Supplementary Information

Bioactive compounds identified by BABM with novel anti-EBOV, anti-ZIKV, or anti-SARS-CoV-2 activities

Anti-EBOV compounds: Arbidol (or umifenovir) (IC_{50}~5 μM) is an antiviral treatment for influenza infection used in Russia and China.[1] A more recent study reported *in vitro* activity of umifenovir at preventing entry of the EBOV Zaïre Kikwit.[2] Difeterol is used as an antihistamine drug in Japan.[3] Difeterol (IC_{50}~3 μM) was also identified as one of the active compounds from a high throughput screen using the EBOV entry assay.[4] Colchicine, podofilox, 7-epi-docetaxel and β-peltatin are all microtubule inhibitors.

Anti-ZIKV compounds: Colchicine is a medication commonly used to treat gout and as an anti-inflammatory drug for some other clinical conditions. Podofilox, or podophyllotoxin, is used as a medical cream to treat genital warts and molluscum contagiosum. 7-Epi-docetaxel is an impurity of docetaxel, a chemotherapy medication used to treat several types of cancer. β-Peltatin is a plant metabolite with antineoplastic properties that belongs to the same structural class as podophyllotoxin.[5] Dolastatin 10, a pentapeptide, is also an inhibitor of microtubule assembly and an investigational drug that has been in clinical trials for its antineoplastic activity.[5] Like colchicine, narciclasine, an amaryllidaceae alkaloid, was intensively investigated as an antitumor compound both *in vitro* and *in vivo* and have shown anti-inflammatory actions *in vivo*.[6] Floxuridine is a nucleoside that belongs to the class of antimetabolites. Floxuridine is an oncology drug most often used in the treatment of colorectal cancer. The other known bioactive compounds include two mycotoxins, diacetoxyscirpenol and T-2 Toxin.

Anti-SARS-CoV-2 compounds: Spiramide (AMI-193) is an experimental antipsychotic that acts as a selective 5-HT2A, 5-HT1A, and D2 receptor antagonist.[7] Ftormetazine is a derivative of the phenothiazine class of antipsychotic drugs.[8] Prenylamine is a calcium channel blocker of the amphetamine chemical class, which was used as a vasodilator in the treatment of angina pectoris and later withdrawn due to cardiotoxicity. [9,10] BEPP [1H-benzimidazole1-ethanol,2,3-dihydro-2-imino-a-(phenoxymethyl)-3-(phenylmethyl)-,monohydrochloride] is a synthetic compound that has been reported to inhibit viral replication by inducing RNA-dependent protein kinase -dependent apoptosis.[11] Triparanol was the first synthetic cholesterol-lowering drug which was introduced in the U.S. in 1960, but was later withdrawn due to severe adverse effects.[12,13] Octoclothepine is a tricyclic antipsychotic drug for the treatment of schizophrenic psychosis with high affinities for the dopamine receptors.[14] Metaphit, the m-isothiocyanate derivative of phencyclidine, is an investigational drug that acts as an

acylator of NMDARAn, sigma and DAT binding sites in the CNS.[15] L-703, 606 is a known antagonist of the neurokinin-1 (NK1) receptor.[16]

Antiviral mechanism of novel anti-SARS-CoV-2 compounds

The PP entry assay detects inhibitors of spike (S) mediated SARS-CoV-2 cell entry. A cell viability counter-screen was also run to filter out inhibition caused by compound cytotoxicity artifacts. Compounds that showed concentration dependent inhibition in the PP entry assay and were inactive or at least 6-fold less potent in the cell viability counter-screen were considered active SARS-CoV-2 cell entry inhibitors. The $3CL^{pro}$, also known as the main protease, of SARS-CoV-2 is a specific viral enzyme that plays an essential role in viral replication. The $3CL^{pro}$ assay is an enzymatic assay that detects inhibitors of this protease. As the $3CL^{pro}$ assay uses a fluorescence readout, a counter screen was run to eliminate fluorescence quenching compounds. Compounds that showed concentration dependent inhibition in the $3CL^{pro}$ assay and were inactive or at least 6-fold less potent in the counter screen were considered active $3CL^{pro}$ inhibitors. Autophagy modulators have been reported to block the SARS-CoV-2 cytopathic effect .[31] The GFP-LC3 assay is a cell-based imaging autophagic flux assay. Data were expressed as three parameters: “% of positive cells”, “Total Spot Area” and “Relative Spot Intensity”. Compounds that increased all three parameters in a concentration dependent manner were considered active autophagy modulators.

