

Table S1 Structure of compounds with different skeletons in the data set

No.	Structure	pEC ₅₀	No.	Structure	pEC ₅₀
001		6.82	009§		9.35
002		8.59	010		6.08
003		8.37	011		6.40
004		8.35	012§		5.96
005		6.82	012		5.81
006		4.60	014		4.68
007		4.70	015§		5.42
008		9.28	016		6.02

To be continued

No.	Structure	pEC ₅₀	No.	Structure	pEC ₅₀
017		3.92	025		7.62
018		7.68	026		8.18
019		6.68	027 [§]		9.30
020		6.48	028		9.85
021		8.14	029		8.31
022		8.36	030		7.64
023		7.76	031		6.30
024 [§]		7.05	032		4.33

To be continued

No.	Structure	pEC ₅₀	No.	Structure	pEC ₅₀
033 [§]		7.14	041		9.25
034		9.46	042		10.22
035		9.64	043		9.15
036 [§]		10.15	044		8.72
037		9.89	045 [§]		8.09
038		10.22	046		9.38
039		7.37	047		8.77
040		9.89	048 [§]		6.58

To be continued

No.	Structure	pEC ₅₀	No.	Structure	pEC ₅₀
049		5.84	057 [§]		5.44
050		5.58	058		5.56
051 [§]		5.17	059		4.76
052		4.86	060 [§]		5.17
053		6.82	061		5.05
054 [§]		6.41	062		4.18
055		4.46	063 [§]		4.52
056		4.98	064		4.52

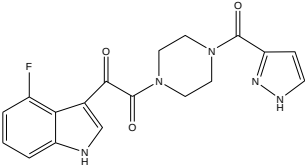
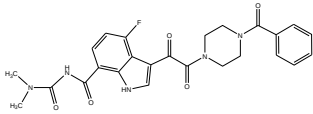
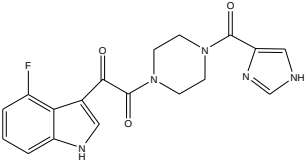
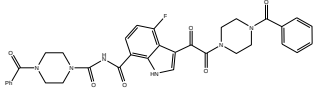
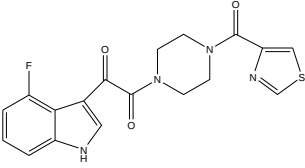
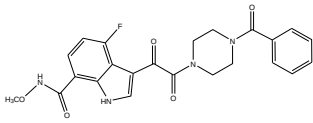
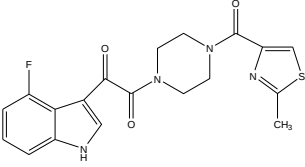
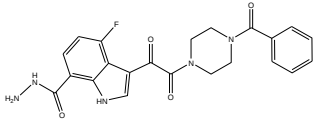
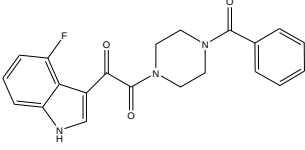
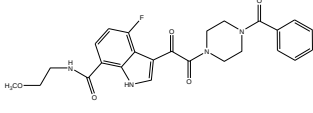
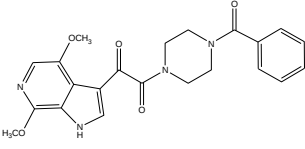
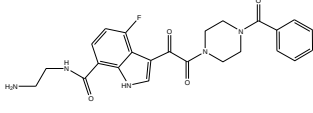
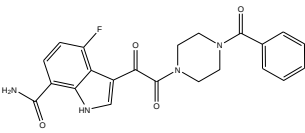
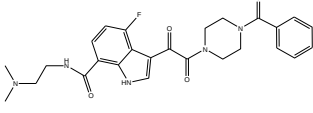
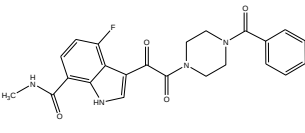
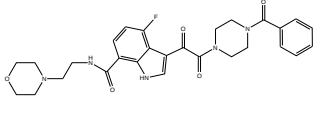
To be continued

