Supplementary data

ACFIS 2.0: an improved web-server for fragment-based drug discovery via a dynamic screening strategy

Xing-Xing Shi[1, †], Zhi-Zheng Wang[1, †], Fan Wang[1, †], Ge-Fei Hao[1, *] and Guang-Fu Yang[1, *]

[†] National Key Laboratory of Green Pesticide, Key Laboratory of Pesticide & Chemical Biology, Ministry of Education, Central China Normal University, Wuhan 430079, P.R. China

* To whom correspondence should be addressed. Email: gfyang@mail.ccnu.edu.cn (Guang-Fu Yang), gefei_hao@foxmail.com (Ge-Fei Hao)

† Joint First Authors

Text S1. Materials and methods

1. Fragment Deconstruction

The fragment deconstruction analysis starts from a protein-ligand complex in PDB format. Firstly, an energy minimization was performed using Amber 16 to optimize the interaction between protein and ligand [1]. The, ligand structure binding in the pocket is deconstructed into fragments according to the retrosynthetic analysis by using DAIM software [2]. Single bond is broken and hydrogen is used to link with heavy atom to make the total charge value of each 'piece' integer.

2. Core Fragment Identification.

After fragment deconstruction, the binding free energy (ΔG) is calculated for each protein–fragment structure using the combination of the MM_PBSA method [3] for the enthalpy and an empirical method for the entropy [4] (Equation 1). Then, ligand efficiency (LE) which is defined as absolute value of ΔG divided by the heavy atom count (HAC) is obtained. All fragments deconstructed from original ligand is sorted based on LE. Fragment with highest LE value is generally identified as core fragment that theoretically has a highly conserved binding conformation and efficient contribution to the entire binding affinity.

$$\Delta G_{bind} = \Delta E_{gas} + \Delta G_{sol} - T\Delta S = \Delta E_{bind} - T\,(\Delta S_{sol} + \Delta S_{conf}) \quad (1)$$

3. Dynamic Fragment Growing.

Dynamic fragment growing begins with the protein-core fragment complex structure. At first, energy minimization and molecular dynamics (MD) simulation are performed on protein-core fragment complex using AMBER16 package. Topology and coordinate files is created in the tleap module based on the Amber ff14SB force field [5] and general AMBER force field (GAFF) [6]. The complex system is solvated in an octahedron box of TIP3P water, with at least 10 Å between the solute and each box edge [7]. Counter ions, Na+ or Cl−, is added to neutralize the net charges of the system. The energy minimization was achieved through three stages: (i) only water molecules and ions are allowed to move, (ii) backbone atoms of protein are restrained and the remaining atoms are allowed to move, (iii) all atoms in the system could move freely. In each stage, 1000 steepest descent steps and 1000 conjugated gradient steps with a convergence criterion of 0.1 kcal mol^{-1} $Å^{-1}$ were carried out in a vacuum. The MD simulation process using explicit-solvent particle mesh Ewald (PME) model is achieved through two stages: (i) the system is heated from 0 to 300K over 500 ps in the NVT ensemble with restraints on the solute (ii) 0.5 ns of MD in the NTP (T = 300 K and P = 1 atm) ensemble is carried out four times successively (4 x 0.5ns = 2ns), saving the structure every 2 ps

Subsequently, RMSD-based clustering is performed on complex structures extracted from entire merged MD trajectories using the cpptraj module in AMBER16. The RMSD cutoff for the neighboring cluster is set as 0.5 Å. A complex structure is randomly picked from each of the top five clusters with the

largest number of conformations, thereby obtaining an ensemble of protein-core fragment complex conformations with large structural difference.

Finally, based on typical protein-core fragment complex structure, new fragment in selected fragment library is linked to the junction of core fragment placed in the binding site using AutoGrow2.0 [8]. The orientation of growing fragment was optimized with minimum steric clashes (overlap volume ≤ 4 $Å^3$) to the surrounding residues. For generated protein-ligand complex structure, energy minimization is performed using the Sander module of Amber16, and binding free energy (ΔG) is calculated using the combination of the MM_PBSA method [3] for the enthalpy and an empirical method for the entropy [4] (Equation 1). Among a set of complex conformations containing the same ligand, the conformation with the lowest ΔG value is used the ultimate complex structure for subsequent ligand comparisons.

4. Molecular Property Evaluation

At the step of comparing generated ligands, a series of molecular properties are predicted for each ligand, including binding affinity, physicochemical properties, drug-likeness, pesticide-likeness and synthetic accessibility. Binding affinity is predicted through binding free energy calculation using the combination of the MM_PBSA method [3] for the enthalpy and an empirical method for the entropy [4] (Equation 1). The physicochemical properties are calculated using Mordred software [9], covering molecular weight (MW), hydrogen bond receptor (HBA), hydrogen bond donor (HBD), water partition coefficient (LogP), and so on. The drug-likeness is evaluated according to physicochemical properties, using qualitative Lipinski's rule of five [10], Ghose's rule [11], as well as the quantitative QED score [12]. The pesticide-likeness was evaluated according to a qualitative rule called Hao's rule [13]. In addition, the synthetic accessibility score is evaluated using RDKit [14].

Text S2. Description of validation protocol for ACFIS 2.0

1. Dataset preparation

As mentioned in the main text, a series of compounds that have been reported with experimental binding affinities or biological activity were collected as test set for the performance validation of ACFIS 2.0. A total of 122 compounds was selected as test objects after literature investigation. Each compound in the test set was required to meet the following restrictions:

1) There is available experimental data on its binding affinity or biological activity.

2) The qualitative conclusion on its affinity or activity could be drawn (e.g. Compound A showed high affinity for X protein, while Compound B exhibited low affinity for X protein) based on the responding reference.

3) The crystal structure of its protein-ligand complex (at least its receptor protein) is available.

4) The core fragment in the compound could be identified based on the responding reference.

Targeting a specific protein, a pair of compounds were generally collected, in which one had high affinity/ activity and the other one showed relatively low affinity/ activity. The case in which the tested compound exhibited relatively high affinity/ activity was classified into the positive sample group, otherwise it is classified as a negative sample group. Overall, the test set was composed of 61 positive samples and 61 negative samples. The data of 122 cases were provided in Table S1.

2. Calculation for each case using ACFIS 2.0

For each case/ sample, a calculation task in CAND_GEN mode was designed to evaluate the predictive performance of ACFIS 2.0. The inputs of a task consisted of a corresponding core fragment-receptor protein complex structure (obtained by deconstructing the protein-ligand complex reported in PDBbind database) and the fragment library built specifically for performance testing (see Figure S3). For each task, ACFIS 2.0 outputted a list of all generated ligands, in which the ligands were sorted by their predicted binding affinities from strong to weak (by their predicted binding free energies from low and high). In this list, the tested compound and its paired compound were designedly contained.

3. Performance evaluation for ACFIS 2.0

In each calculation task (mentioned above), ACFIS 2.0 was expected to reasonably sort all generated ligands and assign the tested compound with an appropriate rank. In principle, the tested compounds that had been experimentally measured to have high affinity or activity should be at the top of the list, while those with low experimental affinity/ activity should be ranked after its paired compound with higher affinity/ activity.

