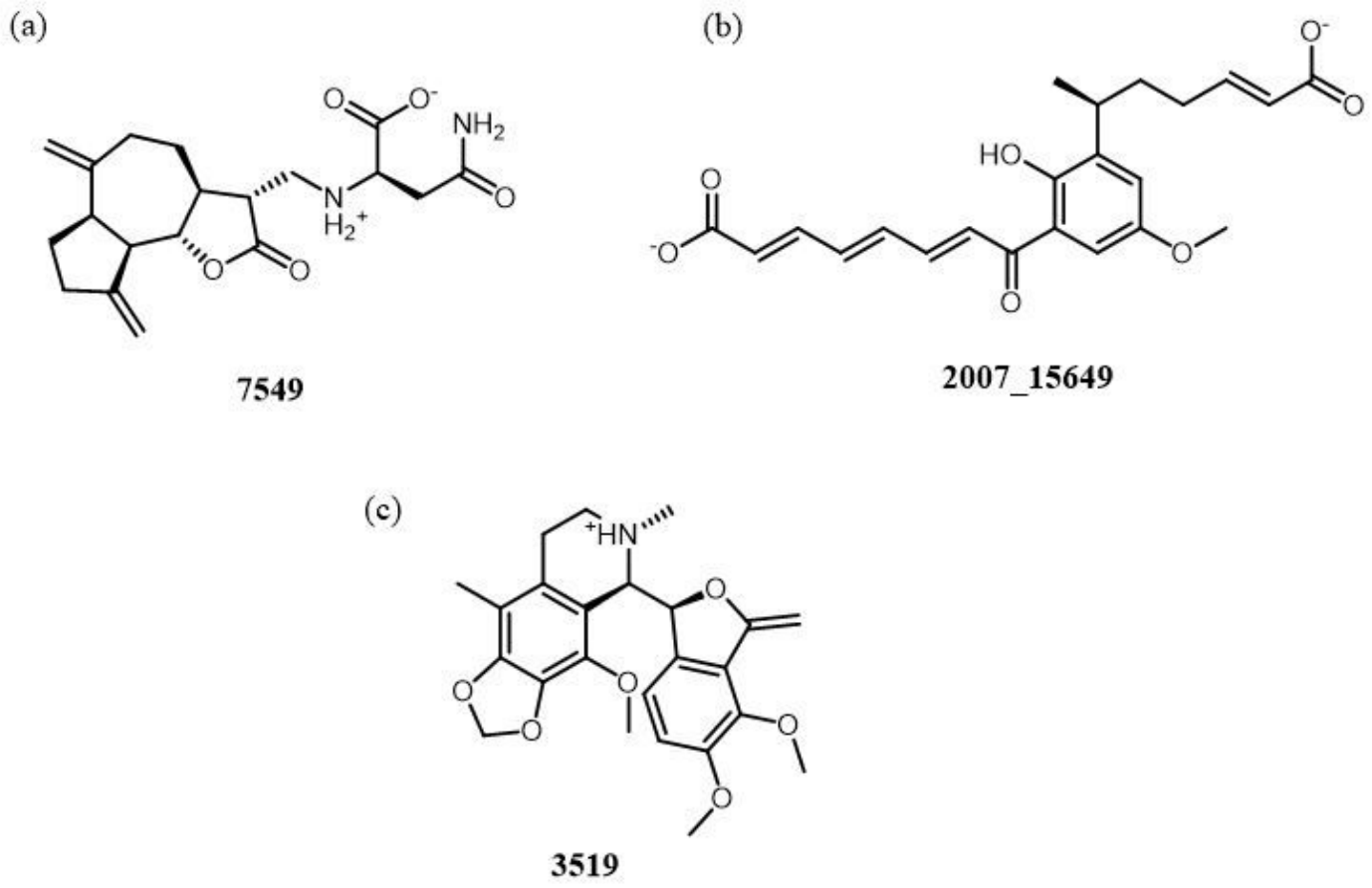


Figure 8

Chemical structures of (a) 7549, (b) 2007_15649 and (c) 3519.