Comparison of structural spaces covered by BABM and SBM

BABM was applied to ZIKV NS1 models as a proof-of-concept comparison to traditional modeling approaches. **Supplementary Figure 2A** shows the 652 experimentally confirmed NS1-expression actives correctly identified by at least one of the three models (SBM, BABM, and CM). Only 43 of the 652 compounds were predicted as active by all three models. BABM identified 82%, while the CM and SBM identified 76% and 28% of 652 active compounds, respectively. BABM and CM identified the majority of the actives, most of which (448 out of 534) were also shared by both models. SBM identified a much smaller number of actives, most of which (118 out of 182) were also different from those identified by the BABM and CM. Out of the 118 actives picked up by the SBM only, 106 were not predicted by the BABM or CM because no assay activity profile data were available on these compounds. The distributions of the 85 experimentally confirmed anti-SARS-CoV-2 compounds among the three types of models (SBM, BABM, and CM) showed trends similar to those of ZIKV NS1 (**Supplementary Figure 2B**). The compounds identified by the BABM is a subset of the ones identified by the CM (45%). Only a small

fraction of the compounds (12%) was identified by the SM. The BABM and SM identified nearly completely different sets of compounds with only one overlapping. As structure-based models rely on structure similarity to make predictions, these models were not reliable in predicting compounds with completely new scaffolds that were not already represented in the training set. Activity-based models which rely on activity profile data for training, in contrast, are not restricted by structure similarity and thus are potentially more powerful in discovering new scaffolds. Models for all three viral targets in this study identified compounds with very low structural similarity to active compounds in the training set (e.g., average Tanimoto similarity scores to the training set as low as 0.04-0.07; **Table 1**). Compared to traditional QSAR models built with chemical structure data alone, the BABM identified compounds that are structurally distinct from the training set and the compounds identified by the SBM (**Supplementary Figure 3**), demonstrating the advantage of the BABM in discovering new chemical types.

Table 1. Model performances. P-values were calculated using the Fisher's exact test.

SARS-CoV2 Model	AUC-ROC (Test Set)	Cherry Pick Validation			Training Set Active Rate*	P
		TP	FP	PPV		
SBM**	0.71±0.01	12	26	31.58%	11.46%	8.97×10^{-4}
BABM-S	0.75±0.02	28	45	38.36%	19.55%	4.64×10^{-4}
BABM-M	0.79±0.02	25	44	36.23%	17.54%	3.47×10^{-4}
CM-S	0.77±0.02	83	176	32.05%	19.57%	2.66×10^{-5}
CM-M	0.81±0.02	60	100	37.50%	17.56%	3.84×10^{-8}

*The training set active rate for the SBM is close to the SARS-CoV2 CPE qHTS assay hit rate. The training set active rates for the activity-based models are higher because the NPC screened for this assay was recently updated with many new drugs not in the older version. Profile data in other assays were not available for these new drugs, most of which were inactive in the CPE assay, thus they were not included in training the activity-based models.

**Not used to select compounds for experimental validation.

NS1 Model	AUC-ROC (Test Set)	Cherry Pick Validation			Training Set Active Rate	P
		TP	FP	PPV		
SBM	0.82±0.01	182	419	30.28%	1.10%	$<10^{-20}$
BABM-S	0.82±0.02	532	692	43.46%	15.90%	$<10^{-20}$
CM-S	0.86±0.01	493	617	44.41%	15.90%	$<10^{-20}$

EBOV Model	AUC-ROC (Test Set)	Cherry Pick Validation*			Training Set Active Rate	P
		TP	FP	PPV		
SBM	0.66±0.02	N/A	N/A	N/A	7.47%	N/A
BABM-S	0.80±0.02	48	12	80.00%	11.79%	$<10^{-20}$
BABM-G	0.70±0.02	13	2	86.67%	13.41%	7.05×10^{-10}
CM-S	0.83±0.01	46	12	79.31%	11.81%	$<10^{-20}$
CM-G	0.78±0.03	16	2	88.89%	13.30%	3.04×10^{-12}

*Cherry picked 96 compounds, 34 of which killed cells at 30 µM and not tested for Ebola. The remaining 62 compounds were used to evaluate model performance.

