

Article

3D-QSAR, molecular docking, and molecular dynamics simulation study of thieno[3,2-b]pyrrole-5-carboxamide derivatives as LSD1 inhibitors

Yongtao Xu^{1,2,3,*}, Zihao He^{1,2,3}, Min Yang^{1,2,3}, Hongyi Liu^{1,2,3}, Xiaoyuan Lu¹, Songjie Zhang^{1,2,3} and Meilan Huang^{4,*}

¹ School of Biomedical Engineering, Xinxiang Medical University, Xinxiang, Henan, 453003, China

² Xinxiang key laboratory of biomedical information research

³ Henan engineering laboratory of combinatorial technique for clinical and biomedical big data

⁴ School of Chemistry and Chemical Engineering, Queen's University Belfast, David Keir Building, Stranmillis Road, Belfast, Northern Ireland, United Kingdom

* Correspondence: email: yxu@xxmu.edu.cn (Y.X.); Tel.: +0086 15560156065 (Y.X.); m.huang@qub.ac.uk (M.H.); Tel.: +0044 028 9097 4698 (M.H.)

Abstract: Histone Lysine Specific Demethylase 1 (LSD1) is overexpressed in many cancers and become a new target for anticancer drugs. In recent years, the small molecule inhibitors with various structures targeting LSD1 have been reported. Here we report the binding interaction modes of a series of thieno[3,2-b]pyrrole-5-carboxamides LSD1 inhibitors using molecular docking, three dimensional quantitative structure-activity relationship (3D-QSAR). Comparative molecular field analysis (CoMFA $q^2=0.783$, $r^2=0.944$, $r^2_{\text{pred}}=0.851$) and Comparative molecular similarity indices analysis (CoMSIA $q^2=0.728$, $r^2=0.982$, $r^2_{\text{pred}}=0.814$) were used to establish 3D-QSAR models, which had good verification and prediction capabilities. Based on the contour maps and the information of molecular docking, 8 novel small molecules were designed in silico, among which compounds D4, D5 and D8 with high predictive activity were subjected to further molecular dynamics simulations (MD), and their possible binding modes were explored. It was found that Asn535 plays a crucial role in stabilizing the inhibitors. Furthermore, the ADME and bioavailability prediction for D4, D5 and D8 were carried out. The results would provide valuable guidance for designing new reversible LSD1 inhibitors in the future.

Keywords: LSD1; molecular inhibitors; thieno[3,2-b]pyrrole-5-carboxamide derivatives; Molecular docking ;3D-QSAR; Molecular dynamics simulations

1. Introduction

Histone modification is an important epigenetic modification including acetylation, methylation and phosphorylation[1]. Gene expression is regulated by controlling conformational transformation of chromosomes between heterochromatin with transcriptional inhibition and euchromatin with transcriptional activation [2]. Histone methylation was considered to be an irreversible process until the discovery of LSD1(also known as KDM1A) in 2004, revealing that histone methylation was a dynamically reversible process [3]. LSD1 is a highly conserved flavin adenine dinucleotide (FAD) dependent amine oxidase that can specifically remove the single-dimethylation of histone lysine H3K4 and H3K9 [4]. LSD1 can also remove methyl of many proteins, such as p53, DNA methyltransferase 1(DNMT1), signal transducer and activator of transcription 3 (STAT3) , E2F transcription factor 1 and so on, thereby regulating the physiological functions of downstream cell [5-9]. Studies have shown that LSD1 is highly expressed in a variety of tumor cells, for example breast cancer, neuroblastoma, lung cancer, gastric cancer, acute myeloid leukemia and so on [10-14]. It have been shown in many animal experiments that the use of small-molecule inhibitors to inhibit the expression of LSD1 can effectively inhibit the differentiation, proliferation, invasion and metastasis

of cancer cells, as well as the growth of tumors [15-17]. These studies indicated that LSD1 not only had important biological significance, but also provided a new method for cancer treatment with LSD1 small molecule inhibitors.

The research on LSD1 inhibitors has attracted increasing interests since the discovery of LSD1. As a member of the monoamine oxidase (MAO) family, LSD1 has a high sequence similarity with homologous protein MAOs, so MAOs inhibitors can be used to inhibit the activity of LSD1 [18]. MAOs inhibitors such as Tranylcypromine (TCPA, Fig.1A), pargyline (Fig.1B), phenelzine (Fig.1C) had low inhibitory activity and poor selectivity for LSD1, but a series of high active and high selective inhibitors have been optimized or designed based on them [19]. Among the TCPA-based series inhibitors, ORY-1001 (Fig.1D) entered phase II clinical trials about acute myelogenous leukemia (AML) in 2013. GSK2879552 (Fig.1E) also entered I period clinical used in the treatment of relapsed or refractory small cell lung cancer (SCLC)[20, 21]. In addition to these irreversible inhibitors covalently bound with FAD, some reversible inhibitors with non-covalently bound were also designed. The IC_{50} of the reversible inhibitor GSK-354 (Fig.1F) reached 90nM [22]. SP-2509 (Fig.1G) was highly active and selective, and its IC_{50} value in vitro reached 13nM [23]. In 2017, thieno[3,2-b]pyrrop-5-carboxamides (Fig.1H) class reversible inhibitors were reported. The highest IC_{50} value of these compounds was 7.8nM, which displayed a remarkable effect on MLL-AF9 human leukemia cells [24, 25].

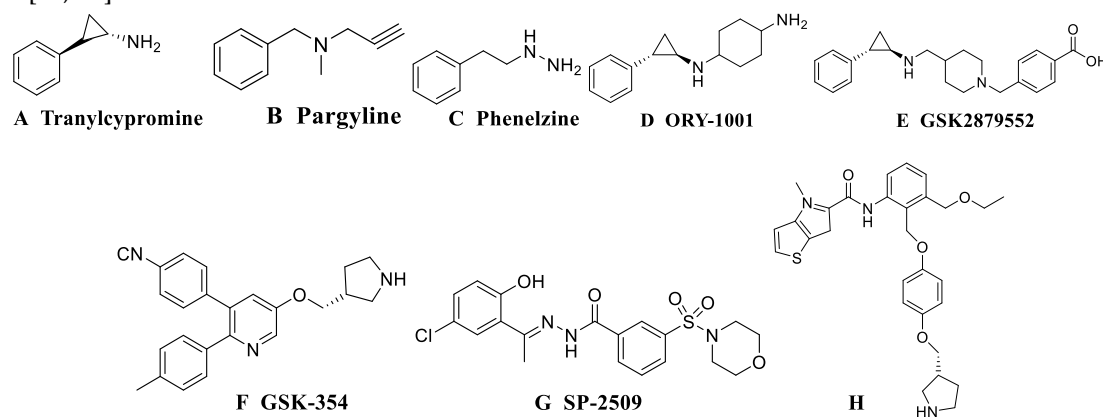


Figure 1. Structures of several reported LSD1 inhibitors.

In vitro activity determination of small molecule inhibitors is time-consuming and costly. The computer-aided drug design (CADD) can not only preliminarily predict the activity of inhibitors, but also save experimental costs and provide guidance for designing more effective inhibitors by exploring the reaction mechanism of inhibitors at the molecular level [26]. A series of 1,2,4-triazine derivatives as new h-HAAO inhibitors using 3D-QSAR, docking and MD and three of these inhibitors were tested in silico associated with relatively high activities [27]. 3D-QSAR method was also used to explain flavonoids inhibitors for Escherichia coli, and molecular docking was performed to predict the binding mode [28]. Ma et al analyzed pyrazolo[3,4-d]pyrimidine derivatives as the TgCDPK1 inhibitors using the 3D-QSAR, docking, dynamics simulation, and then designed some new inhibitors [29]. Therefore, the application of CADD methods into the research of LSD1 inhibitors should be efficiency for designing new and highly active LSD1 inhibitors.

In this study, 55 small molecules were selected from a series of thieno[3,2-b] pyrrop-5-carboxamide (H) compounds [24, 25]. The crystal structures of five small molecules with LSD1 have been reported, but the binding modes of the remaining 50 small molecules with LSD1 were still uncertain. In order to explain the difference of biology activity by docking those small molecules, and explore the possible optimization direction for such inhibitors, 3D-QSAR model was developed based on the docking poses. Contour maps could provide theoretical guidance for the designing new inhibitors. According to the molecular docking results and contour maps, 8 new molecule inhibitors with high predictive activity were designed. The binding mode of 3 new molecule inhibitors with the

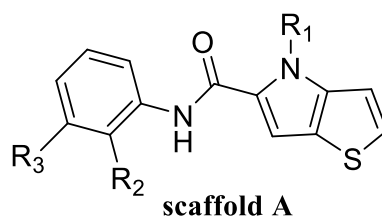
highest activity was explored by molecular docking and molecular dynamics simulations. The present research would provide valuable guidance for the design of new LSD1 inhibitors.

2. Materials and Methods

2.1. Data sets

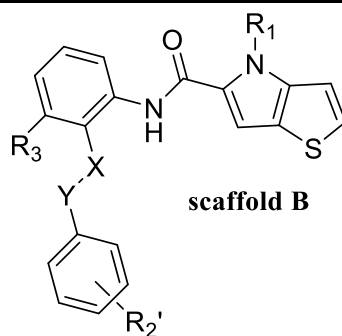
55 compounds that were previously assessed for their LSD1 inhibition activities were used as the data set (Table 1) [24, 25]. These compounds share a common thieno[3,2-b]pyrrop-5-carboxamide scaffold and introducing a substituted chiral pyrrolidine group at the C13 position of the parent scaffold resulted in significantly improved activities, i. e. **51**, **54** and **55**. CoMFA[30] and CoMISA[31] were used to establish 3D-QSAR models. In these two models, different molecular properties were used as independent variables to calculate the relationship between their structure and biological activities, while pIC_{50} ($-\log IC_{50}$) was also used as the dependent variable. The range of pIC_{50} values was from 4.046 to 8.174 in all data sets. 43 compounds were selected as training sets for the construction of CoMFA and CoMSIA models, and the remaining 12 compounds were selected as independent test set for validating reliability of the model. Data set division depended on substituent diversity and the biological data of the test set was evenly distributed within the overall range of activity, the details were shown in Fig. S1. The structures and activity data of all compounds are shown in Table 1.

Table 1. The structures, the actual and predicted activities by CoMFA & CoMSIA models of LSD1 inhibitors.

