

## Prioritisation of Compounds for 3CL<sup>pro</sup> Inhibitor Development on SARS-CoV-2 Variants

Marko Jukič <sup>1,2</sup>, Blaž Škrlj <sup>3</sup>, Gašper Tomšič <sup>4</sup>, Sebastian Pleško <sup>5</sup>, Črtomir Podlipnik <sup>6\*</sup> and Urban Bren <sup>1,2\*</sup>

<sup>1</sup> Laboratory of Physical Chemistry and Chemical Thermodynamics, Faculty of Chemistry and Chemical Engineering, University of Maribor, , Smetanova ulica 17, SI-2000 Maribor, Slovenia; marko.jukic@um.si

<sup>2</sup> Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska, Glagoljaška 8, SI-6000 Koper, Slovenia

<sup>3</sup> Institute Jožef Stefan, Jamova cesta 39, SI-1000 Ljubljana, Slovenia; blaz.skrj@ijs.si

<sup>4</sup> Independent Researcher, Cesta Cirila Kosmača 66, SI-1000 Ljubljana, Slovenia; gasper.tomsic@icloud.com

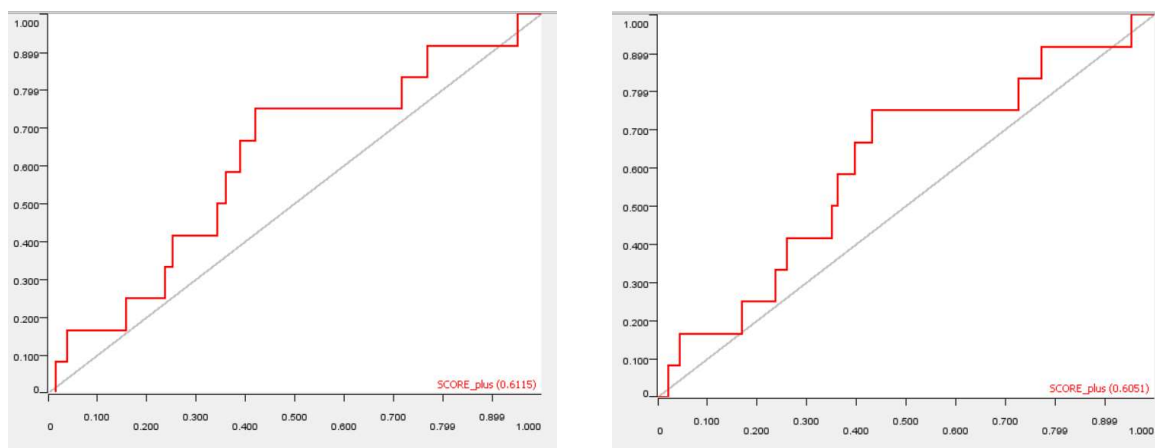
<sup>5</sup> Erudio, Litostrojska cesta 40, SI-1000 Ljubljana, Slovenia; sebastian.plesko@erudio.si

<sup>6</sup> Faculty of Chemistry and Chemical Technology, University of Ljubljana, Večna pot 113, SI-1000 Ljubljana, Slovenia

\* Correspondence: crtomir.podlipnik@fkkt.uni-lj.si (Č.P.); urban.bren@um.si (U.B.); Tel: +386-41-440-198 (Č.P.); Tel: +386-2-22-94-421 (U.B.)

Furthermore, we employed the activity data from PostEra Covid Moonshot project ([https://covid.postera.ai/covid/activity\\_data](https://covid.postera.ai/covid/activity_data), accessed on 8.5.2021). We selected compounds with  $\text{pIC}_{50}$  above 7 as true actives and compounds with  $\text{pIC}_{50}$  up to 4.00436 as inactives or experimental decoys (compounds with no data were left out). When using 2% (due to

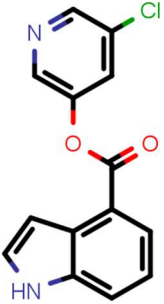
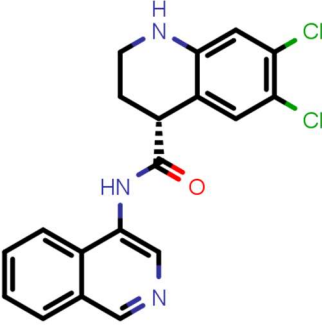
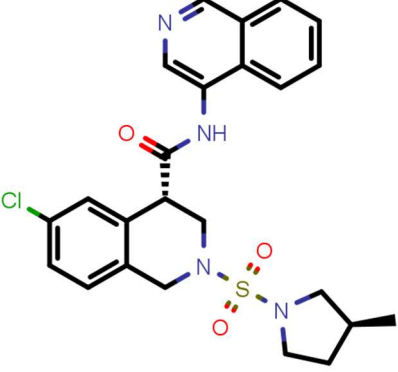
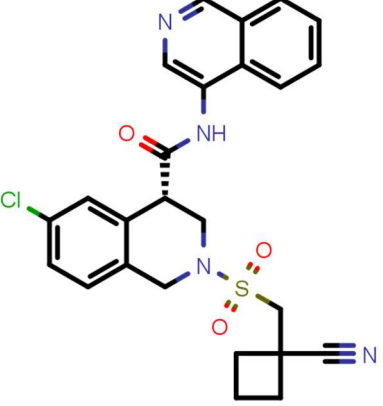
availability of experimental decoys; 446) and 10% of actives in the test set, we obtained a ROC AUC of 0.605 and 0.6115, respectively, indicating the docking protocol can identify active compounds and produce enriched libraries.

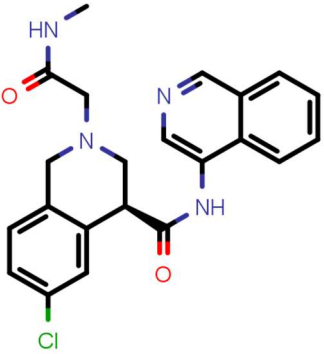
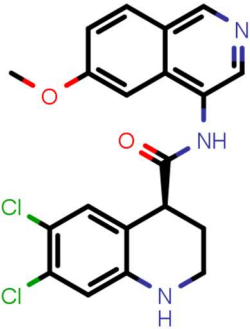
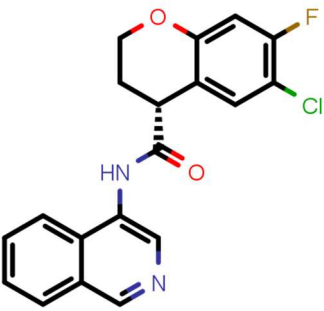


**Figure S2: ROC curves for 2% (left) and 10% (right) actives in the complete test set (PostEra COVID Moonshot project).**

*Postera true actives* ([https://covid.postera.ai/covid/activity\\_data](https://covid.postera.ai/covid/activity_data)):

STRUCTURE	NAME (PostEra), SMILES
	MSK-46727e7b-1  <chem>CC(C)C[C@H](NC(=O)OCc1ccccc1)C(=O)N[C@@H](C[C@H]1CCNC1=O)[C@H](O)S(=O)(=O)[O-]</chem>
	MAT-POS-e69ad64a-2  <chem>C=CC(=O)N(C(=O)[C@@H]1COc2ccc(Cl)cc21)c1cncc2ccccc12</chem>

	<p>ALP-POS-c59291d4-5</p> <p><chem>O=C(Oc1cncc(Cl)c1)c1cccc2[nH]ccc12</chem></p>
	<p>MAT-POS-f9802937-8</p> <p><chem>O=C(Nc1cncc2ccccc12)[C@@H]1CCNc2cc(Cl)c(Cl)cc21</chem></p>
	<p>MAT-POS-4223bc15-11</p> <p><chem>C[C@H]1CCN(S(=O)(=O)N2Cc3ccc(Cl)cc3[C@H](C(=O)Nc3cncc4ccccc34)C2)C1</chem></p>
	<p>MAT-POS-4223bc15-12</p> <p><chem>N#CC1(CS(=O)(=O)N2Cc3ccc(Cl)cc3[C@H](C(=O)Nc3cncc4ccccc34)C2)CCC1</chem></p>

	<p>MAT-POS-4223bc15-23</p> <p><chem>CNC(=O)CN1Cc2ccc(Cl)cc2[C@H](C(=O)Nc2cncc3ccccc23)C1</chem></p>
	<p>MAT-POS-ec6d90b7-6</p> <p><chem>COc1ccc2cncc(NC(=O)[C@H]3CCNc4cc(Cl)c(Cl)cc43)c2c1</chem></p>
	<p>MAT-POS-a13804f0-3</p> <p><chem>O=C(Nc1cncc2ccccc12)[C@@H]1CCOc2cc(F)c(Cl)cc21</chem></p>

### 3. HTVS results

Supplied smiles, all other data is in the supplied xls file.

no	smiles	score
1	<chem>O=C(c1ccc2c(c1)C(=O)N(C2=O)c1c(C)n(n(c1=O)c1cccc1)C)c1ccc2c(c1)C(=O)N(C2=O)c1c(C)n(n(c1=O)c1cccc1)C</chem>	22,8 019
2	<chem>Clc1ccc(cc1S(=O)(=O)Nc1c(C)n(n(c1=O)c1cccc1)C)C(=O)Nc1ccc(cc1)C(C)(C)C</chem>	21,9 941
3	<chem>O=C(c1ccc(c(c1)S(=O)(=O)Nc1c(=O)n(n(c1C)C)c1cccc1)Cl)Nc1ccc2c(c1)cccc2</chem>	21,0 602
4	<chem>COc1ccc(cc1S(=O)(=O)Nc1c(C)n(n(c1=O)c1cccc1)C)c1nnnn1CC</chem>	20,8 952
5	<chem>Cn1c(nnc1[C@@H]1C[C@@H](C1)[NH3+])Cn1c(nc2c1cccc2)N1CCCC1</chem>	20,5 798

