High-Throughput Virtual Screening Identifies Novel N'-(1-phenylethylidene)-benzohydrazides as Potent, Specific, and Reversible LSD1 inhibitors

Venkataswamy Sorna[a†], Emily R. Theisen[a,b†], Bret Stephens[a], Steven L. Warner[a], David J. Bearss[a], Hariprasad Vankayalapati[a,c] and Sunil Sharma[a,c,d*]

a. Center for Investigational Therapeutics (CIT), Huntsman Cancer Institute, University of Utah, 2000 Circle of Hope, and Salt Lake City, UT 84112, USA

b. Department of Pharmaceutics and Pharmaceutical Chemistry, College of Pharmacy, University of Utah, 301 Skaggs Hall, Salt Lake City, UT 84112, USA

c. School of Medicine, University of Utah, 30 N. 1900 E., Salt Lake City, UT 84132, USA

d. Division of Medical Oncology, Huntsman Cancer Institute, University of Utah, 2000 Circle of Hope, and Salt Lake City, UT 84112, USA

***Corresponding Author:** Tel: +1-(801)587-5559, Fax: +1-(801)585-0101, E-mail: sunil.sharma@hci.utah.edu, Huntsman Cancer Institute - Room 3360, 2000 Circle of Hope, Salt Lake City, UT 84112, USA

Section S1. Detailed Virtual Screening Methods

S1.1 Preparation of the Binding Site Model and definition of active site

There were several X-ray crystal complex structures of LSD1 (PDB ID: 2Y48, 3BAT, 3BAU, 2XAF, 2XAG, 2XAH, 2XAJ, 2XAQ and 2XAS) at the beginning of our work, we used LSD1 complex model with an X-ray crystal structure (PDB code 2Z5U) protein coordinates used for fragment-based and structure-based virtual screening. Water molecules were then removed and the missing bond order and geometries were edited. Hydrogen atoms were added and the combined complex structure was submitted for protein preparation and energy minimization calculation using ICM and Schrodinger. The fully refined structure with bound ligand molecule was further submitted for grids calculation to define the active site as the collection of amino acids enclosed within a 12 Å radius sphere centered on the bound ligand (**Figure S1**). The target LSD1was optimized using Monte Carlo simulation and energy optimizations.

S1.2 Pre-processing of three dimensional ligand databases

The external source database in the form of sdf format was processed using the ligand preparation tools. The final coordinates were stored in multi sdf files. Custom filters included Lipinski's rule-of-five (Ro5) and manual filtering to remove very large molecules, dimers, polymers, molecules containing unusual heteroatoms, and highly reactive functional groups. Each of the databases were combined together with a final library of ~2 million molecules commercially available from 26 vendors was considered for virtual screening using Glide SP/XP, ICM, and GOLD.

S1.3 Virtual Screening Method

A flow scheme indicating the steps of the Virtual Screening (VS) process is shown in **Figure S2.** The database of 13 million library compounds was curated using Ligprep, the filters from **Section S1.2**, and Glide HTVS methods to attain the set of ~ 2 million compounds screened against the prepared target. Grid potentials were rapidly generated which accounted for shape of the binding pocket, hydrophobicity, electrostatic potentials, and hydrogen-bonding profile. The compounds were screened using our own workflow (**Figure S2**) for LSD1 binding properties using a rigid target and flexible ligands in the internal coordinate's space. The compounds experimentally confirmed as LSD1 inhibitors were used for regular docking into LSD1. Docking calculations of the LSD1 inhibitors were performed using the ICM and Glide docking module with default setup and re-scoring with GOLD. The structures of the active compounds were energy minimized in the same environment and saved in PDB format. These energy-minimized inhibitors were then reposed into ICM and converted into ICM object, and MMFF charges were assigned for each of the ligand. Docking took an average of 3-4 min/molecule on a four AMD 64-bit processors RedHat linux server with 4 GB of RAM. The speed for each compound was dependent on the number of torsional degrees of freedom.

S1.4 Post processing and Compound Selection Criteria

Compounds having desired scores, hydrogen bond formation and hydrophobic interactions were estimated by interatomic distances for further analysis. The conformational stability of each candidate was also estimated by force field energy difference between the complexes conformation and freely minimized conformation, and the top-scoring candidates from this category were selected for further analysis. Compounds in each of the three categories were visually inspected to eliminate candidates

without ideal hydrogen bond geometry, hydrophobic molecular surfaces, or torsion angles. The resulting 121 focused screening structures were further analyzed using molecular property filters in QikProp, with calculated log S, permeability (Caco2 and MDCK) and Lipinski like criteria.

The commercially available hit compounds given in **Table 1** (**1-6**) and negative hits (**7-10**) was purchased from ChemBridge http://www.hit2lead.com and Enamine http://www.enamine.net. Their characterizations were confirmed using ^{1}H NMR and Mass and their purity was determined by HPLC.