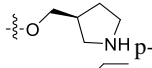
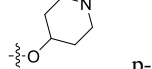
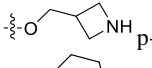
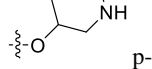
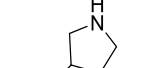
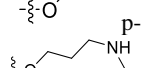
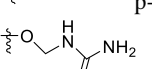
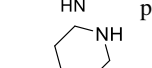
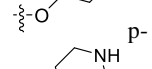
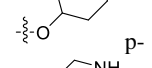
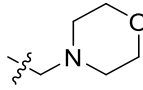
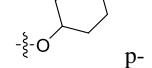
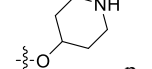
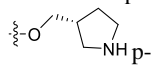
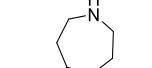
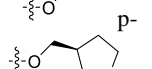
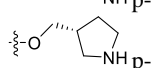


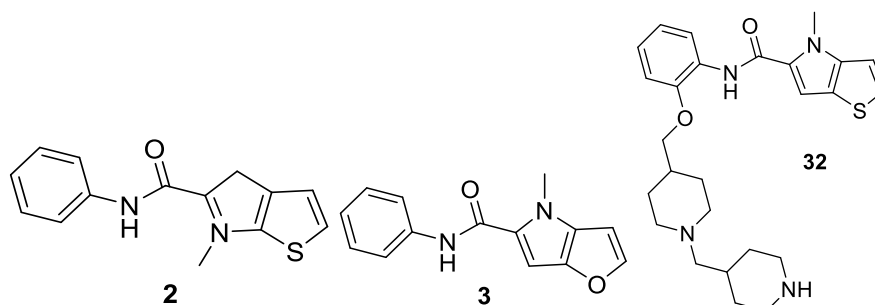
No.	R ₁	R ₂	R ₃	IC ₅₀ (μ M)	pIC ₅₀	CoMFA		CoMSIA	
						Pred.	Res.	Pred.	Res.
1*	CH ₃	H	H	25.7	4.59	4.325	0.2652	4.518	0.0723
4	Et	H	H	9.3	5.032	4.401	0.6315	4.607	0.4249
5	-CH ₂ CH ₂ NH ₂	H	H	48.1	4.318	4.075	0.2432	4.434	-0.1158
6*	Pr	H	H	53.2	4.274	4.274	-0.0005	4.633	-0.3595
7	CH ₃	H	-Cl	18.7	4.728	4.545	0.1826	4.509	0.2186
8	CH ₃	H		21.5	4.668	4.85	-0.1818	4.753	-0.0847
9	CH ₃	H	-F	22.2	4.654	4.439	0.2147	4.526	0.128
10	CH ₃	H	-OCH ₃	26.5	4.577	4.611	-0.0344	4.637	-0.0605
11	CH ₃	H	-N(Me) ₂	29.6	4.529	4.781	-0.2519	4.662	-0.1329
12*	CH ₃	H	-CONH ₂	97.3	4.012	4.604	-0.592	4.601	-0.5885
13	CH ₃	H		2.2	5.658	5.44	0.218	5.702	-0.0441
14	CH ₃	H	-CH ₂ SMe	7.4	5.131	4.892	0.2391	4.855	0.2762
15*	CH ₃	H		11.1	4.955	4.704	0.251	4.804	0.1515
16	CH ₃	H	-CH ₂ OMe	13.4	4.873	4.786	0.0866	4.854	0.0186
17	CH ₃	H	-CH ₂ N(Me) ₂	16.7	4.777	4.415	0.3615	4.748	0.0288
18	CH ₃	H		25.2	4.599	4.745	-0.1462	4.607	-0.0082
19	CH ₃	H	-CH ₂ OH	52.3	4.281	4.701	-0.4202	4.569	-0.2879

20*	CH ₃	H		7.2	5.143	4.687	0.4556	4.258	0.8848
21	CH ₃	H		9.4	5.027	4.765	0.2624	4.903	0.1238
22	CH ₃	H		89.9	4.046	4.605	-0.5588	4.252	-0.2055
23	CH ₃	H		0.162	6.79	7.09	-0.3003	6.735	0.0547
24	CH ₃	H		0.442	6.355	6.567	-0.2116	6.386	-0.0312



No.	X	Y	R ₁	R ₂ '	R ₃	IC ₅₀ (μM)	pIC ₅₀	CoMFA		CoMSIA	
								Pred.	Res.	Pred.	Res.
25	C	O	CH ₃		H	0.31	6.509	6.668	-0.1592	6.515	-0.0063
26*	C	O	CH ₃		H	0.09	7.046	6.523	0.523	6.55	0.4962
27*	C	O	CH ₃		H	1	6	5.755	0.2446	5.481	0.5185
28	C	O	CH ₃		H	68.2	4.166	4.048	0.1178	4.045	0.1205
29	C	C	CH ₃		H	0.177	6.752	6.63	0.1223	6.664	0.0884
30*	O	C	CH ₃		H	0.336	6.474	6.741	-0.2673	6.621	-0.1473
31	C	O	CH ₃		H	1.5	5.824	6.304	-0.4805	5.817	0.0075
33	carbonyl	NH	CH ₃		H	2.2	5.658	5.947	-0.2887	5.631	0.0274
34	C	O	CH ₃		H	2.5	5.602	6.162	-0.5596	5.563	0.0393
35	C	O	CH ₃		H	2.9	5.538	5.812	-0.2741	5.487	0.0506
36	C	O	CH ₃		H	0.018	7.745	7.192	0.5529	7.401	0.3438
37	C	O	CH ₃		H	0.064	7.194	6.873	0.3209	6.974	0.2202
38	C	O	CH ₃		H	0.121	6.917	6.813	0.1039	6.987	-0.0704
39	C	O	CH ₃		H	0.149	6.827	6.98	-0.1532	7.23	-0.4031

40	C	O	CH ₃		H	0.209	6.68	6.378	0.3017	6.79	-0.1097
41	C	O	CH ₃		H	0.451	6.346	6.156	0.19	6.242	0.1036
42	C	O	CH ₃		H	0.618	6.209	6.272	-0.0625	6.271	-0.0616
43	C	O	CH ₃		H	0.674	6.171	6.523	-0.3516	6.262	-0.0908
44	C	O	CH ₃		H	0.804	6.095	6.012	0.0826	6.29	-0.1947
45	C	O	CH ₃		H	1.3	5.886	5.995	-0.1091	5.879	0.0065
46*	C	O	CH ₃		H	3.2	5.495	4.571	0.924	5.319	0.1758
47*	C	O	CH ₃		-COMe	0.075	7.125	6.522	0.6034	6.959	0.1659
48	C	O	CH ₃		-COEt	0.056	7.252	7.182	0.0705	7.313	-0.0613
49	C	O	CH ₃			0.065	7.187	7.031	0.1564	7.098	0.089
50	C	O	CH ₃		-COiPr	0.066	7.18	7.317	-0.1368	7.106	0.0743
51*	C	O	CH ₃		-COMe	0.0084	8.076	7.829	0.2469	8.29	-0.2142
52	C	O	CH ₃		-COMe	0.0292	7.535	7.649	-0.1143	7.597	-0.0619
53	C	O	CH ₃		-COMe	0.0857	7.067	7.14	-0.0733	7.07	-0.003
54	C	O	Et		-COMe	0.0067	8.174	7.856	0.3177	8.175	-0.0008
55	C	O	Me		-COEt	0.0078	8.108	7.752	0.3556	8.148	-0.0404



No.	IC ₅₀ (μM)	pIC ₅₀	CoMFA		CoMSIA	
			Pred.	Res.	Pred.	Res.
2	38.8	4.411	4.453	-0.0422	4.459	-0.0483
3	89.5	4.048	4.269	-0.2213	4.369	-0.3213
32*	1.7	5.77	5.321	0.4492	4.632	1.1379

Pred. = predictor pIC₅₀; Res. = residual; o- = ortho-, m- = meta-, p- = para-; * compounds in test set.

2.2. Molecular docking

Molecular docking and the establishment of CoMFA and CoMSIA models were completed in SYBYL-X2.0 [32]. The 3D structures of 55 small molecules were obtained in Sketch module. Energy minimization was performed using Powell gradient algorithm with a maximum of 1000 iterations, the convergence criterion was limited to $0.001 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$. All compounds were calculated by Gastieger-Huckel charges using Tripos force field. The structure of compound **54** with the highest activity in this series is shown in Fig.2A. When selecting docking results, it was often unreliable to blindly select the pose with the best score. Essential Chemical Interactions Described for Analogue Ligands (ECIDALs) should be defined from previous literature or crystal structure, and then selected the pose with the best score in accordance with ECIDALs was reasonable [33]. The reported 5 crystal structures for complexes of ligand and LSD1 were obtained from Protein Data Bank (pdb codes: 5LGT, 5LGN, 5LGU, 5LHH and 5LHI). By observing the superposition of these crystal structures, we found that thieno[3,2-b]pyrrop-5-carboxamide structures almost overlapped in the same position, as shown in Fig. 2B, those features were defined as the ECIDALs of a series of small molecules. Therefore, when choosing the docking results, we gave priority to selecting the pose consistent with orientation of small molecules in Fig. 2B. The crystal structure of LSD 1(PDBcode: 5LHH, the ligand with the highest activity in this series of compounds) was retrieved from Protein Data bank. Crystal water, metal ions and the original ligand were removed and hydrogen atoms were added before molecular docking. Each small molecule produced 20 docking poses, and the optimal pose was selected for further study by combining ECIDALs and docking score.

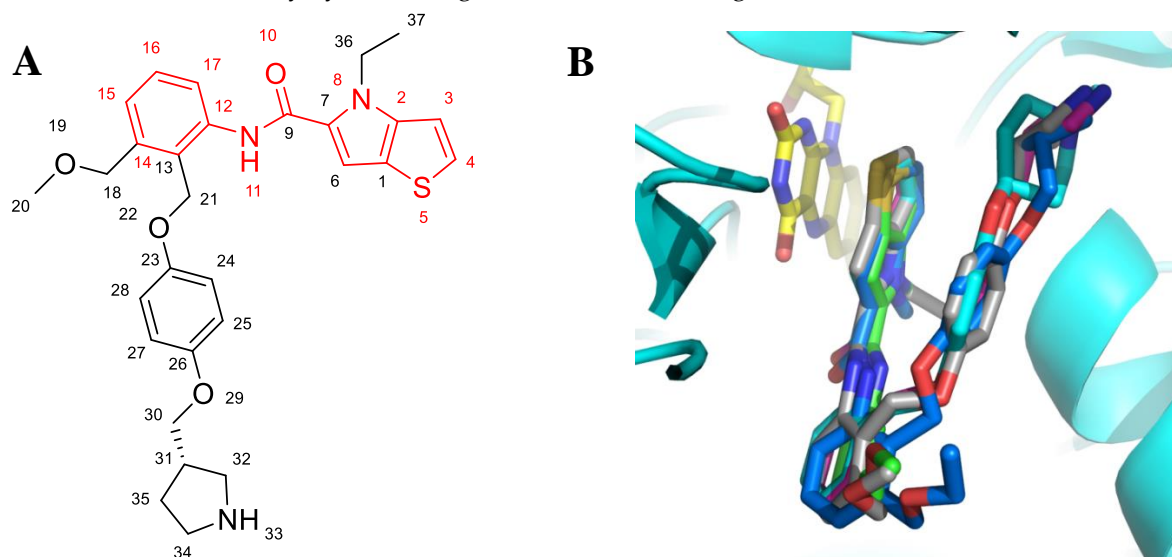


Figure 2. (A) Compound **54**, red denoted thieno[3,2-b]pyrrop-5-carboxamide structure, the common structure of the series of small molecules. (B) superimposing of the Ligands from 5 LSD1 crystal structures (PDB: 5LGT (cyan), 5LGN (green), 5LGU (violet), 5LHH (gray) and 5LHI (blue)).

2.3. Molecular alignment

The quality of molecular alignment was considered as a key factor for the robustness and predictive power of CoMFA and CoMSIA models [34]. There are three major approaches to alignment based on common skeletal alignment, pharmacophore alignment and molecular docking alignment, respectively. Both of the former two methods are based on the energy minimized conformation of small molecules, they could not reflect the ligand-receptor binding pattern correctly. So we chose molecular docking alignment to build 3D-QSAR Models, the optimal docking poses of each small molecule were saved and used in the establishment of QSAR model. The alignment result of all molecules docking-based pose was shown in Fig.3. The alignment results of 43 training set compounds based on docking conformation were shown in Fig.S2.

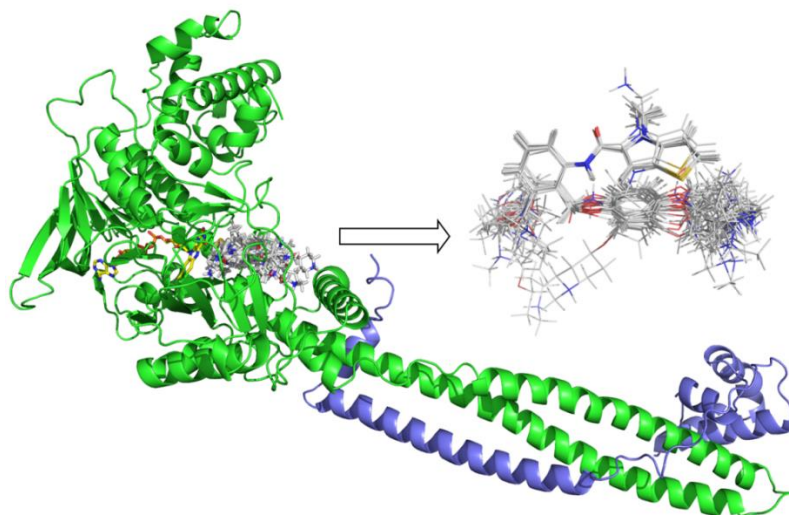


Figure 3. Docking results of small molecules in the active pocket and the alignment based on docking conformation for all ligands. FAD was shown in yellow and Corest was shown in blue.