		-
6	<chem>Cc1cn(c(=O)nc1N1CCCCC1)c1nc(oc1c1cccc1)c1cccc1</chem>	21,0 488
		-
7	<chem>Cn1c(=O)n(C)c(=O)c2c1nc(n2CCSc1nc2c(s1)cccc2)N1CCN(CC1)c1cccc1</chem>	25,6 926
		-
8	<chem>Cn1c(nnc1[C@@H]1C[C@H](C1)[NH3+])Cn1c(nc2c1cccc2)N1CCCC1</chem>	22,5 395
		-
9	<chem>Nc1cccc1c1scc(n1)c1cc(n(c1C)c1ncn[nH]1)C</chem>	22,4 204
		-
10	<chem>[S-]c1nnc(n1c1cnn(c1)Cc1cccc1)C1CCN(CC1)S(=O)(=O)c1cccc1</chem>	25,6 427
		-
11	<chem>COc1cc(ccc1O)c1c/c(=N/[C@H](C(=O)[O-])Cc2nc[nH]c2)/c2c(o1)cc(c(c2)Cl)C</chem>	22,0 924
		-
12	<chem>Clc1cccc(c1)N1CCN(CC1)c1nc2c(n1CCSc1nc3c([nH]1)cccc3)c(=O)[nH]c(=O)n2C</chem>	21,0 076
		-
13	<chem>Clc1cccc(c1)N1CCN(CC1)c1nc2c(n1CCSc1nc3c(o1)cccc3)c(=O)[nH]c(=O)n2C</chem>	23,3 288
		-
14	<chem>CCOC(=O)c1cccc1NC(=O)N[C@H](c1c(sc2c1CCCC2)n1cccc1)C</chem>	24,0 728
		-
15	<chem>Clc1cccc(c1)N1CCN(CC1)c1nc2c(n1CCSc1nnnn1c1cccc1)c(=O)[nH]c(=O)n2C</chem>	20,8 048
		-
16	<chem>[S-]c1nnc(n1c1cnn(c1)Cc1cccc1)[C@H]1CCCN(C1)S(=O)(=O)c1cccc1</chem>	22,9 086
		-
17	<chem>Clc1ccc(c(c1)Cl)Cn1ncc(c1)n1c([S-])nnc1C1CCN(CC1)S(=O)(=O)C</chem>	24,2 461
		-
18	<chem>Cc1nn(c(c1N1CCN(CC1)C(=O)c1[nH]nc(c1)c1cccc1)C)c1cccc1</chem>	22,6 734
		-
19	<chem>CCN(S(=O)(=O)c1ccc(cc1)C(=O)N1CCN(CC1)C1=NS(=O)(=O)c2c1cccc2)CC</chem>	23,9 761
		-
20	<chem>[S-]c1nnc(n1c1cnn(c1)Cc1cccc1)[C@H](n1nccc1C)C</chem>	23,2 749
		-
21	<chem>Cc1nn(c(c1N1CCN(CC1)C(=O)c1[nH]nc(c1)c1cccs1)C)[C@H]1CCS(=O)(=O)C1</chem>	23,0 223
		-
22	<chem>CC[C@H](c1c(sc2c1CC[N@@H+](C2)CC)n1cccc1)NC(=O)Nc1ccc(c(c1)C)C</chem>	21,6 719
		-
23	<chem>Cc1cnn(c1)[C@@H](c1nnc(n1c1cnn(c1)Cc1cccc1)[S-])C</chem>	21,3 904
		-
24	<chem>Clc1cccc1Cn1c(nc2c1c(=O)n(C)c(=O)n2C)N1CC[NH2+][CC1]</chem>	20,5 788
		-
25	<chem>c1ccc(cc1)c1[nH+]c(cn1c1sc(nc1c1cccc1)c1cccc1)c1cccc1</chem>	23,4 83
		-
26	<chem>Clc1ccc(cc1)Cn1ncc(c1)n1c([S-])nnc1[C@@H]1CCCN(C1)S(=O)(=O)C</chem>	21,0 243
		-
27	<chem>COCc1cc(C)nc2c1c(n1cccc1)c(s2)c1ccc(cc1)C#N</chem>	20,5 256

		-
		23,2
28	[O-]C(=O)[C@@H]1CN(C[C@H]1c1ccncc1)c1nnnn1c1cccc1	589
		-
		20,7
29	O=C(c1c(C)nn(c1n1cccn1)C)NC[C@H]1[NH2+]CCc2c1cccc2	071
		-
		22,7
30	O=C(c1cnn(c1n1cccc1)c1cccc1F)Nc1ccc2c(c1)OCO2	862
		-
		20,8
31	O=C(Nc1cc(C)cc(c1)C)CSc1nnc(n1Cc1cccc1)n1cccc1	099
		-
		21,0
32	Cc1cnn(c1)[C@H](c1nnc(n1c1cnn(c1)Cc1cccc1)[S-])C	524
		-
		24,7
33	Clc1ccc2c(c1)c(=O)nc(n2C1=CCN(CC1)C(=O)Cc1nonc1C)C	638
		-
		20,8
34	O=C(c1noc(c1)c1ccco1)N1CCN(CC1)c1c(C)nn(c1C)c1cccc1	548
		-
		21,6
35	O=C(N[C@H](c1c(sc2c1CCCC2)n1cccc1)C)Nc1cccc(c1)Br	433
		-
		22,0
36	Cc1ccc(cc1)c1[nH+]c(c1n1c1sc(nc1c1cccc1)c1cccc1)c1cccc1	427
		-
		22,0
37	CCc1nc(=O)c2c(n1C1=CCN(CC1)C(=O)C[C@H]1C(=O)NCC[N@H+]1C)cccc2	768
		-
		21,0
38	C[C@@H]1CCCN(C1)C(=O)[C@@H]1CCN(CC1)C1=NS(=O)(=O)C(=C1C)c1ccc(c(c1)C)C	378
		-
		21,1
39	CCS(=O)(=O)c1sc(c(n1)S(=O)(=O)c1ccc(cc1)Cl)N1CCCCC1	373
		-
		20,8
40	Clc1ccc(cc1)S(=O)(=O)NCc1c(n2cccc2)n(c(c1C)C)Cc1cccc1	832
		-
		25,5
41	Cc1nnc(o1)[C@]12COC[C@@H]2CN(C1)c1nc2c(n1C)cccc2	332
		-
		21,8
42	O=C(N1CCN(CC1)c1c(C)nn(c1C)[C@@H]1CCS(=O)(=O)C1)CCc1c(C)noc1C	208
		-
		20,9
43	CC(=O)Nc1ccc(cc1)NC(=O)c1cnn(c1n1cccc1)c1cccc1	442
		-
		20,9
44	O=C(Cn1c(=O)n(c2c1cccc2)C1=CCCC1)NCc1ccc(cc1)NC(=O)C	659
		-
		21,0
45	O=C(Nc1ccc2c(c1)OCO2)CSc1nnc(n1c1ccc(cc1)C)N1CCCC1	771
		-
		21,5
46	NC(=O)C1CCN(CC1)C(=O)Cn1nc(C)c2c(c1=O)c(n1cccc1)n(n2)c1ccc(cc1)Cl	932
		-
		20,5
47	COc1cc(C)c(c(c1)C)S(=O)(=O)Nc1c(C)n(n(c1=O)c1cccc1)C	389
		-
		24,0
48	S=C(NC(=O)c1cccc2c1cccc2)Nc1c(cnn1c1ccc(cc1)S(=O)(=O)N)C(=O)N	096
		-
		21,6
49	O=C(N=c1[nH]c2c([nH]1)cccc2)NC[C@@H]1CCC[C@H](C1)CNC(=O)N=c1[nH]c2c([nH]1)cccc2	642

		-
		23,3
50	<chem>O=C(c1c(/N=C/c2c3ccccc3cc3c2cccc3)sc2c1CCCC2)NCc1cccc1</chem>	325
		-
		20,6
51	<chem>O=C(c1ccc(cc1)S(=O)(=O)N1CCOCC1)N1CCN(CC1)C1=NS(=O)(=O)c2c1cccc2</chem>	105
		-
		22,4
52	<chem>CCOC(=O)N1CCC(CC1)NC(=O)c1nnn(c1n1cccc1)Cc1cccc1</chem>	857
		-
		21,2
53	<chem>O=C(c1cccc(c1)S(=O)(=O)Nc1c(C)n(n(c1=O)c1cccc1)C)Nc1cccc1C</chem>	497
		-
		22,8
54	<chem>c1ccc(nc1)c1noc(n1)[C@@H]1CCCN1c1nnnn1c1cccc1</chem>	455
		-
		23,2
55	<chem>O=C([C@H]1CCCN(C1)c1nnc(s1)n1c(C)ccc1C)NCc1cccc1Cl</chem>	297
		-
		21,0
56	<chem>O=C(c1nnn(c1n1cccc1)c1cccc1)Nc1cnccc1C</chem>	582
		-
		22,4
57	<chem>O[C@H]1CCN(C1)c1nc2c(n1C)ccc(c2)C(=O)NCCc1ccc(o1)C</chem>	271
		-
		22,2
58	<chem>S=C1N[C@@H]([C@@H](N1c1ccc(cc1)OC1CCCC1)c1cc(n(c1C)c1nccs1)C)c1ccccn1</chem>	097
		-
		22,7
59	<chem>Cc1ccc(cc1)n1c(c2ccccc2)c(cc1c1cccc1)/C=N/c1sc2c(c1C(=O)NC1CCCCC1)CCCC2</chem>	301
		-
		29,0
60	<chem>[NH3+][C][C@@H]1O[C@H](O[C@@H]2[C@H]([NH3+])C[C@@H]([C@@H]([C@@H]2O)O[C@H]2O[C@@H](CO)[C@H]([C@@H]([C@H]2O)[NH3+])O)[NH3+])[C@@H]([C@@H]([C@@H]1O)O)O</chem>	171
		-
		23,2
61	<chem>COc1ccc(cc1)c1nc(sc1C)n1c(C)c2c(c1C)c(C)nnc2C</chem>	024
		-
		23,9
62	<chem>Cc1cccc(c1)N1C(=S)N[C@@H]([C@@H]1c1cc(n(c1C)c1nccs1)C)c1ccccn1</chem>	945
		-
		26,6
63	<chem>S=C(Nc1c(C)nn(c1C)Cc1cccc1C)N[C@@H]1C[C@H]2CC[C@@H](C1)[N@@H+]2C</chem>	451
		-
		20,5
64	<chem>OCCNc1nc(=O)n(cc1C)c1nc(sc1c1cccc1)c1cccc1</chem>	596
		-
		23,1
65	<chem>O=C(Nc1ccc2c(c1)OCO2)CSc1nnc(n1Cc1cccc1)N1CCOCC1</chem>	729
		-
		29,3
66	<chem>Cc1sc(nc1c1ccc(cc1)Oc1cccc1)n1c(C)c2c(c1C)c(C)nnc2C</chem>	019
		-
		22,9
67	<chem>CCOc1ccc(cc1)S(=O)(=O)N(c1c(C)n(n(c1=O)c1cccc1)C)Cc1ccc(cc1)C</chem>	799
		-
		22,5
68	<chem>O=C(C1CCN(CC1)c1nnc(s1)n1c(C)ccc1C)NCc1cccc(c1)Cl</chem>	213
		-
		23,5
69	<chem>Clc1ccc(cc1)N1C(=S)N[C@@H]([C@@H]1c1cc(n(c1C)c1nccs1)C)c1ccccn1</chem>	559
		-
		30,9
70	<chem>Clc1ccc(cc1)n1nc(/c(=N/C(=O)c2nnn(c2c2ccccc2)c2ccccc2)/s1)C(=O)N1CCOCC1</chem>	678
		-
		22,5
71	<chem>COc1cc(ccc1O)c1nc2n(c1Nc1cccc(c1)Cl)c(C)ccc2</chem>	591