No.	Structure	pEC ₅₀	No.	Structure	pEC ₅₀
065		5.60	073		4.26
066 [§]		5.08	074		4.74
067		4.09	075 [§]		4.43
068		5.05	076		4.84
069 [§]		5.22	077		5.46
070		5.08	078		5.81
071		5.17	079		5.38
072 [§]		4.44	080		4.06

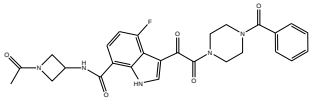
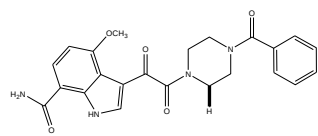
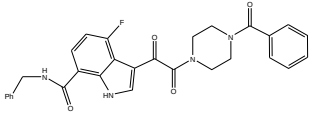
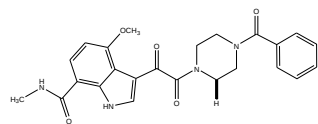
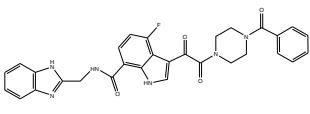
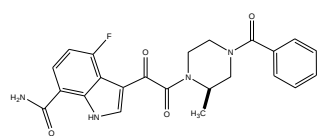
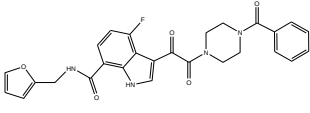
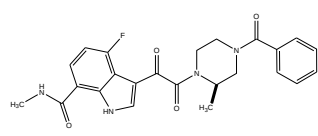
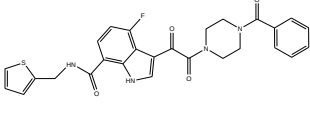
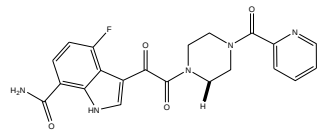
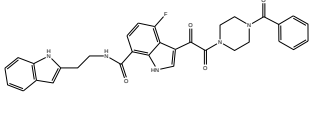
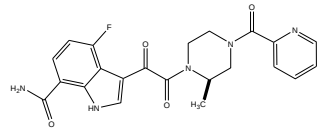
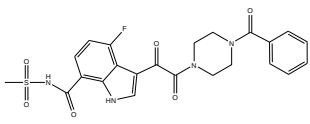
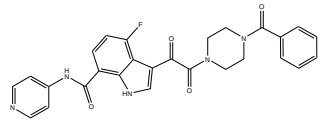
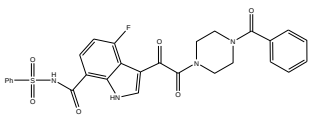
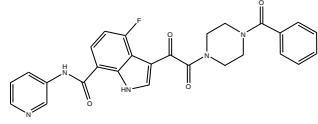
To be continued

No.	Structure	pEC ₅₀	No.	Structure	pEC ₅₀
081 [§]		5.46	089		8.15
082		8.59	090 [§]		7.69
083		8.35	091		9.15
084 [§]		6.81	092		7.76
085		7.63	093		7.92
086		8.27	094		8.49
087 [§]		8.51	095		9.40
088		8.85	096 [§]		6.81

To be continued

No.	Structure	pEC ₅₀	No.	Structure	pEC ₅₀
097		7.14	105§		6.39
098		5.91	106		8.57
099		8.06	107		8.42
100		7.64	108		8.57
101		8.91	109		7.73
102§		9.12	110		7.89
103		8.69	111		6.50
104		9.28	112		8.10

To be continued

No.	Structure	pEC ₅₀	No.	Structure	pEC ₅₀
113		7.95	121		9.12
114		6.96	122		9.12
115		8.17	123		9.85
116		8.89	124		9.12
117 [§]		8.97	125		8.12
118		8.97	126		8.85
119		8.58	127		8.47
120 [§]		9.54	128		8.04

