Supporting Information

Pose Filter-based Ensemble Learning Enables Discovery of Orally Active, Nonsteroidal Farnesoid X Receptor Agonists

Jie Xia[╠], Zhenyi Wang[◇☼], Yi Huan[◎], Wenjie Xue[╠], Xing Wang[◎], Yuxi Wang[●], Zhenming Liu[●], Jui-Hua Hsieh[§], Liangren Zhang[●],Song Wu[╠*], Zhufang Shen[◎*], Hongmin Zhang[☼*], Xiang Simon Wang[Δ*]

[╠]*State Key Laboratory of Bioactive Substance and Function of Natural Medicines, Department of New Drug Research and Development, Institute of Materia Medica, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100050, China;* [◇] *Hefei National Laboratory for Physical Sciences at the Microscale and School of Life Sciences, University of Science and Technology of China, Hefei, Anhui, 230026, China;* [☼]*Department of Biology, Guangdong Provincial Key Laboratory of Cell Microenvironment and Disease Research, Shenzhen Key Laboratory of Cell Microenvironment and SUSTech-HKU Joint Laboratories for Matrix Biology, Southern University of Science and Technology, Shenzhen 518055, China;* [◎] *State Key Laboratory of Bioactive Substance and Function of Natural Medicines, Department of Pharmacology, Institute of Materia Medica, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100050, China;* [●]*State Key Laboratory of Natural and Biomimetic Drugs, School of Pharmaceutical Sciences, Peking University, Beijing 100191, China*; [§]*Kelly Government Solutions, Research Triangle Park, North Carolina 27709, U.S.A.* [Δ]*Artificial Intelligence and Drug Discovery Core Laboratory for District of Columbia Center for AIDS Research (DC CFAR), Department of*

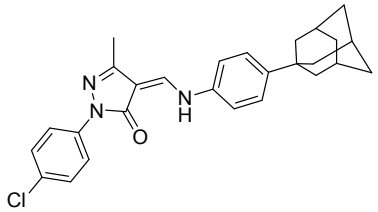
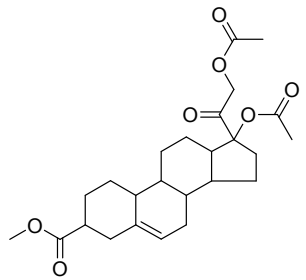
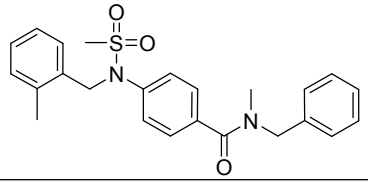
Pharmaceutical Sciences, College of Pharmacy, Howard University, Washington, DC 20059, U.S.A.; [ll] *These authors contributed equally to this work.*

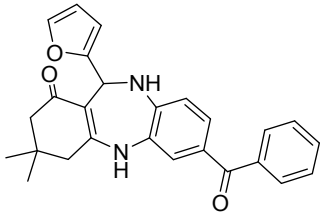
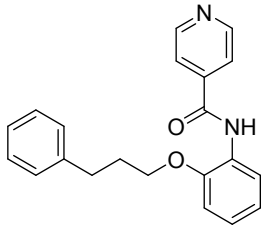
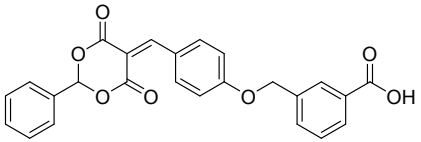
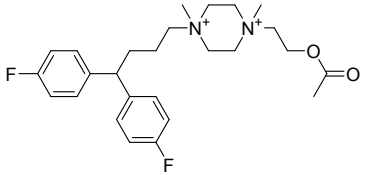
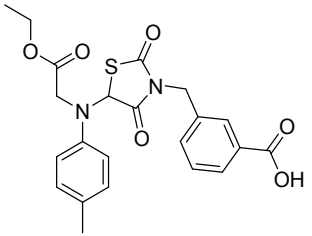
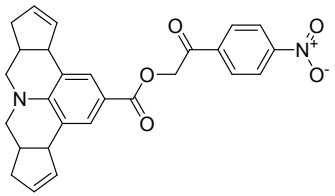
[**]Emails: X.S.W (x.simon.wang@gmail.com) or H-M.Z (zhanghm@sustech.edu.cn) or Z-F.S (shenzhf@imm.ac.cn) or S.W. (ws@imm.ac.cn)

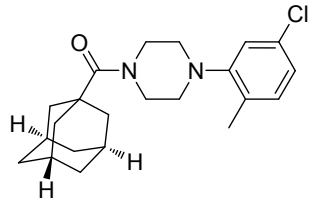
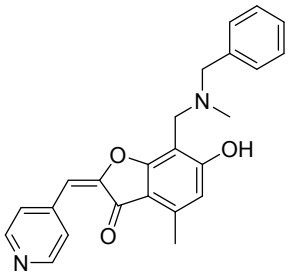
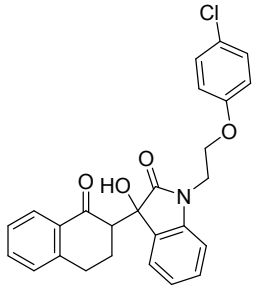
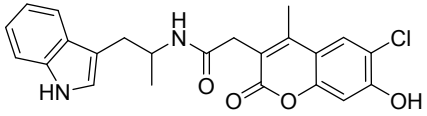
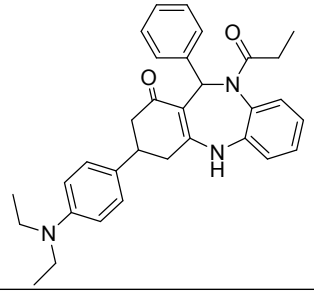
Table S1. ROC enrichments of 24 SBVS approaches based on different crystal structures of hFXR and the PFE-coupling SBVS approach.