2.4. CoMFA and CoMSIA models

In CoMFA analysis, the steric field and electrostatic field were calculated by Lennard Jones and Coulombic potential functions respectively. In the calculation process, the compound was placed in a cubic lattice, which had a grid space of 2 Å and a probe atom of sp^3 hybridized carbon atom. The van der Waals radius of the atom was 1.52 Å and the net charge was +1.0. The energy cut-off value was set to 30kcal/mol and the default value was adopted for other parameters [35]. For CoMSIA, in addition to calculating steric field and electrostatic field parameters, hydrophobic field, hydrogen bond acceptor field and hydrogen bond donor field parameters were also calculated. Therefore, in addition to the van der Waals radius of 1.52 Å, the net charge of +1.0 and probe with charge of +1.0, while hydrophobic parameters, hydrogen bond acceptor parameters and hydrogen bond donor parameters also should be set to +1 for calculate the compound characteristics [36]. The attenuation coefficient was set to 0.3 in default. In CoMSIA model, Gaussian functions were used to determine the distance between molecule atoms and probe atoms [37].

2.5. CoMFA and CoMSIA statistical analysis

The partial least square (PLS) method was used to analyze CoMFA and CoMSIA models. Cross-validation analysis was carried out with leave-one-out (LOO) to obtain the cross-validation coefficient q^2 and the optimal number of components (ONC) [38]. The cross-validation coefficient q^2 could be get as follows [39]:

$$q^2 = 1 - \frac{\text{PRESS}}{\text{PRESS}_0} \quad (1)$$

In the formula, PRESS represents the sum of the squares of the difference between the predicted and the actual pIC_{50} values and PRESS_0 is the actual pIC_{50} value. The non-cross-validated correlation coefficient (r^2), F-statistic values (F), standard error of estimate (SEE) and contributions of each field were calculated [40]. The predictive correlation coefficient (r^2_{pred}) value could be obtained to measure the predictive capability of the 3D-QSAR models. It can be obtained by the formula[41] :

$$r^2_{\text{pred}} = 1 - \frac{\text{PRESS}}{\text{SD}} \quad (2)$$

In the formula, SD is the sum of the squared deviations between the pIC_{50} values in test set and mean pIC_{50} in training set; PRESS stands for the sum of squared deviations between predicted pIC_{50} values and actual pIC_{50} values in test set.

In order to further evaluate the true predictive abilities, external validation parameters R^2 , k , k' , R_0^2 , $R_0'^2$, r_m^2 need to be calculated. R^2 is the correlation coefficient between the observed and predicted activities in the test set with the (0,0) intercept. R_0^2 and k are the correlation coefficient between the experimental (x) versus predicted activities (y) for the test set through origin and the corresponding slope of regression line. $R_0'^2$ and k' are the correlation coefficients between the predicted (y) versus experimental activities (x) for the test set through origin and the corresponding slope of regression line. The calculation formula are as follows[42]:

$$R^2 = \frac{[\sum(Y_{obs} - \bar{Y}_{obs})(Y_{pred} - \bar{Y}_{pred})]^2}{\sum(Y_{pred} - \bar{Y}_{pred})^2 \times \sum(Y_{obs} - \bar{Y}_{obs})^2} \quad (3)$$

$$k = \frac{\sum(Y_{obs} \times Y_{pred})}{\sum(Y_{pred})^2} \quad (4)$$

$$k' = \frac{\sum(Y_{obs} \times Y_{pred})}{\sum(Y_{obs})^2} \quad (5)$$

$$R_0^2 = 1 - \frac{\sum(Y_{obs} - k \times Y_{pred})^2}{\sum(Y_{obs} - \bar{Y}_{obs})^2} \quad (6)$$

$$R_0'^2 = 1 - \frac{\sum(Y_{pred} - k' \times Y_{obs})^2}{\sum(Y_{pred} - \bar{Y}_{pred})^2} \quad (7)$$

$$r_m^2 = R^2 \times (1 - \sqrt{R^2 - R_0^2}) \quad (8)$$

Among them, Y_{obs} and Y_{pred} represent the experimental and predicted activities in the test set, \bar{Y}_{obs} and \bar{Y}_{pred} are the average value of the experimental or predicted activities in the test set.

The model is acceptable if the parameters in the 3D-QSAR model meet the following criteria [43]:

$$\begin{aligned} q^2 &> 0.5 \\ r_{pred}^2 &> 0.6 \\ \frac{(R^2 - R_0^2)}{R^2} < 0.1 \text{ or } \frac{(R^2 - R_0'^2)}{R^2} < 0.1 \\ 0.85 &\leq k \leq 1.15 \text{ or } 0.85 \leq k' \leq 1.15 \\ |R_0^2 - R_0'^2| &< 0.3 \\ r_m^2 &> 0.5 \end{aligned}$$

Y-randomization test is usually used to evaluate the robustness of the model [44]. The dependent variable y is randomly shuffled while the independent variable x matrix is kept unaltered. The process is repeated many times and for each run, a set of new q^2 and r^2 is generated. If the q^2 and r^2 of the models are low, it indicates the good calibration result is not due to chance correlation and the QSAR model is robust.

2.6. Molecular dynamics simulations

In order to prove the reliability of the docking results, the binding mode between the series of compounds and LSD1 protein need to be further investigated, and molecular dynamics simulation was carried out using AMBER14 software package [45]. Using the docking results of ligands with LSD1 as the initial conformation, the parameter file of ligands was generated by Antechamber module. Amberff10 force field was used for receptor protein and GAFF force field was used for ligand molecules [46]. The water box adopted TIP3P water model with a margin distance of 8 Å. After energy minimization, the complex was heated from 0K to 300K during 250ps in NVT ensemble, the constant pressure of 1 atm was equilibrated at 300k for another 50ps. Finally, for compound 54 and newly designed compounds with high predictive activity, 50ns MD was performed under NPT ensemble with the pressure of 1 atm and 300k.

2.7. Binding free energy calculations

MM/GBSA method was used to calculate binding free energy. 200 snapshots were received from the last 2ns trajectory file for calculation. The binding free energy formula for protein and ligand were as follows [27]:

$$\Delta G_{bind} = \Delta G_{complex} - \Delta G_{protein} - \Delta G_{ligand}$$

$$\begin{aligned}
 &= \Delta E_{MM} + \Delta G_{sol} - T\Delta S \\
 &= \Delta E_{vdw} + \Delta E_{ele} + \Delta G_{GB} + \Delta G_{SA} - T\Delta S
 \end{aligned}$$

In the formula, $\Delta G_{complex}$ represents the total free energy of protein-ligand complex, $\Delta G_{protein}$ is the total binding energy of protein in solvents and ΔG_{ligand} means the total binding energy of ligand in solvents. ΔE_{MM} is the interaction energy of protein and ligand under gas-phase, and the energy could be obtained by calculating van der Waals energy ΔE_{vdw} and electrostatic energy ΔE_{ele} . ΔG_{sol} stands for the free energy of solvation, which could be achieved by computing polar solvation energy ΔG_{GB} and non-polar solvation energy ΔG_{SA} . $T\Delta S$ stands for entropy contribution, which is generally ignored, because it consumes a large amount of computing resources and has a weak impact on the results. After calculating the binding free energy, we decomposed the energy onto each residue to obtain the key amino acids which have a great impact on it.

2.8. Prediction of ADME and bioavailability

Nowadays, the speed of drug research and development is accelerating, and the number of candidate compounds is increasing. It would waste a lot of resources to put them into experiments directly. Therefore, it is necessary to use computational modeling method to evaluate their bioavailability and pharmacokinetics. SwissADME web tool (<http://www.swissadme.ch>) [47] was used to predict our new LSD1 inhibitors.

Bioavailability evaluation includes the following aspects: lipophilicity, molecular weight, polarity, saturation. The absorption, distribution, metabolism and excretion (ADME) evaluation includes: human gastrointestinal absorption (HIA), blood-brain barrier (BBB) permeation, cytochrome P450-3A4 (CYP3A4) enzyme inhibition and skin permeation (log K_p). Finally, drug-likeness evaluation of our new compounds was conducted according to Lipinski's rule.

3. Results and Discussion

3.1. Validation of docking reliability

Both 3D-QSAR modeling and MD were based on the results of molecular docking, so it was necessary to verify the reliability of docking results: the structures of five compounds **16**, **25**, **36**, **54** and **55** from the docking results, and the crystal structures of those compounds were received from protein data bank (PDB code: 5LGN, 5LGT, 5LGU, 5LHH and 5LHI, respectively). The docking pose of five compounds and their corresponding crystal poses were superimposed, and it can be seen that the crystal pose and the docked pose were almost in the same position, having similar spatial orientations (Fig. S3), with the RMSDs being 0.78 Å, 1.13 Å, 0.62 Å, 0.69 Å, 1.07 Å respectively. The result suggested that the docking result was reasonable and could be used for further simulation and analysis.

3.2. Statistical results of CoMFA and CoMISA

The stepwise development of CoMFA and possible CoMSIA models using different fields were presented in Table 2[48]. The internal prediction ability (q^2) and the external prediction ability (r_{pred}^2) were important criteria for measuring the QSAR model. [41] CoMFA-S and CoMFA-E were separately modeled using steric field and electrostatic field, and their q^2 values were acceptable, the r_{pred}^2 of CoMFA-S was also acceptable, but the $r_{pred}^2 < 0.6$ in CoMFA-E indicated that the external prediction ability of the model did not up to the standard. When combined to CoMFA-SE, both q^2 and r_{pred}^2 improved. The result of CoMFA-SE model gave a cross-validated q^2 of 0.783 with an ONC value of 4, a non-cross-validated r^2 of 0.944, a predictive correlation coefficient r_{pred}^2 of 0.851, a SEE value of 0.3, and an F value of 160.128. The contribution of the steric field was 39.4% and 60.6% belonged to electrostatic field.

Table 2. Statistical parameters of CoMFA and CoMSIA models based on docked pose.

	q^2	ONC	r^2	r_{pred}^2	SEE	F value	Contributions				
							Steric	Electrostatic	Hydrophobic	H-bond acceptor	H-bond donor
CoMFA-S	0.682	3	0.928	0.740	0.335	168.134	1	-	-	-	-
CoMFA-E	0.688	5	0.945	0.574	0.301	127.014	-	1	-	-	-
CoMFA-SE	0.783	4	0.944	0.851	0.3	160.128	0.394	0.606	-	-	-
CoMSIA-EHDA	0.717	6	0.989	0.812	0.135	552.841	-	0.341	0.219	0.141	0.3
CoMSIA-SHDA	0.724	3	0.958	0.811	0.257	295.325	0.242	-	0.259	0.142	0.357
CoMSIA-SEDA	0.721	4	0.972	0.791	0.213	325.942	0.201	0.365	-	0.129	0.305
CoMSIA-SEHD	0.722	3	0.942	0.807	0.302	209.989	0.179	0.356	0.197	-	0.268
CoMSIA-SEHA	0.727	2	0.977	0.785	0.194	309.365	0.182	0.417	0.244	0.156	-
CoMSIA-ALL	0.728	5	0.982	0.814	0.170	412.470	0.144	0.292	0.179	0.125	0.259

The CoMSIA-SEHA and CoMSIA-ALL models generated similar q^2 (0.727 and 0.728, respectively). However, the r_{pred}^2 of CoMSIA-ALL was higher than that of CoMSIA-SEHA (0.814 and 0.785, respectively), so we choose CoMSIA-ALL model with stronger external prediction ability as the final CoMSIA model. CoMSIA-ALL model had the highest cross-validated q^2 of 0.728 with an ONC value of 5, non-cross-validated r^2 of 0.982, the highest predictive correlation coefficient r_{pred}^2 of 0.814, a SEE value was 0.170, and an F value was 412.470. The contributions of steric, electrostatic, hydrophobic, H-donor, and acceptor field were 14.4%, 29.2%, 17.9%, 25.9% and 12.5%, respectively. The result illustrated that the electrostatic, H-bond donor and hydrophobic fields played important roles in the model. And among these three fields, the electrostatic field was the most important interaction between the ligand and the receptor protein.