Section S2. Purchased Hits – Analytical Data

The commercially available hit compounds given in **Table 1** (**1-10**) was purchased from ChemBridge http://www.hit2lead.com and Enamine http://www.enamine.net. Their characterizations were confirmed using ^{1}H NMR and Mass and their purity was determined by HPLC.

S2.1 (E)-4-hydroxy-N'-(2-hydroxybenzylidene)benzohydrazide (1): ^{1}H NMR (400 MHz, DMSO-d_6): δ 11.84 (s, 1H), 11.39 (s, 1H), 10.11 (s, 1H), 8.54 (s, 1H), 7.87 (m, 2H), 7.49 (d, 1H, J = 8.4 Hz), 7.28 (t, 1H, J = 8.4 Hz), 6.91 (m, 4H). ESI-MS: 256.2 $[M+H]^+$.

S2.2 (E)-N'-(5-chloro-2-hydroxybenzylidene)-4-hydroxybenzohydrazide (2): ^{1}H NMR (400 MHz, DMSO-d_6): δ 11.89 (s, 1H), 11.32 (bs, 1H), 10.04 (s, 1H), 8.57 (s, 1H),7.82 (m, 2H), 7.61 (d, 1H, J = 2.4 Hz), 7.29 (dd, 1H, J = 2.4 & 8.8 Hz), 6.95 (d, 1H, J = 8.4 Hz), 6.88 (m, 2H). ESI-MS: 290.7 $[M+H]^+$.

S2.3 (E)-4-hydroxy-N'-(1-(2-hydroxyphenyl)ethylidene)benzohydrazide (3): ^{1}H NMR (400 MHz, DMSO-d_6): δ 10.94 (s, 1H), 10.03 (bs, 1H), 7.84 (m, 2H), 7.62 (dd, 1H, J = 1.6 & 8.0 Hz), 7.29 (t, 1H, J = 8.4 Hz), 6.90 (m, 4H), 2.47 (s, 3H). ESI-MS: 270.28 $[M+H]^+$.

S2.4 (E)-4-bromo-N'-(2-hydroxybenzylidene)benzohydrazide (4): ^{1}H NMR (400 MHz, DMSO-d_6): δ 12.01 (s, 1H), 11.15 (s, 1H), 8.62 (s, 1H), 7.88 (m, 2H), 7.75 (m, 2H), 7.53 (d, 1H, J = 8.8 Hz), 7.31 (t, 1H, J = 8.8 Hz), 6.93 (m, 2H). ESI-MS: 319.16 $[M+H]^+$.

S2.5 (E)-3-chloro-N'-(2-hydroxybenzylidene)benzohydrazide (5): ^{1}H NMR (400 MHz, DMSO-d_6): δ 12.09 (s, 1H), 11.12 (s, 1H), 8.64 (s, 1H), 7.98 (s, 1H), 7.90 (m, 1H), 7.67 (m, 1H), 7.57 (m, 2H), 7.31 (t, 1H, J = 7.6 Hz), 6.93 (m, 2H). ESI-MS: 274.70 $[M+H]^+$.

S2.6 (E)-N'-(1-(2-hydroxyphenyl)ethylidene)-3-(morpholinosulfonyl)benzohydrazide (6): ESI-MS: 403.4 $[M+H]^+$, purity by HPLC 97.25%.

S2.7 (E)-3-(morpholinosulfonyl)-N'-(1-(naphthalen-1-yl)ethylidene)benzohydrazide (7): ESI-MS: 437.5 $[M+H]^+$.

S2.8 5-chloro-N'-(2-fluoro-5-(morpholinosulfonyl)benzoyl)-2-methoxybenzohydrazide (8): ^{1}H NMR (400 MHz, DMSO-d_6): δ 7.90 (m, 3H), 7.51 (m, 2H), 7.22 (m, 1H), 4.02 (s, 3H), 3.71 (m, 4H), 2.97 (m, 4H). ESI-MS: 471.8 $[M+H]^+$.

S2.9 N'-(3-chlorobenzoyl)-2-fluoro-5-(morpholinosulfonyl)benzohydrazide (9): ^{1}H NMR (400 MHz, DMSO-d_6): δ 8.01 (m, 2H), 7.92 (m, 2H), 7.53 (m, 3H), 3.69 (m, 4H), 2.96 (m, 4H). ESI-MS: 441.8 $[M+H]^+$.

S2.10 (E)-N,N-diethyl-3-(2-(1-(p-tolyl)ethylidene)hydrazinecarbonyl)benzenesulfonamide (10): ESI-MS: 387.5 $[M+H]^+$, purity by HPLC 94%.