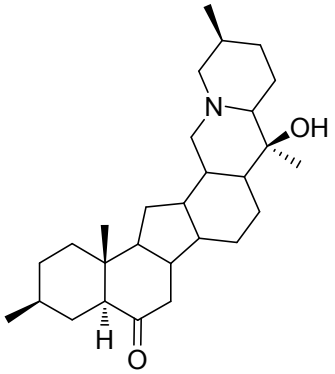
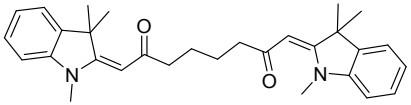
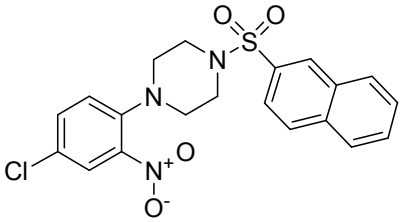
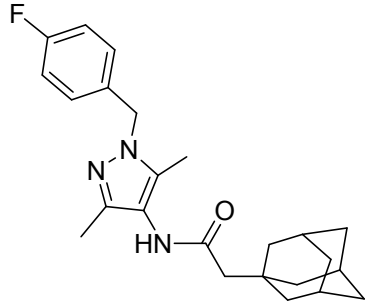
Index	PDB code	$ROCE_{0.5\%}$	$ROCE_{1\%}$	$ROCE_{2\%}$	$ROCE_{5\%}$
1	1OSH	0.68	0.34	0.17	0.54
2	3BEJ	2.70	2.70	2.03	1.69
3	3DCT	16.22	12.84	10.30	5.61
4	3DCU	14.87	10.92	10.98	6.42
5	3FLI	3.38	3.14	2.36	1.22
6	3FXV	23.65	13.18	7.60	3.99
7	3GD2	33.11	21.62	13.01	5.95
8	3HC5	34.48	20.27	12.16	6.22
9	3HC6	20.27	19.59	13.01	6.49
10	3L1B	10.81	7.43	4.32	2.50
11	3OKH	3.38	2.03	1.18	0.68
12	3OKI	11.49	6.42	3.72	2.16
13	3OLF	14.87	7.77	4.39	2.30
14	3OMK	14.19	7.77	4.22	2.03
15	3OMM	16.89	8.78	5.07	2.57
16	3OOF	16.22	8.45	4.39	2.09
17	3OOK	14.19	7.77	4.56	2.23
18	3P88	27.03	17.57	10.98	5.81
19	3P89	31.08	18.92	11.99	6.49
20	**3RUT**	**41.22**	**23.65**	**13.68**	**7.03**
21	3RUU	35.85	23.65	14.02	7.09
22	3RVF	35.81	20.27	12.67	5.95
23	4QE6	22.30	12.50	7.43	4.32
24	4WVD	1.35	1.35	1.35	1.08
25	**PFE_3RUT**	**53.38**	**30.41**	**18.41**	**9.12**

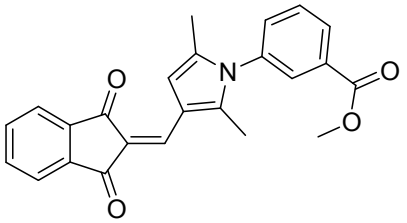
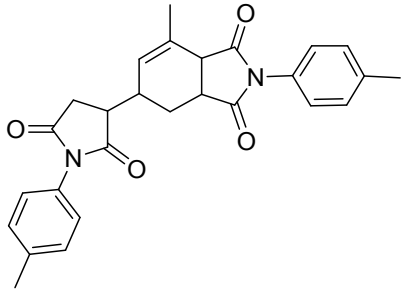
Table S2. 20 potential hits selected from Specs chemical library by the 3RUT-based SBVS approach with the hFXR-specific PFE, and their FXR agonistic activity (fold change) tested by the FXR transactivation assay in HEK293T cells.

Cluster	ID number	Chemical structure	FRED Chemgauss4 score	Tanimoto coefficient (To nearest neighbor)	FXR transactivation[a]	
					50 μM (mean±sem)[b]	10 μM (mean±sem)
1	AG-690/13505038		-19.26	0.63	0.91±±0.16	n.d.[c]
2	AK-967/37080019		-20.29	0.73	**1.11±0.05**	1.02±0.12
3	AO-081/41755482		-19.06	0.65	0.63±0.21	n.d.

4	AG-690/13705907		-18.72	0.59	0.87±0.20	n.d.
5	AN-329/43449676		-18.88	0.54	0.93±0.01	n.d.
6	AM-879/15041526		-19.67	0.71	0.98±0.10	n.d.
7	AE-641/30115025		-19.20	0.61	0.74±0.24	n.d.
8	AP-893/40872500		-19.69	0.63	**1.06±0.08**	1.11±0.23
9	AG-205/36713003		-19.48	0.56	0.47±0.11	n.d.

10	AO-990/37423025 (compound XJ034)		-18.76	0.62	**2.19±0.15**	**4.34±0.74**
11	AO-022/43453279		-18.69	0.64	0.47±0.31	n.d.
12	AK-778/43420935		-19.68	0.70	0.30±0.07	n.d.
13	AO-022/43452495		-19.00	0.67	**1.20±0.02**	1.11±0.35
14	AK-968/15361047		-19.21	0.69	0.80±0.16	n.d.

15	AQ-152/42730387		-19.79	0.70	0.66±0.19	n.d.
16	AE-473/31198008		-18.81	0.65	0.67±0.16	n.d.
17	AG-670/15543045		-18.89	0.58	**1.08±0.06**	1.13±0.17
18	AK-968/40730372		-18.85	0.61	0.88±0.23	n.d.

