Computer-Aided Discovery and Characterization of Novel Ebola Virus Inhibitors

Stephen J. Capuzzi[1,#], **Wei Sun**[2,#], **Eugene N. Muratov**[1,3], **Carles Martínez-Romero**[4,5], **Shihua He**[6], **Wenjun Zhu**[6,7], **Hao Li**[2], **Gregory Tawa**[2], **Ethan G. Fisher**[2], **Miao Xu**[2], **Paul Shinn**[2], **Xiangguo Qiu**[6,7], **Adolfo García-Sastre**[4,5,8], **Wei Zheng**[2,*], and **Alexander Tropsha**[1,*]

[1] Laboratory for Molecular Modeling, Division of Chemical Biology and Medicinal Chemistry, UNC Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, USA.

[2] National Center for Advancing Translational Sciences, National Institutes of Health, Bethesda, MD 20892, USA.

[3] Department of Chemical Technology, Odessa National Polytechnic University, Odessa, 65000, Ukraine.

[4] Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.

[5] Global Health and Emerging Pathogens Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.

[6] Special Pathogens Program, National Microbiology Laboratory, Public Health Agency of Canada, 1015 Arlington Street, Winnipeg, MB R3E 3R2 Canada.

[7] Department of Medical Microbiology, University of Manitoba, 745 Bannatyne Avenue, Winnipeg, Manitoba R3E 0J9, Canada

*****Corresponding Authors** Alexander Tropsha; Laboratory for Molecular Modeling, Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC, 27599, USA; Tel: +19199662955; Fax: +19199660102; alex_tropsha@unc.edu. Wei Zheng; Division of Pre-Clinical Innovation, National Center for Advancing Translational Sciences, National Institutes of Health, 9800 Medical Center Drive, Bethesda, MD 20892, USA. Tel: +13012175251; Fax: +13012176018; wzheng@mail.nih.gov.
[#]Contributed equally

Author Contributions:
SJC, and WS contributed equally.

ASSOCIATED CONTENT

Supporting Information

Csv versions of Tables 1 and 2 containing molecular formula strings are provided as Supplementary Files 16 and 17, respectively. All QSAR models from Chembench, HiT QSAR, and GUSAR are available for download from Figshare (https://figshare.com/s/d394aa911c37b08bcc64) in Supplementary Files 11–13, respectively. Additionally, all the datasets and Chembench models are provided in and on the Chembench Web-Portal (https://chembench.mml.unc.edu/), which provides public access and use of data and models used in this study. The P1, P2, HEK, and HeLa training sets are publicly indexed as "Ebola_SM1" and "Ebola_PCM4", "151105_Ebola_Toxicity_HEK", and "151105_Ebola_Toxicity_HELA", respectively. The Chembench P1, P2, and HeLa models are publicly indexed as "153004_ebola_Strict_Model1_166_DragonH", "151305_ebola_1224_PCM4", and "151105_ebola_tox_HeLa", respectively.

All other supplementary data, i.e., supplementary Figures and Tables, readme file with brief description of supplemental files, and Supplementary Files 1–10 and 14–18 are provided in Supplementary_data.zip. This material is available free of charge via the Internet at http://pubs.acs.org.

[8]Department of Medicine, Division of Infectious Diseases, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

Abstract

The Ebola virus (EBOV) causes severe human infection that lacks effective treatment. A recent screen identified a series of compounds that block EBOV-like particle entry into human cells. Using data from this screen, Quantitative Structure-Activity Relationship (QSAR) models were built and employed for virtual screening of a ~17 million compound library. Experimental testing of 102 hits yielded 14 compounds with IC_{50} values under 10 μM, including several sub-micromolar inhibitors, and more than 10-fold selectivity against host cytotoxicity. These confirmed hits include FDA-approved drugs and clinical candidates with non-antiviral indications, as well as compounds with novel scaffolds and no previously known bioactivity. Five selected hits inhibited BSL-4 live-EBOV infection in a dose-dependent manner, including vindesine (0.34 μM). Additional studies of these novel anti-EBOV compounds revealed their mechanisms of action, including the inhibition of NPC1 protein, cathepsin B/L, and lysosomal function. Compounds identified in this study are among the most potent and well-characterized anti-EBOV inhibitors reported to date.

Graphical Abstract
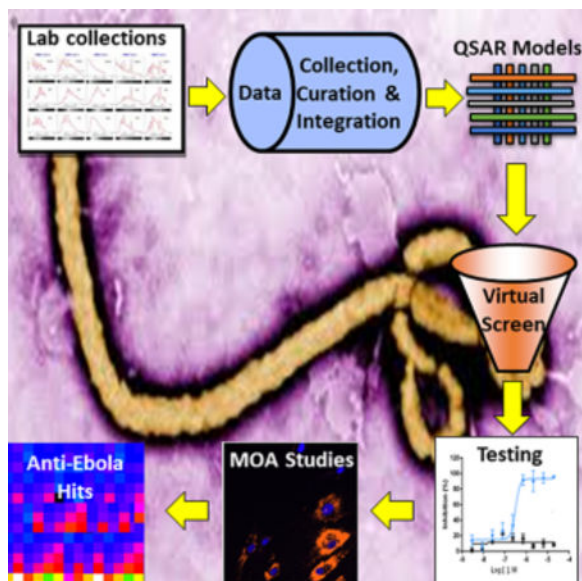
Keywords

EBOV; antiviral drug screening; QSAR; computer-aided drug design

INTRODUCTION

The 2014 Ebola outbreak was the largest and most persistent since the discovery of the Ebola virus (EBOV) in 1976. Alarmingly, a new EBOV outbreak was confirmed in the Democratic Republic of Congo in May 2017.[1] Though advances in the research and

