Supporting Information

Modeling the Structure-Activity-Relationship of Arbidol Derivatives and other SARS-CoV-2 Fusion Inhibitors Targeting the S2 Segment of the Spike Protein.

Matthew R. Freidel,[1] and Roger S. Armen[1*]

[1]Department of Pharmaceutical Sciences, College of Pharmacy, Thomas Jefferson University, 901 Walnut St. Suite 918, Philadelphia, Pennsylvania, 19170.

*Corresponding author, E-mail: roger.armen@jefferson.edu

Table S1. Virtual screening data for ARB derivatives. Compounds are ranked by predicted free energy of binding (ΔG_{bind}) in (kcal/mol). Predicted physiochemical properties of molecular weight (MW) and cLogP are also provided along with Smiles strings describing the 2D structure of each small-molecule derivative.

	ΔG	MW	cLogP	Smiles
1	-9.89	633.6	7.3	CCOC(=O)C1=C(C)N(CC2=CC=CC=C2)C2=CC(Br)=C(O)C(CN3CCC(CC3)(C(=O)OCC)C3=CC=CC=C3)=C12
2	-9.88	633.6	7.7	CCOC(=O)C1=C(C)N(C2=CC(Br)=C(O)C(CN3CCC(CC3)(C(=O)OCC)C3=CC=CC=C3)=C12)C1=CC=CC(C)=C1
3	-9.19	604.8	3.9	CCOC(=O)C1=C(CSC2=CC=CC=C2)N(C)C2=C1C(CN1CCN(CCOCCO)CC1)=C(O)C(=C2)C1=CN=CC=C1
4	-8.19	605.6	6.8	CCOC(=O)C1=C(CN2CCN(CC2)C2=CC=CC=C2C)N(C2=CC(Br)=C(O)C(CN(C)C)=C12)C1=CC=CC=C1
5	-7.87	513.6	6.3	CCOC(=O)C1=C(C)N(C2=C1C(CN1CCN(CC1)C1=CC=C(C)C=C1)=C(O)C=C2)C1=CC=C(OC)C=C1
6	-7.82	501.4	5.4	CCOC(=O)C1=C(COC2=CC=CC=C2)N(C)C2=CC(Br)=C(O)C(CN3CCCCC3)=C12
7	-7.80	573.9	7.3	CCOC(=O)C1=C(CSC2=CC=CC=C2)N(C2=CC(Br)=C(O)C(CN(C)C)=C12)C1=CC=C(Cl)C=C1
8	-7.75	641.4	7.9	CCOC(=O)C1=C(C)N(C2=CC(Br)=C(O)C(CN3CCN(CC3)C3=CC=CC=C3C)=C12)C1=CC(Br)=CC=C1
9	-7.65	553.5	6.8	CCOC(=O)C1=C(CSC2=CC=CC=C2)N(CC2=CC=CC=C2)C2=CC(Br)=C(O)C(CN(C)C)=C12
10	-7.49	582.3	5.1	CCOC(=O)C1=C(COC2=CC=C(Br)C=C2)N(C)C2=CC(Br)=C(O)C(CN3CCOCC3)=C12
11	-7.46	521.5	5.3	CCOC(=O)C1=C(CSC2=CC=C(OCC)C=C2)N(C)C2=CC(Br)=C(O)C(CN(C)C)=C12
12	-7.41	428.5	5.6	CCOC(=O)C1=C(N(CC2=CC=CC=C2)C2=C1C(CN(C)C)=C(O)C=C2)C1=CC=CC=C1