19	AO-081/40681111		-18.83	0.61	0.76±0.11	n.d.
20	AG-690/37048062		-18.73	0.65	**1.04±0.34**	0.95±0.30
Vehicle	DMSO				1.00±0.00	1.00±0.00
Positive	CDCA				13.94±1.40	1.43±0.18

[a]: Fold change in the luciferase activity after the compound treatment with respect to that after the vehicle treatment. A compound with a fold change greater than 1 is regarded as an agonist in the FXR transactivation assay.

[b]: Mean: the average of triplicates; sem: the standard error of the mean.

[c]: Not determined.

Table S3. Top 0.5% (i.e. 115) compounds ranked at the top of the compound list: comparison between 3RUT-based SBVS with the PFE and 3RUT-based SBVS without the PFE.

rank	3RUT-based SBVS with the PFE		3RUT-based SBVS without the PFE	
	IDNUMBER	Chemgauss4 score	IDNUMBER	Chemgauss4 score
1	AG-777/36179021	-21.5951	AG-777/36179021	-21.9201
2	AN-988/14856342	-21.2006	AN-988/14609093	-21.645
3	AE-641/15486007	-20.9758	AG-690/10384015	-21.2039
4	AO-022/43453912	-20.5684	AN-988/14856342	-21.2006
5	AG-690/12892831	-20.5267	AN-329/37380003	-21.0949
6	AG-690/12868962	-20.4958	AN-319/40684809	-21.0893
7	AK-968/15360950	-20.4699	AE-842/32538013	-20.9899
8	AL-281/15562107	-20.3537	AG-205/36713003	-20.9877
9	AG-205/36712046	-20.3457	AE-641/15486007	-20.9758
10	AN-329/43448767	-20.3355	AN-698/40781751	-20.8797
11	AK-967/37080019	-20.2944	AG-205/05873030	-20.7426
12	AK-918/12271022	-20.1976	AK-778/43420923	-20.7191
13	AK-918/12087276	-19.9139	AG-205/40649516	-20.6985
14	AK-918/40864029	-19.9036	AK-918/12271022	-20.6619
15	AG-690/12889814	-19.8673	AG-690/10376017	-20.6367
16	AQ-152/42730387	-19.7917	AG-205/36712046	-20.6287
17	AG-690/40720730	-19.7342	AK-918/41676992	-20.6185
18	AG-205/36626058	-19.7016	AH-487/41660565	-20.6095
19	AP-893/40872500	-19.6947	AT-057/43486225	-20.596
20	AK-778/43420935	-19.6825	AO-022/42600809	-20.5931
21	AM-879/15041526	-19.6667	AO-022/43453912	-20.5684
22	AG-690/12509040	-19.6055	AK-967/37080019	-20.5386
23	AG-690/11307155	-19.5929	AG-690/12892831	-20.5267
24	AG-205/14194070	-19.571	AG-690/12868962	-20.4958
25	AK-968/15359311	-19.5439	AK-778/43206443	-20.4954
26	AG-205/12145059	-19.5131	AG-690/11307053	-20.492
27	AG-205/36713003	-19.4796	AM-879/15327266	-20.484
28	AP-970/43029468	-19.4512	AK-968/15360950	-20.4699
29	AG-690/12892898	-19.4395	AK-968/15008640	-20.4562
30	AG-205/37007036	-19.4366	AK-778/41050967	-20.4501
31	AM-879/12211009	-19.428	AT-057/43486171	-20.4415
32	AK-918/40678977	-19.4185	AG-205/12145059	-20.4357
33	AP-970/42444942	-19.4102	AG-205/33660050	-20.378
34	AM-879/15327266	-19.3524	AH-487/14758265	-20.3756
35	AN-698/40781751	-19.3221	AM-879/15041526	-20.3714
36	AH-487/15582312	-19.2824	AG-205/34704040	-20.3651
37	AJ-292/41686181	-19.2785	AN-329/43448767	-20.3619
38	AP-970/42444952	-19.2632	AL-281/15562107	-20.3537
39	AG-690/13505038	-19.2614	AK-968/12100100	-20.3498
40	AG-690/10378010	-19.2162	AO-476/43415731	-20.33
41	AH-262/02658045	-19.2085	AE-848/15342037	-20.3176
42	AK-968/15361047	-19.2076	AH-487/14756898	-20.3131
43	AE-641/30115025	-19.1993	AE-848/10478006	-20.3053