For positive samples, the top 20% was manually set as the ranking threshold. If a test compound ranks in the top 20% of the outputted complete list, it is computationally considered to have high affinity/

activity, and the corresponding case is predicted as positive. Otherwise, the corresponding case is predicted as negative.

For negative samples, if a test compound ranks after its paired compound in the outputted list, the corresponding case is predicted as negative. Otherwise, the corresponding case is predicted as positive.

After all testing tasks were completed, calculation results were compared with experimental findings. According to our statistics, 55 was true positive (TP), 8 was false positive (FP), 6 was false negative (FN), and 53 was true negative (TN). Here are the definitions of these elements:

1) TP refers to the number of positive samples correctly predicted as positive.

2) FP refers to the number of negative samples incorrectly predicted as positive.

3) FN refers to the number of positive samples incorrectly predicted as negative.

4) TN refers to the number of negative samples incorrectly predicted as negative.

At last, the accuracy, precision, sensitivity, specificity, and negative positive value (NPV) of ACFIS 2.0 in this performance testing were measured.

Accuracy = (TP+TN)/(TP+FP+FN+TN) = 0.885

Precision = TP/(TP+FP) = 0.873

Sensitivity = TP/(TP+FN) = 0.901

Specificity = TN/(FP+TN) = 0.868

NPV = TN/(TN+FN) = 0.898

		Experimental		
		Positive	Negative	
Predicted	Positive	45 (TP[1])	14 (FP[2])	Precision 0.762
	Negative	16 (FN[3])	47 (TN[4])	NPV[5] 0.746
		Sensitivity[6]	Specificity	Accuracy
		0.738	0.770	0.754

[1] True Positive [2] False Positive [3] False Negative
[4] True Negative [5] Negative Predictive Value [6] Also called "Recall"

Figure S1. The performance of ACFIS (old version) when using the same test set as ACFIS 2.0