Supplementary Tables

Table 1. Docking scores of compounds **1-10**

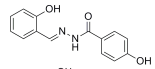
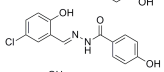
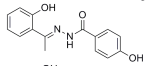
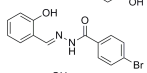
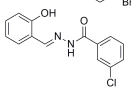
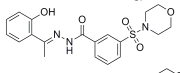
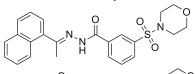
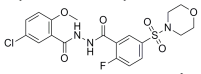
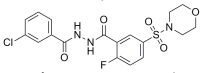
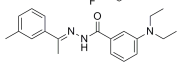
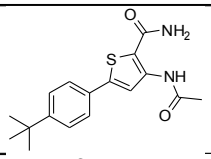
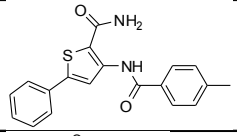
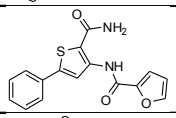
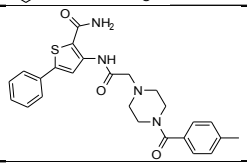
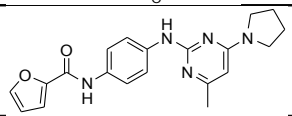
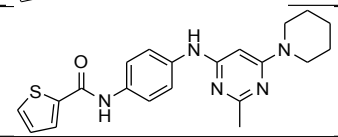
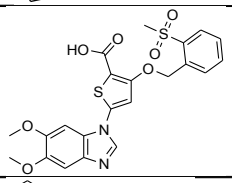
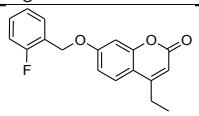
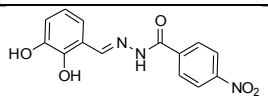
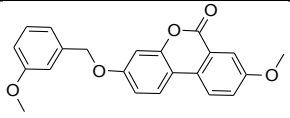
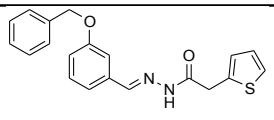
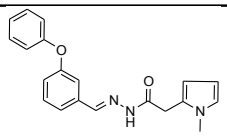
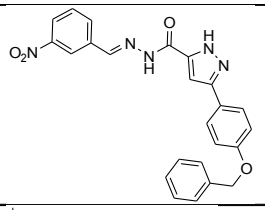
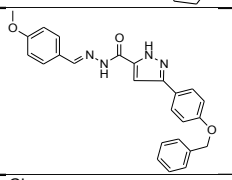
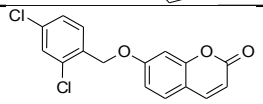
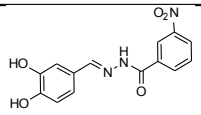
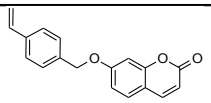
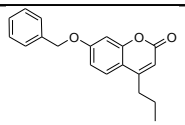
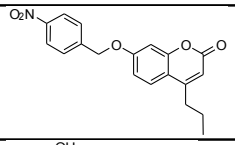
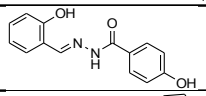
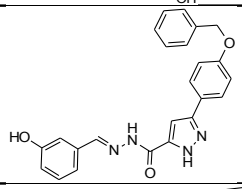
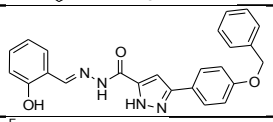
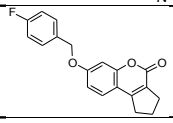
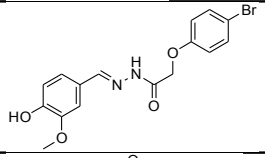
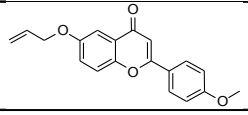
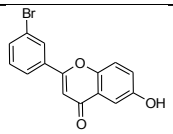
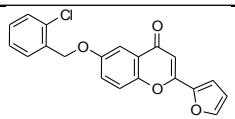
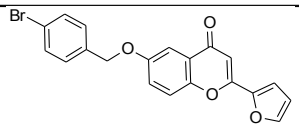
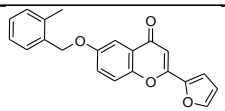
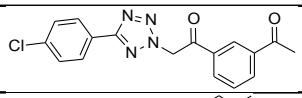
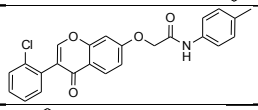
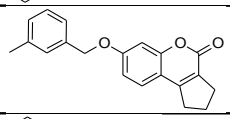
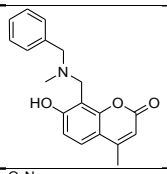
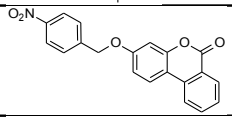
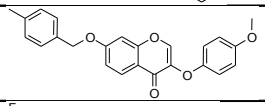
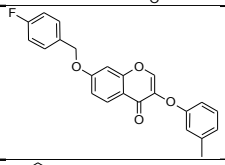
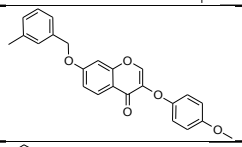
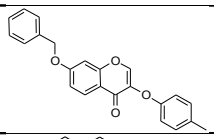
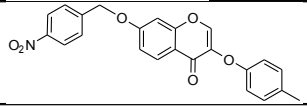
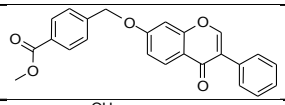
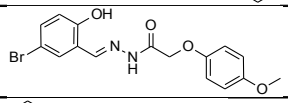
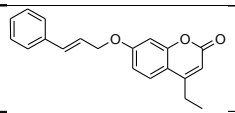
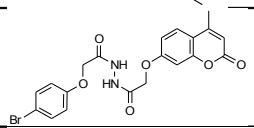
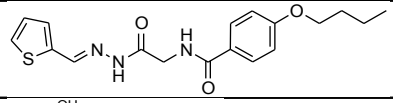
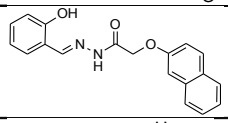
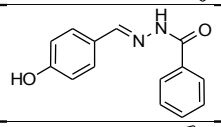
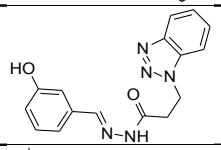
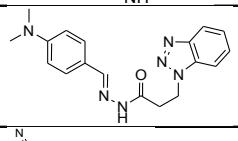
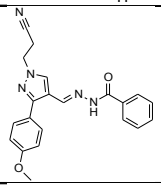
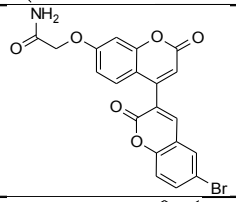
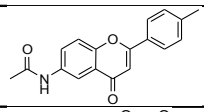
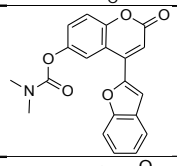
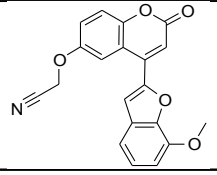
S. No	Structure	ICM score	Glide score	Gold fitness score
1		-42.25	-8.14	56.26
2		-42.25	-7.92	58.21
3		-21.91	-7.87	51.29
4		-37.77	-8.64	57.69
5		-36.3	-8.84	47.98
6		-24	-6.26	43.26
7		-20.97	-6.14	46.64
8		-18.39	-6.63	49.93
9		-8.16	-7.21	41.86
10		-8.5	-6.81	52.19