44	AP-263/42611384	-19.1977	AG-205/15156183	-20.2731
45	AT-057/43486118	-19.1964	AN-153/12400022	-20.2383
46	AG-205/37245006	-19.1647	AT-057/43486103	-20.219
47	AO-365/15161007	-19.1448	AN-023/14771018	-20.2138
48	AT-057/43486171	-19.143	AC-907/43493849	-20.2113
49	AN-584/43447539	-19.1157	AG-690/15444517	-20.206
50	AG-690/12509042	-19.1076	AS-871/43475179	-20.1945
51	AK-918/12349066	-19.1067	AP-518/40848445	-20.1885
52	AG-205/12668002	-19.0924	AG-690/40719370	-20.1806
53	AE-562/12222821	-19.0862	AG-205/13022176	-20.1771
54	AG-205/13888077	-19.0664	AE-848/15342064	-20.1723
55	AO-081/41755482	-19.0554	AG-205/08691017	-20.1681
56	AK-968/14004841	-19.053	AK-918/40909831	-20.1618
57	AG-205/06946035	-19.0348	AH-487/41658122	-20.1487
58	AG-690/10384025	-19.0083	AH-487/14758433	-20.1409
59	AO-022/43452495	-18.9998	AG-690/13157313	-20.1297
60	AK-918/43446434	-18.9994	AK-918/40864029	-20.1239
61	AG-690/13774270	-18.9955	AO-080/42577309	-20.1145
62	AH-262/02658047	-18.9946	AK-968/37129072	-20.1084
63	AG-205/13629068	-18.9839	AG-205/40649572	-20.0892
64	AK-778/43464903	-18.983	AN-698/41889858	-20.0608
65	AK-778/43413658	-18.9711	AK-778/15446027	-20.0587
66	AG-690/36596052	-18.9692	AO-022/43453448	-20.0352
67	AO-022/43454466	-18.9461	AH-487/41657703	-20.0299
68	AG-690/15444517	-18.9334	AO-080/43441870	-20.0201
69	AG-690/15439551	-18.927	AN-988/14610123	-19.9871
70	AK-778/43206314	-18.927	AO-022/43454102	-19.9831
71	AK-778/41507472	-18.9238	AQ-152/42730387	-19.9725
72	AM-879/41892999	-18.9157	AM-879/13491009	-19.9699
73	AH-487/15274135	-18.9106	AG-690/13115213	-19.9648
74	AK-918/11755215	-18.9056	AT-057/43319011	-19.9557
75	AG-690/40639019	-18.8884	AT-057/43319158	-19.9455
76	AK-918/11940266	-18.887	AO-080/43378945	-19.9344
77	AG-670/15543045	-18.8863	AG-690/12889814	-19.9258
78	AN-329/43449676	-18.8806	AG-690/12509042	-19.9218
79	AH-487/42197034	-18.8802	AK-918/12087276	-19.9139
80	AG-690/36718052	-18.8662	AO-365/43403179	-19.9037
81	AK-918/40656509	-18.8633	AG-205/40649630	-19.9037
82	AN-153/43311934	-18.86	AS-871/43469786	-19.8996
83	AK-968/40730372	-18.8506	AN-698/41607033	-19.8989
84	AO-476/43250190	-18.8474	AK-918/11940266	-19.8859
85	AN-329/43211155	-18.8349	AK-968/15342060	-19.8802
86	AO-081/40681111	-18.8347	AH-487/42144988	-19.876
87	AP-970/42444941	-18.8289	AK-918/40656512	-19.8694
88	AK-918/41676992	-18.8283	AT-057/43468158	-19.8689
89	AM-879/41892539	-18.8187	AO-022/43454125	-19.8623
90	AE-473/31198008	-18.8073	AK-968/36799025	-19.8588
91	AO-080/43342612	-18.8018	AJ-292/41686181	-19.8437

92	AM-879/12211045	-18.7997	AG-205/36712061	-19.8431
93	AT-057/43468158	-18.7881	AT-057/43469785	-19.8398
94	AT-057/43486236	-18.7813	AN-584/43447539	-19.8329
95	AH-487/15149443	-18.7722	AN-329/43211155	-19.8304
96	AO-990/37423025	-18.7587	AH-487/14758193	-19.8127
97	AP-893/40872498	-18.7505	AO-081/15571113	-19.8119
98	AK-918/40863864	-18.7496	AG-690/13507396	-19.8117
99	AG-205/32539001	-18.7465	AM-879/12211045	-19.8104
100	AK-968/12100100	-18.7455	AO-081/40738375	-19.808
101	AK-778/15446027	-18.7422	AG-205/12365032	-19.7983
102	AI-031/40912018	-18.7415	AI-204/31683043	-19.793
103	AG-670/34811010	-18.7375	AG-690/11665463	-19.791
104	AG-690/37048062	-18.7347	AG-205/12365035	-19.7858
105	AG-690/12620687	-18.7314	AO-081/41888628	-19.7828
106	AG-690/13705907	-18.7187	AG-690/11307155	-19.771
107	AT-057/43469237	-18.7164	AK-918/40810327	-19.7663
108	AN-329/43448662	-18.7133	AG-205/06946035	-19.7631
109	AK-778/15446017	-18.7112	AN-648/14912027	-19.7613
110	AP-124/42855905	-18.7059	AK-968/15360974	-19.7587
111	AO-022/43452702	-18.7043	AO-081/41196620	-19.7531
112	AQ-149/42126241	-18.6955	AH-262/02658045	-19.7516
113	AH-487/11681003	-18.6937	AP-518/42582911	-19.7465
114	AG-690/11307030	-18.6923	AG-690/40720730	-19.7342
115	AO-022/43453279	-18.6867	AT-057/43486124	-19.7289

Table S4. The nearest neighbor of compound XJ034 (AO-990/37423025) identified from both ChEMBL24 and NRList BDB by MACCS fingerprints-based similarity search.

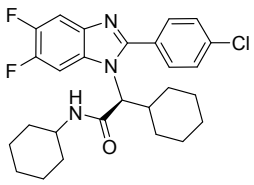
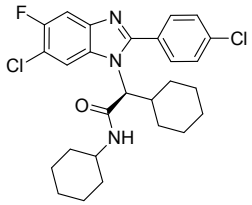
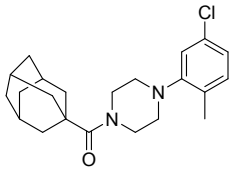
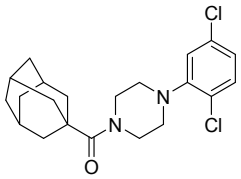
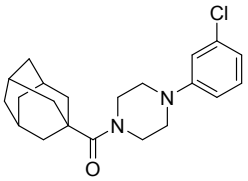
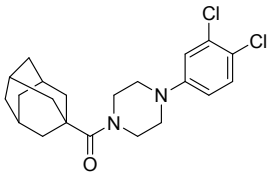
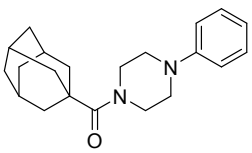
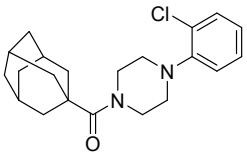
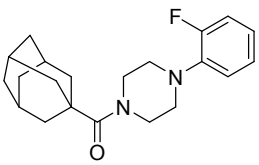
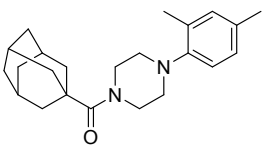
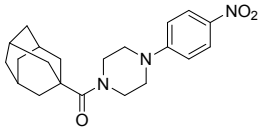
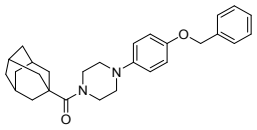
Items	Nearest neighbor	
Tanimoto coefficient	0.62	
CMPD_CHEMBLID	CHEMBL1668238	CHEMBL1642362
Chemical structure		
EC_{50}	50 nM	100 nM
Assay description	Agonist activity at Gal4-fused human FXR by luciferase reporter gene transactivation assay[1]	