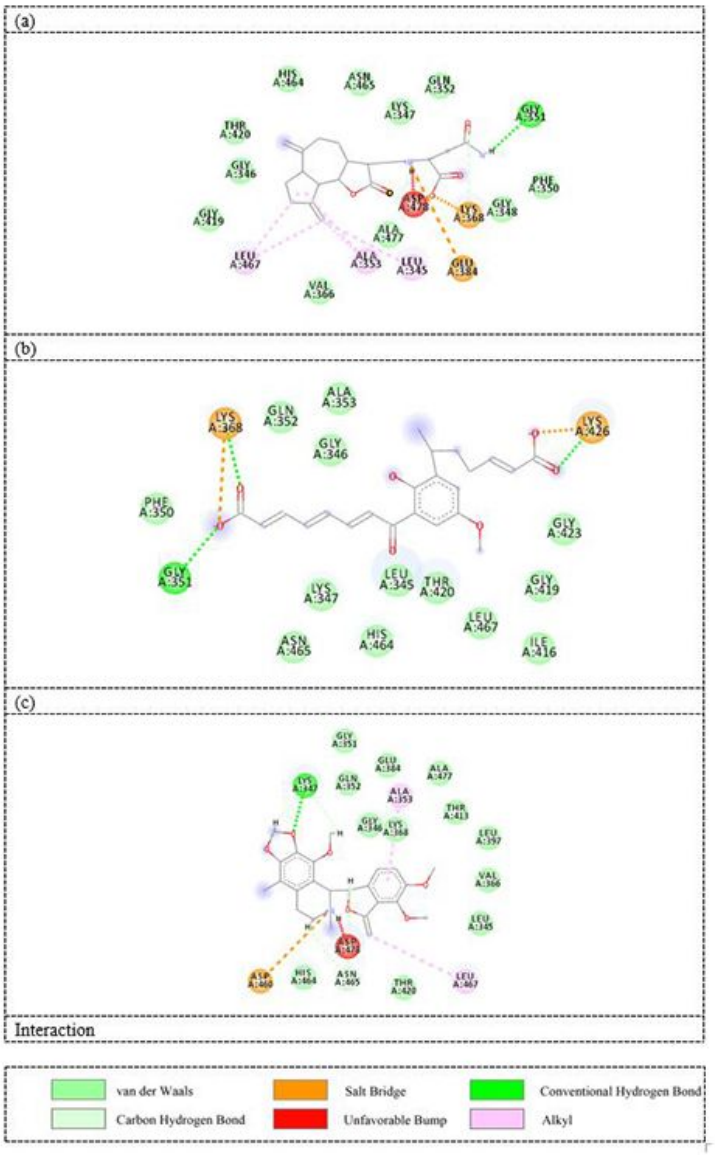


Figure 9

Two-Dimensional diagrams of three candidates' docking results. (a) 7549, (b) 2007_15649 and (c) 3519.