development of Ebola therapeutics have been made,[2–4] Ebola drug discovery endeavors are hindered due to the high virulence of the EBOV and its biosafety level 4 (BSL-4) classification.[5] Recently, a biosafety level 2 (BSL-2) Ebola virus-like particle (VLP) entry assay was developed and utilized for a drug repurposing screen of Food and Drug Administration (FDA)-approved drugs.[6–8] The Ebola VLP contains glycoprotein (GP) and the matrix protein VP40 fused to a beta-lactamase reporter for monitoring of VLP entry into cells. Although this BSL-2 Ebola VLP assay enables rapid compound screening, it requires a centrifugation step for assay plates at 1,500 g for 45 minutes at 4 °C that limits its screening throughput. Computational approaches that leverage generated data can be used to design or select small sets of compounds for lead identification in order to reduce the time and costs of high throughput screening.

Using the existing data from the Ebola VLP entry assay as well as cytotoxicity data, QSAR models[9] can be built and then employed for virtual screening of large chemical libraries to predict active compounds against EBOV infection with low expected toxicity. Indeed, QSAR modeling approaches have been previously employed for identification of compounds with efficacy against EBOV.[10,11] Herein, we describe a study that relied on synergistic combination of statistical data modeling and experimental testing for both antiviral inhibitor potency and host cell cytotoxicity (Figure 1). Our study utilized both BSL-2 and BSL-4 assays to experimentally validate hits identified computationally.

To identify compounds with anti-EBOV activity and limited host cell cytotoxicity, we designed an integrated QSAR modeling system for virtual compound screening that is combined with experimental testing on a focused set of predicted compounds. In this study, existing antiviral activity and compound cytotoxicity data were collected and carefully curated; respective QSAR models were built and rigorously validated; these models were employed for virtual screening of a large chemical library (~17 million compounds), resulting in 102 hits prioritized for experimental testing; the anti-EBOV activity in the Ebola VLP assay and cytotoxicity in host cells of these hits were determined experimentally in BSL-2 and BSL-4 assays; and the mechanisms of anti-EBOV activity for confirmed hits were identified. Ultimately, 14 potent hits with activity ranging between 0.272 μM and 9.65 μM as well as more than 10-fold selectivity over compound cytotoxicity in host cells were confirmed. Next, five selected hits were shown to inhibit BSL-4 live-EBOV infection in a dose-dependent manner. Two of these hits possessed novel scaffolds, making them candidates for further medicinal chemistry optimization as potential anti-EBOV agents. This study presents the first example of computationally-driven prioritization and experimental discovery of novel potent anti-Ebola compounds with high therapeutic windows in the published literature.

RESULTS

Model Performance

Prior to the modeling, MODIs of 0.69 and 0.68 were calculated for the P1 and P2 datasets, respectively. For each protocol, three separate software packages (Chembench, HiT QSAR, and GUSAR) employing different descriptors and different machine learning techniques (MLTs) were utilized for model building. In total, six individual models were built and

rigorously validated. Results of 5-fold external cross-validation are presented in Supplementary Table 1. In order to demonstrate that the predictive power of the models was not due to random correlation between bioactivity and chemical descriptors, 1000 rounds of Y-randomization was performed. No Y-randomized models had a CCR above 0.60.

For P1, models built with HiT QSAR and GUSAR had the highest predictive accuracy, irrespective of the use of different chemical descriptors and MLTs. For P2, HiT QSAR again showed the best performance. Additionally, the CCR of the Chembench model improved by ~7% for P2 over P1. All models were deemed robust and statistically valid (see Supplementary Table 1).

For HeLa and HEK cell lines, MODI of 0.65 and 0.70 were obtained, respectively. For HeLa cytotoxicity, GUSAR yielded the best overall model. Chembench and GUSAR had inverse sensitivity and specificity profiles, indicating that Chembench could better identify toxic compounds, while GUSAR could better identity non-toxic compound. This observation highlights the reciprocal benefit of consensus modeling, *i.e.*, utilizing all the models for VS. No Y-randomized models had a CCR in access of 0.60. All models were deemed robust and statistically valid. For HEK cytotoxicity, GUSAR again proved to be the best overall model. On the other hand, Chembench and HiT QSAR were not statistically validated, as several metrics fell below the 0.60 threshold. Thus, only GUSAR was used for prediction of HEK cytotoxicity. Y-randomized models for GUSAR did not exceeded a CCR of 0.60. A summary of all model performance can be found in Supplementary Table 1. Original HTS data are available in Supplementary Files 1–4. All datasets and developed models are provided in Supplementary Files 5–13. Datasets and Chembench models are also publicly available on the Chembench Web-Portal (https://chembench.mml.unc.edu/).