13	-7.38	424.5	3.8	CCOC(=O)C1=C(C)N(C2=C1C(CN1CCOCC1)=C(O)C=C2)C1=CC=C(OC)C=C1
14	-7.35	602.6	3.9	CCOC(=O)C1=C(CN2CCN(CC#N)CC2)N(C2CCCCC2)C2=CC(Br)=C(O)C(CN3CCOCC3)=C12
15	-7.35	606.6	4.2	CCOC(=O)C1=C(CSC2=CC=CC=C2)N(C)C2=CC(Br)=C(O)C(CN3CCN(CCOCCO)CC3)=C12
16	-7.33	509.4	3.7	CCOC(=O)C1=C(CS(=O)(=O)C2=CC=CC=C2)N(C)C2=CC(Br)=C(O)C(CN(C)C)=C12
17	-7.24	556.5	5.2	CCOC(=O)C1=C(CN2CCCCC2)N(C2=CC(Br)=C(O)C(CN3CCOCC3)=C12)C1=CC=CC=C1
18	-7.22	471.4	5.8	CCOC(=O)C1=C(C)N(C2=CC(Br)=C(O)C(CN3CCCCC3)=C12)C1=CC=CC=C1
19	-7.18	531.5	5.0	CCOC(=O)C1=C(CSC2=C(C#N)C(C)=CC(C)=N2)N(C)C2=CC(Br)=C(O)C(CN(C)C)=C12
20	-7.16	461.4	4.6	CCOC(=O)C1=C(COC2=CC=CC=C2)N(C)C2=CC(Br)=C(O)C(CN(C)C)=C12
21	-7.11	548.5	4.9	CCOC(=O)C1=C(CN2CCCC2)N(C2CCCCC2)C2=CC(Br)=C(O)C(CN3CCOCC3)=C12
22	-7.10	543.9	4.3	CCOC(=O)C1=C(CS(=O)(=O)C2=CC=C(Cl)C=C2)N(C)C2=CC(Br)=C(O)C(CN(C)C)=C12
23	-7.10	446.6	5.4	CCOC(=O)C1=C(CSC2=CC=CC=C2)N(C)C2=C1C1=C(OCN(C)C1)C1=CC=CC=C21
24	-7.06	485.4	5.5	CCOC(=O)C1=C(C)N(CC2=CC=CC=C2)C2=CC(Br)=C(O)C(CN3CCCCC3)=C12
25	-7.04	491.4	5.6	CCOC(=O)C1=C(CSC2=CC=C(C)C=C2)N(C)C2=CC(Br)=C(O)C(CN(C)C)=C12
26	-7.03	491.4	4.4	CCOC(=O)C1=C(COC2=CC=C(OC)C=C2)N(C)C2=CC(Br)=C(O)C(CN(C)C)=C12
27	-6.99	477.4	5.1	CCOC(=O)C1=C(CSC2=CC=CC=C2)N(C)C2=CC(Br)=C(O)C(CN(C)C)=C12
28	-6.97	545.5	6.8	CCOC(=O)C1=C(CSC2=CC=CC=C2)N(C2CCCCC2)C2=CC(Br)=C(O)C(CN(C)C)=C12
29	-6.95	440.6	4.1	CCOC(=O)C1=C(CSC2=CC=CC=C2)N(C)C2=C1C(CN1CCOCC1)=C(O)C=C2
30	-6.90	390.4	1.9	CCOC(=O)C1=C(C)N(CC(=O)OC)C2=C1C(CN1CCOCC1)=C(O)C=C2
31	-6.87	514.5	5.4	CCOC(=O)C1=C(CN2CCCCC2)N(C2=CC(Br)=C(O)C(CN(C)C)=C12)C1=CC=CC=C1
32	-6.83	539.5	6.7	CCOC(=O)C1=C(CSC2=CC=CC=C2)N(C2=CC(Br)=C(O)C(CN(C)C)=C12)C1=CC=CC=C1
33	-6.83	596.5	3.8	CCOC(=O)C1=C(CN2CCN(CC#N)CC2)N(C2=CC(Br)=C(O)C(CN3CCOCC3)=C12)C1=CC=CC=C1
34	-6.81	437.4	5.0	CCCCN1C(C)=C(C(=O)OCC)C2=C(CN3CCCC3)C(O)=C(Br)C=C12
35	-6.81	437.5	3.8	CCOC(=O)C1=C(C)N(C2=C1C(CN1CCN(C)CC1)=C(O)C=C2)C1=CC=C(OC)C=C1