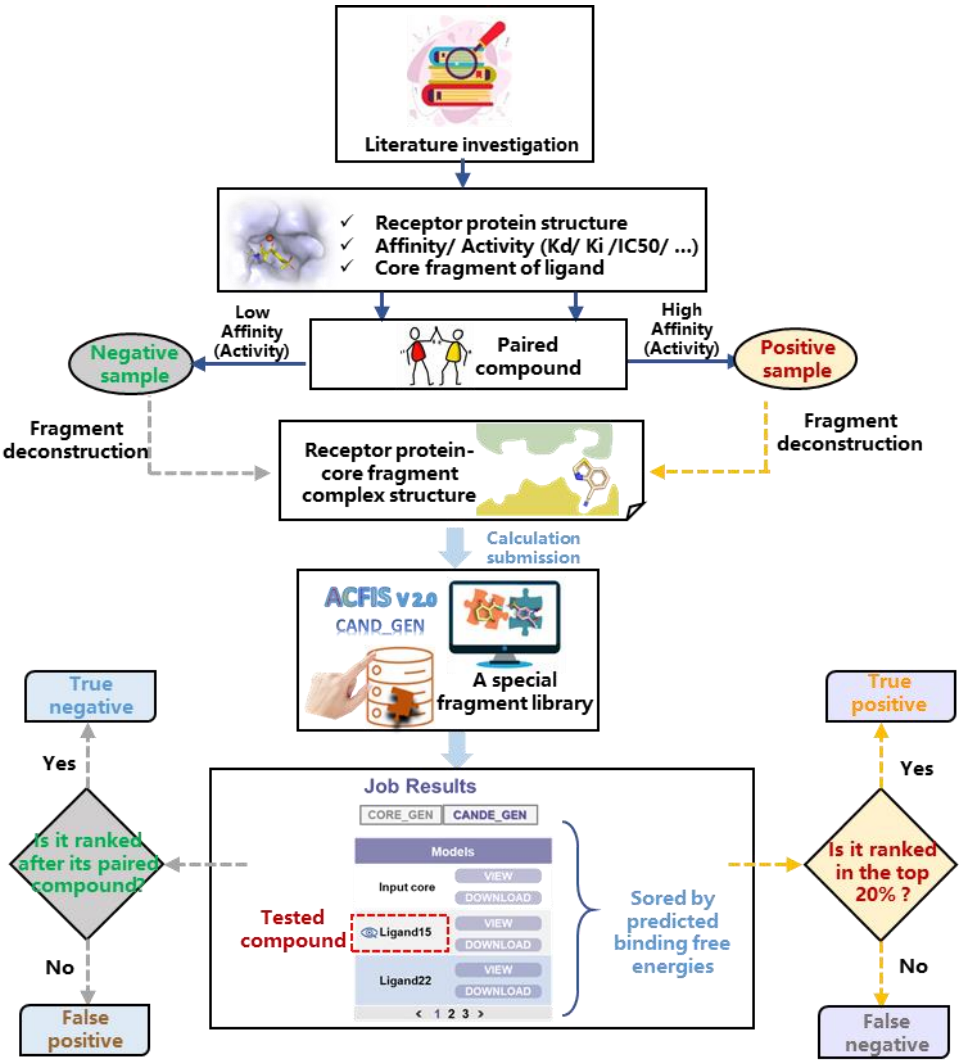


Figure S2. Validation workflow for ACFIS 2.0

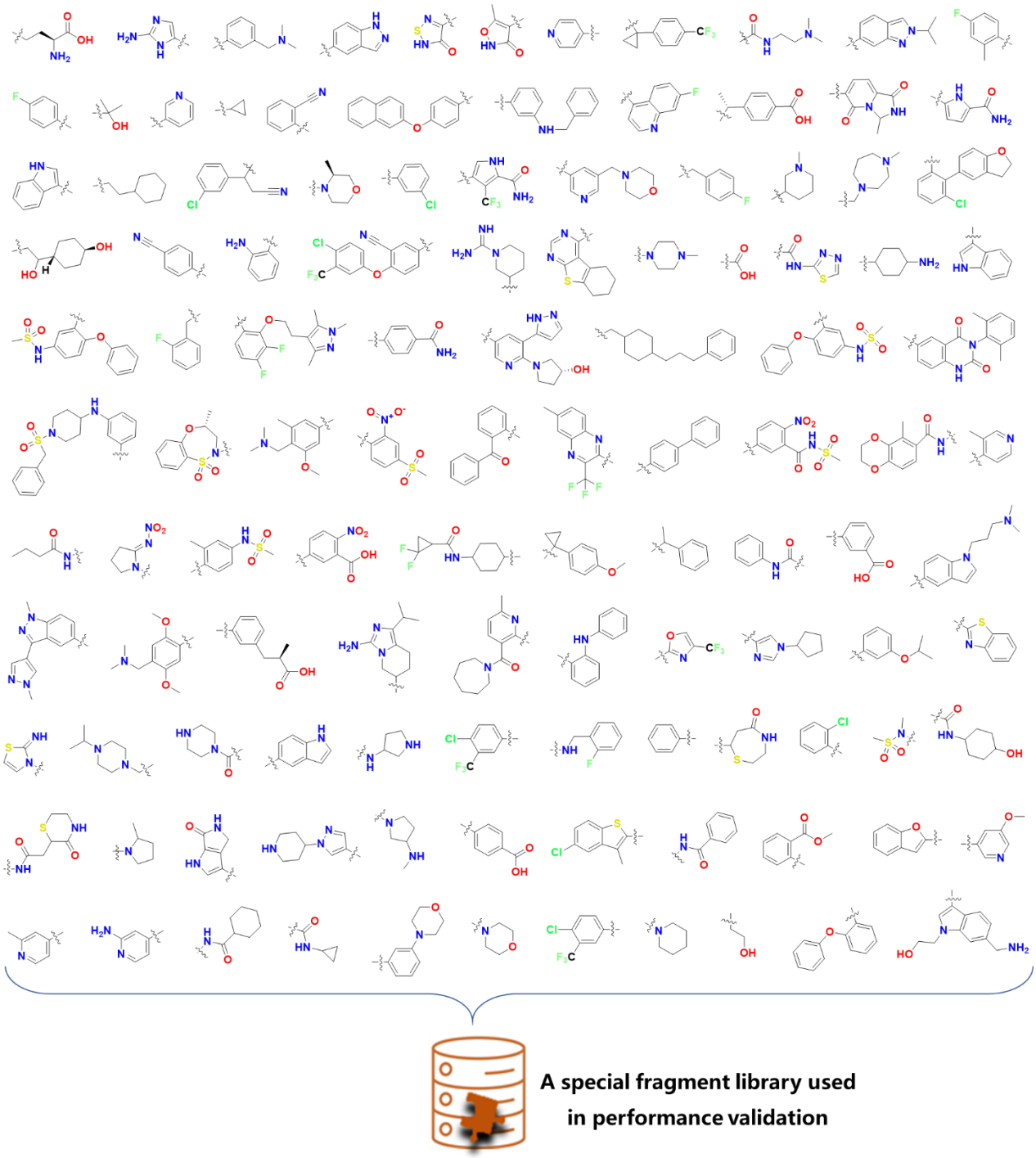
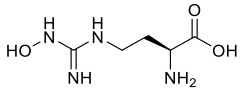
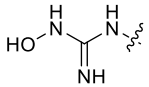
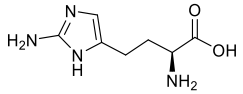
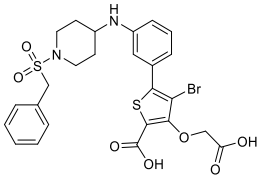
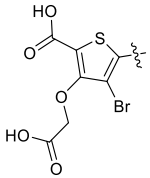
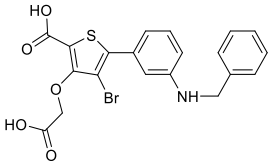
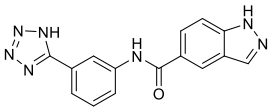
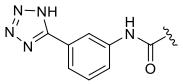
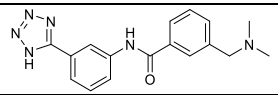
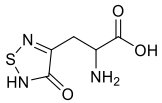
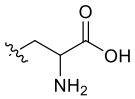
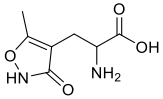
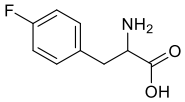
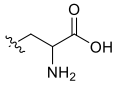
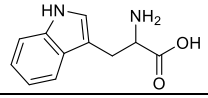
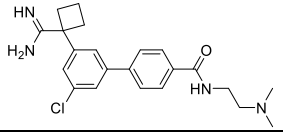
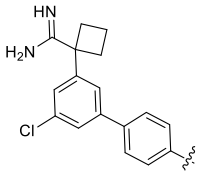
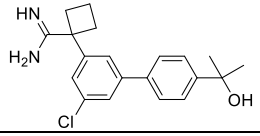
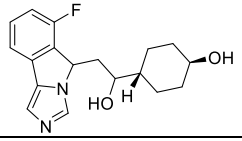
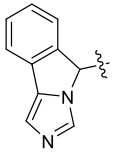
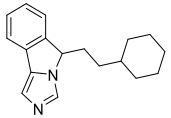
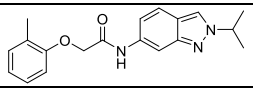
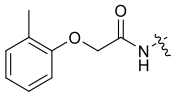
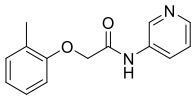
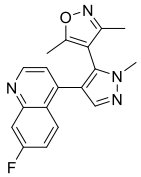
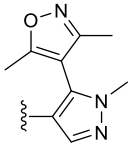
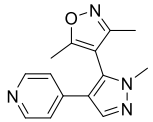
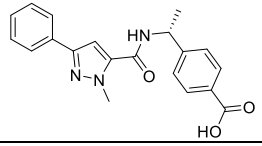
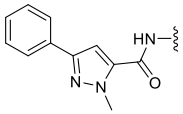
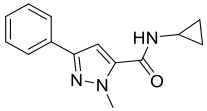
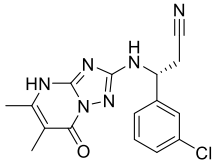
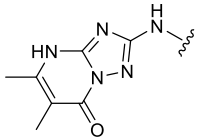
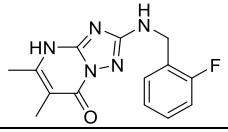
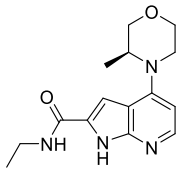
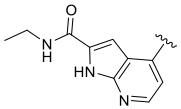
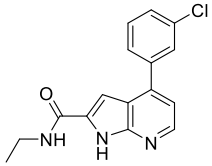
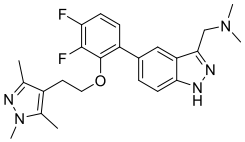
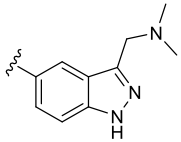


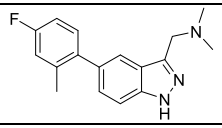
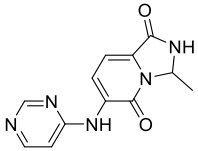
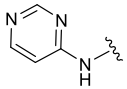
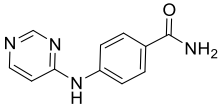
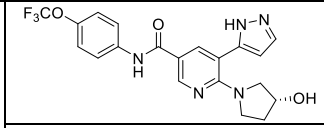
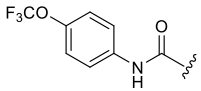
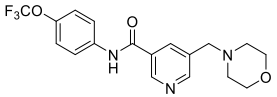
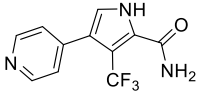
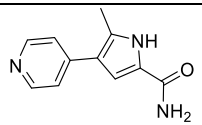
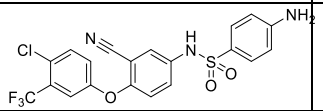
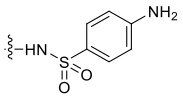
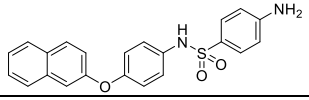
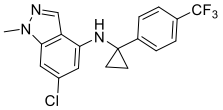
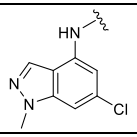
Figure S3. Fragment library used in ACFIS 2.0 performance validation