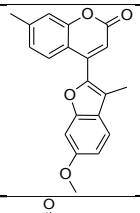
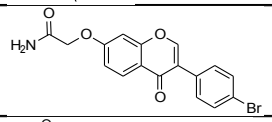
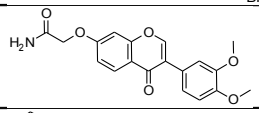
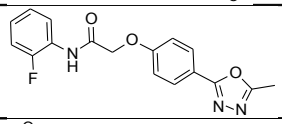
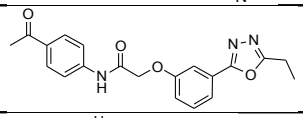
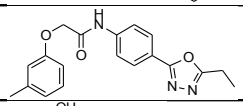
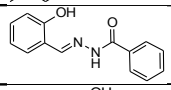
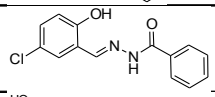
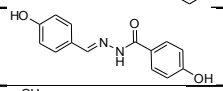
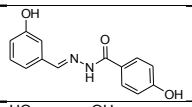
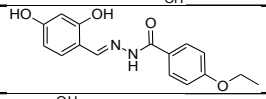
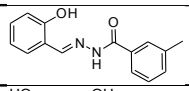
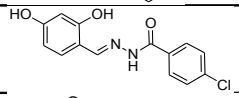
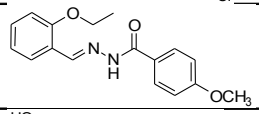
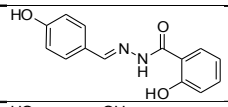
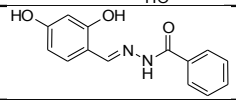
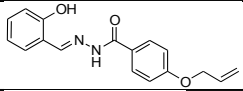
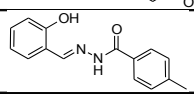
Table S2. Commercially available LSD1 hits (111) from the list of 121 compounds selected.