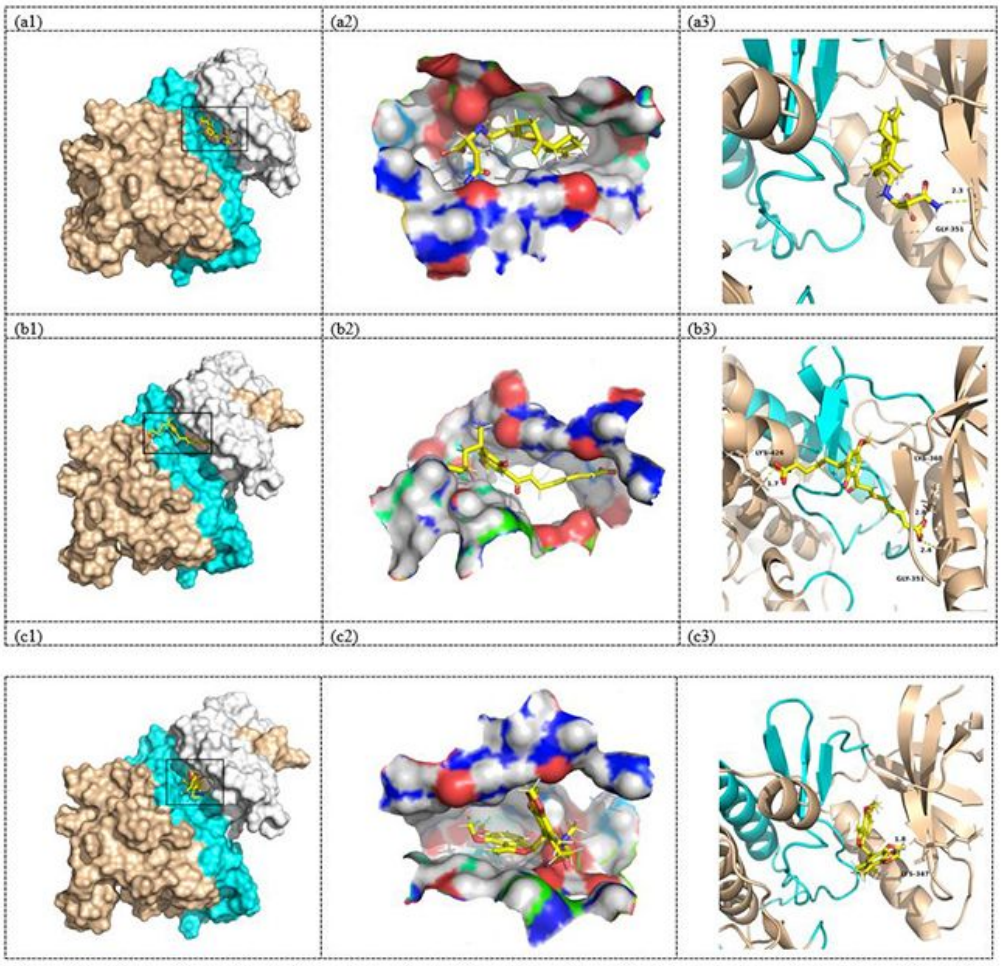


Figure 10

Docking pose of (a) 7549, (b) 2007_15649 and (c) 3519 with Limk1. In (3), the yellow dash lines stand for H-bonds.