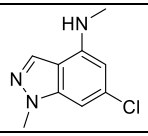
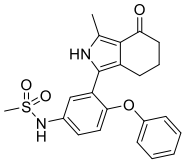
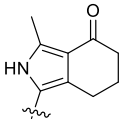
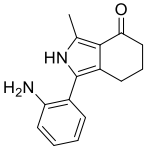
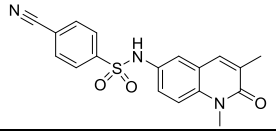
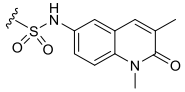
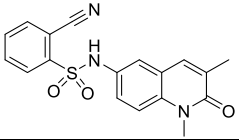
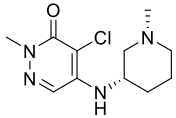
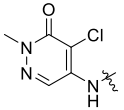
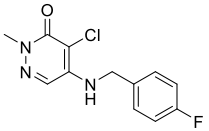
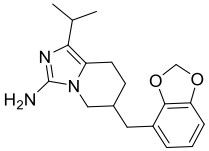
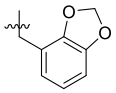
Table S1. **The test set of 122 cases used for performance validation and the predictive results of ACFIS 2.0/ ACFIS for these cases**

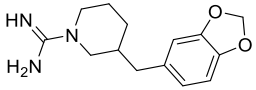
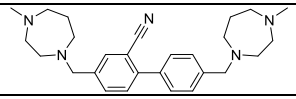
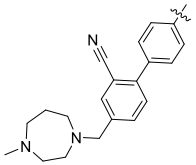
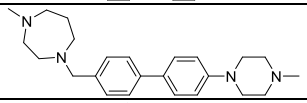
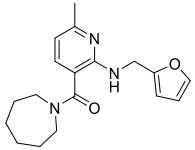
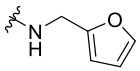
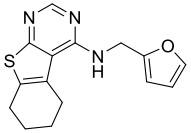
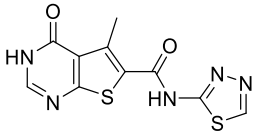
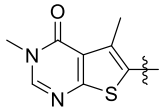
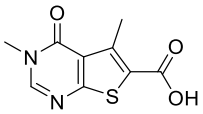
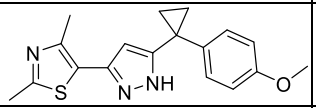
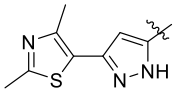
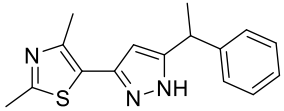
No.	Protein Name	Tested Compound	Experimental Data				Inputted Core Structure	Computational Data (ACFIS 2.0)			Computational Data (ACFIS 1.0)		
			Affinity / Activity	PDB of its complex structure	Ref (DOI)	Sample Group		Rank of Test Com.	Qualitative Classification	Fact	Rank of Test Com.	Qualitative Classification	Fact
1	Arginase-1		*K*d =47.51 nM (High Affinity)	3KV2	10.1016/j.abb.2010.02.004	Positive		17	Positive	True	18	Positive	True
2			*K*d =3 nM (Low Affinity)	3MFV	10.1021/jm100306a	Negative		68	Negative	True	10	Positive	False
3	Protein-tyrosine phosphatase 1b		*K*i =0.004 uM (High Activity)	2QBP	10.1021/jm0702478	Positive		14	Positive	True	11	Positive	True
4			*K*i =0.47 uM (Low Activity)	2QBP		Negative		26	Negative	True	43	Negative	True
5	β-lactamase		*Ki* =1.1 uM (High Activity)	4DE1	10.1021/jm2014138	Positive		15	Positive	True	56	Negative	False
6			*Ki* =76.0 uM (Low Activity)	4DE2		Negative		86	Negative	True	1	Positive	False
7	Glutamate receptor 2		*Ki* =531 nM (High Activity)	3BFU	10.1021/jm701126w	Positive		12	Positive	True	14	Positive	True
8			*Kd* =12.8 uM (Low Affinity)	1P1Q	10.1073/pnas.1037393100	Negative		39	Negative	True	80	Negative	True

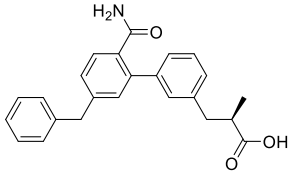
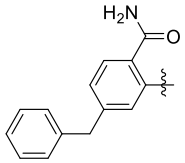
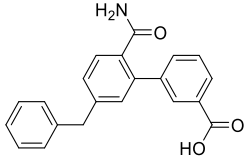
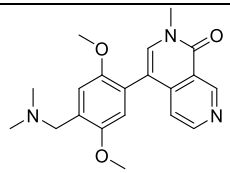
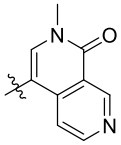
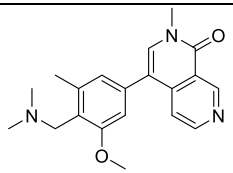
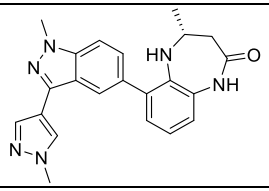
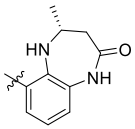
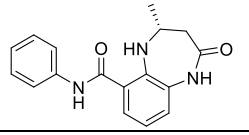
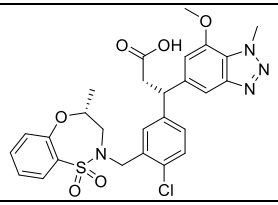
9	Transporter		*Ki* =950 nM (High Activity)	3F3C	10.1126/science.1166777	Positive		16	Positive	True	77	Negative	False
10			*Ki* =64.8 uM (High Activity)	3F3A		Negative		77	Negative	True	39	Positive	False
11	Apolipoprotein E4		*Kd* < 5 uM (High Affinity)		10.1021/acs.jmedchem.9b00178	Positive		51	Negative	False	15	Positive	True
12			*Kd* =30 uM (Low Affinity)	6NCO		Negative		49	Positive	False	62	Negative	True
13	Indoleamine 2,3-dioxygenase 1		IC_{50} =0.028 uM (High Activity)	6O3I	10.1021/acs.jmedchem.9b00662	Positive		12	Positive	True	6	Positive	True
14			IC_{50} =0.135 uM (Low Activity)			Negative		11	Positive	False	43	Negative	True
15	Palmitoleoyl-protein carboxylesterase NOTUM		IC_{50} =0.032 uM (High Activity)	6R8Q	10.1039/c9md00096h	Positive		3	Positive	True	45	Negative	False
16			IC_{50} =33 uM (Low Activity)	6G25		Negative		28	Negative	True	90	Negative	True
17	Histone-lysine N-methyltransferase NSD3		IC_{50} = 1.9 uM (High Activity)	6G2F	10.1038/s41589-019-0310-x	Positive		17	Positive	True	10	Positive	True