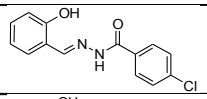
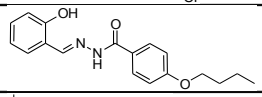
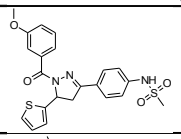
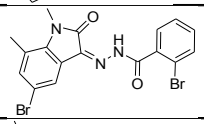
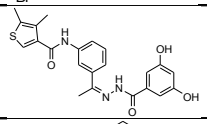
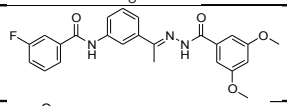
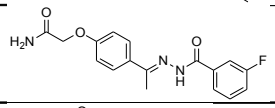
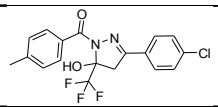
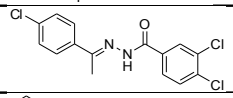
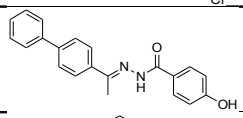
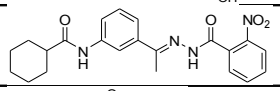
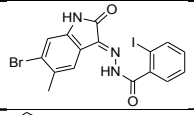
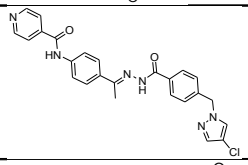
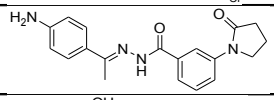
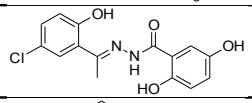
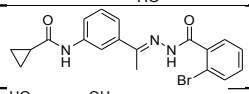
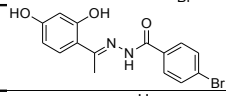
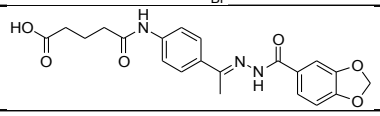
S. No	Structure	IC50 (μM) LSD1	ICM Score (kcal/mol)	GLIDE Score (kcal/mol)	GOLD Fitness Score (kcal/mol)
11		>100	-16.89	-4.87	27.21
12		>100	-16.34	-4.89	29.21
13		>100	-16.21	-4.76	24.21
14		>100	-21.21	-5.27	28.23
15		12.2	-17.22	-5.12	18.21
16		18.6	-26.81	-6.96	28.21
17		67.3	-27.28	-5.23	29.81
18		>100	-17.79	-7.43	22.74
19		>100	-14.34	-5.99	31.04
20		>100	-17.24	-4.76	20.17
21		>100	-28.21	-6.29	30.61
22		>100	-26.24	-7.14	34.82

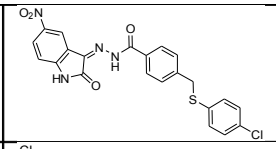
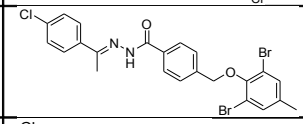
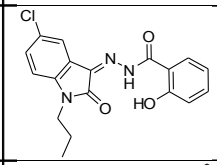
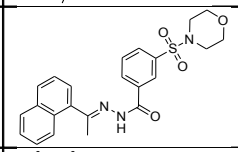
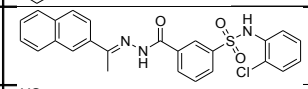
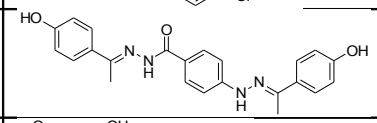
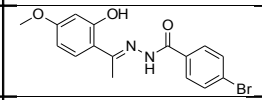
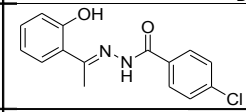
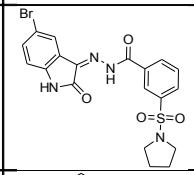
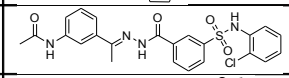
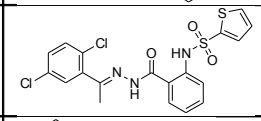
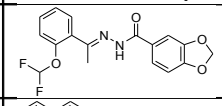
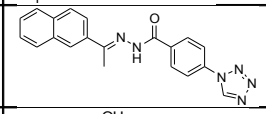
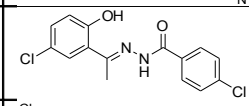
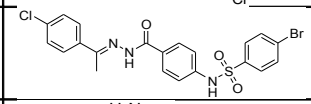
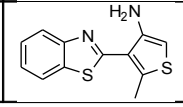
23		>100	-23.23	-5.86	32.76
24		>100	-21.28	-6.13	32.52
25		>100	-14.93	-4.21	20.12
26		>100	-13.34	-7.24	32.61
27		>100	-11.21	-6.21	25.21
28		>100	-11.29	-5.34	23.78
29		>100	-16.25	-4.88	30.22
30		0.196	-42.25	-8.14	56.26
31		22	-27.29	-6.77	31.55
32		>100	-21.89	-6.69	32.31
33		>100	-13.29	-6.86	39.03
34		>100	-22.95	-6.29	37.19
35		37	-21.38	-7.22	25.94