		-
		22,9
72	<chem>S=C1N[C@H]([C@@H](N1c1ccc(c(c1)C)C)c1cc(n(c1C)c1nccs1)C)c1ccccc1</chem>	884
		-
		21,5
73	<chem>COc1cc(ccc1NS(=O)(=O)C)N1C(=S)N[C@H]([C@@H]1c1cc(n(c1C)c1nccs1)C)c1ccccc1</chem>	062
		-
		24,5
74	<chem>COc1ccc(cc1)c1[nH]n/c(=N/C(=O)Cc2ccccc2)/c1)[C@H]1CCS(=O)(=O)C1</chem>	127
		-
		23,0
75	<chem>Cc1ccc(cc1)N1C(=S)N[C@@H]([C@@H]1c1cc(n(c1C)c1nccs1)C)c1ccccc1</chem>	373
		-
		21,0
76	<chem>COc1cccc1Nc1c(nc2n1c(C)ccc2)c1ccc(cc1)O</chem>	136
		-
		23,0
77	<chem>Fc1ccc(cc1)[C@H]1C(=C(C(=O)N1c1c(C)n(n(c1=O)c1ccccc1)C)O)C(=O)c1ccc(cc1)C</chem>	111
		-
		22,4
78	<chem>S=C1N[C@H]([C@@H](N1c1ccc(c(c1)C)Br)c1cc(n(c1C)c1nccs1)C)c1ccccc1</chem>	476
		-
		20,8
79	<chem>COc1cc(ccc1OC)C(=O)/N=c/1\cc([nH]n1[C@H]1CCS(=O)(=O)C1)c1ccccc1</chem>	044
		-
		22,6
80	<chem>N#Cc1c(/N=C/c2c(OCc3ccc(cc3)Cl)ccc3c2cccc3)sc2c1CCCC2</chem>	305
		-
		22,2
81	<chem>Cc1ccc(cc1)N1C(=S)N[C@H]([C@@H]1c1cc(n(c1C)c1nccs1)C)c1ccccc1</chem>	832
		-
		22,4
82	<chem>Oc1cccc1N1C(=S)N[C@H]([C@H]1c1cc(n(c1C)c1nccs1)C)c1ccccc1</chem>	993
		-
		22,1
83	<chem>S=C1N[C@H]([C@@H](N1c1cccc1)c1cc(n(c1C)c1nccs1)C)c1ccccc1</chem>	985
		-
		22,1
84	<chem>S=C1N[C@H]([C@@H](N1c1ccc(c(c1)C)F)c1cc(n(c1C)c1nccs1)C)c1ccccc1</chem>	438
		-
		23,0
85	<chem>S=C1N[C@H]([C@@H](N1C1CCCC1)c1cc(n(c1C)c1nccs1)C)c1ccccc1</chem>	523
		-
		21,6
86	<chem>S=C(Nc1c(cnn1c1ccc(cc1C)C)C(=O)N)NC(=O)c1ccc(c(c1)S(=O)(=O)N1CCOCC1)Cl</chem>	727
		-
		20,6
87	<chem>COc1ccc(cc1)N1C(=S)N[C@H]([C@@H]1c1cc(n(c1C)c1nccs1)C)c1ccccc1</chem>	732
		-
		23,6
88	<chem>Fc1ccc(cc1)N1C(=S)N[C@@H]([C@H]1c1cc(n(c1C)c1nccs1)C)c1ccccc1</chem>	303
		-
		23,7
89	<chem>S=C1N[C@@H]([C@@H](N1c1ccc(cc1)Oc1ccccc1)c1cc(n(c1C)c1nccs1)C)c1ccccc1</chem>	284
		-
		22,5
90	<chem>Cc1sc(nc1c1ccc(c(c1)C)C)n1c(C)c2c(c1C)c(C)nnc2C</chem>	38
		-
		20,9
91	<chem>Cc1cccc(c1)c1nn(cc1C(=O)N(c1nc2c(s1)cccc2)C[C@H]1CCCCO1)c1ccccc1</chem>	917
		-
		23,8
92	<chem>COc1cc(ccc1OC)c1c/c(=N\c2sc(c(n2)c2cccs2)C)/c2c(o1)cccc2</chem>	736
		-
		23,9
93	<chem>N#Cc1c(/N=C/c2cn(c3c2cccc3)Cc2ccc3c(c2)cccc3)sc2c1CC[N@@H+](C2)C</chem>	072

		-
		21,4
94	<chem>Cc1ccnc(n1)Sc1oc(cc1Br)/C=N/c1sc2c(c1C(=O)Nc1ccccc1)CCCC2</chem>	688
		-
		25,8
95	<chem>O=C(Nc1cc(C)ccc1C)CCN1C(=S)N[C@H]([C@@H]1c1cc(n(c1C)c1nccs1)C)c1ccccc1</chem>	391
		-
		23,9
96	<chem>Oc1ccccc1N1C(=S)N[C@@H]([C@H]1c1cc(n(c1C)c1nccs1)C)c1ccccc1</chem>	707
		-
		20,7
97	<chem>COCC[NH+]1CCN(CC1)C(=O)c1cccc(c1)c1nn(c(n1)N(C)C)C</chem>	663
		-
		20,9
98	<chem>COc1ccc(cc1)c1cc(nn1[C@H]1CCS(=O)(=O)C1)C(=O)N=c1[nH]cc[nH]1</chem>	064
		-
		25,3
99	<chem>O=C1N[C@@H](C(=O)N1CC(=O)c1cc(n(c1C)c1nccs1)C)Cc1c[nH]c2c1cccc2</chem>	222
		-
		21,0
100	<chem>S=C1N[C@H]([C@@H](N1c1ccc2c(c1)OCO2)c1cc(n(c1C)c1nccs1)C)c1ccccc1</chem>	808
		-
		22,1
101	<chem>Brc1ccc(cc1)S(=O)(=O)Nc1nc2cccc2nc1n1ccc(=NC(=O)c2cccc2)cc1</chem>	721
		-
		20,9
102	<chem>Clc1ccc(cc1)N1C(=S)N[C@H]([C@H]1c1cc(n(c1C)c1nccs1)C)c1ccccc1</chem>	23
		-
		21,8
103	<chem>Oc1ccccc1N1C(=S)N[C@@H]([C@H]1c1cc(n(c1C)c1nccs1)C)c1ccccc1</chem>	087
		-
		25,7
104	<chem>Clc1ccc(cc1)NC(=O)c1c(/N=C/c2ccc3c(c2)Sc2c(N3C)cccc2)sc2c1CC[C@H](C2)C(C)(C)C</chem>	525
		-
		21,3
105	<chem>COc1cccc(c1)N(C(=O)c1ccc(cc1)NC(=O)C)c1nc2c(s1)cccc2</chem>	127
		-
		24,6
106	<chem>O=C(Nc1ccc(cc1)C)CCN1C(=S)N[C@H]([C@@H]1c1cc(n(c1C)c1nccs1)C)c1ccccc1</chem>	782
		-
		23,6
107	<chem>COc1cccc(c1O)/C=N/c1c(C)n(n(c1=O)c1ccccc1)C</chem>	146
		-
		21,9
108	<chem>Clc1ccc(c(c1)Cl)COc1ccccc1/C=N/c1sc2c(c1C(=O)NCc1ccccc1)CCCC2</chem>	44
		-
		22,0
109	<chem>Oc1ccccc1/C=N/C1=CC=C(CC1)Cc1ccc(cc1)/N=C/c1ccccc1O</chem>	927
		-
		21,7
110	<chem>Clc1ccc2c(c1)oc(=O)n2CC(=O)c1cc(n(c1C)c1nccs1)C</chem>	132
		-
		22,9
111	<chem>O=C(N=c1[nH]c2c([nH]1)cccc2)CSCc1nc2sc(cc2c(=O)[nH]1)C</chem>	56
		-
112	<chem>C[NH2+][C@@H]1[C@@H](O[C@H]2[C@H]([C@H]1O)O[C@H]([C@H](C2)[NH3+])O[C@@H]1[C@H]([NH3+])C[C@@H]([C@@H]([C@@H]1O)O)[NH3+])O[C@@H]1O[C@H](CO)[C@@H]([C@H]([C@@H]1O)O)[NH3+]</chem>	31,2 203
		-
		26,7
113	<chem>O=C(N1CCN(CC1)S(=O)(=O)c1ccc(cc1)F)CCC(=O)N=c1[nH]c2c([nH]1)cccc2</chem>	982
		-
		21,3
114	<chem>N#Cc1c(/N=C/c2ccc(s2)c2cccs2)sc2c1CC[C@H](C2)C</chem>	807
		-
		21,4
115	<chem>S=C1N[C@H]([C@@H](N1c1ccc(cc1)Oc1ccccc1)c1cc(n(c1C)c1nccs1)C)c1ccccc1</chem>	658