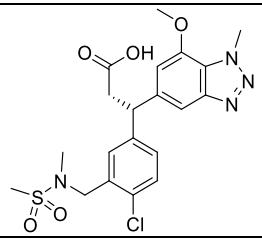
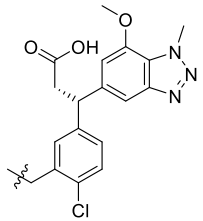
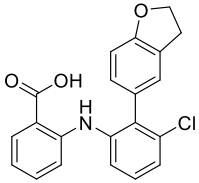
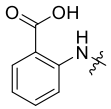
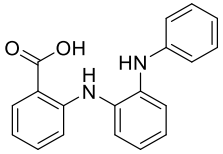
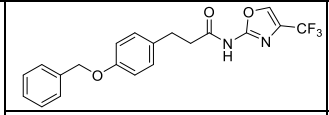
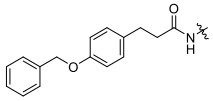
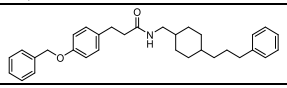
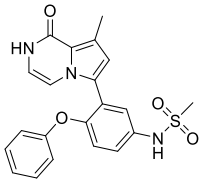
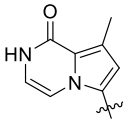
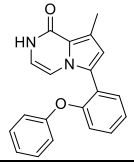
18			IC_{50} = 13 uM (Low Activity)	6G25		Negative		36	Negative	True	58	Negative	True
19	D-3-phosphoglycerate dehydrogenase		IC_{50} =0.4 uM (High Activity)	6RJ3	10.1021/acs.jmedchem.9b00718	Positive		1	Positive	True	9	Positive	True
20			K_d =100 uM (Low Affinity)	6RIH		Negative		95	Negative	True	63	Negative	True
21	Phosphopantetheine adenylyltransferase		IC_{50}=0.037 uM (High Activity)	6CCK	10.1021/acs.jmedchem.7b01691	Positive		6	Positive	True	19	Positive	True
22			IC_{50} = 31uM (Low Activity)	6CCM		Negative		57	Negative	True	103	Negative	True
23	7,8-dihydro-8-oxoguanine triphosphatase		IC_{50} =8 nM (High Activity)	6F22	10.1021/acs.jmedchem.7b01884	Positive		9	Positive	True	8	Positive	True
24			IC_{50} =0.02 uM (Low Activity			Negative		43	Negative	True	67	Negative	True
25	Human N-myristoyltransferase		IC_{50} <1 nM (High Activity)	5MU6	10.1038/s41557-018-0039-2	Positive		2	Positive	True	10	Positive	True

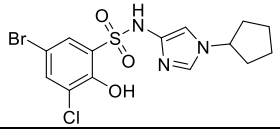
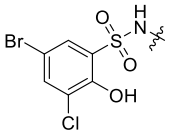
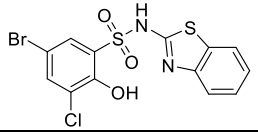
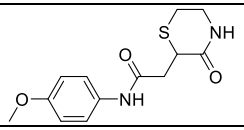
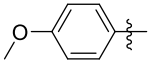
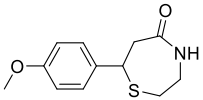
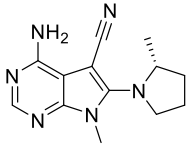
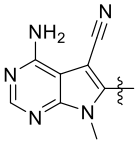
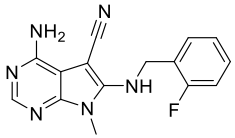
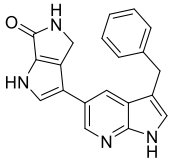
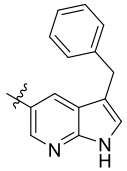
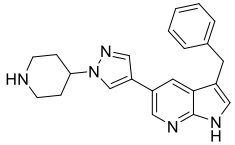
26			IC_{50} =20 uM (Low Activity)	5O4V		Negative		28	Negative	True	60	Negative	True
27	MNK1/2		IC_{50}= 0.028 uM (High Activity)	6CK3	10.1021/acs.jmedchem.7b01795	Positive		11	Positive	True	17	Positive	True
28			IC_{50} = 0.69 uM (Low Activity)	6CJE		Negative		16	Negative	True	77	Negative	True
29	BCR-ABL1		IC_{50} =0.5 uM (High Activity)	5MO4	10.1038/nature21702	Positive		5	Positive	True	16	Positive	True
30			*Kd* =2 uM (Low Activity)			Negative		28	Negative	True	70	Negative	True
31	Cyclin-dependent kinase 8		IC_{50} =3 nM (High Activity)	5XS2	10.1016/j.bmcl.2017.07.080	Positive		23	Positive	True	39	Negative	False
32			IC_{50} = 0.24 uM (Low Activity)			Negative		27	Negative	True	25	Positive	False
33	Platelet-activating factor acetylhydrolase		IC_{50}=14 nM (High Activity)	5YEA	10.1021/acs.jmedchem.7b01530	Positive		8	Positive	True	17	Positive	True
34			IC_{50} = 3.41uM (Low Activity)	5YE7		Negative		29	Negative	True	12	Positive	False
35	cGMP-dependent 3',5'-cyclic		*Ki* =14 nM (High Activity)	6B96	10.1016/j.bmcl.2017.10.054	Positive		7	Positive	True	21	Positive	True

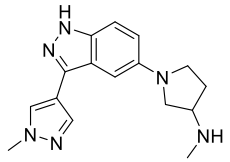
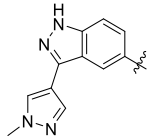
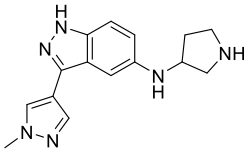
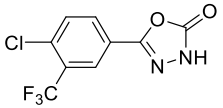
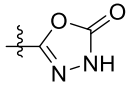
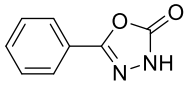
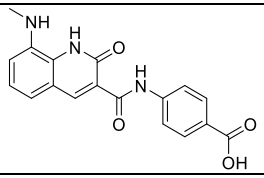
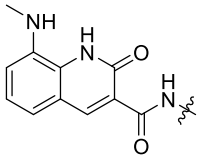
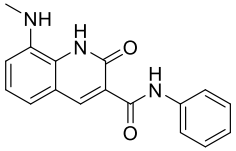
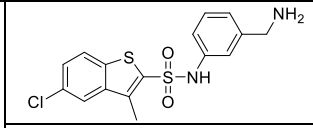
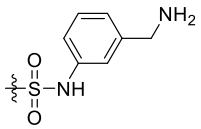
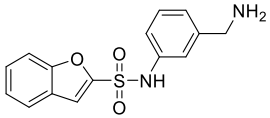
36	phosphodiesterase		*Ki* =22 uM (Low Activity)	6B98		Negative		74	Negative	True	98	Negative	True
37	Bromodomain-containing protein 4		*Ki* = 33 nM (High Activity)	5UER	-	Positive		12	Positive	True	15	Positive	True
38			*Ki*= 9.5 uM (Low Activity)		-	Negative		61	Negative	True	92	Negative	True
39	Peregrin (BRPF family)		IC_{50} =7.9 nM (High Activity)	5T4V	10.1021/acs.jmedchem.6b01583	Positive		14	Positive	True	7	Positive	True
40			IC_{50} =0.43uM (Low Activity)			Negative		27	Negative	True	51	Negative	True
41	PCAF/GCN5		IC_{50} =0.86 uM (High Activity)	5ML0	10.1021/acs.jmedchem.6b01566	Positive		41	Negative	False	26	Negative	False
42			IC_{50} =16 uM (Low Activity)	5MKX		Negative		86	Negative	True	25	Positive	False
43	PPC2/EED		IC_{50} = 0.43uM (High Activity)	5U62	10.1021/acs.jmedchem.6b01473	Positive		13	Positive	True	19	Positive	True