36		17	-19.29	-8.66	29.93
37		>100	-13.12	-5.13	22.14
38		>100	-17.37	-4.77	25.65
39		>100	-18.58	-4.86	25.92
40		>100	-16.43	-5.16	22.74
41		>100	-19.99	-6.16	21.04
42		>100	-16.19	-4.42	20.17
43		>100	-17.23	-5.66	20.61
44		>100	-13.87	-3.33	24.82
45		>100	-11.81	-5.77	22.76
46		>100	-17.99	-5.99	22.52
47		>100	-14.39	-4.96	20.12
48		>100	-53.29	-4.77	22.61
49		>100	-17.64	-4.52	25.21

50		>100	-17.31	-4.86	23.78
51		>100	-21.28	-7.33	30.22
52		>100	-19.73	-4.97	31.55
53		>100	-17.34	-3.59	22.31
54		>100	-20.21	-6.76	39.03
55		>100	-26.29	-5.23	37.19
56		>100	-26.25	-6.22	35.94
57		67	-21.28	-6.66	39.93
58		>100	-19.33	-6.33	32.14
59		>100	-29.84	-5.67	25.65
60		>100	-16.23	-4.19	25.92
61		>100	-11.97	-3.16	19.22
62		>100	-14.27	-3.17	20.71
63		18	-21.88	-4.42	22.23

64		>100	-17.13	-4.22	22.07
65		>100	-16.55	-5.13	26.62
66		>100	-17.11	-4.37	30.49
67		>100	-19.39	-2.79	33.71
68		>100	-16.87	-4.69	31.58
69		32	-21.88	-3.17	30.98
70		>1 uM	-24.43	-6.52	30.62
71		>1 uM	-23.94	-6.33	30.97
72		>1 uM	-21.41	-7.23	31.28
73		>1 uM	-21.99	-9.47	32.23
74		>1 uM	-26.25	-8.99	35.26
75		>1 uM	-29.18	-7.79	36.75
76		>1 uM	-24.23	-7.17	30.42
77		>1 uM	-23.37	-7.43	38.68
78		>1 uM	-21.81	-7.46	30.59
79		>1 uM	-26.54	-7.13	30.29
80		>1 uM	-26.45	-8.17	35.82
81		>1 uM	-27.31	-8.21	38.72

82		>1 uM	-26.99	-7.06	31.62
83		>1 uM	-26.35	-6.20	30.01
84		>10 uM	-28.18	-6.42	31.67
84		>10 uM	-22.33	-8.93	31.27
86		>10 uM	-26.39	-8.13	34.82
87		>10 uM	-31.96	-6.17	30.89
88		>10 uM	-29.64	-6.86	31.04
89		>10 uM	-21.75	-7.36	33.14
90		>10 uM	-29.81	-7.77	32.02
91		>10 uM	-26.79	-7.44	33.02
92		>10 uM	-32.55	-8.16	33.69
93		>10 uM	-19.28	-6.29	34.21
94		>10 uM	-25.66	-7.67	31.48
95		>10 uM	-21.77	-8.16	37.94
96		>10 uM	-23.61	-7.16	33.75
97		>10 uM	-29.59	-7.97	30.41
98		>10 uM	-29.435	-8.89	32.92
99		>10 uM	-31.41	-6.16	33.19

100		>10 uM	-31.89	-.923	32.41
101		>10 uM	-36.29	-7.87	34.16
102		>10 uM	-24.19	-7.22	33.67
103		>10 uM	-27.11	-7.66	34.92
104		>10 uM	-22.17	-7.17	31.97
105		>10 uM	-32.77	-6.95	31.62
106		>10 uM	-36.74	-6.16	30.52
107		>10 uM	-36.75	-6.14	30.65
108		>10 uM	-31.56	-7.27	41.62
109		>10 uM	-21.87	-8.19	31.56
110		>10 uM	-34.21	-8.79	32.61
111		>10 uM	-39.88	-8.29	33.41
112		>10 uM	-34.13	-7.42	37.79
113		>10 uM	-33.39	-6.76	32.73
114		>10 uM	-31.21	-8.29	32.51
115		>100 uM	-16.44	-4.19	21.99

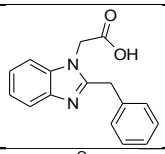
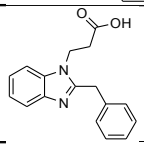
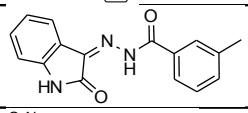
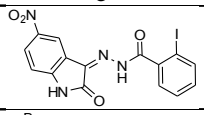
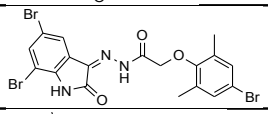
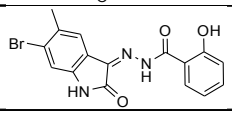
116		>10 uM	-24.21	-7.39	32.08
117		>10 uM	-29.78	-6.49	35.05
118		>10 uM	-24.43	-6.41	30.12
119		>10 uM	-23.89	-7.99	32.45
120		>10 uM	-21.29	-6.16	22.08
121		>10 uM	-16.74	-5.19	25.05