		-
		21,2
116	<chem>N#Cc1c(/N=C/c2ccc(s2)c2cccs2)sc2c1CC[C@@H](C2)C</chem>	289
		-
		22,0
117	<chem>CC(=O)N[C@H](C(=O)NC1CCN(CC1)c1nnnn1c1cccc1)Cc1cccc1</chem>	681
		-
		22,1
118	<chem>O=C(c1snnc1C)/N=c/1\cc([nH]n1[C@H]1CCS(=O)(=O)C1)c1cccc1</chem>	781
		-
		22,9
119	<chem>Clc1ccc(cc1)NC(=O)c1c(/N=C/c2ccc(o2)N2CCOCC2)sc2c1CC[C@H](C2)C(C)(C)C</chem>	963
		-
		21,1
120	<chem>N#Cc1cccc1COc1ccc(cc1/C=N/c1sc2c(c1C(=O)Nc1ccc(cc1)Cl)CC[C@H](C2)C(C)(C)C)Br</chem>	32
		-
		24,1
121	<chem>C[C@H]1C[C@H](C)CN(C1)S(=O)(=O)c1cccc(c1)C(=O)N=C(c1ccc(cc1)N(C)C)c1ccc(cc1)N(C)C</chem>	544
		-
		22,8
122	<chem>O=C(Nc1ccc(cc1)C)CCN1C(=S)N[C@@H]([C@H]1c1cc(n(c1C)c1nccs1)C)c1ccccn1</chem>	567
		-
		20,5
123	<chem>N#Cc1c(/N=C/c2c(OCc3ccc(cc3)Br)ccc3c2cccc3)sc2c1CC[C@H](C2)C(C)(C)C</chem>	024
		-
		23,0
124	<chem>COc1ccc(cc1)c1c/c(=N\c2sc(c(n2)c2cccc2)C)/c2c(o1)cccc2</chem>	855
		-
	<chem>C[NH2+][C@H]([C@@H]1CC[C@@H]([C@@H](O1)O[C@@H]1[C@H]([NH3+])C[C@@H]([C@@H]([C@@H]1O)O[C@H]1OC[C@]([C@H]([C@H]1O)[NH2+])C(O)[NH3+])[NH3+])C</chem>	21,1
125		49
		-
		21,4
126	<chem>Clc1ccc(cc1)S(=O)(=O)c1ccc(o1)C(=O)/N=c\1/s[nH]c(n1)c1cccc1</chem>	012
		-
		21,3
127	<chem>O=C(Nc1ccc(cc1)C)CCN1C(=S)N[C@H]([C@H]1c1cc(n(c1C)c1nccs1)C)c1ccccn1</chem>	416
		-
		23,1
128	<chem>O=C(c1nn(C)c(=O)c2c1cccc2)/N=c/1\cc([nH]n1[C@@H]1CCS(=O)(=O)C1)C</chem>	844
		-
		21,3
129	<chem>COCC(=O)Nc1ccc(cc1Cl)N1C(=S)N[C@H]([C@@H]1c1cc(n(c1C)c1nccs1)C)c1ccccn1</chem>	439
		-
		25,5
130	<chem>O=C(c1cc(=O)n(c2c1cccc2)C)/N=c/1\cc([nH]n1[C@@H]1CCS(=O)(=O)C1)C</chem>	837
		-
	<chem>[NH3+]C[C@@H]1O[C@H](O[C@H]2[C@H]([NH3+])C[C@H]([C@H]([C@H]2O)O[C@H]2O[C@@H](CO)[C@H]([C@H]([C@H]2O)[NH3+])O)[NH3+])[C@@H](C[C@@H]1O)[NH3+]</chem>	27,7
131		312
		-
		21,7
132	<chem>O=C(c1sc(c(c1)c1cccc1)N1CCOCC1)Nc1ccc(cc1)n1cnnn1</chem>	378
		-
		20,6
133	<chem>CCCS(=O)(=O)c1nnc(s1)/N=C(/COc1ccc(cc1)Br)\O</chem>	256
		-
		20,8
134	<chem>Clc1ccc(cc1)NC(=O)c1c(/N=C/c2c(OCc3ccc(cc3)Br)ccc3c2cccc3)sc2c1CC[C@H](C2)C(C)(C)C</chem>	778
		-
		20,9
135	<chem>O=C1N(c2ccc3c(c2)sc(n3)Sc2ccc(o2)/C=N/c2c(C)n(n(c2=O)c2cccc2)C)C(=O)c2c1cccc2</chem>	96
		-
	<chem>[NH3+]C[C@@H]1O[C@@H](O[C@@H]2[C@H]([NH3+])C[C@@H]([C@H]([C@H]2O)O[C@@H]2O[C@H](CO)[C@@H]([C@@H]([C@@H]2O)[NH3+])O)[NH3+])[C@@H]([C@H]([C@H]1O)O)O</chem>	27,5
136		638
		-
		22,3
137	<chem>O=C(c1cccc(c1)Nc1ncccn1)/N=c/1\cc([nH]n1[C@@H]1CCS(=O)(=O)C1)C</chem>	551

		-
		21,8
138	<chem>Fc1ccc2c(c1)cc(n2C)C(=O)/N=c/1\cc([nH]n1[C@H]1CCS(=O)(=O)C1)C</chem>	077
		-
		20,7
139	<chem>CCOc1ccc(cc1)N1C(=S)N[C@H]([C@H]1c1cc(n(c1C)c1nccs1)C)c1ccccc1</chem>	452
		-
		21,0
140	<chem>O1CCN(CC1)c1nnc(n1Cc1ccccc1)SCc1noc(c1)c1ccco1</chem>	275
		-
		22,4
141	<chem>CCOc1cc(/C=N/c2c(C)n(n(c2=O)c2ccccc2)C)ccc1O</chem>	696
		-
		20,7
142	<chem>O=C(c1c(/N=C/c2c(O)ccc3c2cccc3)sc2c1CCCC2)NCc1ccccc1</chem>	485
		-
		21,2
143	<chem>Oc1ccc(cc1)n1nnnc1SCC(=O)c1cc(n(c1C)c1nccs1)C</chem>	242
		-
		23,8
144	<chem>COCCn1ccc2c1cccc2C(=O)/N=c/1\cc([nH]n1[C@H]1CCS(=O)(=O)C1)C</chem>	236
		-
		20,6
145	<chem>S=C(Nc1nnc(c(=O)[nH]1)C)/N=C/c1ccc(cc1)N(C)C</chem>	842
		-
		20,5
146	<chem>N#Cc1ccccc1NC(=O)CSc1nc2cccc2c(=O)n1c1[n-]cnn1</chem>	326
		-
		20,6
147	<chem>Brc1cc(oc1Sc1ccccc1)/C=N/c1sc2c(c1C(=O)NCc1ccco1)CCCC2</chem>	364
		-
		20,8
148	<chem>COc1cccc(c1)N(C(=O)c1ccncc1)c1nc2c(s1)cccc2</chem>	889
		-
		25,6
149	<chem>COc1cc(ccc1O)c1nc2n(c1/N=C/c1cn(c3c1cccc3)Cc1ccccc1Cl)cccc2</chem>	91
		-
		32,5
150	<chem>O=C(c1cccc(c1)N1CCCS1(=O)=O)Nc1cccc(c1)c1nnc2n1CCCCC2</chem>	13
		-
		23,5
151	<chem>C#CCn1/c(=N/C(=O)c2ccc3c(c2)CCCC3)/sc2c1ccc(c2)S(=O)(=O)N</chem>	295
		-
		22,5
152	<chem>Oc1cc(C)c(c(c1O)/C=N/c1c(=O)n(n(c1C)C)c1ccccc1</chem>	703
		-
		22,7
153	<chem>COc1ccc2c(c1)s/c(=N\C(=O)CCCN1sc3c(c1=O)cccc3)/[nH]2</chem>	957
		-
		24,5
154	<chem>Clc1ccc(cc1)NC(=O)c1c(/N=C/c2cc(c(o2)Sc2ncccn2)Br)sc2c1CCCC2</chem>	014
		-
		21,0
155	<chem>Clc1cccc(n1)C(=O)/N=c/1\cc([nH]n1[C@H]1CCS(=O)(=O)C1)c1ccccc1</chem>	956
		-
		22,2
156	<chem>COC(=O)c1c(NC(=S)NCc2cn(nc2c2cccc2)c2cccc2)sc2c1CCCC2</chem>	075
		-
		21,9
157	<chem>Cc1ccc(cc1)c1n[nH]cc1/C=N/c1sccc1C(=O)Nc1ccccc1C</chem>	494
		-
		20,9
158	<chem>O=C([C@H]1CCCN(C1)c1nnc(s1)n1c(C)ccc1C)NCCC[NH+]1CCCCC1</chem>	396
		-
		20,7
159	<chem>O=C(Nc1ccccc1C)CCN1C(=S)N[C@H]([C@H]1c1cc(n(c1C)c1nccs1)C)c1ccccc1</chem>	708

		-
		22,8
160	<chem>Cc1ccc(cc1)N1[NH2+][C@](NC1=S)(C)c1ccccc1Br</chem>	459
		-
		29,0
161	<chem>COc1ccc(cc1S(=O)(=O)NCC1ccccc1)C(=O)Nc1c(Cl)cccc1Cl</chem>	161
		-
		24,7
162	<chem>O[C@H](C[NH+](C1CCCCC1)C1CCCCC1)Cn1c(c2ccccc2)c(c2c1c1ccccc1cc2)c1ccccc1</chem>	942
		-
		22,0
163	<chem>O=C(c1cc(nc2c1c(C)no2)C(C)C)N=c1[nH]c2c([nH]1)cccc2</chem>	301
		-
		25,1
164	<chem>O=C(c1c(/N=C/c2c[nH]c3c2cccc3)sc2c1CCCC2)Nc1ccccc1</chem>	821
		-
		20,7
165	<chem>OCCCN1C(=S)N[C@H]([C@@H]1c1cc(n(c1C)c1nccs1)C)c1ccccc1</chem>	201
		-
		20,5
166	<chem>Clc1ccc(cc1)c1noc(n1)CCCS1nnc(n1Cc1ccccc1)N1CCOCC1</chem>	987
		-
		22,0
167	<chem>COc1cc(ccc1O)c1nc2n(c1/N=C/c1ccc(o1)c1cccc(c1C)Cl)cccc2</chem>	229
		-
		21,7
168	<chem>O=C(C1CCN(CC1)c1ccc(nn1)Cl)N=c1[nH]c2c([nH]1)cccc2</chem>	947
		-
		21,5
169	<chem>O=C(c1sc2c(c1N)cccn2)/N=c/1\sc2c([nH]1)cccc2</chem>	002
		-
		25,0
170	<chem>Clc1ccc(cc1)c1ccc(=O)n(n1)CC(=O)/N=c/1\sc2c([nH]1)cccc2</chem>	465
		-
		21,3
171	<chem>O=C(c1sc(c(c1N)c1sc2c(n1)cccc2)Nc1ccccc1)c1ccccc1</chem>	489
		-
		21,8
172	<chem>Clc1ccc(cc1)C(=O)c1sc(c(c1N)c1nc2c(s1)cccc2)Nc1cc(C)ccc1C</chem>	297
		-
		21,4
173	<chem>O/C(=N\c1scc(n1)C[N@H+]1CCc2c(C1)cccc2)/c1cc(=O)c2c(o1)cccc2</chem>	759
		-
		22,1
174	<chem>CCOC(=O)c1ccc(s1)NC(=O)c1cccc(c1)S(=O)(=O)N1CCCc2c1cccc2</chem>	124
		-
		23,0
175	<chem>O=C1Nc2nc(nn2[C@H](C1)c1ccc(cc1)Cl)/N=C(/COc1ccc2c(c1)cccc2)\O</chem>	783
		-
		22,2
176	<chem>Fc1ccc(cc1)NC(=S)Nc1sc2c(c1C(=O)N)CCCC2</chem>	44
		-
		21,5
177	<chem>C[C@@H]1CCc2c(C1)sc(c2C(=O)N)NC(=S)NC(=O)c1sc2c(c1Cl)ccc(c2)Cl</chem>	284
		-
		22,4
178	<chem>COc1ccc(cc1)C(=O)/N=c/1\sn(nc1C(=O)Nc1ccccc1)c1ccc(cc1)Cl</chem>	954
		-
		21,1
179	<chem>CC(NC(=O)c1cccc(c1)Cn1/c(=N/C(=O)c2ccccc2F)/sc(c1C)C)C</chem>	394
		-
		23,0
180	<chem>Cc1cccc(c1)C(=O)NC(=S)Nc1sc2c(c1C(=O)N)CCCCC2</chem>	945
		-
		22,9
181	<chem>COc1ccc(cc1)c1[nH]n(/c(=N/C(=O)C(C)C)/c1)[C@H]1CCS(=O)(=O)C1</chem>	94