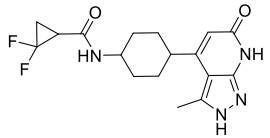
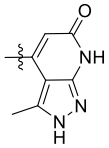
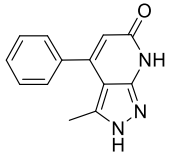
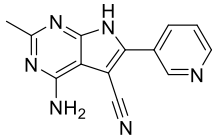
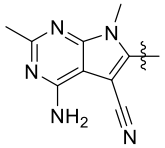
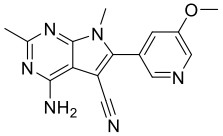
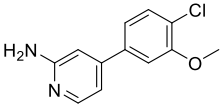
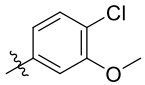
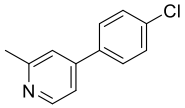
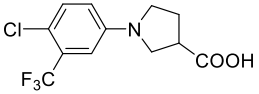
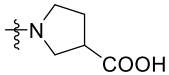
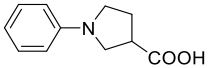
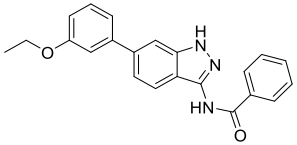
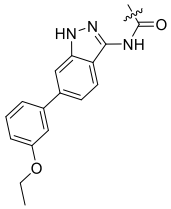
44			IC_{50} =3.9 uM (Low Activity)	5U5T		Negative		43	Negative	True	14	Positive	False
45	Choline Kinase α		*K*d=0.01uM (High Affinity)	5EQY	10.1021/acs.jmedchem.5b01552	Positive		31	Negative	False	10	Positive	True
46			*K*d =0.769uM (Low Affinity)			Negative		28	Positive	False	104	Negative	True
47	Renin		IC_{50} =38 nM (High Activity)	5SZ9	10.1016/j.bmc.2016.09.030	Positive		16	Positive	True	5	Positive	True
48			IC_{50} =43 uM (Low Activity)	5SY3		Negative		57	Negative	True	21	Negative	True
49	BCATm		IC_{50}=0.2 uM (High Activity)	5I5X	10.1021/acs.jmedchem.5b01607	Positive		35	Negative	False	8	Positive	True
50			IC_{50}=63 uM (Low Activity)	5I5V		Negative		94	Negative	True	111	Negative	True
51	Catechol O-Methyltransferase		IC_{50} =75 nM (High Activity)	5K0L	10.1021/acs.jmedchem.6b00927	Positive		10	Positive	True	34	Negative	False
52			IC_{50} =0.9 uM (Low Activity)	5K0B		Negative		56	Negative	True	77	Negative	True

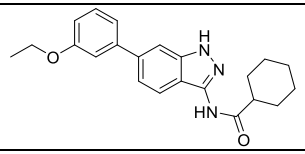
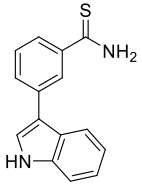
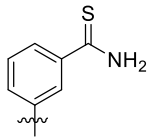
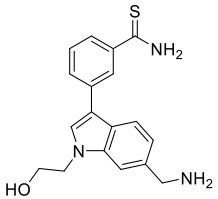
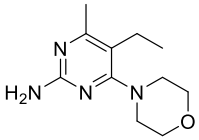
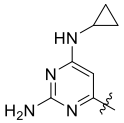
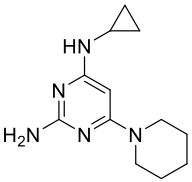
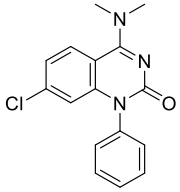
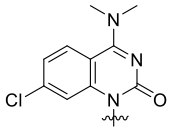
53	Phospholipase A2		IC_{50} = 0.01 uM (High Activity)	5G3N	10.1021/acsmedchemlett.6b00188	Positive		5	Positive	True	9	Positive	True
54			IC_{50} =0.91 uM (Low Activity)	-		Negative		16	Negative	True	60	Negative	True
55	BRD9		IC_{50} =14 nM (High Activity)	5F1H	10.1021/acs.jmedchem.5b01865	Positive		16	Positive	True	15	Positive	True
56			IC_{50}=0.656 uM (Low Activity)			Negative		36	Negative	True	17	Negative	True
57	CBP/EP300 bromodomain		IC_{50} =0.03 uM (High Activity)	5I8G	10.1021/acsmedchemlett.6b00075	Positive		13	Positive	True	11	Positive	True
58			IC_{50} =0.53 uM (Low Activity)	5I86		Negative		18	Negative	True	60	Negative	True
59	KEAP1		IC_{50} =1.3 nM (High Activity)	5FNU	10.1021/acs.jmedchem.6b00228	Positive		20	Positive	True	5	Positive	True

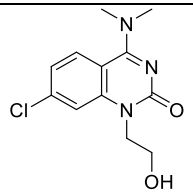
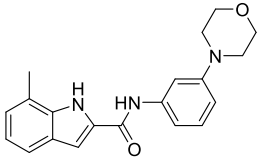
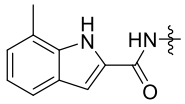
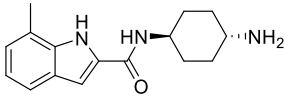
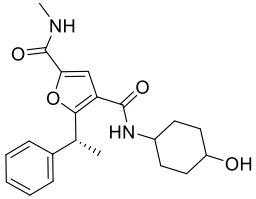
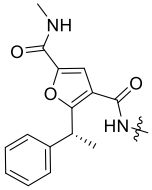
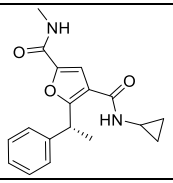
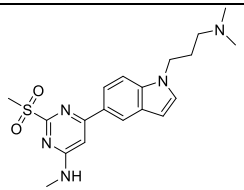
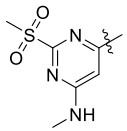
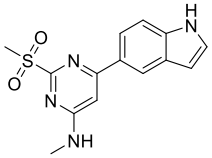
60			IC_{50} =3.4 uM (Low Activity)	5FNS		Negative		31	Negative	True	74	Negative	True
61	FABP4		Kd =37.4 nM (High Activity)	6LJV	10.1021/acs.jmedchem.9b02107	Positive		16	Positive	True	10	Positive	True
62			Kd =3741.9 nM (Low Activity)	-		Negative		59	Negative	True	42	Negative	True
63	LTA4 Hydrolase		IC_{50} =0.57 uM (High Activity)	4Y2T	10.1016/j.bmc.2015.03.083	Positive		39	Negative	False	18	Positive	True
64			IC_{50} =18.3 uM (Low Activity)	-		Negative		73	Negative	True	9	Positive	False
65	BRD4		IC_{50} =15 nM (High Activity)	6KEF	-	Positive		15	Positive	True	12	Positive	True
66			IC_{50} =4.5 uM (Low Activity)	-	-	Negative		82	Negative	True	41	Negative	True