To be continued

No.	Structure	pEC ₅₀	No.	Structure	pEC ₅₀
129 [§]		9.19	137		9.49
130		9.40	138 [§]		10.40
131		9.72	139		9.68
132		8.51	140		9.37
133		9.03	141		10.40
134		10.70	142		9.92
135 [§]		10.24	143		10.22
136		11.24	144		10.22

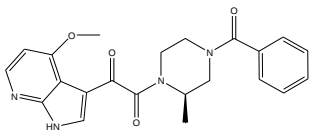
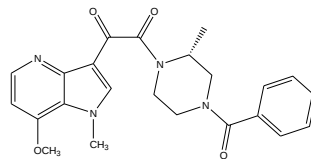
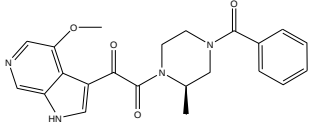
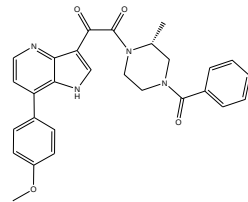
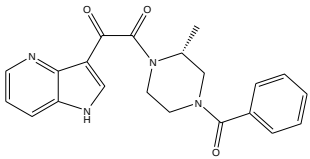
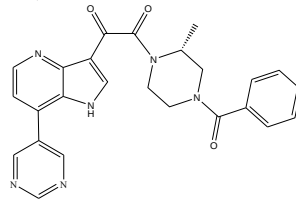
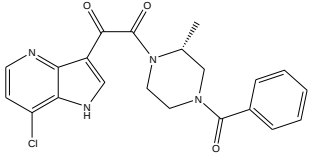
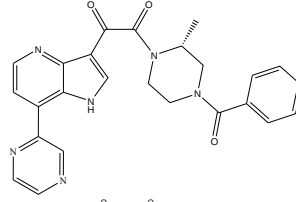
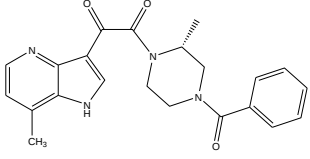
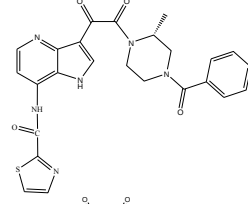
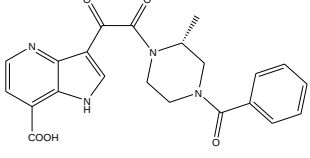
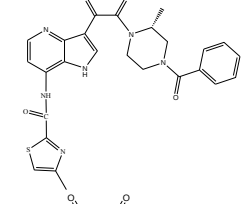
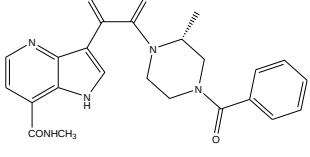
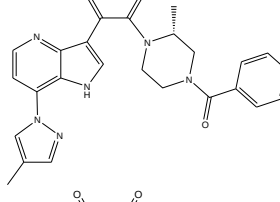
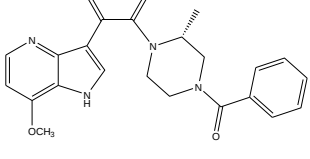
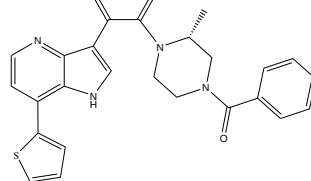
To be continued

No.	Structure	pEC ₅₀	No.	Structure	pEC ₅₀
145		11.00	153 [§]		10.10
146		11.17	154		9.06
147 [§]		10.30	155		9.28
148		10.22	156 [§]		10.22
149		11.27	157		10.03
150 [§]		10.30	158		8.45
151		10.30	159		11.09
152		10.52	160		10.40