Table S5. The structure-activity relationship of compound XJ034.

Name	ID Number	Chemical structure	FXR transactivation[a]			
			50 μM (Mean±sem)[b]	10 μM (Mean±sem)	EC_{50} [μM]	Fold max
XJ034	AO990/37423025		1.63±0.06	4.34±0.74	5.1	7.0
XJ034-1	AP-263/41967830		2.01±0.31	1.64±0.12	3.8	5.9
XJ034-2	AK-968/12164410		1.64±0.31	5.84±1.54	9.7	7.7
XJ034-3	AP-263/41670570		0.95±0.24	n.d.[c]	n.d.	n.d.
XJ034-4	AO-365/40118306		1.09±0.18	n.d.	n.d.	n.d.
XJ034-5	AP-263/10926012		1.16±0.42	0.97±0.31	n.d.	n.d.
XJ034-6	AO-990/37423026		1.07±0.08	n.d.	n.d.	n.d.
XJ034-7	AP-263/41967828		0.99±0.16	n.d.	n.d.	n.d.

XJ034-8	AG-690/34551047		1.04±0.31	n.d.	n.d.	n.d.
XJ034-9	AK-968/41017697		1.43±0.15	1.18±0.23	n.d.	n.d.
Vehicle			1.00±0.00	1.00±0.00	n.d.	n.d.

[a]: Fold change in the luciferase activity after the compound treatment with respect to that after the vehicle treatment.
[b]: Mean: the average of triplicates; Sem: the standard error of the mean.
[c]: Not determined.

Table S6. The available ligand-bound hFXR crystal structures retrieved from Protein Data Bank (PDB, accessed Feb. 2016).

Index	PDB code	Organism(s)	Functional effect	Coactivator/corepressor/peptide
1	1OSH	homo sapiens	Agonist	none
2	3BEJ	homo sapiens	Agonist	co-activator peptide
3	3DCT	homo sapiens	Agonist	Nuclear receptor coactivator 1
4	3DCU	homo sapiens	Agonist	Nuclear receptor coactivator 1
5	3FLI	homo sapiens	Agonist	none
6	3FXV	homo sapiens	Agonist	12-meric peptide from Nuclear receptor coactivator 1
7	3GD2	homo sapiens	Agonist	activator peptide
8	3HC5	homo sapiens	Agonist	Nuclear receptor coactivator 1
9	3HC6	homo sapiens	Agonist	Nuclear receptor coactivator 1
10	3L1B	homo sapiens	Agonist	none
11	3OKH	homo sapiens	Agonist	peptide of Nuclear receptor coactivator 1
12	3OKI	homo sapiens	Agonist	peptide of Nuclear receptor coactivator 1
13	3OLF	homo sapiens	Agonist	peptide of Nuclear receptor coactivator 1
14	3OMK	homo sapiens	Agonist	peptide of Nuclear receptor coactivator 1
15	3OMM	homo sapiens	Agonist	peptide of Nuclear receptor coactivator 1
16	3OOF	homo sapiens	Agonist	peptide of Nuclear receptor coactivator 1
17	3OOK	homo sapiens	Agonist	peptide of Nuclear receptor coactivator 1
18	3P88	homo sapiens	Agonist	Nuclear receptor coactivator 1
19	3P89	homo sapiens	Agonist	Nuclear receptor coactivator 1
20	3RUT	homo sapiens	Agonist	Nuclear receptor coactivator 1
21	3RUU	homo sapiens	Agonist	Nuclear receptor coactivator 1
22	3RVF	homo sapiens	Agonist	Nuclear receptor coactivator 1
23	4QE6	homo sapiens	Agonist	Nuclear receptor coactivator 2
24	4QE8	homo sapiens	Agonist	Nuclear receptor coactivator 2
25	4WVD	homo sapiens	Agonist	Nuclear receptor corepressor 1
26	1OSV	Rattus norvegicus, Mus musculus	Agonist	Nuclear receptor coactivator 2
27	1OT7	Rattus norvegicus	Agonist	dodecamer peptide fragment of RPGR-interacting protein 1
28	4OIV	homo sapiens	Antagonist	none

Table S7. The numbers of native-like poses and pose decoys in the initial data sets and the data sets for modeling.