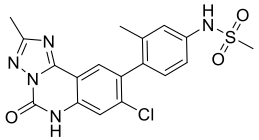
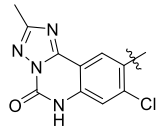
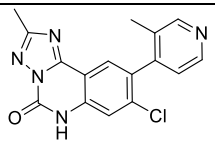
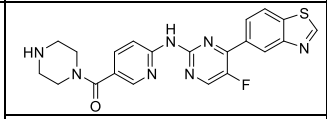
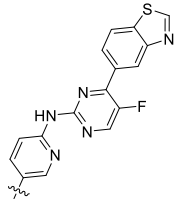
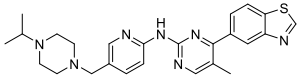
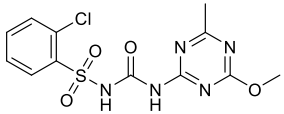
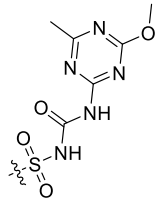
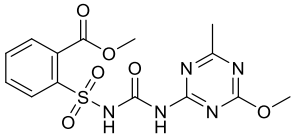
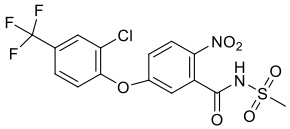
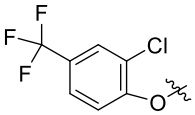
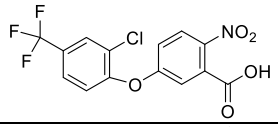
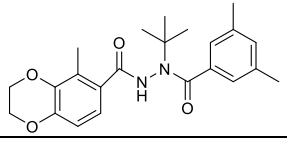
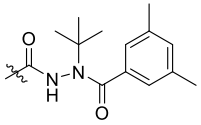
67	WDR5		*Kd*= 0.4 uM (High Affinity)	6UJL	10.1021/acs.jmedchem.0c00224	Positive		10	Positive	True	13	Positive	True
68			*Kd* =32 uM (Low Affinity)	-		Negative		90	Negative	True	85	Negative	True
69	Mek1		IC_{50} =0.33 uM (High Activity)	7B9L	10.1021/acsmedchemlett.0c00563	Positive		17	Positive	True	37	Negative	False
70			IC_{50} =450 uM (Low Activity)	-		Negative		89	Negative	True	75	Negative	True
71	LLRK2		IC_{50} =11 nM (High Activity)	7BJR	10.1021/acs.jmedchem.1c00720	Positive		15	Positive	True	75	Negative	True
72			IC_{50} =1.73 uM (Low Activity)	7BJM		Negative		34	Negative	True	46	Positive	False
73	c-MET		IC_{50} =44 nM (High Activity)	-	10.1021/acsmedchemlett.0c00392	Positive		16	Positive	True	5	Positive	True
74			IC_{50} = 0.66 uM (Low Activity)	7B41		Negative		31	Negative	True	59	Negative	True

75	Axl		IC_{50} =37 nM (High Activity)	-	10.1074/jbc.M116.771485	Positive		19	Positive	True	13	Positive	True
76			IC_{50} =0.432 uM	5U6B		Negative		39	Negative	True	11	Positive	False
77	Notum Carboxylesterase		IC_{50} =31 nM (High Activity)	6ZVL	10.1021/acs.jmedchem.0c01391	Positive		36	Negative	False	55	Negative	False
78			IC_{50}=41 uM (Low Activity)			Negative		61	Negative	True	85	Negative	True
79	DNA Gyrase		Kd= 17 nM (High Affinity)	6KZZ	10.1021/acsomega.0c00865	Positive		8	Positive	True	13	Positive	True
80			Kd=0.24uM (Low Affinity)			Negative		19	Negative	True	53	Negative	True
81	Enoyl-acyl-carrier-protein reductase		IC_{50} = 2.2 uM (High Activity)	6SQL	10.1021/acs.jmedchem.0c00007	Positive		17	Positive	True	31	Negative	False
82			IC_{50} =22 uM (Low Activity)			Negative		18	Positive	False	38	Negative	True

83	JAK1		IC_{50} =64 nM (High Activity)	6TPF	10.1021/acs.jmedchem.0c00359	Positive		8	Positive	True	16	Positive	True
84			IC_{50} =2.2 uM (Low Activity)	6TPD		Negative		39	Negative	True	42	Negative	True
85	DYRK1B		IC_{50}= 65 nM (High Activity)	7A5D	10.1021/acs.jmedchem.1c00024	Positive		11	Positive	True	43	Negative	False
86			IC_{50} =3 uM (Low Activity)	-		Negative		9	Positive	False	26	Positive	False
87	DYRK1A		IC_{50} =218 nM (High Activity)	7AJ4	10.1021/acs.jmedchem.1c00023	Positive		18	Positive	True	16	Positive	True
88			IC_{50}>10uM (Low Activity)			Negative		63	Negative	True	9	Positive	False
89	Notum Carboxylesteras		IC_{50} =0.11 uM (High Activity)	6YSK	10.1021/acs.jmedchem.0c00660	Positive		15	Positive	True	48	Negative	False
90			IC_{50} =48 uM (Low Activity)	6YV2		Negative		71	Negative	True	50	Negative	True
91	FGFR2		IC_{50} =0.46 uM (High Activity)	7OZF	10.1021/acs.jmedchem.1c01163	Positive		14	Positive	True	17	Positive	True

92			IC_{50} >10 uM (Low Activity)			Negative		10	Positive	False	57	Negative	True
93	ASH1L		IC_{50} =4 uM (High Activity)	6WZW	10.1038/s41467-021-23152-6	Positive		18	Positive	True	35	Negative	False
94			IC_{50} =50.5 uM (Low Activity)			Negative		84	Negative	True	19	Positive	False
95	MTH1		IC_{50} =6 nM (High Activity)	6JVP	10.1016/j.bioorg.2021.104813	Positive		6	Positive	True	56	Negative	False
96			IC_{50}= 1.5 uM (Low Activity)	6JVG		Negative		25	Negative	True	41	Positive	False
97	MAT2A		IC_{50}= 22 nM (High Activity)	7BHV	10.1021/acs.jmedchem.1c00067	Positive		8	Positive	True	18	Positive	True

98			IC_{50} =7.2 uM (Low Activity)	7BHU		Negative		30	Negative	True	46	Negative	True
99	SETD2		IC_{50} =0.818 uM (High Activity)	-	10.1021/acsmedchemlett.1c00272	Positive		6	Positive	True	14	Positive	True
100			IC_{50} =170 uM (Low Activity)	7LZB		Negative		91	Negative	True	54	Negative	True
101	BRD4		PIC_{50} =7.7 (High Activity)	7OE6	10.1021/acs.jmedchem.1c00365	Positive		14	Positive	True	44	Negative	False
102			PIC_{50} =4.9 (Low Activity)			Negative		61	Negative	True	95	Negative	True
103	BPTF		IC_{50} =32 nM (High Activity)	-	-	Positive		16	Positive	True	9	Positive	True
104			IC_{50} =0.698 uM (Low Activity)	7F5D	-	Negative		28	Negative	True	37	Negative	True