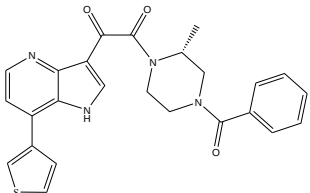
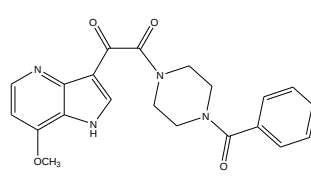
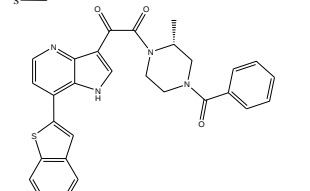
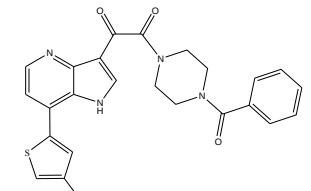
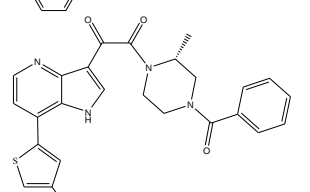
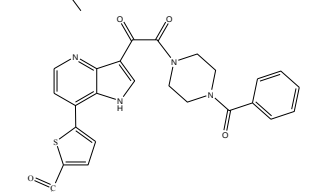
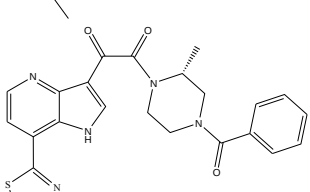
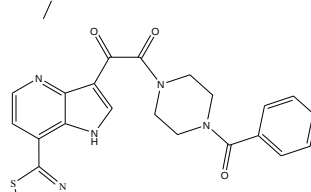
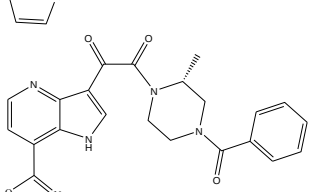
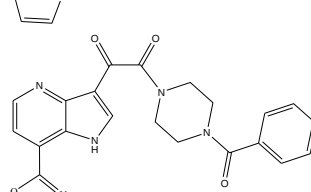
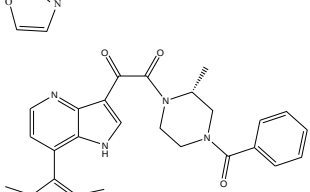
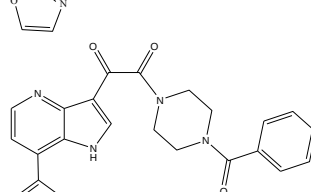
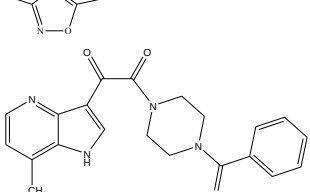
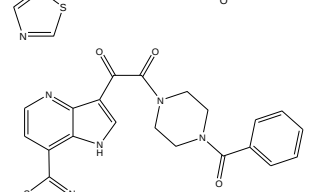
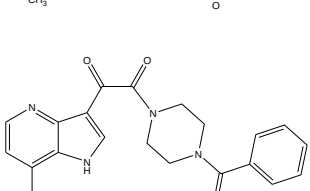
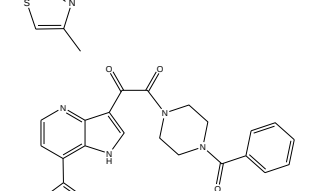
To be continued

No.	Structure	pEC ₅₀	No.	Structure	pEC ₅₀
161		9.52	169		6.72
162 [§]		10.00	170		7.70
163		10.00	171 [§]		8.31
164		9.80	172		8.80
165 [§]		9.80	173		8.77
166		10.40	174		6.24
167		10.15	175		7.66
168 [§]		7.22	176		9.10

To be continued

No.	Structure	pEC ₅₀	No.	Structure	pEC ₅₀
177 [§]		9.00	185		8.13
178		8.89	186 [§]		9.49
179		8.81	187		9.17
180		8.91	188		9.57
181		8.51	189 [§]		9.37
182		8.47	190		8.47
183 [§]		8.61	191		7.21
184		9.20	192		10.70