Index	PDB code	No. of poses generated by redocking			No. of poses for modeling	
		native-like	decoys	balanced Yes/No	native-like	decoys
1	1OSH	777*	223	No	295	223
2	3BEJ	722*	278	No	385	278
3	3DCT	813*	187	No	269	187
4	3DCU	827*	173	No	160	173
5	3FLI	265	735*	No	265	249
6	3FXV	360	640	Yes	360	640
7	3GD2	731*	269	No	249	269
8	3HC5	889*	111	No	86	111
9	3HC6	813*	187	No	156	187
10	3L1B	504	496	Yes	504	496
11	3OKH	274	726*	No	274	298
12	3OKI	232	768*	No	232	218
13	3OLF	223	567*	No	223	253
14	3OMK	137	863*	No	137	134
15	3OMM	258	441	Yes	258	441
16	3OOF	252	748*	No	252	236
17	3OOK	56	98	Yes	56	98
18	3P88	972*	28	No	37	28
19	3P89	987*	13	No	7	13
20	3RUT	972*	28	No	26	28
21	3RUU	945*	55	No	58	55
22	3RVF	967*	33	No	29	33
23	4QE6	967*	33	No	58	33
24	4WVD	874*	126	No	165	126

*Overrepresented pose class that requires downsampling.

Table S8. The statistics of each SVM-based PF including content of the training set, the parameters (C, γ) used for model building, cross-validation (CV) accuracy for the training set, content of the test set for model validation and prediction accuracy (PA) for the test set.

Index	PDB code	Training set		RBF model		CV accuracy%	Test set		Prediction accuracy%
		Native-like poses (no.)	Pose decoys (no.)	*C*	*γ(10^{-4})*		Native-like poses (no.)	Pose decoys (no.)	
1	1OSH	244	171	2	78.1	99.3	51	52	97.1
2	3BEJ	314	217	8	19.5	98.5	71	61	100.0
3	3DCT	215	150	8	78.1	94.8	54	37	91.2
4	3DCU	132	135	8	78.1	94.4	28	38	89.4
5	3FLI	212	200	2	78.1	99.5	53	49	100.0
6	3FXV	284	516	128	0.3	97.1	76	124	96.0
7	3GD2	195	220	32	1.2	97.1	54	49	95.1
8	3HC5	66	92	32	4.9	83.5	20	19	76.9
9	3HC6	124	151	8	78.1	89.1	32	36	88.2
10	3L1B	404	396	8	78.1	97.6	100	100	97.0
11	3OKH	213	245	8	19.5	99.6	61	53	100.0
12	3OKI	184	176	2	78.1	98.3	48	42	97.8
13	3OLF	180	201	2	78.1	97.9	43	52	98.9
14	3OMK	110	107	32	19.5	97.7	27	27	98.1
15	3OMM	209	351	8	19.5	99.5	49	90	99.3
16	3OOF	205	186	2	78.1	99.2	47	50	100.0
17	3OOK	49	75	2	78.1	97.6	7	23	96.7
18	3P88	28	24	2	312.5	78.8	9	4	61.5
19	3P89	5	11	8	78.1	81.3	2	2	50.0
20	3RUT	23	21	512	1.2	59.1	3	7	80.0
21	3RUU	47	44	8	4.9	81.3	11	11	81.8

22	3RVF	22	28	32	19.5	84.0	7	5	75.0
23	4QE6	48	25	8	78.1	93.2	10	8	100.0
24	4WVD	138	95	2	1.2	98.7	27	31	100.0

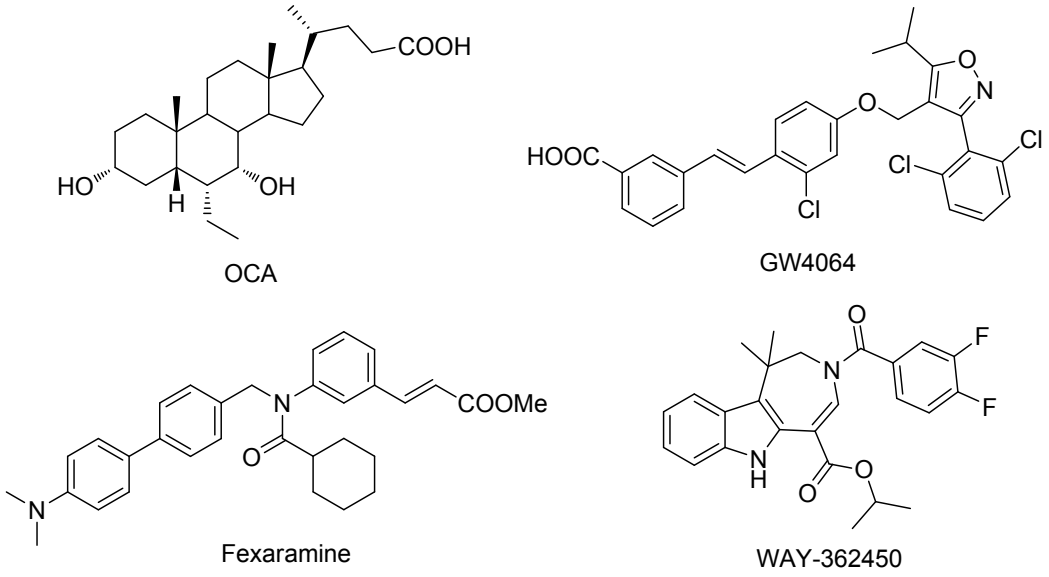


Figure S1. The chemical structures of representative FXR agonists.