		-
182	<chem>C[C@@H]1CCN(CC1)c1ccc(o1)/C=N/c1sc2c(c1C(=O)Nc1cccc1)CCCC2</chem>	21,9 718
		-
183	<chem>CCn1/c(=N/C(=O)c2ccc(cc2)S(=O)(=O)N2C[C@H](C)O[C@H](C2)C)/sc2c1ccc(c2)Cl</chem>	21,7 689
		-
184	<chem>O=C(c1ccoc1C)/N=c/1\[nH]c2c(n1CC[NH+])1CCCCC1)cccc2</chem>	24,1 678
		-
185	<chem>C1CC[C@H](C[NH2+])1Cc1cnc/c(=N\c2ncc(s2)c2cccc2)/[nH]1</chem>	21,8 052
		-
186	<chem>Clc1cccc(c1)c1n[nH]/c(=N/C(=O)NCc2n(C)ncc2C2CC2)/s1</chem>	21,2 698
		-
187	<chem>COC(=O)Nc1c(nc(n1c1sc(nc1c1cccc1)c1cccc1)c1cccc1)c1cccc1</chem>	20,6 814
		-
188	<chem>Fc1ccc(cc1)c1c[nH]/c(=N\C(=O)c2[nH]nc3c2C[C@H](CC3)C(C)(C)C)/n1C</chem>	20,6 693
		-
189	<chem>Cc1nn(c(c1)C)c1cccc(c1)C(=O)/N=c/1\cc([nH]n1[C@@H]1CCS(=O)(=O)C1)C</chem>	22,0 209
		-
190	<chem>CC[N@@H+]1CCc2c(C1)sc(c2C(=O)N)NC(=S)NC(=O)c1c(Br)cnn1C</chem>	20,5 992
		-
191	<chem>COc1ccc2c(c1)sc(n2)N1C(=O)c2c([C@H]1c1ccc(c(c1)OC)O)c(=O)c1c(o2)cccc1</chem>	23,2 908
		-
192	<chem>Clc1ccc(cc1)COC1cccc(c1)/C=N/c1sc2c(c1C(=O)Nc1ccc(cc1)Cl)CC[C@@H](C2)C(C)(C)C</chem>	21,1 232
		-
193	<chem>COC1cc(OC)ccc1C(=O)/N=c/1\sc2c(n1C)ccc(c2)S(=O)(=O)C</chem>	20,8 435
		-
194	<chem>O=C(/N=c/1\cc([nH]n1[C@H]1CCS(=O)(=O)C1)C)CCc1c(C)n[nH]c1C</chem>	20,5 321
		-
195	<chem>O1CCN(CC1)c1nnc(n1Cc1cccc1)SCc1onc(n1)c1cscc1</chem>	20,6 485
		-
196	<chem>Clc1ccc2c(c1)s/c(=N\C(=O)c1ccc(cc1)S(=O)(=O)N1CCCc3c1cccc3)/n2C</chem>	23,7 146
		-
197	<chem>CC(=O)/N=C\1/CC(=CC(=C1)[NH3+])CCc1cccc1</chem>	20,6 249
		-
198	<chem>O=C(Cc1c(C)noc1C)N=c1[nH]c2c([nH]1)cccc2</chem>	21,8 791
		-
199	<chem>[NH3+]C[C@H]1O[C@H](O[C@@H]2[C@@H]([NH3+])C[C@H]([C@H]([C@@H]2O)O[C@@H]2O[C@@H](CO)[C@H]([C@@H]([C@@H]2O)[NH3+])O)[NH3+])[C@H](C[C@@H]1O)[NH3+]</chem>	22,2 492
		-
200	<chem>[NH3+]C[C@@H]1O[C@H](O[C@@H]2[C@@H]([NH3+])C[C@H]([C@H]([C@@H]2O)O[C@H]2O[C@@H](CO)[C@H]([C@@H]([C@@H]2O)[NH3+])O)[NH3+])[C@H]([C@@H]([C@@H]([C@@H]1O)O)O</chem>	24,5 993

#### 4. DL Classification

Chompounds from ChEMBL with 3CL<sup>Pro</sup> activity data in csv format (SMILES and ChEMBL Standard value (IC<sub>50</sub> in nM))

Cc1nc(S(=O)(=O)c2ccccc2)c(C#N)c(C)c1[N+](=O)[O-],13000  
 CCOC(=O)/C=C/[C@H](C[C@@H]1CCNC1=O)NC(=O)[C@H](CC=C(C)C)CC(=O)[C@@H](NC(=O)c1cc(C)on1)  
 C(C)C,70000  
 CSc1[nH]nc(NC(=O)c2cccs2)c1S(=O)(=O)c1ccccc1,15000  
 O=S(=O)(Cc1[nH]c(-c2ccc(Cl)s2)c[s+](1)c1cccs1.[Br-],18000  
 Cc1oc(C(C)(C)C)cc1-c1cc(NS(=O)(=O)c2cccs2)[nH]n1,10000  
 O=C(Oc1cncc(Cl)c1)c1ccco1,60  
 CCOC(=O)/C(C#N)=C/Nc1ccc(S(=O)(=O)c2ccc(/N=C/C(C#N)=C(\O)OCC)cc2)cc1,16000  
 CSc1nn(-c2c([N+](=O)[O-])c(C)nn2C)c(-c2cccs2)c1C#N,18000  
 O=[N+](/[O-])c1ccc(S(=O)(=O)c2ccc(Cl)cc2)[n+](/[O-])c1,15000  
 O=C(O)c1ccc(S(=O)(=O)c2cc(Br)c(O)c(Br)c2)cc1,16000  
 CCOC(=O)C(=CNc1ccc(S(=O)(=O)c2ccc(/N=C/C(C(=O)OCC)cc2)cc1)C(=O)OCC,32000  
 Cc1noc(NC(=O)c2ccc(-c3cc(C(F)(F)F)nn3C)s2)c1[N+](=O)[O-],5000  
 CC(Sc1nc(O)c(C#N)c(-c2ccccc2)n1)C(=O)Nc1ccc(Cl)cc1,300000  
 Cc1cc(O)nc(SCC(=O)Nc2ccc(Oc3ccc(Cl)cc3)cc2)n1,200000  
 CCCc1cc(O)nc(SCC(=O)Nc2ccc(Cl)cc2)n1,30000  
 O=C1C(=O)N(CC2COc3ccccc3O2)c2ccc(I)cc21,13500  
 O=C1C(=O)N(Cc2cc3ccccc3s2)c2c1cccc2[N+](=O)[O-],2000  
 Cc1nn(C)c(NCc2ccc(-c3cccs3)s2)c1[N+](=O)[O-],20000  
 CC(C)c1ccc(NC(=O)CSc2nccc(-c3csc(COc4ccccc4Cl)n3)n2)cc1,200000  
 Cc1nc(-c2nc(-c3ccnc(SCC(=O)Nc4ccc(Cl)cc4)n3)cs2)cs1,70794578438413.7  
 CSc1sc(-c2nc(C)cs2)c(C)c1-c1ccnc(SCC(=O)Nc2ccc(Cl)cc2)n1,11000  
 Cn1nc(-c2ccc(-c3ccnc(SCC(=O)Nc4ccc(Cl)cc4)n3)s2)cc1C(F)(F)F,1000000000000000  
 O=C(CSc1nccc(-c2csc(-c3ccccc3)n2)n1)Nc1cc(Cl)cc(Cl)c1,3000  
 COC(OC)c1cc(O)nc(SCC(=O)Nc2ccc(C(F)(F)F)cc2)n1,4897788193684.46  
 COC(OC)c1cc(O)nc(SCC(=O)Nc2ccc(C(F)(F)F)cc2)n1,200000  
 Cc1cc(O)nc(SCC(=O)Nc2cc(Cl)ccc2Oc2ccccc2)n1,1000000000000000  
 CSc1sc(-c2nc(C)cs2)c(C)c1-c1ccnc(SCC(=O)Nc2ccccc2Cl)n1,12000  
 Cc1cc(O)nc(SCC(=O)Nc2ccc(Oc3ccc(Cl)cc3)cc2)n1,4897788193684.46  
 Cc1noc(C)c1CN1C(=O)C(=O)c2cc(C#N)ccc21,7200  
 CC(C)(C)c1ccc(NC(=O)CSc2nc(O)c(C#N)c(C3CCCC3)n2)cc1,2818382931264.45  
 O=C(CSc1nccc(-c2cc(-c3ccccc3)no2)n1)Nc1ccc(Cl)cc1,66069344800759.7  
 COc1ccc(-c2nc(SCC(=O)Nc3ccc([N+](=O)[O-])cc3)nc(O)c2C#N)c1,300000  
 O=C(CSc1nccc(-c2cc(-c3ccccc3Cl)no2)n1)Nc1ccc(Cl)cc1,66069344800759.7  
 CCCc1cc(O)nc(SCC(=O)Nc2ccc(Cl)cc2)n1,33113112148259.1  
 O=C(Oc1cncc(Cl)c1)c1cscn1,270  
 O=C(Oc1cncc(Cl)c1)c1ccc(-c2ccc(Cl)cc2)o1,63  
 O=C(Nc1ccc(Cl)cc1)c1ccc(CN2C(=O)C(=O)c3cc(I)ccc32)s1,12570  
 CCCCN1C(=O)C(=O)c2cc(C(N)=O)ccc21,19000  
 O=C1C(=O)N(Cc2cc3ccccc3s2)c2ccccc21,13110  
 O=C1C(=O)N(Cc2cc3ccccc3s2)c2ccccc(Cl)c21,11200  
 CCCCN1C(=O)C(=O)c2cc(I)ccc21,66000  
 O=C(Oc1cncc(Br)c1)c1ccco1,50  
 O=C(Oc1cncc(Cl)c1)c1cc2ccccc2o1,170  
 C[C@H]1COC2=C1C(=O)C(=O)c1c2ccc2c1CCCC2(C)C,10100  
 Cc1coc2c1C(=O)C(=O)c1c-2ccc2c1CCCC2(C)CO,52000  
 O=C(Oc1cncc(Cl)c1)c1cc2ccccc2s1,95  
 C[C@H]1COC2=C1C(=O)C(=O)c1c2ccc2c1CCCC2(C)C,226700  
 COc1cccc(C(=O)Oc2cncc(Cl)c2)c1,340  
 COc1cccc(-c2nc(SCC(=O)Nc3ccc([N+](=O)[O-])cc3)nc(O)c2C#N)c1,3311311214825.91  
 COc1ccc(NC(=O)CSc2nc(O)cc(-c3ccccc3)n2)cc1OC,21877616239495.5  
 O=C(CSc1nccc(-c2cc(-c3ccccc3Cl)no2)n1)Nc1ccc(Cl)cc1,15000  
 Cc1nc(SCC(=O)Nc2ccc([N+](=O)[O-])cc2)nc(O)c1C,10000000000000  
 COc1cccc(-c2nc(SCC(=O)Nc3ccc(C(C)=O)cc3)nc(O)c2C#N)c1,16595869074375.6  
 O=C1C(=O)N(Cc2cc3ccccc3s2)c2c(Br)cccc21,980  
 O=[N+](/[O-])c1ccc(S(=O)(=O)c2ccc([N+](=O)[O-])cc2)cc1,25000  
 Cc1cccc2c3c(ccc12)C1=C(C(=O)C3=O)[C@@H](C)CO1,14400  
 Cc1coc2c1C(=O)C(=O)c1c-2ccc2c1CCCC2(C)C,1600  
 CCC(Sc1nc(O)c(C#N)c(-c2ccc(OC)c2)n1)C(=O)Nc1ccc(C(C)=O)cc1,16595869074375.6  
 CSc1sc(-c2nc(C)cs2)c(C)c1-c1ccnc(SCC(=O)Nc2ccccc2Cl)n1,83176377110267.1  
 COc1ccc(-c2ccnc(SCC(=O)Nc3ccc(C(C)C)cc3)n2)cc1,2454708915685.03  
 CCOC1ccc(N2C(=O)CC(Sc3nc(C)cc(C)n3)C2=O)cc1,300000  
 O=C(CSc1nccc(-c2cc(-c3ccccc3)no2)n1)Nc1ccc(Cl)cc1,15000  
 Cc1nc(-c2nc(-c3ccnc(SCC(=O)Nc4ccc(Cl)cc4)n3)cs2)cs1,14000  
 CSc1sc(-c2nc(C)cs2)c(C)c1-c1ccnc(SCC(=O)Nc2ccc(Cl)cc2)n1,89125093813374.4  
 O=C1C(=O)N(Cc2ccc(C(=O)N3CCCC3)s2)c2ccc(I)cc21,17500  
 O=C1C(=O)N(Cc2cc3ccccc3s2)c2ccc(I)cc21,950