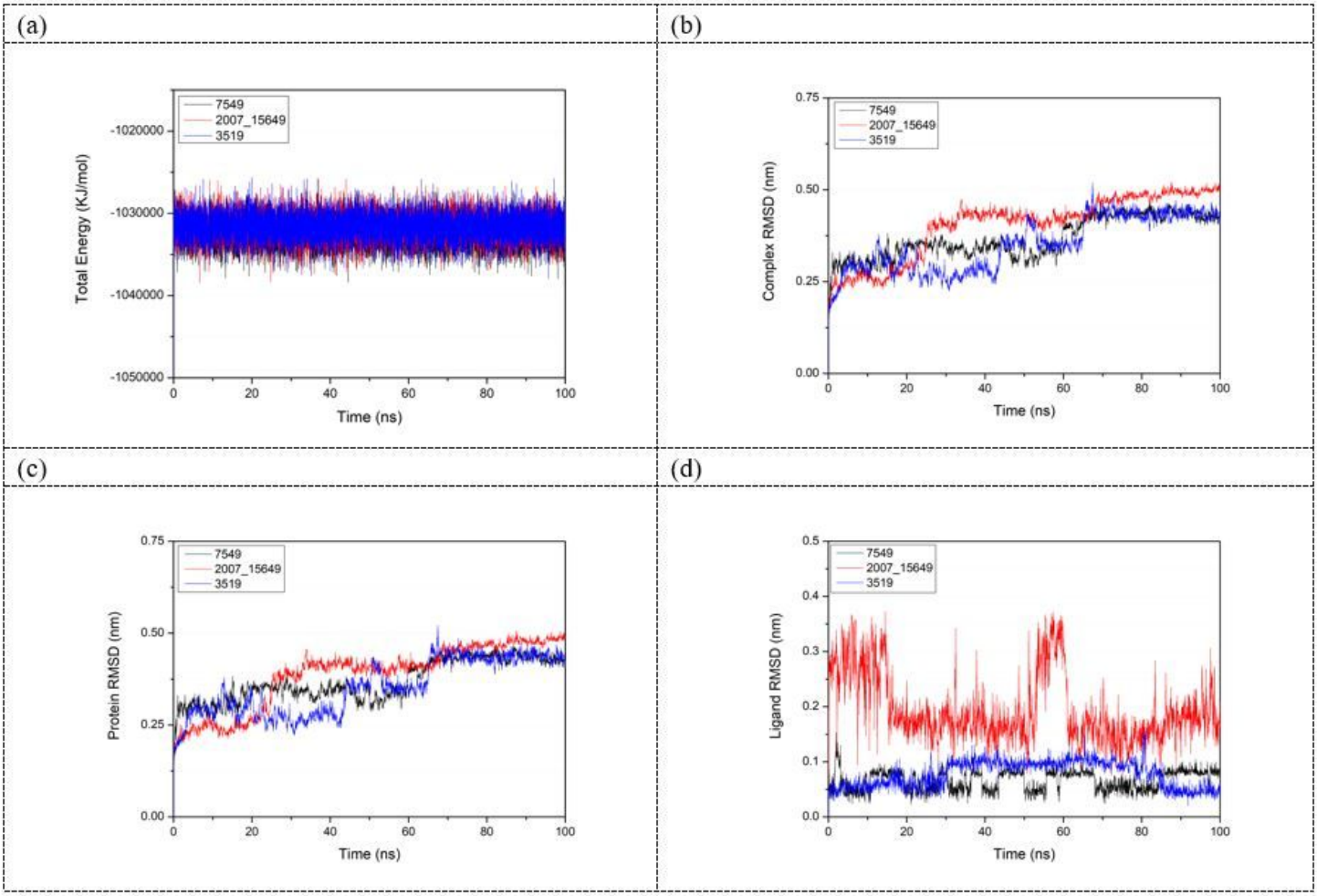


Figure 11

Changes of total energy and RMSD in molecular dynamics simulation. (a) Total energy changes between protein and ligands, (b) RMSD changes of complexs, (c) RMSD changes of protein, (d) RMSD changes of ligands.