Table 3 showed the results of external validation of the CoMFA-SE and CoMSIA-ALL models (hereinafter, CoMFA-SE was called CoMFA and CoMSIA-ALL was called CoMSIA). Although both CoMFA and CoMSIA models had good r_{pred}^2 values (Table 2), however, according to Tropsha[49], only high r_{pred}^2 did not show the true predictive ability of the model, only when the model satisfied condition 1, condition 2a or 2b, condition 3a or 3b, condition 4a or 4b, condition 5 and condition 6, the models had good external prediction ability. It could be seen from the table 3 that the parameters of CoMFA and CoMSIA models satisfied all above requirements. Among them, r_m^2 was an important indicator to measure the approximation between experimental activity and predicted activity in test set.

Table 3. Results of external validation parameters for CoMFA and CoMSIA.

Condition	Parameters	Threshold Value	CoMFA	CoMSIA
1	R^2	>0.6	0.897	0.846
2a	R_0^2	Close to value of R^2	0.896	0.831
2b	$R_0'^2$	Close to value of R^2	0.892	0.845
3a	k	$0.85 < k < 1.15$	1.046	1.029
3b	k'	$0.85 < k' < 1.15$	0.952	0.965
4a	$(R^2 - R_0^2)/R^2$	<0.1	0.001	0.018
4b	$(R^2 - R_0'^2)/R^2$	<0.1	0.005	0.001
5	$ R_0^2 - R_0'^2 $	<0.3	0.004	0.014
6	r_m^2	>0.5	0.868	0.743

The plots of actual pIC_{50} against predicted pIC_{50} by the CoMFA and CoMSIA model were shown in Fig 4. It can be seen that the black solid point and the red solid point were close to the line $Y = X$, which indicated that the actual and predicted activities of the whole data set had a strong linear relationship. The Y value of the hollow circle represented leave-one-out (LOO) cross-validated predictions.

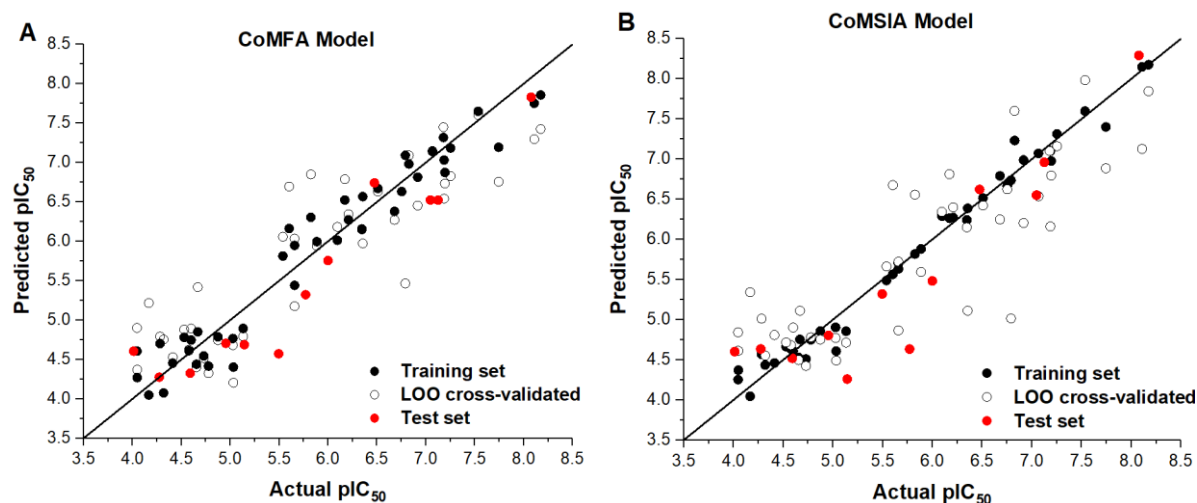


Figure 4. Plots of experimental activities against predicted activities by the optimal of CoMFA model (A) and CoMSIA model(B). Hollow circle indicated a scatter plot of LOO cross-validated.

In addition, the results of 10 random shuffles for Y-randomization test were shown in Table 4. These results were obtained by using the dependent variable (biological activity) randomly shuffled and then using the original independent variable matrix to build new QSAR models. The q^2 and r^2 values of new QSAR models were really low. Therefore, the possibility of random correlations was ruled out. The randomly arranged bioactivities used for the test were shown in Table S1.

Table 4. q^2 and r^2 values after several Y-randomization test.

Iteration	CoMFA		CoMSIA	
	q^2	r^2	q^2	r^2
Random_1	0.001	0.152	-0.012	0.235
Random_2	-0.181	0.268	-0.225	0.445
Random_3	-0.097	0.105	-0.242	0.34
Random_4	0.237	0.345	0.219	0.373
Random_5	0.29	0.309	0.264	0.424
Random_6	0.175	0.304	0.216	0.359
Random_7	0.05	0.189	0.082	0.271
Random_8	-0.036	0.217	0.015	0.408
Random_9	0.166	0.283	0.176	0.337
Random_10	-0.031	0.134	-0.062	0.283

3.3. CoMFA contour maps

The predicted values of the activity of small molecules in the 3D-QSAR model were shown in Table 1. The contour maps were provided to explain the relationship between the structures and

activities of compounds. The modification of small molecules according to the tips given by the contour maps would be conducive to the improvement of the activity of small molecules. The contour maps were shown by StDev*Coeff type, and the most active compound **54** was put into the contour maps as a reference for explanation.

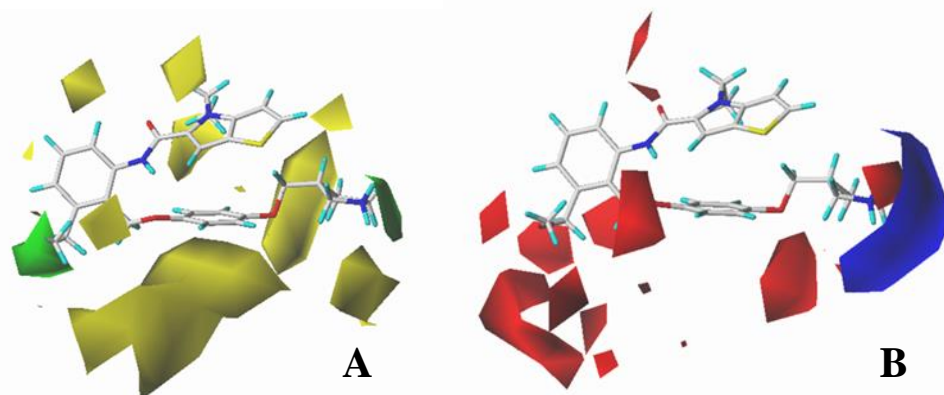


Figure 5. CoMFA contour maps were based on compound **54** as the template. (A) Steric contour maps: Green and yellow displayed the favorable and unfavorable region. (B) electrostatic contour maps: Blue and red indicated favorable and unfavorable region.

Fig. 5A was the contour maps of the CoMFA steric field. The green contour maps mean that the bulky groups were beneficial for improving activity, while the yellow contour maps mean the bulky groups were disfavored. Around C18 and O19 of compound **54**, there was a green contour near R₃ group, indicating that increasing the volume at this location was conducive to improve the activity. For example, compound **13** (pIC₅₀=5.658), **14** (pIC₅₀=5.131, R₃=CH₂SMe), **20** (pIC₅₀=5.143) and **21** (pIC₅₀=5.027) introduced medium-sized substituents in R₃, and their activity improved a little compared with compound **1** (pIC₅₀=4.59, R₃=H). There was a green contour at N33 of compound **54**, illustrating substituent R₂ length increases to R'₂ appropriately, which was beneficial to the improvement of the compounds activity, as the activities of compounds **26** (pIC₅₀=7.046), **36** (pIC₅₀=7.745), **37** (pIC₅₀=7.149) were greater than that of compounds with no substituent group at R₂ such as compounds **1** (pIC₅₀=4.59). R₂ of compound **31** (pIC₅₀=5.824) had substituent, but it did not reach the green block at N33 of the compound, so its activity was not significantly improved. A yellow contour appeared at the position of R₁ substituent, indicating that the excessive substituent volume at this position was not conducive to the improvement of activity, such as the activities arrangement of **6** (pIC₅₀=4.274, R₁=Pr) < **5** (pIC₅₀=4.318, R₁=CH₂CH₂NH₂) < **4** (pIC₅₀=5.032, R₁=Et). Another large yellow contour appeared below hydroquinone in compound **54**, indicating that increasing the volume here was not useful for the improvement of activity. Such as R'₂ substituent on ortho substitution of the benzene ring of compound **28** (pIC₅₀=4.166) touched the yellow color contour, R'₂ substituent on of meta substitution of the benzene ring of compound **27** (pIC₅₀=6) also contacted the yellow color contour, therefore compounds **27** and **28** reduced significantly in activity. There was a yellow color contour in C30 and C31, substituent R'₂ volume should not be too large, such as compounds **45** (pIC₅₀=5.886) and **46** (pIC₅₀=5.495) reduced in activity.

The contour maps of CoMFA electrostatic field was shown in Fig.5B. The red contour maps mean that the electronegative groups were beneficial for improving activity, while the blue contour maps mean the electropositive groups were favorable. There were two red contours around O22 and O29 of template molecule **54**, which indicated that the addition of electronegative groups can improve the activity. For example, the activity of compound **26** (pIC₅₀=7.046) was higher than that of compound **29** (pIC₅₀=6.752), because the oxygen atom was at the Y of compound **26** and the carbon atom was at the Y of compound **29**. For **31** (pIC₅₀=5.824), **34** (pIC₅₀=5.602) and **35** (pIC₅₀=5.538) compounds, the electronegativity of R'₂ substituents was weak, but they touched the position O29 in red color contour, so the activities of these compounds were weaker. The presence of red contour around O19 indicated

that R_3 should replace groups with strong electronegativity, such as compound **7** ($pIC_{50}=4.728$, $R_3=Cl$) and **9** ($pIC_{50}=4.654$, $R_3=F$), which were more active than compound **1** ($pIC_{50}=4.59$, $R_3=H$). Blue color block appeared at N33 illustrated that the increase of positive group at R'_2 position was beneficial to the increase of activity. The compounds **51** ($pIC_{50}=8.076$), **54** ($pIC_{50}=8.174$) and **55** ($pIC_{50}=8.108$) with the electropositive strong substituents at R'_2 position were the most active in all series of compounds.

3.4. CoMSIA contour maps

Fig. 6A and 6B were respectively contour maps of the steric field and electrostatic field of CoMSIA model, which were different from those of CoMFA model in some aspects, but the analysis results were consistent. For example, in the steric field, a green contour appeared at the C23 and C31 position of molecule **54**, which also expressed the need to increase the length of R_2 substituents. Fig. 6C was the contour maps of the hydrophobic field. The yellow contour maps mean that the hydrophobic groups were beneficial for improving activity, while the white contour maps mean the hydrophilic groups were favorable.