To be continued

No.	Structure	pEC ₅₀	No.	Structure	pEC ₅₀
193		9.74	201§		7.91
194		10.00	202		10.70
195§		10.40	203		9.54
196		10.00	204§		9.68
197		9.74	205		9.41
198		8.98	206		10.05
199		7.33	207§		9.70
200		7.74	208		9.28

To be continued

No.	Structure	pEC ₅₀	No.	Structure	pEC ₅₀
209		9.03	217		7.63
210 [§]		8.77	218		7.97
211		8.13	219 [§]		7.94
212		9.49	220		9.62
213 [§]		9.19	221		9.68
214		9.06	222 [§]		9.47
215		9.05	223		9.51
216 [§]		9.68	224		9.15

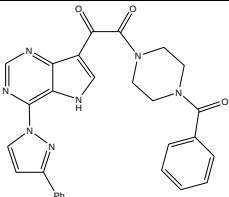
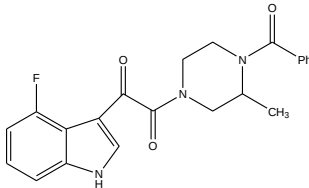
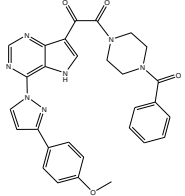
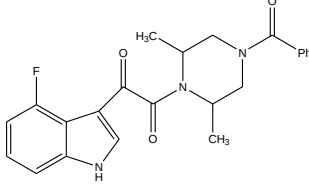
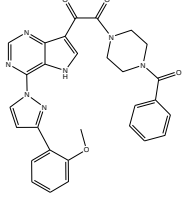
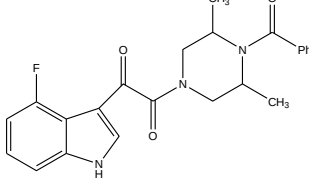
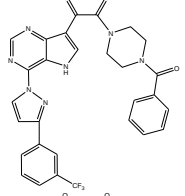
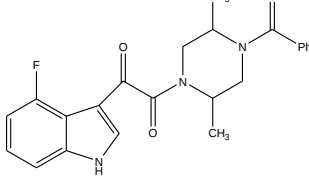
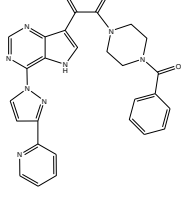
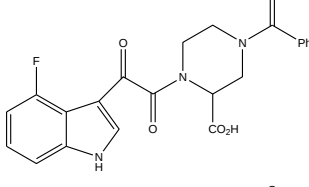
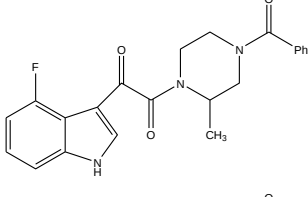
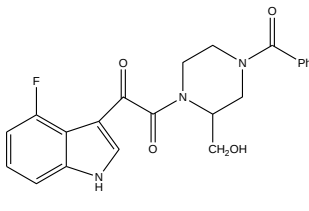
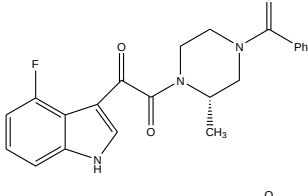
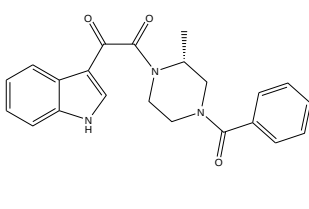
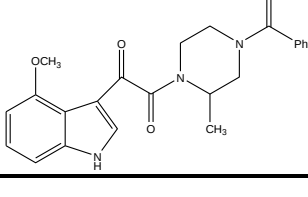
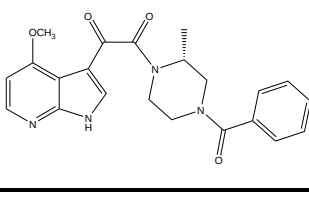
To be continued

No.	Structure	pEC ₅₀	No.	Structure	pEC ₅₀
225 [§]		9.72	233		9.00
226		9.36	234 [§]		9.00
227		9.34	235		7.96
228 [§]		8.80	236		8.39
229		9.51	237		7.92
230		8.27	238		7.36
231 [§]		9.23	239		7.74
232		7.97	240		9.47