Table 2. Training dataset compositions for models

SARS-CoV2 Data Set	Feature Count	Active Count	Inactive Count	Prediction Set
SBM	729	279	2155	617,947
BABM-S	130	209	860	109,291
BABM-M	225	200	940	99,660
CM-S	859	209	859	103,400
CM-M	954	200	939	96,142

NS1 Data Set	Feature Count	Active Count	Inactive Count	Prediction Set
SBM	729	1,023	91,643	527,715
BABM-S	130	281	1,486	108,593
CM-S	859	281	1,486	102,701

EBOV Data Set	Feature Count	Active Count	Inactive Count	Prediction Set
SBM	729	148	1,833	618,400
BABM-S	130	128	958	109,274
BABM-G	29	110	710	70,932
CM-S	859	128	956	103,384
CM-G	758	108	704	64,397

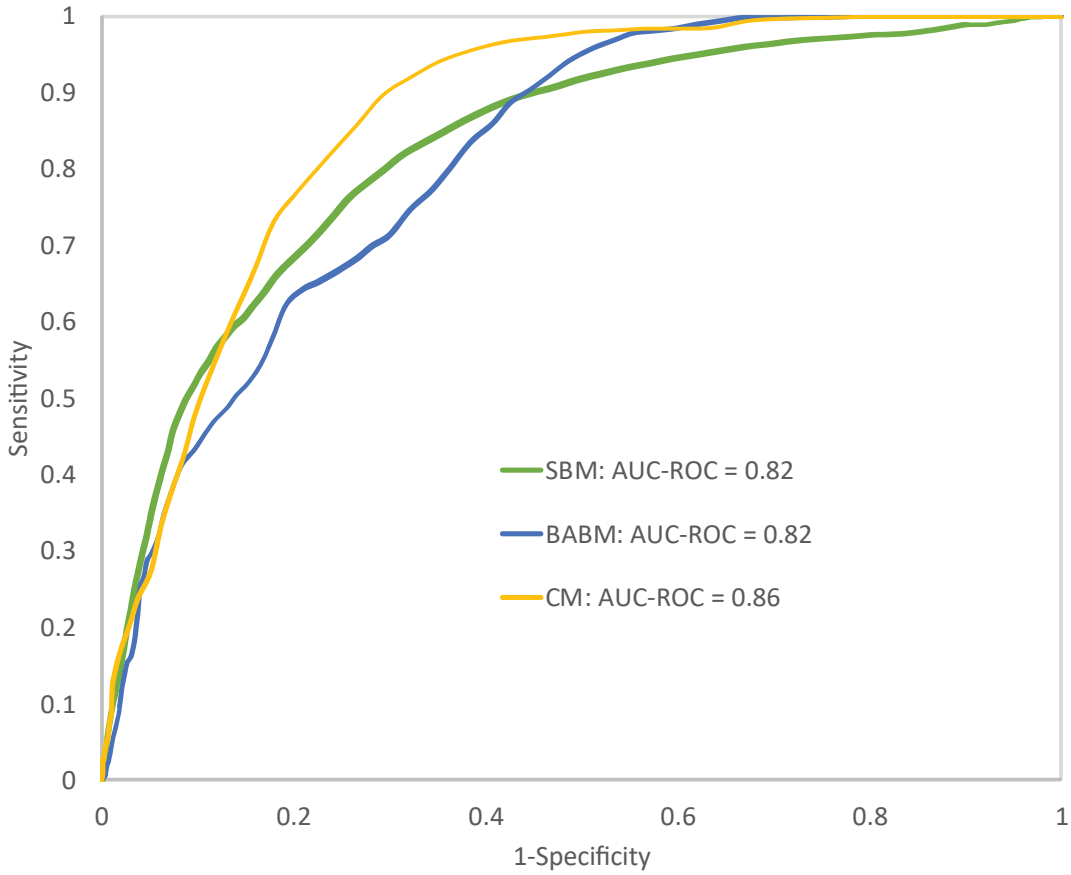
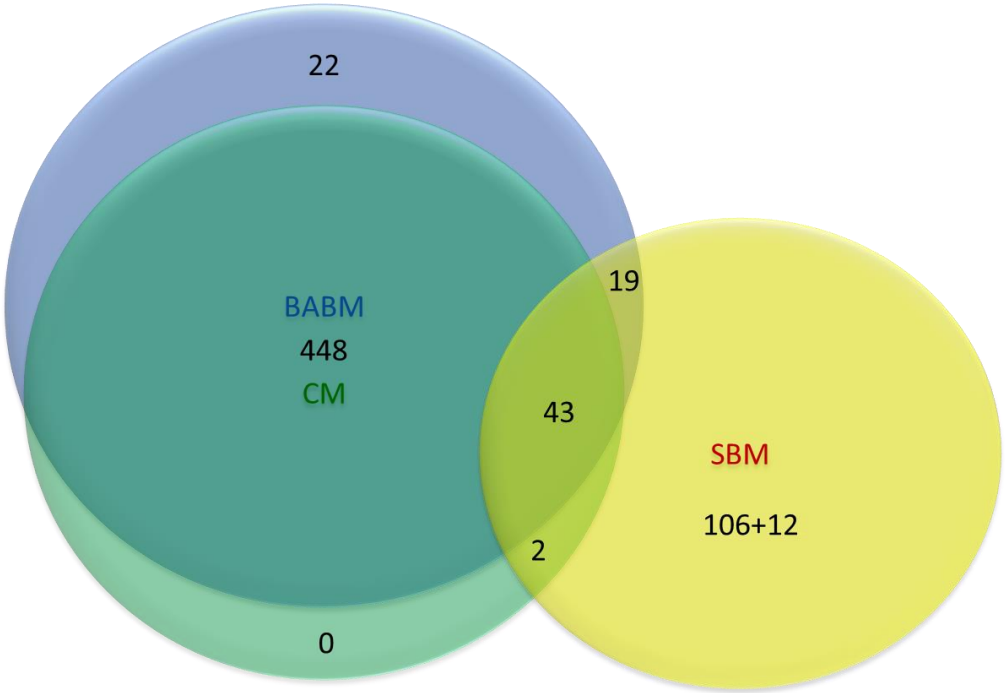