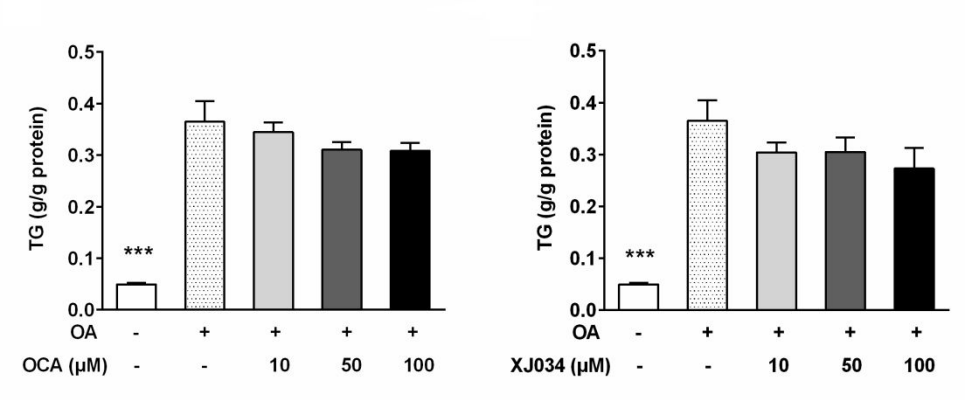


Figure S2. The effect of compound XJ034 on the intracellular triglyceride (TG) in human hepatic L02 cells with fatty accumulation (treated with 0.2 mM oleic acid). **Left:** Treatment with 0.2 mM oleic acid and the indicated concentrations of OCA; **Right:** Treatment with 0.2 mM oleic acid and the indicated concentrations of compound XJ034. Data are represented as mean ± SD. n = 4 for each group. ***$P < 0.001$ compared with the oleic acid (OA) treatment group.

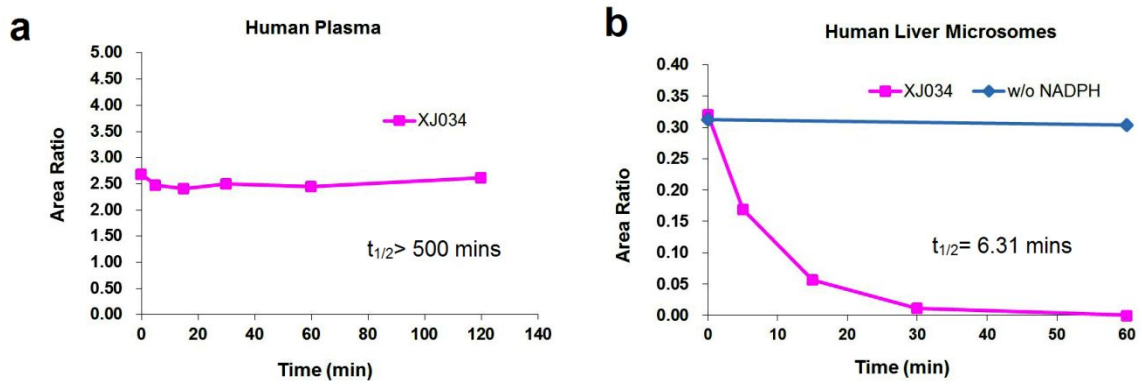


Figure S3. *In vitro* stability of compound XJ034. (**a**) peak area ratio as a function of incubation time in human plasma. The half-life ($t_{1/2}$) of compound XJ034 was estimated as greater than 500 minutes. (**b**) peak area ratio as a function of incubation time in human liver microsomes. The half-life ($t_{1/2}$) of compound XJ034 was determined as 6.31 minutes.

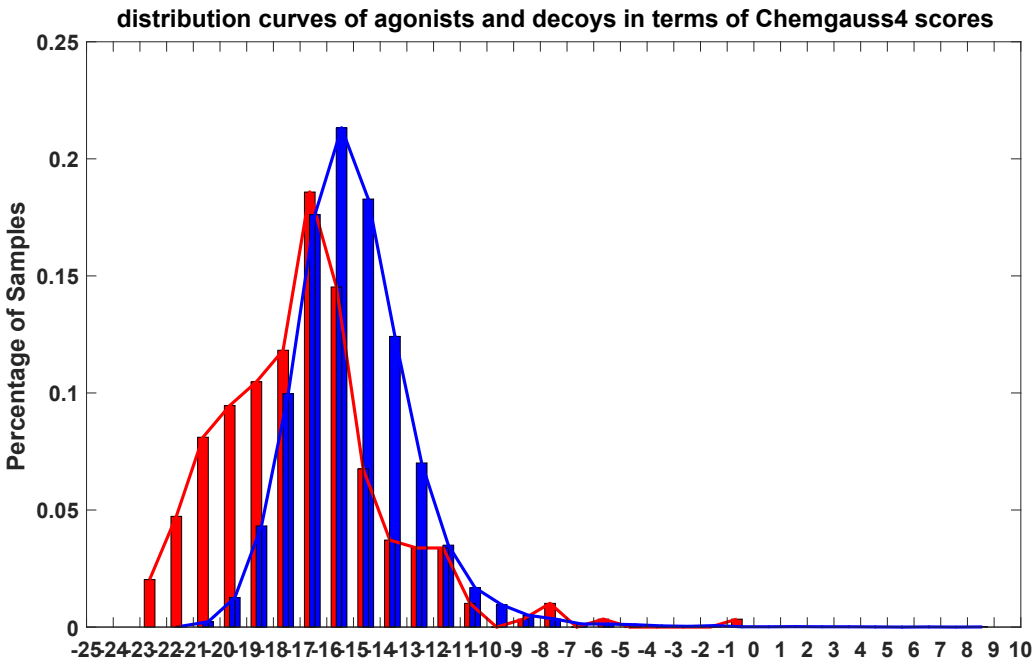


Figure S4. Distribution curves of 296 agonists and 16092 binding decoys from NRList BDB in terms of Chemgauss4 scores. Color codes: red, agonsits; blue, binding decoys. A threshold of -16 in terms of Chemgauss4 score was used to preliminarily filter chemical structures after molecular docking.