36	-6.80	473.4	4.7	CCOC(=O)C1=C(C)N(C2=CC(Br)=C(O)C(CN3CCOCC3)=C12)C1=CC=CC=C1
37	-6.76	358.4	4.2	CCCCN1C(C)=C(C(=O)OCC)C2=C1C=CC(O)=C2CN1CCCC1
38	-6.74	487.4	5.2	CCOC(=O)C1=C(C)N(C2=CC(Br)=C(O)C(CN3CCCC3)=C12)C1=CC=C(OC)C=C1
39	-6.73	516.4	4.4	CCOC(=O)C1=C(CN2CCOCC2)N(C2=CC(Br)=C(O)C(CN(C)C)=C12)C1=CC=CC=C1
40	-6.72	374.5	3.6	CCCCN1C(C)=C(C(=O)OCC)C2=C1C=CC(O)=C2CN1CCOCC1
41	-6.72	474.6	6.0	CCOC(=O)C1=C(CSC2=CC=CC=C2)N(CC2=CC=CC=C2)C2=C1C(CN(C)C)=C(O)C=C2
42	-6.65	477.4	5.1	CCOC(=O)C1=C(CSC2=CC=CC=C2)N(C)C2=CC(O)=C(Br)C(CN(C)C)=C12
43	-6.65	434.5	3.1	CCOC(=O)C1=C(CN2CCCC2)N(C)C2=C1C1=C(OCN(C)C1)C(=C2)C1=CN=CC=C1
44	-6.61	500.4	5.0	CCOC(=O)C1=C(CN2CCCC2)N(C2=CC(Br)=C(O)C(CN(C)C)=C12)C1=CC=CC=C1
45	-6.60	398.5	4.3	CCOC(=O)C1=C(CSC2=CC=CC=C2)N(C)C2=C1C(CN(C)C)=C(O)C=C2
46	-6.57	504.4	5.2	CCOC(=O)C1=C(CN2CCCC2)N(C2CCCCC2)C2=CC(Br)=C3OCN(C)CC3=C12
47	-6.50	372.5	4.7	CCCCN1C(C)=C(C(=O)OCC)C2=C1C=CC(O)=C2CN1CCCCC1
48	-6.46	511.9	5.7	CCOC(=O)C1=C(CSC2=CC=C(Cl)C=C2)N(C)C2=CC(Br)=C(O)C(CN(C)C)=C12
49	-6.45	428.9	4.5	CCOC(=O)C1=C(C)N(C2=C1C(CN1CCOCC1)=C(O)C=C2)C1=CC=C(Cl)C=C1
50	-6.40	453.4	4.4	CCCCN1C(C)=C(C(=O)OCC)C2=C(CN3CCOCC3)C(O)=C(Br)C=C12
51	-6.39	562.5	4.2	CCOC(=O)C1=C(CSC2=CC=CC=C2)N(C)C2=CC(Br)=C(O)C(CN3CCN(CCO)CC3)=C12
52	-6.37	506.5	5.1	CCOC(=O)C1=C(CN2CCCC2)N(C2CCCCC2)C2=CC(Br)=C(O)C(CN(C)C)=C12
53	-6.32	348.4	2.1	CCOC(=O)C1=C(C)N(CC(=O)OC)C2=C1C(CN(C)C)=C(O)C=C2
54	-6.30	387.5	3.7	CCCCN1C(C)=C(C(=O)OCC)C2=C1C=CC(O)=C2CN1CCN(C)CC1
55	-6.30	423.5	4.5	CCOC(=O)C1=C(CN(CC)CC)N(C2=C1C(CN(C)C)=C(O)C=C2)C1=CC=CC=C1