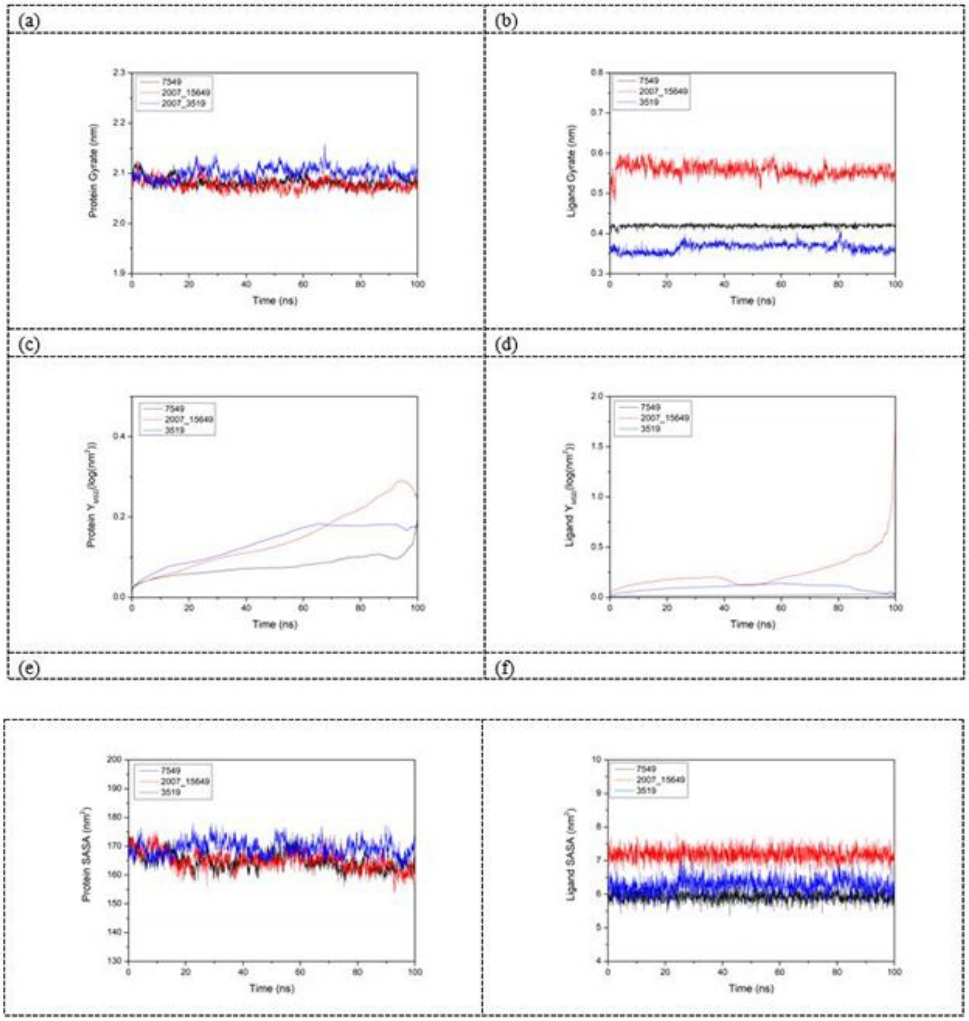


Figure 12

Gyrate, MSD, and SASA changes of proteins and ligands in molecular dynamics simulations. (a) Gyrate of protein, (b) gyrate of ligands, (c) MSD of protein, (d) MSD of ligands, (e) SASA of protein, (f) SASA of ligands.