Figure 1. Example ROC curves on the test set. Curves were selected from the NS1 models. BABM = biological activity-based model; SBM = structure-based model; CM = combined model.

A.



B.

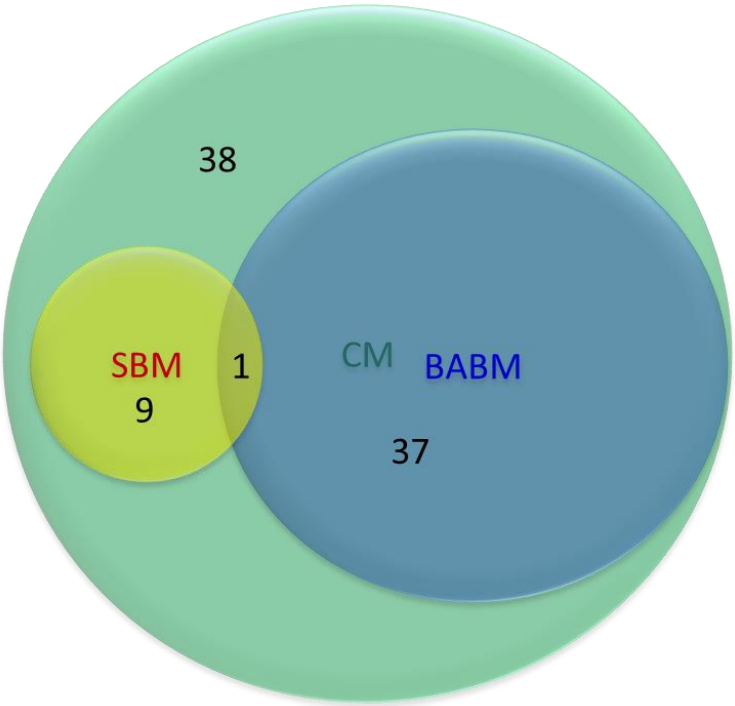


Figure 2. Confirmed active ZIKA NS1 inhibitors (A) and anti-SARS-CoV-2 compounds (B)correctly identified by each model. BABM = biological activity-based model; SBM = structure-based model; CM = combined model.

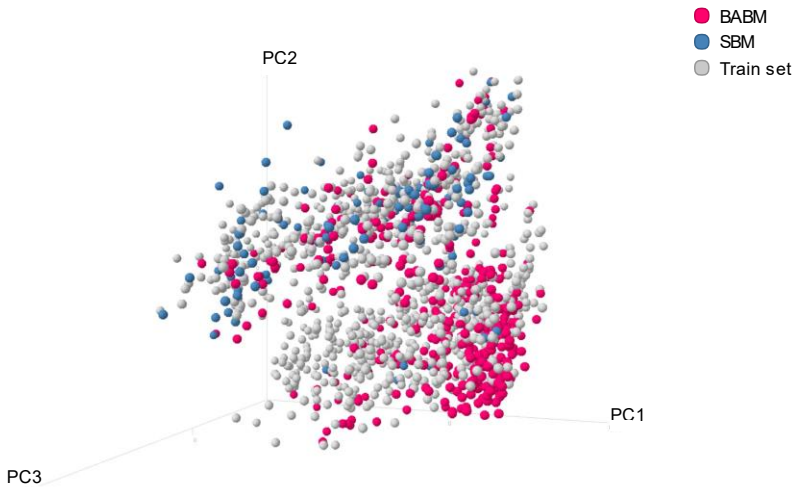


Figure 3. Chemical structure space occupied by compounds predicted as active by the BABM and SBM NS1 models, and active compounds in the training set, i.e., the original NS1 screen. The plot shows that the compounds identified by the activity-based model occupied structure spaces distinct from those identified by the structure-based model.