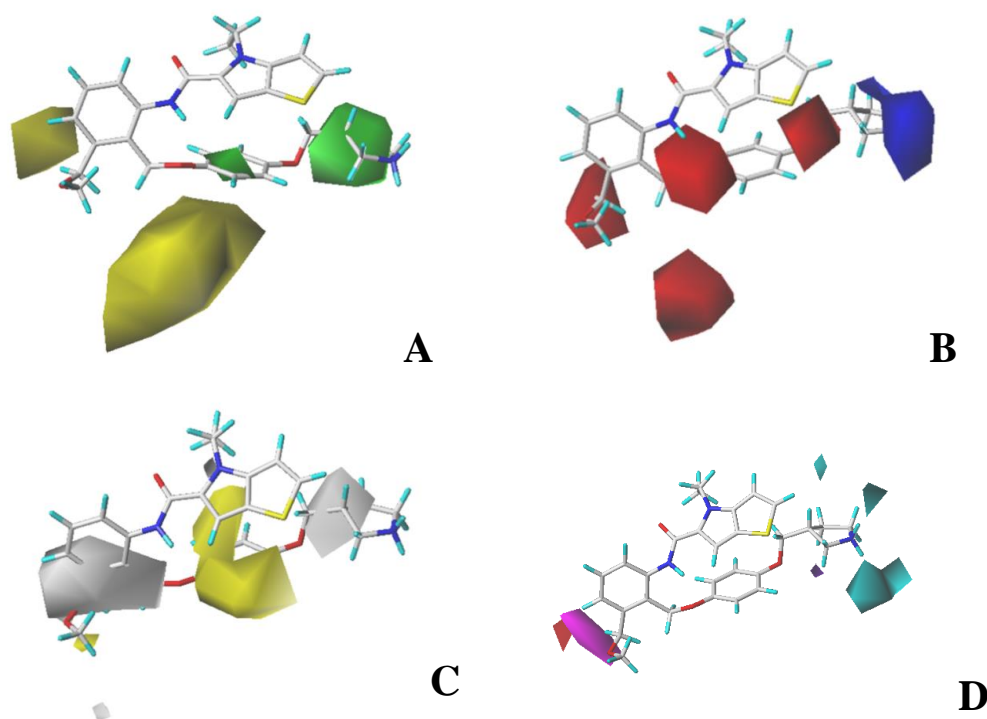


Figure 6. CoMSIA contour maps were based on compound **54** as template. (A) Steric field: Green and yellow displayed the favorable and unfavorable region. (B) Electrostatic fields: Blue and red indicated favorable and unfavorable region. (C) Hydrophobic field: yellow and grey represented favorable and unfavorable region. (D) Hydrogen bond donor: cyan and purple indicated favorable and unfavorable region, and Hydrogen acceptor fields: magenta and red illustrated favorable and unfavorable region.

A large yellow contour was interspersed around C23 and C26 of compound **54**, illustrating that for R_2 substituents, improving the hydrophobicity of the intermediate part would be conducive to improving the activity. For example, in the middle part of R_2 , compound **32** ($pIC_{50}=5.57$) was piperidine ring, while compound **29** ($pIC_{50}=6.752$) was benzene ring. Benzene ring was more hydrophobic than piperidine, so the activity of compound **29** was greater than **32**. A small piece of white color in R'_2 , demonstrated that adding hydrophilic substituent was beneficial to the improvement of the activity, R'_2 of **50** ($pIC_{50} = 7.18$) and **51** ($pIC_{50} = 8.076$) were replaced by piperidine and pyrrolidine, respectively, so they had high activity. There was also a white contour around C14, indicating that increasing the hydrophilicity of R_3 substituents could improve the

activity of the compound. R_3 of Compound **18** ($pIC_{50}=4.599$) contained benzene rings, so the reason for its low activity could be explained. The R_3 of **7** ($pIC_{50}=4.728$, $R_3=Cl$) and **9** ($pIC_{50}=4.654$, $R_3=F$) were hydrophobic halogen substituents, so their activities were not high. Contour maps of hydrogen bond donor and hydrogen bond acceptor were shown in Fig. 6D, the cyan contour represented that adding hydrogen bond donors was beneficial to activity and the purple contour represented that hydrogen bonding donor groups were unfavorable, the magenta color contour expressed hydrogen bond acceptor group had a positive effect in activity and the red color contour implied hydrogen bond receptor groups had a negative impact in activity. A large cyan contour appeared at N33 of compound **54**, indicating that the addition of H-bond donor group here would increase the activity. The docking results also showed that compound **54** formed a hydrogen bond with Asp555 of LSD1 protein. R'_2 of **39** ($pIC_{50}=6.827$) had piperidine as a hydrogen bond donor, and R'_2 end of **41** ($pIC_{50}=6.346$) was ethyl, therefore the activity of **39** was higher than **41**. The magenta contour appeared around C20, indicating that the introduction of hydrogen bond receptor groups in R_3 location would improve the activity, such as **16** ($pIC_{50}=4.873$, $R_3=CH_2OMe$) > **1** ($pIC_{50}=4.59$, $R_3=H$).

3.5. Docking analysis

We hope to explain the difference of activity between small molecules through molecular docking, to understand what binding mechanisms will benefit the interaction between ligand and protein, and to provide ideas for the future design of small molecules. Therefore, three representative compounds were chosen for docking analysis: Compound **1** was the common structure of this series compounds. Compound **28** with extended R_2 substituents was not increased but decreased in activity. The compound **54** had the highest activity.

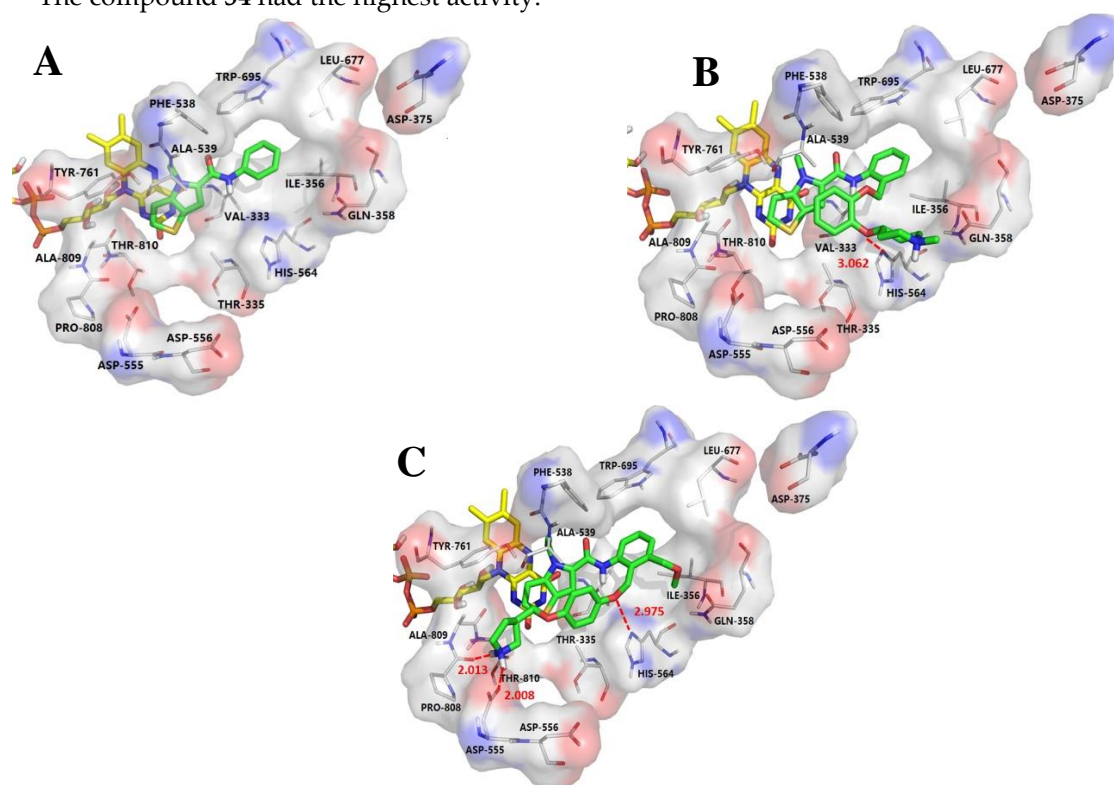


Figure 7. Binding modes between the selected structure of compound **1** (A), **28** (B), **54** (C) and LSD1 (PDB code: 5LHH) with lowest energy, respectively. Compounds are shown in greenstick model. FAD is shown in yellow stick model. The nitrogen, oxygen, sulfur atoms are shown in blue, red, yellow, respectively. Hydrogen bonds are shown in red dash lines.

As shown in Fig. 7A for compound **1**, the thiophene-pyrrole ring was surrounded by Val333, Phe538, Leu677, Trp695, Thr761 and Thr810, and thiophene-pyrrole ring and FAD formed an aromatic-aromatic interaction. Benzamide was surrounded by residues of Val333, Trp695 and

Phe538. The reason why low activity of compound **1** was that there was no hydrogen bond at all. It was observed that Asp555 and Asp556 belonged to electronegative region and was also a hydrogen-bond receptor residue. Therefore, for the series of compounds, introducing positive groups at R₂' and the formation of hydrogen bond with Asp555 and Asp556 were a potential modification method. The conclusion was also confirmed by the results of 3D-QSAR analysis.

Compound **28** was shown in Fig.7B, ortho-substitution of benzene rings in benzamide was performed, and the thiophene-pyrrole ring formed an aromatic–aromatic interaction with FAD. The substituents on the benzamide formed a hydrogen bond with His564 (His564-NH···O, bond length 3.06 Å). Due to the piperidine ring replaced the ortho-position H of the benzene ring, the electropositive substituents did not extend to the electronegative region at Asp555, which was also the reason why the activity of compound **28** was not high enough. Meanwhile, it was verified with QSAR to explain why there were large yellow blocks under hydroquinone in the CoMFA contour maps. Compound **28** was in the cavity, but the small molecule was still some distance away from the electronegativity region at Asp555, indicating that there was still potentiality for further improvement of activity.

For compound **54**, thiophene-pyrrole heterocyclines still interacted with FAD to form an aromatic–aromatic interaction. As shown in Fig.7C, the benzene ring of benzamide was approached by ortho-meta double substitution. The ortho-substituents extended to the electronegative region of Asp555, where the basic N on pyrrolidine formed a hydrogen bond (Asp555-HO···HN, bond length 2.01 Å) with Asp555 and another hydrogen bond with Pro808 (Pro808=O···HN, bond length 2.01 Å). This result was consistent with analysis of electrostatic field and hydrogen bond donor field in QSAR. Furthermore, compound **54** formed a hydrogen bond (His564-NH···O, bond length 2.97 Å) with His564. In addition, compound **54** interacted with the Asp555 electronegative region to form a salt bridge. This analysis explained why compound **54** was the highest in activity. At the same time, we observed that Asp375 was also an electronegative region and a residue of hydrogen bond acceptor, but small molecules were still some distance away from this region. If they interact with this region, the activity of small molecules may be further enhanced, which could be used as the next direction to modify such small molecule inhibitors.

We found that the reason for the difference of inhibitor activity could be reasonably explained, indicating that binding mode might be the actual binding mode of small molecules. For newly designed small molecules, using molecular docking to observe the interaction between small molecules and LSD1 is very important for preliminary prediction of the activity of new small molecule inhibitors. When selecting the docking results of newly designed small molecules, we can rely on the above information for reference.

3.6. Design of novel derivatives

The structure-activity relationship (SAR) information revealed by the above 3D-QSAR contour maps analysis was summarized as shown in Figure 8, which may be helpful in designing new LSD1 inhibitors with high activity. R₁ substituents should not be too large, R₂ substituents should be appropriately increased in order to achieve the length of electronegativity region, and the structure of hydroquinone played an important role in improving the activity. Because this structure not only increased the volume of R₂, but also conformed to the suggestion of improving the activity of electrostatic field. For example, it was required to add electronegative groups at O22 and O29 positions of template compound **54**, which was also of great significance in the hydrophobic field, therefore, this structure was retained when designing new inhibitors. R₂' substituent should introduce electropositive substituent.