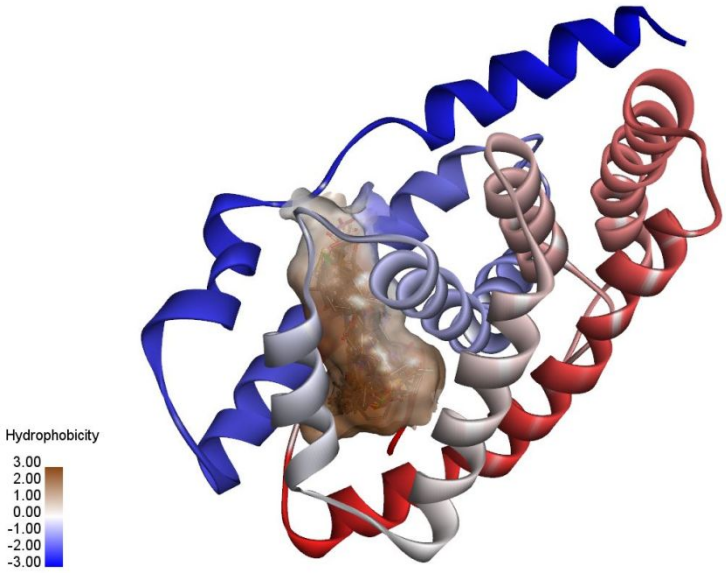
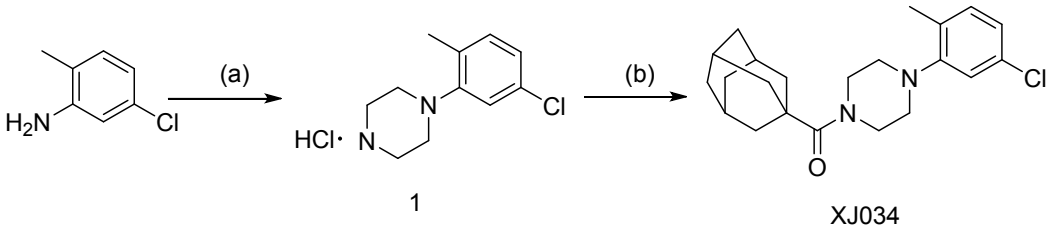


Figure S5. 20 potential hits with plausible binding modes similar to the cognate ligand in the crystal structure with PDB code of 3RUT.

Scheme S1. The synthesis and structural characterization of the compound XJ034.



Reaction conditions. (**a**) bis(2-chloroethyl)amine hydrochloride, diethylene glycol monomethyl ether, 150 °C; (**b**) adamantine carboxylic acid, HBTU, HOBt, DIPEA, DMF, r.t., 47%.

General Methods. All the materials were purchased from commercial suppliers and used without drying or purification, unless when necessary. The ^{1}H NMR (500 MHz) and ^{13}C-NMR spectra (100 MHz) were recorded by Bruker spectrometers (Varian Mercury, USA), with tetramethylsilane (TMS) as an internal standard. High resolution electrospray ionization mass spectra (HRMS) were recorded by Thermo Scientific™ Exactive™ Plus mass spectrometry (Thermo, USA). Reactions were monitored by thin layer chromatography (TLC) on silica gel sheets GF254 (Yantai Chemical Industry Research Institute, China). Spots were detected under ultraviolet light of 254 nm.

*1-(5-chloro-2-methylphenyl)piperazine hydrochloride (**1**)*. A mixture of 5-chloro-2-methylaniline (423 mg, 3 mmol), bis(2-chloroethyl)amine hydrochloride (535.5 mg, 3 mmol) and diethylene glycol monomethyl ether (0.75 mL) was heated and reacted at 150 °C for about 12 h. The reaction mixture was cooled down to room temperature and dissolved in methanol (4 mL), followed by the addition of diethyl ether (150 mL). The resulting precipitate was recovered by filtration and washed with diethyl ether to give 1 as hydrochloride (510 mg, 74%). The hydrochloride was used for the next reaction without further purification.

*1-adamantyl-[4-(5-chloro-2-methylphenyl)piperazin-1-yl]methanone (compound **XJ034**)*. To a solution of adamantanecarboxylic acid (268.5 mg, 1.5 mmol) in dry DMF (10 mL), O-benzotriazol-1-yl-*N*,*N*,N′,*N*′-tetramethyluronium hexafluorophosphate (HBTU, 420 mg, 1.5 mmol) was added, followed by the addition of 1-hydroxybenzotriazole hydrate (HOBt, 153 mg, 1 mmol), diisopropylethylamine (DIPEA, 260 μL, 1.5 mmol) and **1** (246 mg, 1 mmol). The mixture was stirred at room temperature for 15 min, and DIPEA (260 μL, 1.5 mmol) was added again. The reaction mixture was stirred at room temperature for 12 h, poured into water, and extracted with dichloromethane. The organic layer was washed with brine, dried over anhydrous sodium sulfate, and evaporated to dryness. The residue was purified by flash chromatography on silica gel to give **XJ034** as light yellow solids in 47% yield. MS-ESI m/z: 373.2 [M + H]+. ^{1}H-NMR (500 MHz, DMSO-d6): δ 7.22 (d, J = 7.9 Hz, 1H), 7.05(d, J = 9.2 Hz, 2H), 3.76 (s, 4H), 2.84 (t, J = 4.5 Hz, 4H), 2.27 (s, 3H), 2.02 (s, 3H), 1.97 (s, 6H), 1.78-1.67 (m, 6H). ^{13}C-NMR (125 MHz, DMSO-d6): δ 174.88, 152.89, 132.74, 131.19, 131.15, 123.18, 119.63, 51.98, 45.61, 41.43, 38.98, 36.54, 28.40, 17.70. HRMS calcd for $C_{22}H_{30}ClN_2O$ $[M+H]^+$, 373.2037; found, 373.2041.

References

1. Richter HG, Benson GM, Bleicher KH, et al. Optimization of a novel class of benzimidazole-based farnesoid X receptor (FXR) agonists to improve physicochemical and ADME properties. *Bioorg. Med. Chem. Lett*. 2011, 21, 1134–1140.