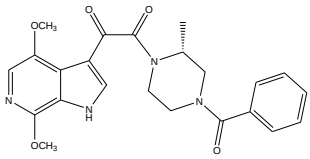
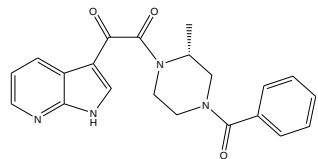
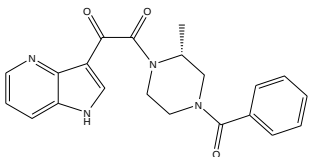
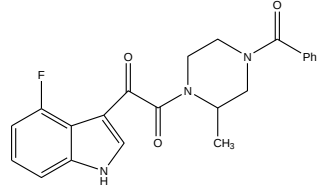
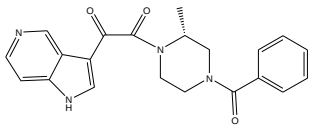
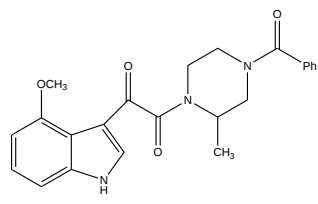
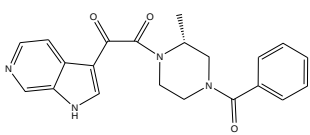
To be continued

No.	Structure	pEC ₅₀	No.	Structure	pEC ₅₀
241		8.89	249§		7.96
242		7.92	250		10.00
243§		9.60	251		9.44
244		8.28	252		8.96
245		8.57	253		9.82
246§		9.51	254		9.12
247		9.06	255§		8.70
248		10.30	256		8.32

To be continued

No.	Structure	pEC ₅₀	No.	Structure	pEC ₅₀
257		9.55	265		8.67
258 [§]		9.62	266		7.63
259		9.44	267 [§]		6.90
260		9.43	268		7.64
261		9.82	269		6.08
262		9.43	270		6.19
263		9.38	271		8.62
264 [§]		9.24	272		8.83

To be continued

No.	Structure	pEC ₅₀	No.	Structure	pEC ₅₀
273 [§]		9.06	277		8.77
274		8.80	278		8.80
275		6.24	279		9.24
276 [§]		7.67			

[§]Test set (GA splitting approach).

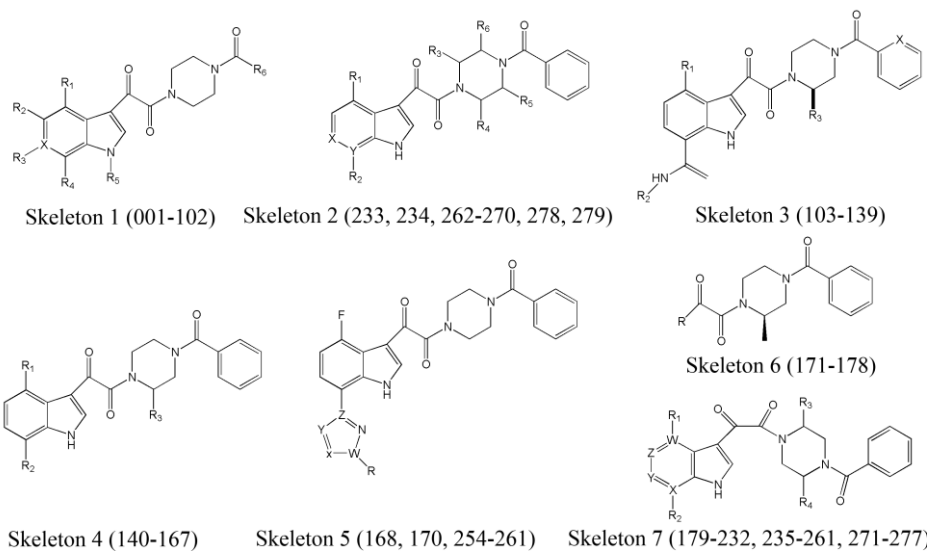


Figure S1. Different skeletons of HIV-1 inhibitors in this work.