O=C(Sc1nnc(C(F)(F)F)[nH]1)c1ccc(C#Cc2ccccc2)o1,3000  
Cc1cc(S(=O)(=O)c2c([N+](=O)[O-])cc(C(F)(F)F)cc2[N+](=O)[O-])c(Cl)cc1Cl,300  
CN1C(=O)C(=O)c2cc(C(N)=O)ccc21,71000  
Cc1coc2c1C(=O)C(=O)c1c-2ccc2c1CCCC2(C)C,89100  
Cc1cccc2c3c(ccc12)C1=C(C(=O)C3=O)[C@@H](C)CO1,1200  
Nc1ncc(S(=O)(=O)c2ccc(Cl)cc2)c(N)n1,6000  
Cc1cc(C(F)(F)F)nc(SCC(=O)Nc2ccc(Cl)cc2F)n1,500000  
CCCCc1ccc(NC(=O)CSc2nc(O)c(C#N)c(C3CCCC3)n2)c(C)c1,4897788193684.46  
O=C(Cc1nccs1)c1nccs1,40000  
Cc1coc2c1C(=O)C(=O)c1c-2ccc2c(C)cccc12,38700  
C[C@H]1COC2=C1C(=O)C(=O)c1c2ccc2c1CCCC2(C)C,800  
O=C(Oc1cncc(Cl)c1)c1cc2ccccc2[nH]1,65  
C[C@H]1COC2=C1C(=O)C(=O)c1c2ccc2c1CCCC2(C)C,87600  
CCOC(=O)/C=C/[C@H](C[C@@H]1CCNC1=O)NC(=O)[C@@H](CC(=O)[C@@H](NC(=O)c1cc(C)on1)C(C)C)Cc  
1cccc1,45000  
CC(C)C1=Cc2ccc3c(c2C(=O)C1=O)CCCC3(C)C,30000  
O=C1C(=O)N(Cc2ccc(F)cc2Cl)c2ccc(I)cc21,9400  
O=C1C(=O)N(Cc2cc3ccccc3s2)c2ccc(F)cc21,4820  
O=[N+][O-]c1cc(C(F)(F)F)ccc1S(=O)(=O)c1ccc(Cl)cc1,12000  
Cc1ccc(S(=O)(=O)c2nc(C)c([N+](=O)[O-])c(C)c2C#N)cc1,13000  
CCOC(=O)c1cncc(SCC(=O)Nc2ccc([N+](=O)[O-])cc2)nc1N,3981071705534.97  
N#Cc1c(O)nc(SCC(=O)Nc2cccc(C(F)(F)F)c2)nc1-c1cccc1,4897788193684.46  
Cn1nc(-c2ccc(-c3ccnc(SCC(=O)Nc4ccc(Cl)cc4)n3)s2)cc1C(F)(F)F,10000  
COc1cccc(-c2nc(SCC(=O)Nc3ccc(S(N)(=O)=O)cc3)nc(O)c2C#N)c1,24547089156850.3  
COc1cccc(-c2nc(SCC(=O)Nc3ccc(C(C)=O)cc3)nc(O)c2C#N)c1,60000  
CCOC(=O)c1cncc(SCC(=O)Nc2ccc([N+](=O)[O-])cc2)nc1N,250000  
CCC(Sc1nc(O)c(C#N)c(-c2cccc(OC)c2)n1)C(=O)Nc1ccc(C(C)=O)cc1,60000  
CC(C)c1ccc(NC(=O)CSc2nccc(-c3cccs3)n2)cc1,24547089156850.3  
O=C(CSc1nccc(-c2cc(-c3ccc(Cl)cc3Cl)no2)n1)Nc1ccc(C(F)(F)F)cc1,66069344800759.7  
O=C(CSc1nccc(-c2cc(-c3ccc(Cl)cc3Cl)no2)n1)Nc1ccc(C(F)(F)F)cc1,15000  
Cc1nc(SCC(=O)Nc2ccc([N+](=O)[O-])cc2)nc(O)c1C,1000000  
CC(C)(C)c1ccc(NC(=O)CSc2nc(O)c(C#N)c(C3CCCC3)n2)cc1,350000  
NC(=O)c1ccc2c(c1)C(=O)C(=O)N2Cc1cccc1,12500  
O=C1C(=O)N(Cc2ccc3ccccc3c2)c2ccc(I)cc21,1100  
COc1ccc(NC(=O)CSc2nc(O)cc(-c3ccccc3)n2)cc1OC,45000  
CCOc1ccc(N2C(=O)CC(Sc3nc(C)cc(C)n3)C2=O)cc1,3311311214825.91  
CC(=O)c1cccc1S(=O)(=O)c1cccc1C(=O)O,16000  
CC1(C)CC(=O)c2c(NCc3ccc3)sc(C#N)c2C1,16000  
Cc1cc(O)nc(SCC(=O)Nc2cc(Cl)ccc2Oc2ccccc2)n1,100000  
CCCCc1ccc(NC(=O)CSc2nc(O)c(C#N)c(C3CCCC3)n2)c(C)c1,200000  
CC(C)C1=Cc2ccc3c(c2C(=O)C1=O)CCCC3(C)C,78900  
O=C(Oc1ccc(S(=O)(=O)c2ccc(OC(=O)C(Cl)=C(Cl)Cl)cc2)cc1)C(Cl)=C(Cl)Cl,900  
NC(=O)c1ccc2c(c1)C(=O)C(=O)N2Cc1ccc2ccccc2c1,370  
CCCN1C(=O)C(=O)c2cc(C(N)=O)ccc21,25000  
O=C1C(=O)N(Cc2ccccc2)c2ccc(I)cc21,50000  
Cc1cccc2c3c(ccc12)C1=C(C(=O)C3=O)[C@@H](C)CO1,4900  
COC(=O)[C@@]1(C)CCc2c1ccc1c2C(=O)C(=O)c2c(C)coc2-1,21100  
O=C1C(=O)N(C/C=C/c2cc3ccccc3s2)c2ccc(I)cc21,23500  
O=C(CSc1nccc(-c2csc(-c3ccccc3)n2)n1)Nc1cc(Cl)cc(Cl)c1,331131121482591  
CC(C)c1ccc(NC(=O)CSc2nccc(-c3cccs3)n2)cc1,40000  
Cc1cc(C(F)(F)F)nc(SCC(=O)Nc2ccc(Cl)cc2F)n1,1995262314968.88  
COc1ccc(-c2ccnc(SCC(=O)Nc3ccc(C(C)C)cc3)n2)cc1,400000  
CC(Sc1nc(O)c(C#N)c(-c2ccccc2)n1)C(=O)Nc1ccc(Cl)cc1,3311311214825.91  
N#Cc1c(O)nc(SCC(=O)Nc2cccc(C(F)(F)F)c2)nc1-c1cccc1,200000  
CC(C)c1ccc(NC(=O)CSc2nccc(-c3csc(COc4ccccc4Cl)n3)n2)cc1,4897788193684.46  
COc1cccc(-c2nc(SCC(=O)Nc3ccc(S(N)(=O)=O)cc3)nc(O)c2C#N)c1,40000  
Cc1coc2c1C(=O)C(=O)c1c-2ccc2c1CCCC2(C)CO,10700  
CC(C)C1=Cc2ccc3c(c2C(=O)C1=O)CCCC3(C)C,21100  
Cc1coc2c1C(=O)C(=O)c1c-2ccc2c(C)cccc12,700  
Cc1coc2c1C(=O)C(=O)c1c-2ccc2c(C)cccc12,11500  
CC(C)C1=Cc2ccc3c(c2C(=O)C1=O)CCCC3(C)C,60200  
Cc1coc2c1C(=O)C(=O)c1c-2ccc2c1CCCC2(C)C,17100  
Cc1cccc2c3c(ccc12)C1=C(C(=O)C3=O)[C@@H](C)CO1,10700  
Cc1coc2c1C(=O)C(=O)c1c-2ccc2c1CCCC2(C)CO,18000  
COC(=O)[C@@]1(C)CCc2c1ccc1c2C(=O)C(=O)c2c(C)coc2-1,10600  
Cc1coc2c1C(=O)C(=O)c1c-2ccc2c(C)cccc12,8800  
Cc1coc2c1C(=O)C(=O)c1c-2ccc2c1CCCC2(C)CO,24800