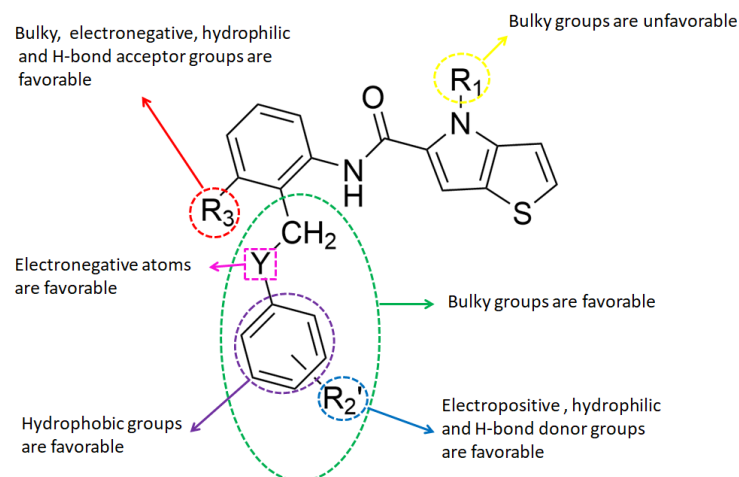


Figure 8. Structure-activity relationship (SAR) information obtained from 3D-QSAR study.

The addition of H-bond donor field at the R_2 showed favorable to increase activity and introduction of hydrogen bond acceptor groups in R_3 could improve the activity. Based on the above analysis, we designed eight new small molecules and used 3D-QSAR model to predict their activity. The structure and prediction results of small molecules were shown in Table 5. As mentioned before, the interaction between small molecules and LSD1 was an important basis for predicting the activity of small molecules. Therefore, we selected **D4**, **D5** and **D8** with high activity for molecular docking. In the process of molecular docking, we preferentially chose binding mode with the same orientation as the binding mode obtained above and higher docking score.

Table 5. Structures and predicted activities for novel designed compounds.

No	X	Y	R ₁	R ₂ '	R ₃	Predict pIC ₅₀	
						CoMFA	CoMSIA
D1	C	O	Et			7.681	7.606
D2	C	S	Et			7.786	8.096
D3	C	O	Et			7.868	7.608
D4	C	O	Et			8.129	8.312
D5	C	O	Et			8.179	8.521
D6	C	O	Et			7.534	8.131
D7	C	O	CH ₃		COH	7.686	7.527
D8	C	O	Et			8.073	8.473

3.7. MD simulations

In order to further analyze the dynamic interactions of ligand-receptor and explore the binding mode of new designed small molecules, we chose the docking results of compound **54**, **D4**, **D5**, **D8** and conducted the molecular dynamics simulation of 50ns. The RMSD plot of C_α for complexes was shown in Fig.9. After a few time, the RSMD fluctuates of the four complexes were all in a very small

range between 1.3Å and 2.4Å, which indicated that the systems had reached a state of stability. Plots of temperature and total-energy versus time were shown in Fig. S4 and Fig. S5.

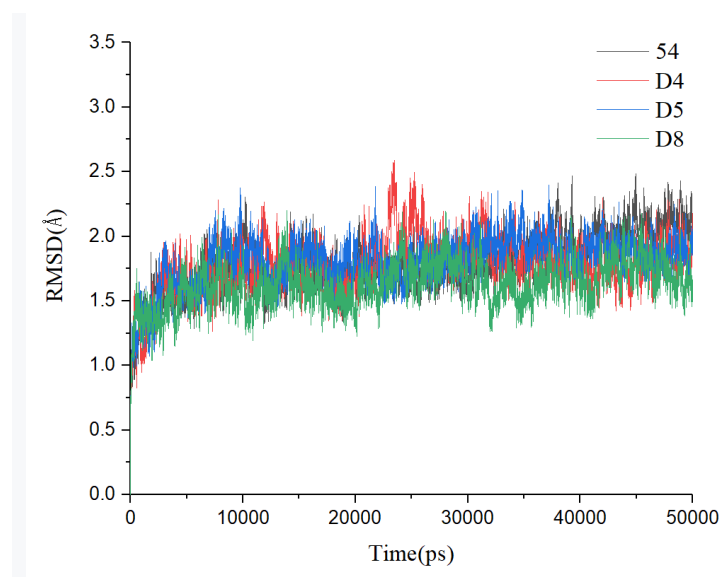


Figure 9. RMSD values of the complexes during 50ns MD simulations.

We compared the LSD1-54 complex conformation between the after MD structure and initial structure, as is shown in Fig10, we found that compound **54** not only remained at the docking site, but also entered the active pocket of histone H3 deeper than the initial structure and blocked the entry of histone H3 into the active site, this prevented further interaction between H3 and FAD. At the same time, these results validated the reliability of the docking results.

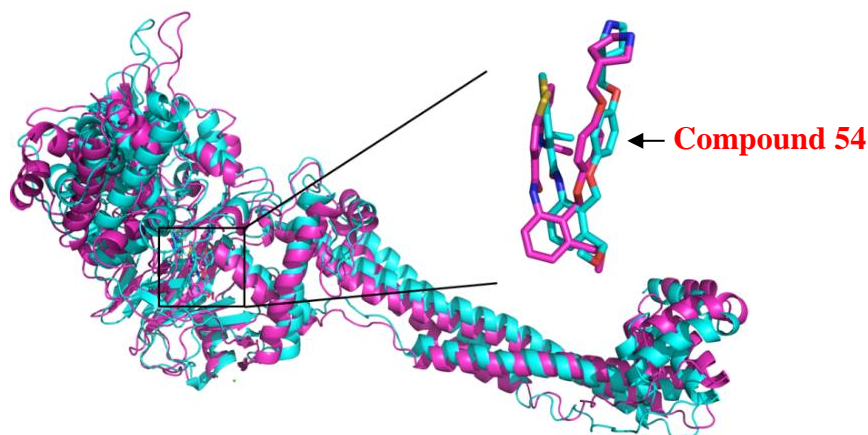


Figure 10. Compound **54** and LSD1 of complex structure alignment between the initial structure (cyan) and MD structure (violet).

The average MD structure was shown in Fig.11. The thiophene-pyrrole ring was more closely connected with FAD, and small molecules formed a salt bridge with Asp555. Although compound **54** had no interaction with Pro808, it formed shorter H bonds with ASP555(Asp555-HO...HN, bond length 1.81 Å) and His564 (His564-NH...O, bond length 2.55 Å). This was the reason why compound **54** had higher activity.

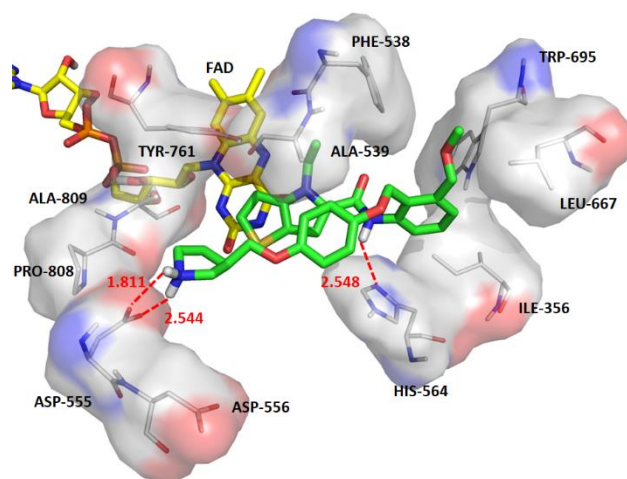
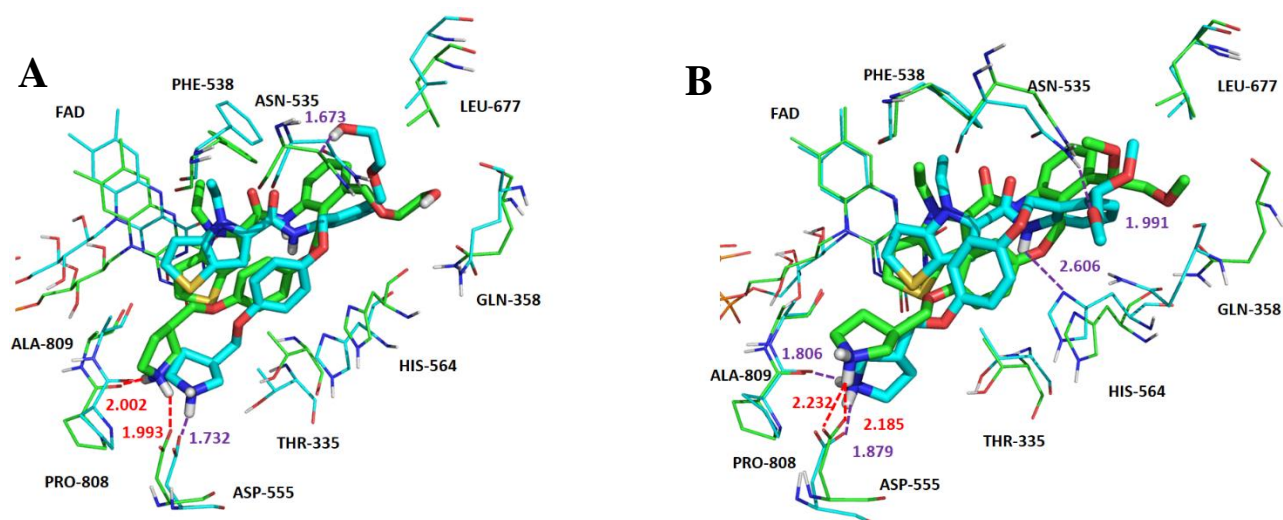


Figure 11. The binding mode of compound 54 with LSD1 after a 50ns MD simulation. Compound 54 and surrounding residues were shown in green stick model and white line model. Hydrogen bonds were shown in dash lines.

We superimposed the average structures of **D4**, **D5**, **D8** complexes during MD equilibrium stage to docking results, which were shown in Fig.12. After 50ns MD, The 3 new small molecules were still in the active pocket, the molecules formed aromatic-aromatic interaction with FAD and salt bridge with Asp555. The **D4** docking results showed that small molecules form hydrogen bonds with Pro808 (Pro808=O...HN, bond length 2.00 Å) and Asp555 (Asp555-HO...HN, bond length 1.99 Å). Although hydrogen bond lost with Pro808 during MD, shorter H-bond Asp555(Asp555-HO...HN, bond length 1.73) and new H-bond (Asn535-NH...O, bond length 1.67Å) were formed. The docking results and MD average structure showed that **D5** and Asp555 established shorter H-bond (Asp555-HO...HN, bond length 1.88) during MD simulation. In addition, small molecules formed 3 new hydrogen bonds with Asn535 (Asn535-NH...O, bond length 1.99), Pro808 (Pro808=O...HN, bond length 1.81 Å) and His564 (His564-NH...O, bond length 2.61 Å), which is attributed the predictive activity associated with **D5**.



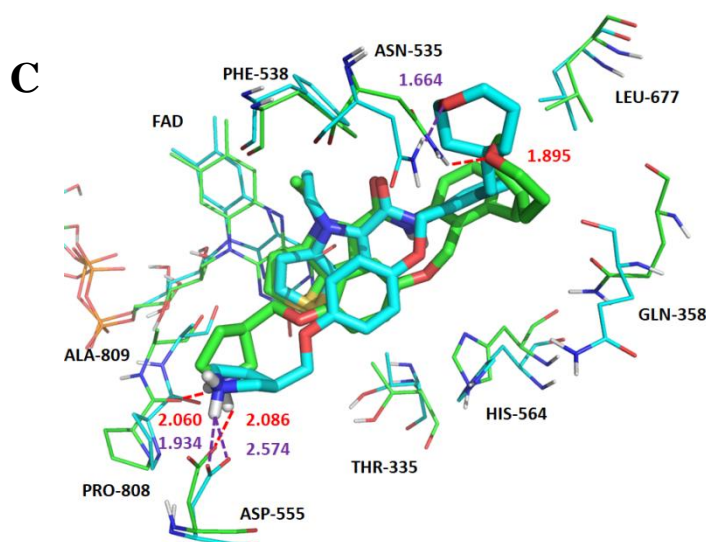


Figure 12. The superposition of the docking structures and MD average structures of compound **D4(A)**, **D5(B)**, **D8(C)**. Carbon atoms of docking result and MD average structure were shown in green and cyan, respectively. H-bond of docking result and 50ns MD result were shown in red dash line and purple dash line, respectively.