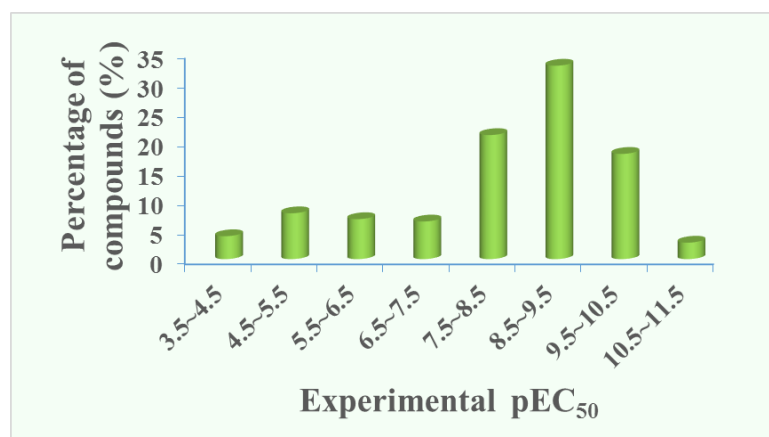


Figure S2. A wide diversity in activity for HIV-1 inhibitors in this work.

The SPL language

The SPL is an interpreted language that supplements the usual commands within SYBYL, whose programs can also be stored and shared by other users. Users have written SPL programs to customize the SYBYL user interface for their site, to structure commands into a logical process for novice users, to implement their own column types in the Molecular Spreadsheet, and to expand and customize SYBYL in many other ways. One use of SPL is to write short, fast “programs” for performing repetitive tasks quickly. The programs at the command line can be typed without saving them for future use. SPL has a wide variety of uses: 1) Create abbreviations for existing SYBYL commands; 2) Modify the behavior of existing SYBYL commands; 3) Add new commands to SYBYL; 4) Call and pass data to external program; 5) Add new menus to SYBYL. Fig. S3 depicts SPL code were written in which energy minimizations were performed using the Tripos force field and Gasteiger-Hückel charges were applied to all the molecules of data set.

```
1 database close
2 setvar dir %prompt(string "Source" "The directary name of your mol2s")
3 if %not(%file_isdir($dir))
4     echo bad input!
5 endif
6 setvar name %prompt(string "Destination" "The name of output database")
7 database create $name
8 tailor set maximin2 maximum_iterations 1000 ||
9 for i in %dir($dir/*.mol2)
10     zap *
11     mol2 in m1 $i
12     charge m1 compute gast_huck |
13     MAXIMIN2 m1 done INTERACTIVE
14     database add m1 replace
15 endfor
16 zap *
17 database close
18 echo
19 echo done    /****Made by YanLi. All rights reserved.****/
```

Figure S3. Energy minimizations were performed using the Tripos force field and Gasteiger-Hückel charges were applied to all the molecules of data set by using SPL code.