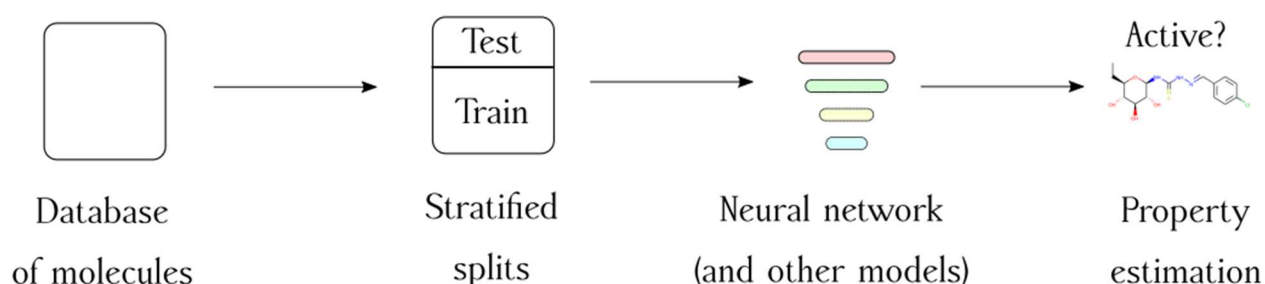
COC(=O)[C@@]1(C)CCCc2c1ccc1c2C(=O)C(=O)c2c(C)coc2-1,9200

Compounds with any activity data on viral cysteine proteases were extracted from ChEMBL (4.11.2020) and the complete file in csv format is supplied as part of supporting information. The decoyes employed were obtained as a standard 1k decoy set from Schrödinger LLC.

chembl\_id,canonical\_smiles,organism,full\_mwt,l1,l2,l3

### 3.1 Machine Learning

Herein, the following steps are of relevance. First, the database of relevant target and decoy molecules is constructed. Next, stratified splits of the data are obtained – these types of data splits ensure approximately equal distributions of different classes (relevant for classification). Multiple splits are obtained, and for each one, a model is trained and evaluated. As the best performing model was a neural network, we emphasize this model next. The overall architecture can be summarized as follows:



**Figure S3: The overview of the learning process.**

The architecture can be understood as a feedforward neural network with additional dropout layers attached. The final layer is either a ReLU – in this case the network performs regression for one target, or sigmoid, in which case it performs binary classification. Feature space was constructed as follows. We used the DeepChem library to produce RDKit descriptors via the “`RDKitDescriptors()`” call on the input SMILES-formatted molecules. The considered descriptors were the following ones and referenced in the section RDKit (<https://www.rdkit.org/docs/GettingStartedInPython.html#list-of-available-descriptors>):

['MaxEStateIndex', 'MinEStateIndex', 'MaxAbsEStateIndex', 'MinAbsEStateIndex', 'qed', 'MolWt', 'HeavyAtomMolWt', 'ExactMolWt', 'NumValenceElectrons', 'NumRadicalElectrons',

'MaxPartialCharge', 'MinPartialCharge', 'MaxAbsPartialCharge', 'MinAbsPartialCharge', 'FpDensityMorgan1', 'FpDensityMorgan2', 'FpDensityMorgan3', 'BCUT2D\_MWHI', 'BCUT2D\_MWLOW', 'BCUT2D\_CHGHI', 'BCUT2D\_CHGLO', 'BCUT2D\_LOGPHI', 'BCUT2D\_LOGPLOW', 'BCUT2D\_MRHI', 'BCUT2D\_MRLOW', 'BalabanJ', 'BertzCT', 'Chi0', 'Chi0n', 'Chi0v', 'Chi1', 'Chi1n', 'Chi1v', 'Chi2n', 'Chi2v', 'Chi3n', 'Chi3v', 'Chi4n', 'Chi4v', 'HallKierAlpha', 'Ipc', 'Kappa1', 'Kappa2', 'Kappa3', 'LabuteASA', 'PEOE\_VSA1', 'PEOE\_VSA10', 'PEOE\_VSA11', 'PEOE\_VSA12', 'PEOE\_VSA13', 'PEOE\_VSA14', 'PEOE\_VSA2', 'PEOE\_VSA3', 'PEOE\_VSA4', 'PEOE\_VSA5', 'PEOE\_VSA6', 'PEOE\_VSA7', 'PEOE\_VSA8', 'PEOE\_VSA9', 'SMR\_VSA1', 'SMR\_VSA10', 'SMR\_VSA2', 'SMR\_VSA3', 'SMR\_VSA4', 'SMR\_VSA5', 'SMR\_VSA6', 'SMR\_VSA7', 'SMR\_VSA8', 'SMR\_VSA9', 'SlogP\_VSA1', 'SlogP\_VSA10', 'SlogP\_VSA11', 'SlogP\_VSA12', 'SlogP\_VSA2', 'SlogP\_VSA3', 'SlogP\_VSA4', 'SlogP\_VSA5', 'SlogP\_VSA6', 'SlogP\_VSA7', 'SlogP\_VSA8', 'SlogP\_VSA9', 'TPSA', 'EState\_VSA1', 'EState\_VSA10', 'EState\_VSA11', 'EState\_VSA2', 'EState\_VSA3', 'EState\_VSA4', 'EState\_VSA5', 'EState\_VSA6', 'EState\_VSA7', 'EState\_VSA8', 'EState\_VSA9', 'VSA\_EState1', 'VSA\_EState10', 'VSA\_EState2', 'VSA\_EState3', 'VSA\_EState4', 'VSA\_EState5', 'VSA\_EState6', 'VSA\_EState7', 'VSA\_EState8', 'VSA\_EState9', 'FractionCSP3', 'HeavyAtomCount', 'NHOHCount', 'NOCount', 'NumAliphaticCarbocycles', 'NumAliphaticHeterocycles', 'NumAliphaticRings', 'NumAromaticCarbocycles', 'NumAromaticHeterocycles', 'NumAromaticRings', 'NumHAcceptors', 'NumHDonors', 'NumHeteroatoms', 'NumRotatableBonds', 'NumSaturatedCarbocycles', 'NumSaturatedHeterocycles', 'NumSaturatedRings', 'RingCount', 'MolLogP', 'MolMR', 'fr\_Al\_COO', 'fr\_Al\_OH', 'fr\_Al\_OH\_noTert', 'fr\_ArN', 'fr\_Ar\_COO', 'fr\_Ar\_N', 'fr\_Ar\_NH', 'fr\_Ar\_OH', 'fr\_COO', 'fr\_COO2', 'fr\_C\_O', 'fr\_C\_O\_noCOO', 'fr\_C\_S', 'fr\_HOCCN', 'fr\_Imine', 'fr\_NH0', 'fr\_NH1', 'fr\_NH2', 'fr\_N\_O', 'fr\_Ndealkylation1', 'fr\_Ndealkylation2', 'fr\_Nhpyrrole', 'fr\_SH', 'fr\_aldehyde', 'fr\_alkyl\_carbamate', 'fr\_alkyl\_halide', 'fr\_allylic\_oxid', 'fr\_amide', 'fr\_amidine', 'fr\_aniline', 'fr\_aryl\_methyl', 'fr\_azide', 'fr\_azo', 'fr\_barbitur', 'fr\_benzene', 'fr\_benzodiazepine', 'fr\_bicyclic', 'fr\_diazo', 'fr\_dihydropyridine', 'fr\_epoxide', 'fr\_ester', 'fr\_ether', 'fr\_furan', 'fr\_guanido', 'fr\_halogen', 'fr\_hdrzine', 'fr\_hdrzone', 'fr\_imidazole', 'fr\_imide', 'fr\_isocyan', 'fr\_isothiocyan', 'fr\_ketone', 'fr\_ketone\_Topliss', 'fr\_lactam', 'fr\_lactone', 'fr\_methoxy', 'fr\_morpholine', 'fr\_nitrile', 'fr\_nitro', 'fr\_nitro\_ arom', 'fr\_nitro\_ arom\_nonortho', 'fr\_nitroso', 'fr\_oxazole', 'fr\_oxime', 'fr\_para\_hydroxylation', 'fr\_phenol', 'fr\_phenol\_noOrthoHbond', 'fr\_phos\_acid', 'fr\_phos\_ester', 'fr\_piperdine', 'fr\_piperzine', 'fr\_priamide', 'fr\_prisulfonamd', 'fr\_pyridine', 'fr\_quatN', 'fr\_sulfide', 'fr\_sulfonamd', 'fr\_sulfone', 'fr\_term\_acetylene', 'fr\_tetrazole', 'fr\_thiazole', 'fr\_thiocyan', 'fr\_thiophene', 'fr\_unbrch\_alkane', 'fr\_urea']

The descriptors are also referenced and described further in the Supporting Information under the section RDKit.

## 5. Physicochemical properties

Compound descriptors were calculated using QikProp by Schrödinger and RDKit ([rdkit.org](https://www.rdkit.org/docs/index.html); <https://www.rdkit.org/docs/index.html>; accessed on 10.1.2021). [63]

### QikProp:

- molecule name

Molecule name taken from the title line in the input structure file. If the title line is blank, the input file name is used.

- #stars

Number of property or descriptor values that fall outside the 95% range of similar values for known drugs. Outlying descriptors and predicted properties are denoted with asterisks (\*) in the .out file. A large number of stars suggests that a molecule is less drug-like than molecules with few stars. The following properties and descriptors are included in the determination of #stars: MW, dipole, IP, EA, SASA, FOSA, FISA, PISA, WPSA, PSA, volume, #rotor, donorHB, accptHB, glob, QPpolrz, QPlogPC16, QPlogPoct, QPlogPw, QPlogPo/w, logS, QPLogKhsa, QPlogBB, #metabol  
0 – 5

- #amine

Number of non-conjugated amine groups.  
0 – 1

- #amidine

Number of amidine and guanidine groups.  
0

- #acid

Number of carboxylic acid groups.  
0 – 1

- #amide

Number of non-conjugated amide groups.  
0 – 1

- #rotor

Number of non-trivial (not CX3), non-hindered (not alkene, amide, small ring) rotatable bonds.  
0 – 15

- #rtvFG

Number of reactive functional groups; the specific groups are listed in the jobname .out file. The presence of these groups can lead to false positives in HTS assays and to decomposition, reactivity, or toxicity problems in vivo. See Appendix A of the QikProp User Manual for a complete list.  
0 – 2

- CNS

Predicted central nervous system activity on a –2 (inactive) to +2 (active) scale.  
–2 to +2

- mol\_MW

Molecular weight of the molecule.  
130.0 – 725.0

- dipole

Computed dipole moment of the molecule.