The docking result of **D8** showed that small molecules form hydrogen bonds with Pro808 (Pro808 = O... HN, bond length 2.06) and Asp555 (Asp555-HO... HN, bond length 2.09), respectively. At the same time, small molecules formed hydrogen bonds with Asn535 (Asn535-NH...O, bond length 1.89). During MD simulations, although H-bond was lost with Pro808, shorter H-bonds were formed with Asp555 (Asp555-HO... HN, bond length 1.93) and Asn535 (Asn535-NH... O, bond length 1.66). MD analysis indicated that **D4**, **D5**, **D8** forms favorable interaction with LSD1. In the meantime, the importance of Asp555 for the activity of these small molecules was proved. In accordance with the previous QSAR model and docking analysis, R_3 was modified according to QSAR contour map (enlarging volume and adding hydrogen bond acceptor groups). After these modifications, we found that the R_3 groups of **D4**, **D5** and **D8** compounds formed hydrogen bonds with Asn535 in the superposition, among these excellent compounds reported, none of them formed H-bond with Asn535, which may be the reason why **D4**, **D5** and **D8** have higher predictive activity than compound **54**. The modification of R_3 and Asn535 deserve the attention of future designers of such inhibitors.

In addition, it was found that small molecules and FAD formed aromatic-aromatic interaction before and after MD, which was of great significance to the binding mode of these small molecules.

3.8. Binding free energy calculation

In order to further verify the binding affinity of **D4**, **D5**, **D8** with LSD1, the binding free energies of the three compounds with LSD1 were calculated by MM/GBSA method, and the binding free energy of compound **54** with LSD1 was also computed as a reference. The results are shown in Table 6.

Table 6. Binding free energies of ligand-protein complex.

No.	ΔE_{ele} kcal mol ⁻¹	ΔE_{vdw} kcal mol ⁻¹	ΔG_{GB} kcal mol ⁻¹	ΔG_{SA} kcal mol ⁻¹	ΔG_{sol} kcal mol ⁻¹	ΔG_{bind} kcal mol ⁻¹	pIC ₅₀ predicted
LSD1- 54	-103.7501	-52.4748	128.9897	-6.6194	122.3703	-33.8537	8.174
LSD1- D4	-134.9198	-49.4678	151.9764	-6.5057	145.4707	-38.9169	8.312
LSD1- D5	-101.5999	-61.0727	120.0679	-7.7538	112.3141	-50.3592	8.521
LSD1- D8	-86.1198	-59.8317	109.082	-7.9847	101.0974	-44.8516	8.473

The binding free energies of compounds **54**, **D4**, **D5**, **D8** with LSD1 were -33.85 kcal mol⁻¹, -38.92 kcal mol⁻¹, -50.36 kcal mol⁻¹, -44.85 kcal mol⁻¹, respectively. It is generally believed that the lower of ΔG_{bind} value, the more stable of the complex. Table 6 showed that ΔG_{bind} values were coincided well

with the predicted pIC₅₀ values of the QSAR model. The electrostatic energy ΔE_{ele} contributed much more to the binding free energy than other energies, indicating that electrostatic interaction played a key role in the complex system, which could be caused by the interaction between basic N in small molecules and Asp555 negative region. Similarly, the van der Waals energy ΔE_{vdw} illustrated that hydrophobic interaction was also important in binding process. The polar solvation energy ΔG_{GB} is positive, indicating that it was not conducive to G_{bind} , the reason was that excessive binding pocket could cause exposure of ligand to solvents. However, ΔG_{SA} values were negative, which means that non-polar solvation energy was beneficial to ΔG_{bind} . Because the newly designed compounds **D4**, **D5** and **D8** had lower ΔG_{bind} values than **54**, they may have stronger inhibitory effect on LSD1.

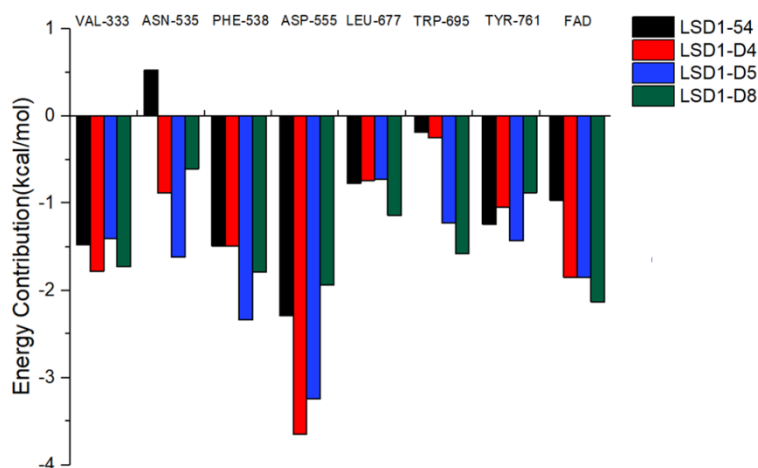


Figure 13. Binding free energy decomposition plots.

In order to analyze which amino acids contributed more to the binding free energy of the system, we decomposed the binding free energy into each amino acid and observed that the amino acids that contribute more to the binding free energy were Val333, Asn535, Phe538, Asp555, Leu677, Trp695, Tyr761 and FAD. Their energy contributions were shown in Fig.13. It was found that Asp555 contributed the most to the four systems, which again showed that Asp555 was closely related to the activities of these small molecules. Furthermore, FAD contributed a lot to the four systems. The docking results showed that thiophene-pyrrole and FAD had aromatic-aromatic interaction, which was important for the binding mode of these small molecules. **D4**, **D5** and **D8** formed hydrogen bond with Asn535 and compound **54** did not form hydrogen bond with it, therefore Asn535 was beneficial to the binding free energy of D4, D5, D8, but has little contribution to the binding with compound **54**. Although Asp555 contributed less to **D5** than **D4**, Asn535, Phe538 and TRP695 contributed more to **D5** than **D4**, which may be the reason why **D5** activity was higher than **D4**.

3.9. ADME and bioavailability analysis

It is necessary to carry out bioavailability and pharmacokinetics prediction on the novel candidate compounds before experiment. We predicted the candidate compounds **D4**, **D5** and **D8**, and compound **54** with the best activity was selected as the control group. The results were shown in Table 7.

Table 7. Bioavailability and pharmacokinetics prediction.

No.	MW (g mol ⁻¹)	Fraction Csp ³	Num.rotatable bonds	TPSA (Å ²)	log P	log S	HIA	BBB	CYP3A4 inhibition	log K _p (cm s ⁻¹)	Drug- likeness Lipinski
54	519.66	0.34	12	101.99	4.44	-5.37	High	No	Yes	-6.43	Yes
D4	549.68	0.37	14	122.22	4.12	-4.97	High	No	Yes	-7.10	Yes
D5	563.71	0.39	14	111.22	4.56	-5.41	High	No	Yes	-6.78	Yes

D8	587.77	0.44	12	101.99	4.90	-6.73	Low	No	Yes	-5.73	Yes
Optimal range	<800	0.25-1	≤10	20-130	-0.7-5	-10-6	-	-	-	-	-

Compound saturation was measured by fraction Csp3 and polarity was calculated by TPSA (topological polar surface area). Lipophilicity used log P to quantify. The log P values of **D4**, **D5** and **D8** were in the best range, which indicated that they had good absorbency. LogS was used to measure solubility, and the log S values of **D4**, **D5** and **D8** were also within a reasonable range, which demonstrated that they had good solubility. The only indicator for **D4**, **D5** and **D8** beyond the best range was num.rotatable bonds, which mean that the molecules had high flexibility in the human body. So we need to minimize the number of rotational bonds in the design of such small molecules in the future. Overall, **D4**, **D5** and **D8** had really high bioavailability. High stands for compounds **54**, **D4** and **D5** with good gastrointestinal absorption ability. Low means that **D8** may not have gastrointestinal absorption ability. No represented that compounds **54**, **D4**, **D5** and **D8** were not brain penetrant. The two properties could be more intuitively reflected from Fig. S6. Yes represented that these compounds had inhibitory effects on CYP3A4 and could be excreted through metabolic biotransformation. Skin permeability was measured by log K_p, and **D8** had highest skin permeability. Finally, Lipinski's rule was used to evaluate the drug-likeness of four compounds. The results were Yes. The bioavailability and ADME analysis showed that the novel designed compounds might become safer and more active LSD1 inhibitors.

4. Conclusions

In this study, we chose the series of thieno[3,2-b]pyrrole-5-carboxamides as reversible LSD1 inhibitors using molecular docking, 3D-QSAR. ECIDALs was followed when selecting docking results[33]. The representative results of molecular docking for small molecules were used to explain the differences in biology activity. The CoMFA ($q^2=0.783$, $r^2=0.944$) and CoMISA models ($q^2=0.728$, $r^2=0.982$) were received to perform molecular modeling study. The results showed that the models had good internal verification ability and external prediction ability. The models were further validated following the criteria given by Tropsha and Roy[42,49], and were determined to be statistically reliable and robust. We could use this model to predict the activity of this series of compounds to reduce the loss of blind investment in the experiment. Likewise, the contour maps also indicated the relationship between the structure and activity of small molecules. The results of molecular docking and the information prompted by contour maps confirmed each other. Based on these 3D QSAR models, 8 new small molecules were designed in silico. Further docking, MD, calculation of binding free energy, ADME and bioavailability prediction were carried out for these designed compounds and the results indicated that **D4**, **D5** and **D8** show good potential to become LSD1 inhibitors with better activity than compound **54**. At the same time, key amino acids affecting the activity of these inhibitors, such as Asp555, Pro808, Asn535, were easy to form H-bonds with small molecules, after R₃ was modified, Asn535 formed H-bond with small molecules. Val333, Phe538, Trp695, Thr761 and FAD could enable inhibitors to maintain stability in binding site. Our study would provide some theoretical guidance and help for the future design of novel reversible LSD1 inhibitor.

Supplementary Materials: The following are available online.

Acknowledgments: The authors are grateful for the financial support from The National Natural Science Foundation of China (Grant No.21603180), China Scholarship Council(Grant No.201708410492), Research Project of Xinxiang Medical University (Grant No.505060, 505144) and Open topic at School of Biomedical Engineering, Xinxiang Medical University (Grant No.2018-BME-KFKT-01, 2018-BME-KFKT-10, 2018-BME-KFKT-11).

The authors also thank Prof. Zheng Jie from Guangdong University of Technology for offering software support and helpful discussion.