56	-6.29	394.5	3.9	CCOC(=O)C1=C(C)N(C2=C1C(CN1CCOCC1)=C(O)C=C2)C1=CC=CC=C1
57	-6.22	426.5	6.0	CCOC(=O)C1=C(C)N(C2=C1C1=C(OCN(CC3=CC=CC=C3)C1)C=C2)C1=CC=CC=C1
58	-6.15	454.4	2.7	CCOC(=O)C1=C(CN2CCOCC2)N(C)C2=CC(Br)=C(O)C(CN(C)C)=C12
59	-6.13	452.4	3.8	CCOC(=O)C1=C(CN2CCCCC2)N(C)C2=CC(Br)=C(O)C(CN(C)C)=C12
60	-6.02	382.4	4.0	CCOC(=O)C1=C(C)N(C2=C1C(CN(C)C)=C(O)C=C2)C1=CC=C(OC)C=C1
61	-6.00	364.4	4.3	CCOC(=O)C1=C(C)N(C)C2=C1C1=C(OCN(CC3=CC=CC=C3)C1)C=C2
62	-5.92	483.4	4.9	CCOC(=O)C1=C(CSC2=CSC=C2)N(C)C2=CC(Br)=C(O)C(CN(C)C)=C12
63	-5.77	381.4	2.0	CCOC(=O)C1=C(CO)N(C)C2=C1C1=C(OCN(C)C1)C(=C2)C1=CN=CC=C1
64	-5.75	330.4	3.3	CCOC(=O)C1=C(C)N(C)C2=C1C(CN1CCCCC1)=C(O)C=C2
65	-5.71	460.6	6.0	CCOC(=O)C1=C(CSC2=CC=CC=C2)N(C2=C1C(CN(C)C)=C(O)C=C2)C1=CC=CC=C1
66	-5.70	347.4	2.4	CCOC(=O)C1=C(C)N(C)C2=C1C(CN(C)C)=C(O)C(CN(C)C)=C2
67	-5.68	474.4	4.6	CCOC(=O)C1=C(CN(C)C)N(C2=CC(Br)=C(O)C(CN(C)C)=C12)C1=CC=CC=C1
68	-5.67	369.3	3.3	CCOC(=O)C1=C(C)N(C)C2=CC(Br)=C(O)C(CN(C)C)=C12
69	-5.67	483.4	5.0	CCOC(=O)C1=C(CSC2=CC=CS2)N(C)C2=CC(Br)=C(O)C(CN(C)C)=C12
70	-5.62	290.4	2.5	CCOC(=O)C1=C(C)N(C)C2=C1C(CN(C)C)=C(O)C=C2
71	-5.62	412.3	2.9	CCOC(=O)C1=C(CN(C)C)N(C)C2=CC(Br)=C(O)C(CN(C)C)=C12
72	-5.61	344.5	3.6	CCOC(=O)C1=C(C)N(C)C2=C1C(CN1CCC(C)CC1)=C(O)C=C2
73	-5.61	453.6	4.2	CCOC(=O)C1=C(CSC2=CC=CC=C2)N(C)C2=C1C(CN1CCN(C)CC1)=C(O)C=C2
74	-5.57	318.4	3.2	CCOC(=O)C1=C(C)N(C)C2=C1C(CN(CC)CC)=C(O)C=C2
75	-5.44	352.4	4.2	CCOC(=O)C1=C(C)N(C2=C1C(CN(C)C)=C(O)C=C2)C1=CC=CC=C1