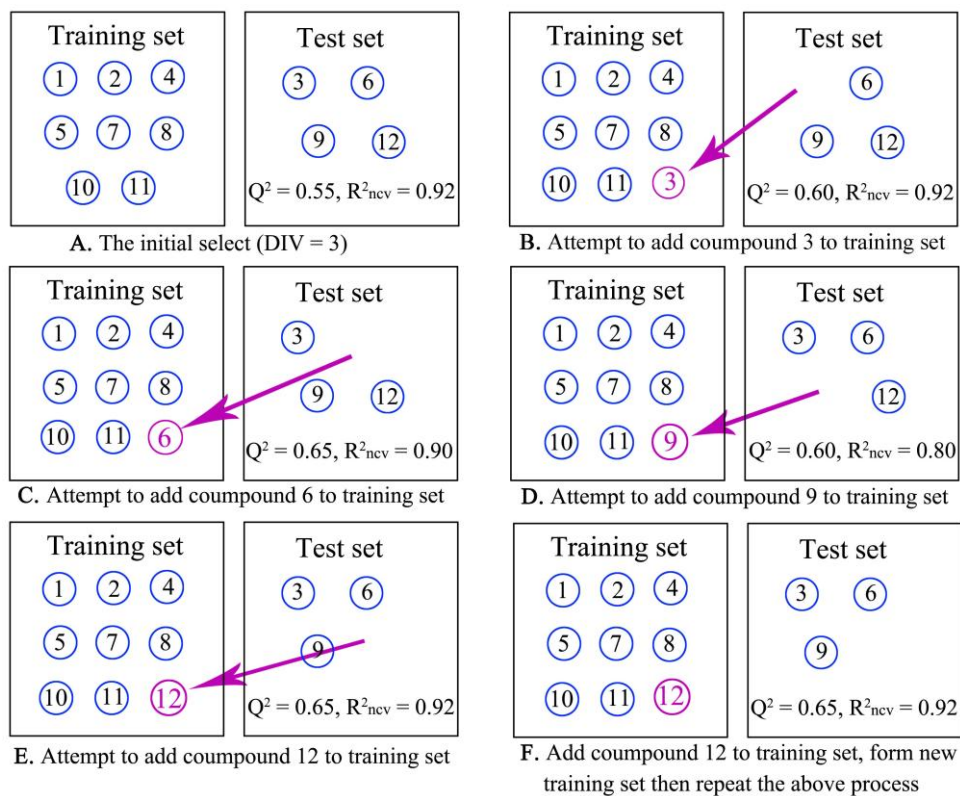


Figure S4. An instance of the 11 compounds divided into training and test sets using GA. Firstly, the whole compounds can be divisible by div (=3) as initial test set, the rest of compounds are as the training test. Then, add each compound in test set respectively to the training set, calculate the model parameters and record. Thirdly, choose the best compound, and go to the next selection. At last, the compound 12 has the best Q^2 and R^2_{nev} , thus the compound 12 was added to the training set.

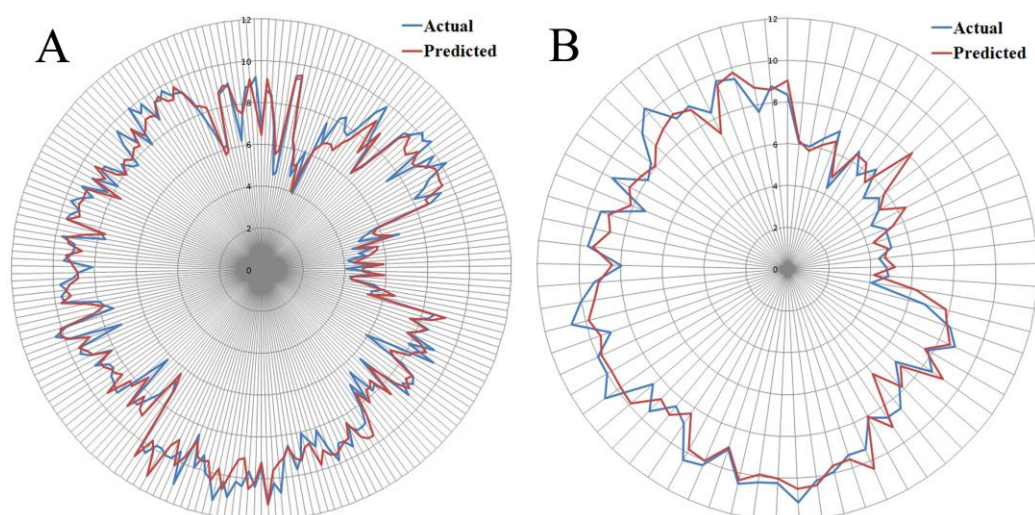


Figure S5. The ligand-based correlation plots of the predicted versus the actual pEC₅₀ values using the training (A) and the test (B) set compounds from the CoMSIA models. The red and blue solid lines are the regression lines for the actual and predicted bioactivities of training and test compounds.