References

1. Strahl, B.D.; Allis, C.D. The language of covalent histone modifications. *Nature* **2000**, *403*, 41-45.
2. Tony, K. Chromatin modifications and their function. *Cell* **2007**, *128*, 693-705.
3. Shi, Y.; Lan, F.; Matson, C.; Mulligan, P.; Whetstine, J.R.; Cole, P.A.; Casero, R.A.; Shi, Y. Histone Demethylation Mediated by the Nuclear Amine Oxidase Homolog LSD1. *Cell* **2004**, *119*, 941-953.
4. Benoit, L.; Lv, R.; Jernej, M.; Kristina, H.; Ryan, F.; Yang, X.; Shichong, L.; Garcia, B.A.; Hao, W.; Feizhen, W. A Specific LSD1/KDM1A Isoform Regulates Neuronal Differentiation through H3K9 Demethylation. *Molecular Cell* **2015**, *57*, 957-970.
5. Garcia-Bassets, I.; Kwon, Y.-S.; Telese, F.; Prefontaine, G.G.; Hutt, K.R.; Cheng, C.S.; Ju, B.-G.; Ohgi, K.A.; Wang, J.; Escoubet-Lozach, L. Histone methylation-dependent mechanisms impose ligand dependency for gene activation by nuclear receptors. *Cell* **2007**, *128*, 505-518.
6. Jing, H.; Roopsha, S.; Espejo, A.B.; Gyu, L.M.; Dorsey, J.A.; Mario, R.; Susanne, O.; Ramin, S.; Bedford, M.T.; Thomas, J. p53 is regulated by the lysine demethylase LSD1. *Nature* **2007**, *449*, 105-108.
7. Jing, W.; Sarah, H.; Kurash, J.K.; Hong, L.; Frédérique, G.; Jeffrey, B.; Hui, S.; Weitao, S.; Hua, C.; Guoliang, X. The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. *Nature Genetics* **2009**, *41*, 125-129.
8. Q, H.; YS, K.; E, N.; MD, C.; KR, H.; KA, O.; I, G.-B.; DW, R.; CK, G.; MG, R., et al. RETRACTED: Nuclear Receptor-Enhanced Transcription Requires Motor- and LSD1-Dependent Gene Networking in Interchromatin Granules. *Cell* **2008**, *105*, 19199-19204.
9. Shi, Y.J.; Matson, C.; Lan, F.; Iwase, S.; Baba, T.; Shi, Y. Regulation of LSD1 Histone Demethylase Activity by Its Associated Factors. *Molecular Cell* **2005**, *19*, 857.
10. Lokken, A.A.; Zeleznik-Le, N.J. Breaking the LSD1/KDM1A addiction: therapeutic targeting of the epigenetic modifier in AML. *Cancer cell* **2012**, *21*, 451-453.
11. Lv, T.; Yuan, D.; Miao, X.; Lv, Y.; Zhan, P.; Shen, X.; Song, Y. Over-expression of LSD1 promotes proliferation, migration and invasion in non-small cell lung cancer. *PloS one* **2012**, *7*, e35065.
12. Magerl, C.; Ellinger, J.; Braunschweig, T.; Kremmer, E.; Koch, L.K.; Höller, T.; Büttner, R.; Lüscher, B.; Gütgemann, I. H3K4 dimethylation in hepatocellular carcinoma is rare compared with other hepatobiliary and gastrointestinal carcinomas and correlates with expression of the methylase Ash2 and the demethylase LSD1. *Human pathology* **2010**, *41*, 181-189.
13. Schulte, J.H.; Lim, S.; Schramm, A.; Friedrichs, N.; Koster, J.; Versteeg, R.; Ora, I.; Pajtler, K.; Klein-Hitpass, L.; Kuhfittig-Kulle, S. Lysine-specific demethylase 1 is strongly expressed in poorly differentiated neuroblastoma: implications for therapy. *Cancer research* **2009**, *69*, 2065-2071.
14. Wang, Y.; Zhang, H.; Chen, Y.; Sun, Y.; Yang, F.; Yu, W.; Liang, J.; Sun, L.; Yang, X.; Shi, L. LSD1 is a subunit of the NuRD complex and targets the metastasis programs in breast cancer. *Cell* **2009**, *138*, 660-672.
15. C Zheng, Y.; Yu, B.; Z Jiang, G.; J Feng, X.; X He, P.; Y Chu, X.; Zhao, W.; M Liu, H. Irreversible LSD1 inhibitors: application of tranylcypromine and its derivatives in cancer treatment. *Current topics in medicinal chemistry* **2016**, *16*, 2179-2188.
16. McAllister, T.E.; England, K.S.; Hopkinson, R.J.; Brennan, P.E.; Kawamura, A.; Schofield, C.J. Recent progress in histone demethylase inhibitors. *Journal of medicinal chemistry* **2016**, *59*, 1308-1329.
17. Zheng, Y.-C.; Yu, B.; Chen, Z.-S.; Liu, Y.; Liu, H.-M. TCPs: privileged scaffolds for identifying potent LSD1 inhibitors for cancer therapy. *Epigenomics* **2016**, *8*, 651-666.
18. Schmidt, D.M.; McCafferty, D.G. trans-2-Phenylcyclopropylamine is a mechanism-based inactivator of the histone demethylase LSD1. *Biochemistry* **2007**, *46*, 4408-4416.
19. Lee, M.G.; Wynder, C.; Schmidt, D.M.; McCafferty, D.G.; Shiekhhattar, R. Histone H3 lysine 4 demethylation is a target of nonselective antidepressive medications. *Chemistry & biology* **2006**, *13*, 563-567.

20. Maes, T.; Mascaró, C.; Tirapu, I.; Estiarte, A.; Ciceri, F.; Lunardi, S.; Guibourt, N.; Perdones, A.; Lufino, M.M.; Somervaille, T.C. ORY-1001, a potent and selective covalent KDM1A inhibitor, for the treatment of acute leukemia. *Cancer cell* **2018**, *33*, 495-511. e412.
21. Stewart, C.A.; Byers, L.A. Altering the course of small cell lung cancer: targeting cancer stem cells via LSD1 inhibition. *Cancer cell* **2015**, *28*, 4-6.
22. Wu, F.; Zhou, C.; Yao, Y.; Wei, L.; Feng, Z.; Deng, L.; Song, Y. 3-(Piperidin-4-ylmethoxy) pyridine containing compounds are potent inhibitors of lysine specific demethylase 1. *Journal of medicinal chemistry* **2015**, *59*, 253-263.
23. Sorna, V.; Theisen, E.R.; Stephens, B.; Warner, S.L.; Bearss, D.J.; Vankayalapati, H.; Sharma, S. High-throughput virtual screening identifies novel N' -(1-phenylethylidene)-benzohydrazides as potent, specific, and reversible LSD1 inhibitors. *Journal of medicinal chemistry* **2013**, *56*, 9496-9508.
24. Sartori, L.; Mercurio, C.; Amigoni, F.; Cappa, A.; Faga, G.; Fattori, R.; Legnaghi, E.; Ciossani, G.; Mattevi, A.; Meroni, G. Thieno [3, 2-b] pyrrole-5-carboxamides as new reversible inhibitors of histone lysine demethylase KDM1A/LSD1. Part 1: high-throughput screening and preliminary exploration. *Journal of Medicinal Chemistry* **2017**, *60*, 1673-1692.
25. Vianello, P.; Sartori, L.; Amigoni, F.; Cappa, A.; Fagá, G.; Fattori, R.; Legnaghi, E.; Ciossani, G.; Mattevi, A.; Meroni, G. Thieno [3, 2-b] pyrrole-5-carboxamides as new reversible inhibitors of histone lysine demethylase KDM1A/LSD1. Part 2: Structure-based drug design and structure-activity relationship. *Journal of Medicinal Chemistry* **2017**, *60*, 1693-1715.
26. Jorgensen, W.L. The many roles of computation in drug discovery. *Science* **2004**, *303*, 1813-1818.
27. Qian, P.P.; Wang, S.; Feng, K.R.; Ren, Y.J. Molecular modeling studies of 1,2,4-triazine derivatives as novel h-DAAO inhibitors by 3D-QSAR, docking and dynamics simulations. *Rsc Advances* **2018**, *8*, 14311-14327.
28. Fang, Y.; Lu, Y.; Zang, X.; Wu, T.; Qi, X.; Pan, S.; Xu, X. 3D-QSAR and docking studies of flavonoids as potent Escherichia coli inhibitors. *Scientific Reports* **2016**, *6*, 23634.
29. Wu, W.J.; Ma, S.; Zhou, S.; Lin, W.; Zhang, R.; Zheng, K. Study of novel pyrazolo[3,4-d] pyrimidine derivatives as selective TgCDPK1 inhibitors: Molecular docking, structure-based 3D-QSAR and molecular dynamics simulation. *Rsc Advances* **2016**, *6*.
30. Cramer, R.D.; Patterson, D.E.; Bunce, J.D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *Journal of the American Chemical Society* **1988**, *110*, 5959-5967.
31. Klebe, G.; Abraham, U.; Mietzner, T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *Journal of medicinal chemistry* **1994**, *37*, 4130-4146.
32. SYBYL-X 2.0 TRIPOS, Assoc. Inc., St. Louis, MO. 2012.
33. Camila MuñozGutierrez.; Francisco AdasmeCarreño.; Fuentes E. et al. Computational study of the binding orientation and affinity of PPAR γ agonists: inclusion of ligand-induced fit by cross-docking. *Rsc Advances* **2016**, *6*, 145-146.
34. Chen, J.; Yu, R.; Shen, B.; Xu, Y.; Liu, Y.; & Zheng, H. et al. Docking-based 3D-QSAR modeling of the inhibitors of imp metallo- β -lactamase. *Medicinal Chemistry Research* **2013**, *22*, 1730-1739.
35. Song, Q.-L.; Sun, P.-H.; Chen, W.-M. Exploring 3D-QSAR for ketolide derivatives as antibacterial agents using CoMFA and CoMSIA. *Letters in Drug Design & Discovery* **2010**, *7*, 149-159.
36. Klebe, G.; Abraham, U. Comparative molecular similarity index analysis (CoMSIA) to study hydrogen-bonding properties and to score combinatorial libraries. *Journal of computer-aided molecular design* **1999**, *13*, 1-10.
37. Balasubramanian, P.K.; Balupuri, A.; Gadhe, C.G.; Cho, S.J. 3D QSAR modeling study on 7-aminofuro [2,

- 3-c] pyridine derivatives as TAK1 inhibitors using CoMFA and CoMSIA. *Medicinal Chemistry Research* **2015**, *24*, 2347-2365.
38. Clark, R.D.; Fox, P.C. Statistical variation in progressive scrambling. *Journal of computer-aided molecular design* **2004**, *18*, 563-576.
39. Ding, L.; Wang, Z.Z.; Sun, X.D.; Yang, J.; Ma, C.Y.; Li, W.; Liu, H.M. 3D-QSAR (CoMFA, CoMSIA), molecular docking and molecular dynamics simulations study of 6-aryl-5-cyano-pyrimidine derivatives to explore the structure requirements of LSD1 inhibitors. *Bioorganic & Medicinal Chemistry Letters* **2017**, *27*, 15.
40. Srivastava, V.; Gupta, S.; Siddiqi, M.; Mishra, B. 3D-QSAR studies on quinazoline antifolate thymidylate synthase inhibitors by CoMFA and CoMSIA models. *European journal of medicinal chemistry* **2010**, *45*, 1560-1571.
41. Jitender Verma, V.M.K.; Coutinho, E.C. 3D-QSAR in Drug Design - A Review. *Current Topics in Medicinal Chemistry* **2010**, *10*, -.
42. Roy K.; Chakraborty P.; Mitra I.; Ojha PK.; Kar S.; Das RN. Some case studies on application of "r(m)²" metrics for judging quality of quantitative structure-activity relationship predictions: emphasis on scaling of response data. *Journal of Computational Chemistry* **2013**, *34*, 1071-1082.
43. Marcos L.; Cesar M V.; Vásquez-Velásquez David. et al. Structure-Activity Relationships Based on 3D-QSAR CoMFA/CoMSIA and Design of Aryloxypropanol-Amine Agonists with Selectivity for the Human β 3-Adrenergic Receptor and Anti-Obesity and Anti-Diabetic Profiles. *Molecules* **2018**, *23*,1191.
44. Christoph Rücker.; Gerta Rücker.; Meringer M. y-Randomization and Its Variants in QSPR/QSAR. *J.Chem.Inf.Model.* **2007**, *47*, 2345-2357.
45. Case, D.A.; Babin, V.; Berryman, J.; Betz, R.; Cai, Q.; Cerutti, D.; Cheatham Iii, T.; Darden, T.; Duke, R.; Gohlke, H. Amber 14. **2014**.
46. Sun, H.; Li, Y.; Shen, M.; Tian, S.; Xu, L.; Pan, P.; Guan, Y.; Hou, T. Assessing the performance of MM/PBSA and MM/GBSA methods. 5. Improved docking performance using high solute dielectric constant MM/GBSA and MM/PBSA rescoring. *Physical Chemistry Chemical Physics* **2014**, *16*, 22035-22045.
47. Daina, A. et al. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific Reports* **2017**, *7*, 42717.
48. Caballero J.; Michael Fernández.; Saavedra M . et al. 2D Autocorrelation, CoMFA, and CoMSIA modeling of protein tyrosine kinases' inhibition by substituted pyrido[2,3-d]pyrimidine derivatives. *Bioorganic & medicinal chemistry* **2008**, *16*, 810-821.
49. Tropsha A. Best Practices for QSAR Model Development, Validation, and Exploitation. *Molecular Informatics* **2010**, *29*, 476-488.