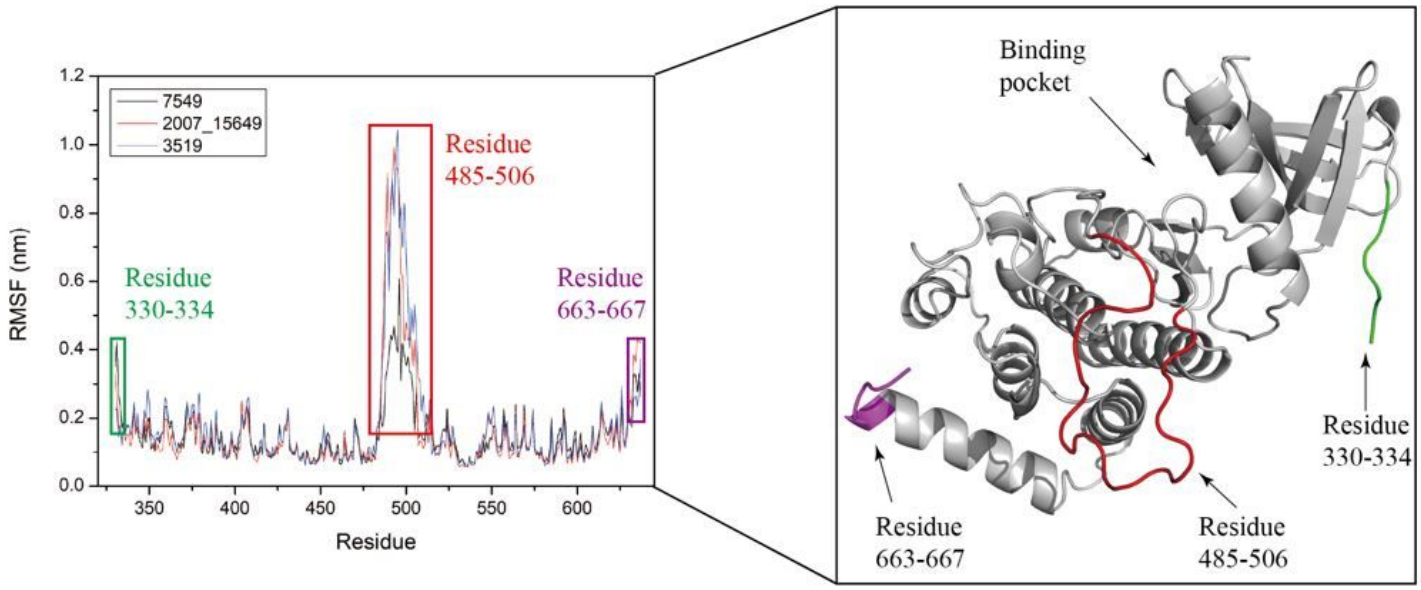


Figure 13

RMSF analysis of Limk1 protein with ligands in molecular dynamics simulations.

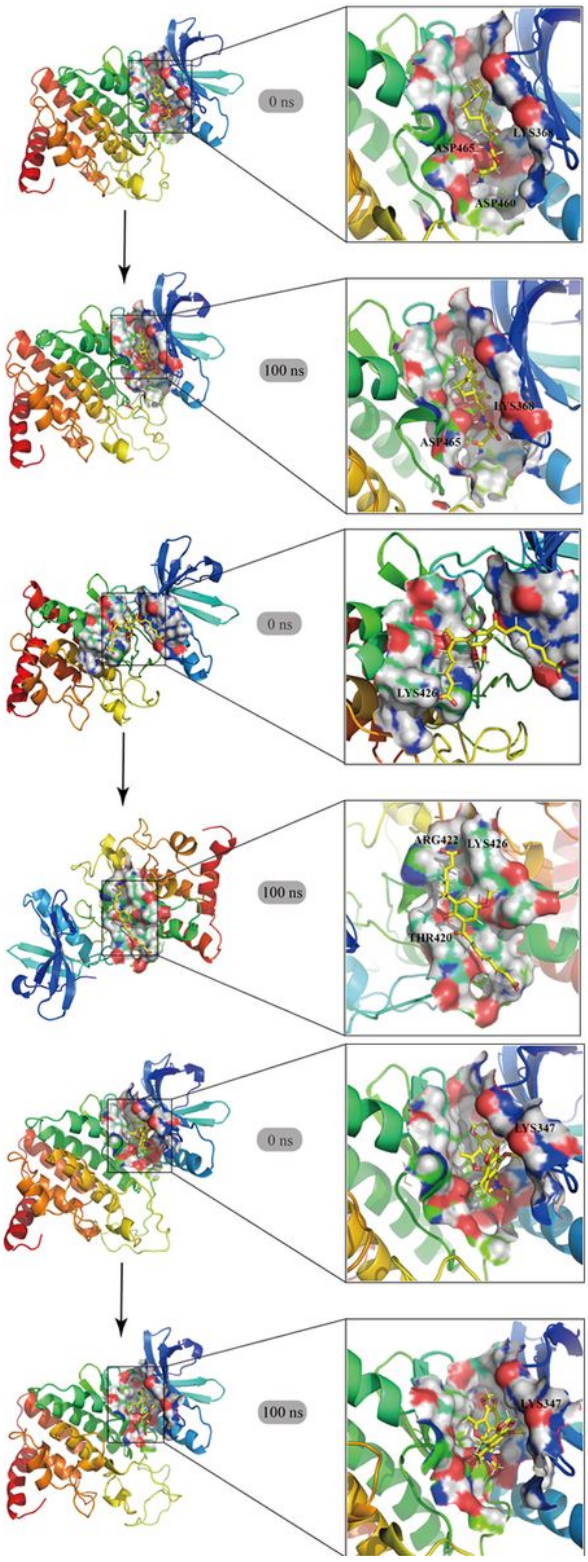


Figure 14

Changes in the initial and final binding posture during molecular dynamics simulations. (a) 7549, (b) 2007_15649, (c) 3519.