76	-5.43	366.5	4.7	CCOC(=O)C1=C(C)N(C2=C1C(CN(C)C)=C(O)C=C2)C1=CC=C(C)C=C1
77	-5.35	366.5	4.2	CCOC(=O)C1=C(C)N(CC2=CC=CC=C2)C2=C1C(CN(C)C)=C(O)C=C2
78	-4.77	315.4	2.4	CCOC(=O)C1=C(C)N(C)C2=C1C(CN(C)C)=C(O)C(=C2)C#N
79	-4.65	345.4	2.3	CCOC(=O)C1=C(C)N(C)C2=C1C(CN1CCN(C)CC1)=C(O)C=C2

Table S2. Cross-Validation datasets randomly generated using a "leave 4 out" strategy.

	Remove for Validation Set (data #)
dataset 1	Remove: 1, 2, 5, 6
dataset 2	Remove: 2, 10, 11, 14
dataset 3	Remove: 3, 8, 11, 13
dataset 4	Remove: 1, 4, 5, 10
dataset 5	Remove: 1, 3, 5, 13
dataset 6	Remove: 6, 9, 12, 14
dataset 7	Remove: 2, 3, 7, 13
dataset 8	Remove: 3, 8, 9, 10
dataset 9	Remove: 4, 8, 9, 12
dataset 10	Remove: 5, 6, 10, 12
dataset 11	Remove: 4, 6, 7, 11
dataset 12	Remove: 7, 9, 11, 12
dataset 13	Remove: 2, 7, 13, 14
dataset 14	Remove: 1, 4, 8, 14

Table S3. Cross-Validation data for linear correlation analysis of OA saponin derivatives predicted free energy of binding (ΔG_{bind}) at Site 1 and Site 2 compared to experiment.

	Site 1 (Test Set)			Site 1 (Validation Set)		
	RMSE (kcal/mol)	R	R^2	RMSE (kcal/mol)	R	R^2
dataset 1	0.94	0.629	0.395	1.52	0.576	0.332
dataset 2	0.95	0.606	0.368	1.51	0.595	0.354
dataset 3	0.96	0.685	0.470	1.46	0.545	0.297
dataset 4	0.91	0.624	0.389	1.63	0.980	0.959
dataset 5	0.93	0.669	0.447	1.59	0.691	0.477
dataset 6	0.96	0.463	0.214	1.47	0.904	0.817

dataset 7	0.96	0.744	0.553	1.44	0.075	0.006
dataset 8	0.96	0.631	0.399	1.46	0.883	0.780
dataset 9	0.96	0.500	0.250	1.47	0.824	0.680
dataset 10	0.93	0.429	0.184	1.55	0.761	0.580
dataset 11	0.96	0.683	0.467	1.43	0.475	0.226
dataset 12	0.96	0.508	0.258	1.43	0.898	0.807
dataset 13	0.95	0.695	0.483	1.50	0.175	0.031
dataset 14	0.94	0.601	0.361	1.52	0.266	0.071
Average	0.95	0.605	0.374	1.50	0.618	0.458
Stdev	0.02	0.095	0.111	0.06	0.287	0.318

	Site 2 (Test Set)			Site 2 (Validation Set)		
	RMSE (kcal/mol)	R	R^2	RMSE (kcal/mol)	R	R^2
dataset 1	0.38	-0.162	0.026	0.57	-0.754	0.568
dataset 2	0.36	-0.312	0.098	0.63	-0.361	0.130
dataset 3	0.36	-0.385	0.148	0.63	-0.033	0.001
dataset 4	0.35	-0.186	0.034	0.68	-0.590	0.348
dataset 5	0.31	-0.123	0.015	0.79	-0.566	0.320
dataset 6	0.38	-0.478	0.229	0.55	-0.055	0.003
dataset 7	0.37	-0.263	0.069	0.61	-0.687	0.471
dataset 8	0.34	-0.468	0.219	0.70	-0.704	0.496
dataset 9	0.41	-0.498	0.248	0.38	0.461	0.213
dataset 10	0.41	-0.622	0.387	0.39	0.503	0.253
dataset 11	0.42	-0.284	0.081	0.36	-0.139	0.019
dataset 12	0.43	-0.513	0.264	0.28	-0.257	0.066
dataset 13	0.38	-0.096	0.009	0.56	-0.803	0.644
dataset 14	0.29	0.043	0.002	0.82	-0.845	0.713
Average	0.37	-0.311	0.131	0.57	-0.345	0.303
Stdev	0.04	0.192	0.120	0.16	0.446	0.245