1.0 – 12.5

- SASA

Total solvent accessible surface area (SASA) in square angstroms using a probe with a 1.4 Å radius.

300.0 – 1000.0

- FOSA

Hydrophobic component of the SASA (saturated carbon and attached hydrogen).

0.0 – 750.0

- FISA

Hydrophilic component of the SASA (SASA on N, O, H on heteroatoms, and carbonyl C).

7.0 – 330.0

- PISA

$\pi$  (carbon and attached hydrogen) component of the SASA.

0.0 – 450.0

- WPSA

Weakly polar component of the SASA (halogens, P, and S).

0.0 – 175.0

- volume

Total solvent-accessible volume in cubic angstroms using a probe with a 1.4 Å radius.

500.0 – 2000.0

- donorHB

Estimated number of hydrogen bonds that would be donated by the solute to water molecules in an aqueous solution. Values are averages taken over a number of configurations, so they can be non-integer.

0.0 – 6.0

- acceptHB

Estimated number of hydrogen bonds that would be accepted by the solute from water molecules in an aqueous solution. Values are averages taken over a number of configurations, so they can be non-integer.

2.0 – 20.0

- $\text{dip}^2/V$

Square of the dipole moment divided by the molecular volume. This is the key term in the Kirkwood-Onsager equation for the free energy of solvation of a dipole with volume  $V$ .

0.0 – 0.13

- $\text{ACxDN}^{.5}/\text{SA}$

Index of cohesive interaction in solids. This term represents the relationship  
; see Bioorg. Med. Chem. Lett. 2000, 10, 1155.

0.0 – 0.05

- glob

Globularity descriptor,

, where  $r$  is the radius of a sphere with a volume equal to the molecular volume. Globularity is 1.0 for a spherical molecule.

0.75 – 0.95

- QPpolrz

Predicted polarizability in cubic angstroms.

13.0 – 70.0

- QPlogPC16

Predicted hexadecane/gas partition coefficient.

4.0 – 18.0

- QPlogPoct

Predicted octanol/gas partition coefficient.

8.0 – 35.0

- QPlogPw

Predicted water/gas partition coefficient.

4.0 – 45.0

- QPlogPo/w

Predicted octanol/water partition coefficient.

–2.0 – 6.5

- QPlogS

Predicted aqueous solubility,  $\log S$ .  $S$  in  $\text{mol dm}^{-3}$  is the concentration of the solute in a saturated solution that is in equilibrium with the crystalline solid.

–6.5 – 0.5

- CIQPlogS

Conformation-independent predicted aqueous solubility,  $\log S$ .  $S$  in  $\text{mol dm}^{-3}$  is the concentration of the solute in a saturated solution that is in equilibrium with the crystalline solid.

–6.5 – 0.5

- QPlogHERG

Predicted IC50 value for blockage of HERG K<sup>+</sup> channels.

concern below –5

- QPPCaco

Predicted apparent Caco-2 cell permeability in nm/sec. Caco-2 cells are a model for the gut-blood barrier. QikProp predictions are for non-active transport.

<25 poor,

>500 great

- QPlogBB

Predicted brain/blood partition coefficient. Note: QikProp predictions are for orally delivered drugs so, for example, dopamine and serotonin are CNS negative because they are too polar to cross the blood-brain barrier

–3.0 – 1.2

- QPPMDCK

Predicted apparent MDCK cell permeability in nm/sec. MDCK cells are considered to be a good mimic for the blood-brain barrier. QikProp predictions are for non-active transport.

<25 poor,

>500 great

- QPlogKp

Predicted skin permeability, log Kp.

–8.0 – –1.0

- IP(ev)

PM3 calculated ionization potential (negative of HOMO energy).

7.9 – 10.5

- EA(eV)

PM3 calculated electron affinity (negative of LUMO energy).

–0.9 – 1.7

- #metab

Number of likely metabolic reactions. See Appendix A of the QikProp User Manual for a complete list of reactions.

1 – 8

- QPlogKhsa

Prediction of binding to human serum albumin.

–1.5 – 1.5

- HumanOralAbsorption

Predicted qualitative human oral absorption: 1, 2, or 3 for low, medium, or high. The text version is reported in the output. The assessment uses a knowledge-based set of rules, including checking for suitable values of PercentHumanOralAbsorption, number of metabolites, number of rotatable bonds, logP, solubility and cell permeability.

- PercentHumanOralAbsorption

Predicted human oral absorption on 0 to 100% scale. The prediction is based on a quantitative multiple linear regression model. This property usually correlates well with HumanOralAbsorption, as both measure the same property.

>80% is high

<25% is poor

- SAFluorine

Solvent-accessible surface area of fluorine atoms.  
0.0 – 100.0

- SAamideO

Solvent-accessible surface area of amide oxygen atoms.  
0.0 – 35.0

- PSA

Van der Waals surface area of polar nitrogen and oxygen atoms and carbonyl carbon atoms.  
7.0 – 200.0

- #NandO

Number of nitrogen and oxygen atoms.  
2 – 15

- RuleOfFive

Number of violations of Lipinski's rule of five. The rules are:  $\text{mol\_MW} < 500$ ,  $\text{QPlogPo/w} < 5$ ,  $\text{donorHB} \leq 5$ ,  $\text{acceptHB} \leq 10$ . Compounds that satisfy these rules are considered drug-like. (The "five" refers to the limits, which are multiples of 5.)  
maximum is 4

- RuleOfThree

Number of violations of Jorgensen's rule of three. The three rules are:  $\text{QPlogS} > -5.7$ ,  $\text{QPCaco} > 22 \text{ nm/s}$ ,  $\text{\# Primary Metabolites} < 7$ . Compounds with fewer (and preferably no) violations of these rules are more likely to be orally available.  
maximum is 3

- #ringatoms

Number of atoms in rings.

- #in34

Number of atoms in 3- or 4-membered rings.

- #in56

Number of atoms in 5- or 6-membered rings.

- #noncon

number of ring atoms not able to form conjugated aromatic systems (e.g.  $\text{sp}^3 \text{ C}$ ).

- #nonHatm

Number of heavy atoms (nonhydrogen atoms).

- Jm

Predicted maximum transdermal transport rate,  $\text{Kp} \times \text{MW} \times \text{S}$  ( $\mu\text{g cm}^{-2} \text{ hr}^{-1}$ ). Kp and S are obtained from the aqueous solubility and skin permeability, QPlogKp and QPlogS. This property is only written to the output file: it is not used in any other calculations.

**RDKit:**

List of Available (2D) Descriptors as in v. 2020.09.1:

Descriptor/Descriptor Family	Notes
Gasteiger/Marsili Partial Charges	<i>Tetrahedron</i> <b>36</b> :3219-28 (1980)
BalabanJ	<i>Chem. Phys. Lett.</i> <b>89</b> :399-404 (1982)
BertzCT	<i>J. Am. Chem. Soc.</i> <b>103</b> :3599-601 (1981)
Ipc	<i>J. Chem. Phys.</i> <b>67</b> :4517-33 (1977)
HallKierAlpha	<i>Rev. Comput. Chem.</i> <b>2</b> :367-422 (1991)
Kappa1 - Kappa3	<i>Rev. Comput. Chem.</i> <b>2</b> :367-422 (1991)
Chi0, Chi1	<i>Rev. Comput. Chem.</i> <b>2</b> :367-422 (1991)
Chi0n - Chi4n	<i>Rev. Comput. Chem.</i> <b>2</b> :367-422 (1991)
Chi0v - Chi4v	<i>Rev. Comput. Chem.</i> <b>2</b> :367-422 (1991)
MolLogP	Wildman and Crippen <i>JCICS</i> <b>39</b> :868-73 (1999)
MolMR	Wildman and Crippen <i>JCICS</i> <b>39</b> :868-73 (1999)
MolWt	
ExactMolWt	
HeavyAtomCount	
HeavyAtomMolWt	
NHOHCount	
NOCCount	
NumHAcceptors	
NumHDonors	
NumHeteroatoms	
NumRotatableBonds	
NumValenceElectrons	
NumAmideBonds	
Num{Aromatic,Saturated,Aliphatic} Rings	
Num{Aromatic,Saturated,Aliphatic} {Hetero,Carbo} cycles	
RingCount	
FractionCSP3	
NumSpiroAtoms	Number of spiro atoms (atoms shared between rings that share exactly one atom)
NumBridgeheadAtoms	Number of bridgehead atoms (atoms shared between rings that share at least two bonds)
TPSA	<i>J. Med. Chem.</i> <b>43</b> :3714-7, (2000) See the section in the RDKit book describing differences to the original publication.
LabuteASA	<i>J. Mol. Graph. Mod.</i> <b>18</b> :464-77 (2000)
PEOE_VSA1 - PEOE_VSA14	MOE-type descriptors using partial charges and surface area



	contributions <a href="http://www.chemcomp.com/journal/vsadesc.htm">http://www.chemcomp.com/journal/vsadesc.htm</a>
SMR_VSA1 - SMR_VSA10	MOE-type descriptors using MR contributions and surface area contributions <a href="http://www.chemcomp.com/journal/vsadesc.htm">http://www.chemcomp.com/journal/vsadesc.htm</a>
SlogP_VSA1 - SlogP_VSA12	MOE-type descriptors using LogP contributions and surface area contributions <a href="http://www.chemcomp.com/journal/vsadesc.htm">http://www.chemcomp.com/journal/vsadesc.htm</a>
EState_VSA1 - EState_VSA11	MOE-type descriptors using EState indices and surface area contributions (developed at RD, not described in the CCG paper)
VSA_EState1 - VSA_EState10	MOE-type descriptors using EState indices and surface area contributions (developed at RD, not described in the CCG paper)
MQNs	Nguyen et al. <i>ChemMedChem</i> <b>4</b> :1803-5 (2009)
Topliss fragments	implemented using a set of SMARTS definitions in \$(RDBASE)/Data/FragmentDescriptors.csv
Autocorr2D	New in 2017.09 release. Todeschini and Consoni “Descriptors from Molecular Geometry” Handbook of Chemoinformatics <a href="https://doi.org/10.1002/9783527618279.ch37">https://doi.org/10.1002/9783527618279.ch37</a>
BCUT2D	New in 2020.09 release. Pearlman and Smith in “3D-QSAR and Drug design: Recent Advances” (1997)

## SUPPORTING INFORMATION REFERENCES

[62] Jukic, M. *et al.* CmDock, <https://gitlab.com/Jukic/cmdock/>; **2021**.

[63] Tosco, P., Stiefl, N., & Landrum, G. (2014). Bringing the MMFF force field to the RDKit: implementation and validation. *Journal of Cheminformatics*, 6(1), 1-4.