Table S3. Tanimoto similarity coefficients comparing compound **12** and known LSD1 inhibitors from **Chart 1**

Compound	Tanimoto Similarity score
A	0.26
B	0.21
C	0.31
D	0.26
E	0.22
F	0.36
G	0.24
H	0.28
I	0.38
J	0.35
K	0.39
L	0.29
M	0.11
N	0.11
O	0.32
P	0.11

Table S4. Off-target inhibition

[12] (nM)	Activity % ± SD					
	CYP1A2	**CYP2C19**	**CYP2C9**	**CYP2D6**	**CYP3A4**	**hERG**
30000						98±4
10000	60±3	53±1	46±2	94±2	21±2	104±2
3333	86±1	91±2	84±1	101±11	46±2	100±6
1111	87±2	105±0.4	98±3	102±10	71±4	102±7
370	96±1	111±3n/a	108±0.2	105±12	87±10	98±3
123	97±8	110±0.1	111±1	110±8	100±3	95±4
41.2	93±3	107±12	107±5	105±12	98±9	98±2
13.7	105±7	114±1	109±1	108±6	98±1	107±0.1
4.57	106±4	106±11	107±8	104±12	99±9	93±2
1.52	111±3	117±3	112±1	109±6	105±1	100±2
0.51	110±2	89±3	107±6	106±2	96±5	

	D-LDH	**GO**
100000	103±7	105±3
30000	103±6	106±3
10000	106±5	108±6
3000	112±5	92±3
1000	121±6	97±0.3
300	118±5	109±3
100	119±5	102±1

Table S5. Comparison of different model fits for enzyme kinetics

Model	Competitive	Noncompetitive	Uncompetitive
v_{max} (F/s) ± SE	635.8±12.83	688.9±10.80	695.8±11.67
K_m (μM) ± SE	0.919±0.1128	1.310±0.0928	1.411±0.1042
k_i (nM) ± SE	4.136±0.7027	39.04±3.046	32.76±2.67
DMSO R^2	0.9033	0.9269	0.9275
1 nM R^2	0.8788	0.8774	0.8758
3 nM R^2	0.9022	0.9077	0.9076
10 nMR^2	0.8492	0.9154	0.9060
30 nM R^2	0.2257	0.9289	0.9146
100 nM R^2	-0.6184	0.6953	0.5652
Global R^2	0.8599	0.9239	0.9198

Supplementary Figures.

Figure S1. Binding Site Model and definition of active site of LSD1 structure.

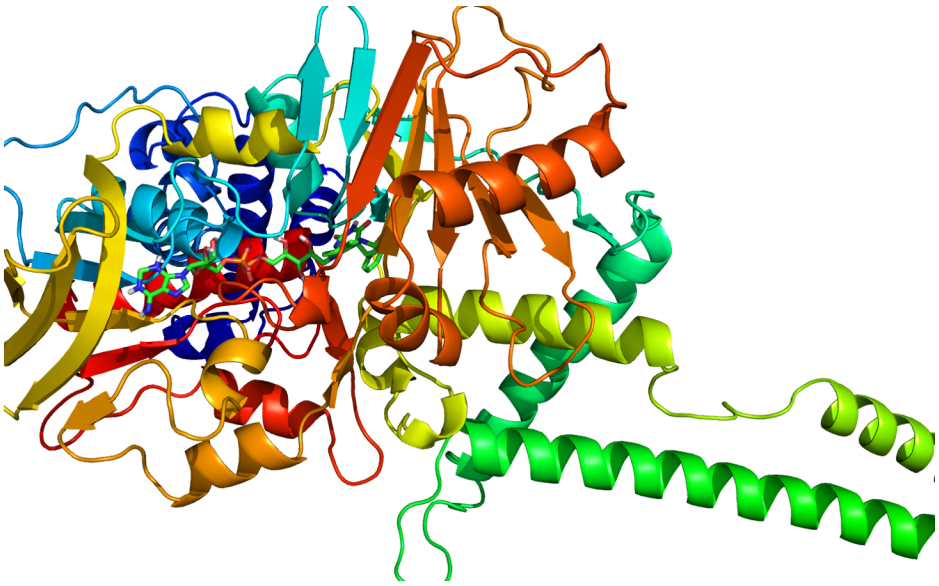


Figure S2. Flow diagram for the Virtual Ligand Screening (VLS) using ICM-VLS, Schrodinger workflow GOLD programs.