References

1 Leneva, I. A., Russell, R. J., Boriskin, Y. S. & Hay, A. J. Characteristics of arbidol-resistant mutants of influenza virus: implications for the mechanism of anti-influenza action of arbidol. *Antiviral Res* **81**, 132-140, doi:10.1016/j.antiviral.2008.10.009 (2009).

2 Pecheur, E. I. *et al.* The Synthetic Antiviral Drug Arbidol Inhibits Globally Prevalent Pathogenic Viruses. *J Virol* **90**, 3086-3092, doi:10.1128/JVI.02077-15 (2016).

3 KEGG. (2019).

4 Kouznetsova, J. *et al.* Identification of 53 compounds that block Ebola virus-like particle entry via a repurposing screen of approved drugs. *Emerg Microbes Infect* **3**, e84, doi:10.1038/emi.2014.88 (2014).

5 Valeriote, F. *et al.* Discovery and development of anticancer agents from plants. *J. Exp. Ther. Oncol.* **2**, 228-236, doi:10.1046/j.1359-4117.2002.01038.x (2002).

6 Furst, R. Narciclasine - an Amaryllidaceae Alkaloid with Potent Antitumor and Anti-Inflammatory Properties. *Planta Med* **82**, 1389-1394, doi:10.1055/s-0042-115034 (2016).

7 Czoty, P. W. & Howell, L. L. Behavioral effects of AMI-193, a 5-HT(2A)- and dopamine D(2)-receptor antagonist, in the squirrel monkey. *Pharmacol Biochem Behav* **67**, 257-264, doi:10.1016/s0091-3057(00)00321-x (2000).
8 Brusova, E. G. Effects of some phenothiazine and dibenzazepine derivatives on the muscarinic cholinergic system. *Byull. Eksp. Biol. Med.* **113**, 60-62 (1992).
9 Upward, J. W., Waller, D. G. & George, C. F. Class II antiarrhythmic agents. *Pharmacol. Ther.* **37**, 81-109, doi:10.1016/0163-7258(88)90021-6 (1988).
10 Bayer, R., Schwarzmaier, J. & Pernice, R. Basic mechanisms underlying prenylamine-induced 'torsade de pointes': differences between prenylamine and fendiline due to basic actions of the isomers. *Curr. Med. Res. Opin.* **11**, 254-272, doi:10.1185/03007998809114244 (1988).
11 Hu, W. *et al.* Double-stranded RNA-dependent protein kinase-dependent apoptosis induction by a novel small compound. *J Pharmacol Exp Ther* **328**, 866-872, doi:10.1124/jpet.108.141754 (2009).
12 Steinberg, D., Avigan, J. & Feigelson, E. B. Effects of Triparanol (Mer-29) on Cholesterol Biosynthesis and on Blood Sterol Levels in Man. *J Clin Invest* **40**, 884-893, doi:10.1172/JCI104323 (1961).
13 Kirby, T. J. Cataracts produced by triparanol. (MER-29). *Trans Am Ophthalmol Soc* **65**, 494-543 (1967).
14 Metysova, J., Metys, J., Dlabac, A., Kazdova, E. & Valchar, M. Pharmacological properties of a potent neuroleptic drug octoclothepin. *Acta Biol Med Ger* **39**, 723-740 (1980).
15 Rafferty, M. F., Mattson, M., Jacobson, A. E. & Rice, K. C. A specific acylating agent for the [3H]phencyclidine receptors in rat brain. *FEBS Lett* **181**, 318-322, doi:10.1016/0014-5793(85)80284-2 (1985).
16 Hamity, M. V., White, S. R. & Hammond, D. L. Effects of neurokinin-1 receptor agonism and antagonism in the rostral ventromedial medulla of rats with acute or persistent inflammatory nociception. *Neuroscience* **165**, 902-913, doi:10.1016/j.neuroscience.2009.10.064 (2010).