105	Tim-3		IC_{50} =0.75 uM (High Activity)	7M3Z	10.1021/acs.jmedchem.1c01336	Positive		17	Positive	True	11	Positive	True
106			IC_{50} =4.9 uM (Low Activity)	7M3Y		Negative		41	Negative	True	39	Negative	True
107	DYRK2		IC_{50} =0.6 nM (High Activity)	8HLT	-	Positive		5	Positive	True	17	Positive	True
108			IC_{50} =1.41 uM (Low Activity)	7EJV	10.1038/s41467-022-30581-4	Negative		27	Negative	True	90	Negative	True
109	Acetohydroxyacid synthase		*Ki* =9.4 nM (High Activity)	-	-	Positive		9	Positive	True	31	Negative	False
110			*Ki* =127 nM (Low Activity)	1YHZ	10.1073/pnas.0508701103	Negative		19	Negative	True	24	Positive	False
111	Protoporphyrinogen Oxidase		IC_{50} =0.28 uM (High Activity)		10.1016/j.jsb.2009.11.012	Positive		9	Positive	True	19	Positive	True
112			IC_{50} =4.00 uM (Low Activity)	3I6D		Negative		6	Positive	False	21	Negative	True
113	Ecdysone receptor		pIC50 > 8.81 (High Activity)	1R20	10.1038/nature02112	Positive		12	Positive	True	20	Positive	True

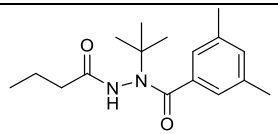
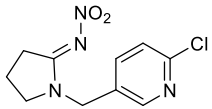
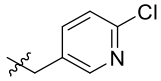
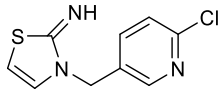
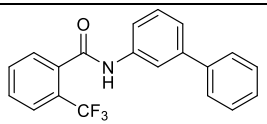
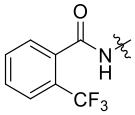
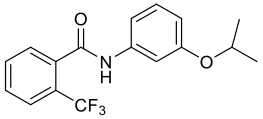
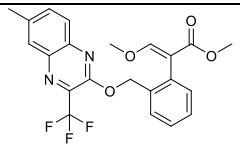
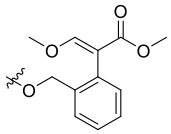
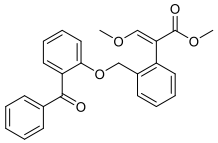
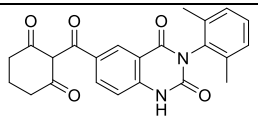
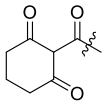
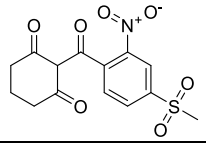
114			pIC50 =5.92 (Low Activity)			Negative		76	Negative	True	92	Negative	True
115	Nicotinic acetylcholine Receptors		Ki =2.2 nM (High Activity)	3C79	10.1073/pnas.0802197105	Positive		6	Positive	True	15	Positive	True
116			Ki =100 nM (Low Activity)			Negative		13	Negative	True	6	Positive	False
117	Complex II		IC_{50} =8.61 µM (High Activity)	-	10.3390/ijms160715287	Positive		17	Positive	True	25	Negative	False
118			IC_{50} =45.9 µM (Low Activity)	4YXD		Negative		12	Positive	False	37	Negative	True
119	Complex III		*Ki*= 83 nM (High Activity)	3TGU	10.1021/ja3001908	Positive		9	Positive	True	3	Positive	True
120			*Ki* > 10µM (Low Activity)	-		Negative		28	Negative	True	36	Negative	True
121	4-hydroxyphenylpyruvate dioxygenase		*Ki* =24 nM (High Activity)	5YY7	-	Positive		13	Positive	True	9	Positive	True
122			*Ki* =0.247 µM (Low Activity)	5YWG	10.1111/febs.14747	Negative		18	Negative	True	104	Negative	True

Table S2. Comparison of ACFIS2 with other computational tools for FBDD in key features.

FBDD Step	Tool	Form	Functionality	Input	Output
Fragment screening	SEED	Software	Fragment docking	Protein structure	Suggested core fragment
Fragment-to-Lead	FragPELE	Software (command line operation)	Fragment growing	Protein–ligand complex file	Protein–suggested ligand complex file
	DeepFrag	Browser app (https://durrantlab.pitt.edu/deepfrag/)	Fragment growing	Receptor and ligand structure file	Suggested-fragments table
	Delinker	Software (command line operation)	Fragment linking	Fragment (SMILES) and receptor structure file	Generated molecules (SMILES)
	SyntaLinker	Software (command line operation)	Fragment linking	Fragment file	Generated molecules
	Autogrow	Software	Fragment growing	Fragment file	Generated molecules
	LigBuilder 2	Software (command line operation)	Fragment growing / linking / mutation	Receptor structure file	Suggested-ligand file
Fragment screening + Fragment-to Lead	ACFIS 2.0	Web server (http://chemyang.ccnu.edu.cn/ccb/server/ACFIS2/)	Fragment hit identification + Fragment growing	Protein–ligand complex file	Protein-suggested ligand / core fragment complex file and ligand file

Reference

[1] Case D A, Cheatham III T E, Darden T, et al. The Amber biomolecular simulation programs. Journal of Computational Chemistry, 2005, 26, 1668-1688.

[2] Kolb P, Caflisch A. Automatic and efficient decomposition of two-dimensional structures of small molecules for fragment-based high-throughput docking. Journal of Medicinal Chemistry, 2006, 49, 7384-7392.

[3] Miller III B R, McGee Jr T D, Swails J M, et al. MMPBSA. py: an efficient program for end-state free energy calculations. Journal of Chemical Theory and Computation, 2012, 8, 3314-3321.

[4] Sun H, Duan L, Chen F, et al. Assessing the performance of MM/PBSA and MM/GBSA methods. 7. Entropy effects on the performance of end-point binding free energy calculation approaches. Physical Chemistry Chemical Physics, 2018, 20, 14450-14460.

[5] Maier J A, Martinez C, Kasavajhala K, et al. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. Journal of Chemical Theory and Computation, 2015, 11, 3696-3713.

[6] Wang J, Wolf R M, Caldwell J W, et al. Development and testing of a general amber force field. Journal of Computational Chemistry, 2004, 25, 1157-1174.

[7] Mark P, Nilsson L. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. The Journal of Physical Chemistry A, 2001, 105, 9954-9960.

[8] Durrant J D, Lindert S, McCammon J A. AutoGrow 3.0: an improved algorithm for chemically tractable, semi-automated protein inhibitor design. Journal of Molecular Graphics and Modelling, 2013, 44, 104-112.

[9] Moriwaki H, Tian Y S, Kawashita N, et al. Mordred: a molecular descriptor calculator. Journal of Cheminformatics, 2018, 10, 1-14.

[10] Lipinski C A, Lombardo F, Dominy B W, et al. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced Drug Delivery Reviews, 2012, 64, 4-17.

[11] Ghose A K, Viswanadhan V N, Wendoloski J J. A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. Journal of Combinatorial Chemistry, 1999, 1, 55-68.

[12] Bickerton G R, Paolini G V, Besnard J, et al. Quantifying the chemical beauty of drugs. Nature Chemistry, 2012, 4, 90-98.

[13] Hao G, Dong Q, Yang G. A comparative study on the constitutive properties of marketed pesticides. Molecular informatics, 2011, 30, 614-622.

[14] Ertl P, Schuffenhauer A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. Journal of Cheminformatics, 2009, 1, 1-11.