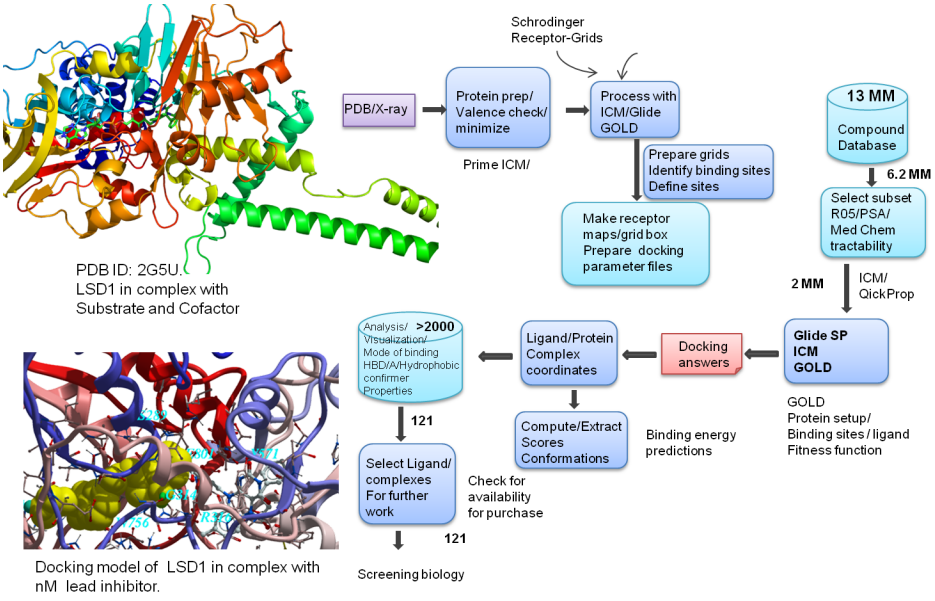


Figure S3. Complete reaction schemes for compounds **11-22**

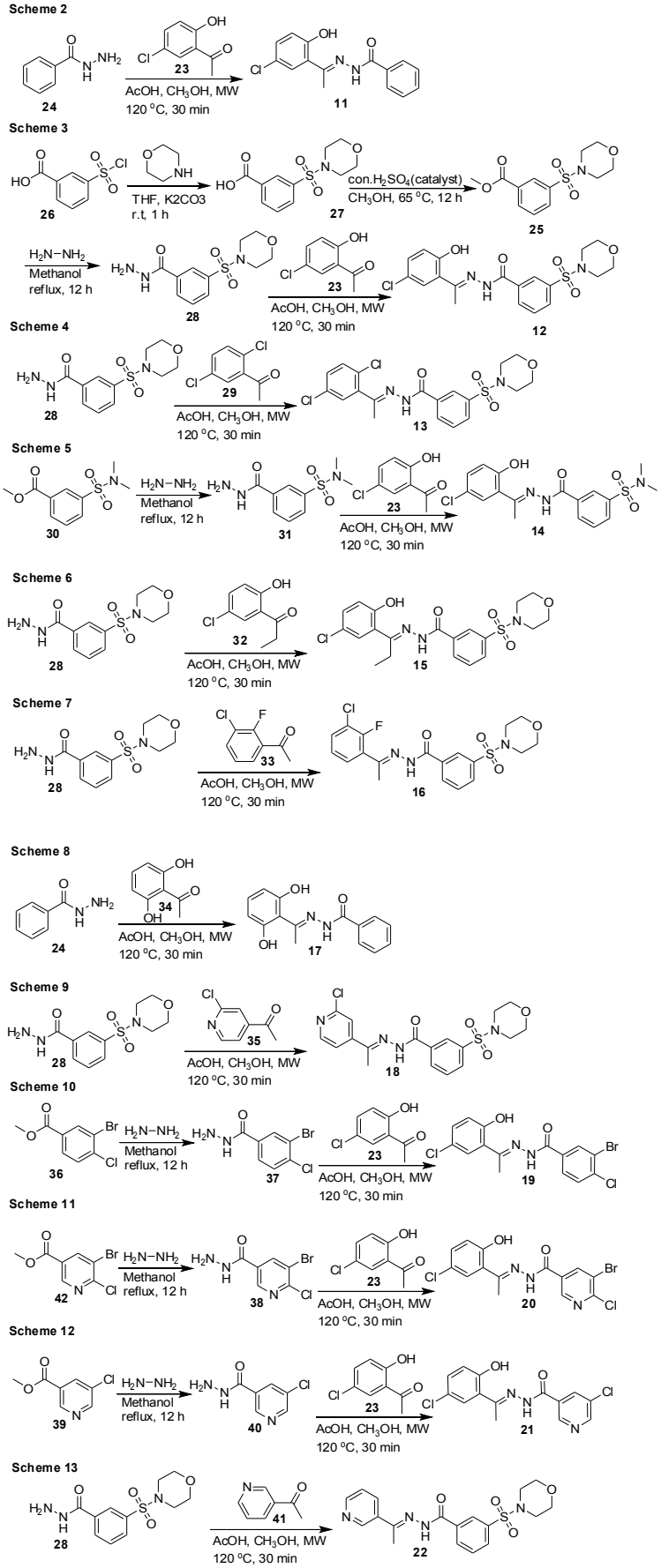


Figure S4. LC-MS Data for Compound 12 (96% purity).