Interpretation of the developed models revealed high impact of lipophilicity on antiviral activity. Lipophilicity affects drug absorption, bioavailability, ligand-receptor interactions, clearance, and toxicity and traditionally plays an important role. Indeed, we have found that the average logP value for training set Ebola inhibitors is 3.8 compared with 2.2 for non-inhibitors. From a design perspective, therefore, given that both logP values do not violate Lipinski's rule for lipophilicity,[12] increasing the logP may be desirable.

QSAR-Based Virtual Screening

QSAR-based virtual screening (VS) was carried out according to the workflow presented in Figure 2. Initially, ~17 million compounds (see Methods) were downloaded, prepared, and screened. As previously stated, "hits" were those compounds that were within the AD of the respective model and predicted by all models to have high antiviral activity and limited host cytotoxicity. In total, 102 VS hits were selected for experimental validation in the Ebola-VLP entry assay. In the light of revealed importance of lipophilicity, we aimed to give a priority to hydrophobic compounds. LogP of the selected hits ranged between 2.63 and 7.22 with the average value of 4.96, which is higher than average of 3.8 for training set Ebola inhibitors.

Experimental Evaluation

Experimental confirmation of Anti-EBOV activity of 14 compounds—Based on the virtual screening results, 102 compounds were purchased and experimentally screened in the Ebola-VLP entry assay in parallel with an ATP content assay to determine compound cytotoxicity in host cells. All compounds were screened at 11 concentration dilutions ranging from 0.001 to 57 μM.[13] Out of 102 compounds tested, 51 showed greater than 50% inhibition, indicating that half of compounds had confirmed anti-EBOV activity. Next, 20 of these 51 compounds exhibited the IC_{50} values under 10 μM. Because the compound cytotoxicity at higher compound concentrations might reduce the Ebola VLP entry in cells, these potential false positive compounds should be deprioritized. After comparing to the compound cytotoxicity data, 14 of these confirmed compounds showed a greater than 10-fold selectivity index (SI) of anti-Ebola VLP entry over compound cytotoxicity. Vindesine and BIX-01294 inhibited the virus in the nanomolar range (Supplementary Figure 1). Moreover, these 14 confirmed compounds, except for ZINC91973695 and ZINC67869167, have known mechanisms of action (MOAs) and therapeutic indications (Table 1), including eight anti-cancer, two antihistamines, and two anti-psychotic and anti-inflammatory agents (Table 1). The remaining two compounds obtained from the ZINC database (ZINC91973695 and ZINC67869167) have no previously reported bioactivities, anti-Ebola or otherwise. Five hits were further validated in a live EBOV infection assay at bio-safety level-4 (BSL-4). All five hits showed dose-response inhibition against EBOV infection (Figure 3). Vindesine was the most potent compounds with an IC_{50} of 0.34 μM. The IC_{50} values of NVP-ADW742, BIX-01294, ZINC67869167, and ZINC91973695 were between 1 μM to 10 μM in the live EBOV infection assays.

Mechanisms of action against EBOV entry

We probed the chemical biology of these hit compounds in both viral and host systems in order to uncover the mechanisms of anti-EBOV action. We examined the potential sites for drug interaction including Niemann Pick C1 (NPC1) protein, lysosomal function, cathepsin B, and cathepsin L,[14–16] as well as the direct binding of these compounds to the Ebola VLP proteins using thermal shift binding assays.[17] The results of chemical biology studies revealed that these compounds may act via one or more these targets/mechanisms.

Direct binding of compounds to Ebola envelop viral proteins—The process of EBOV entry into cells involves binding of viral envelop protein(s) to the cell membrane receptor protein/molecule, endocytosis, movement of endocytic vesicles to early/late endosomes and lysosomes, and ejection of viral RNA into the cytosol.[18] Therefore, inhibition of viral protein binding to cell membrane proteins/binding partners can effectively reduce viral entry and subsequent viral replication in cells. Because the cell surface binding protein/molecule for Ebola viral proteins is still unclear, we determined direct binding of these compounds to recombinant Ebola protein. To examine whether these compounds directly interact with the EBOV, their ability of stabilizing Ebola protein was tested in a thermal shift assay using recombinant Ebola VLP. None of the compounds at 50 uM were able to protect Ebola VLP from thermal denaturation. (Supplementary Figure 2).

Inhibitions of cathepsin B and L activities—Cathepsin B and L are lysosomal endopeptidases that had been reported to prime EBOV proteins in lysosomes before the viral RNAs are injected into the cytosol for virus replication. Inhibition of cathepsin B or L significantly reduces EBOV infection.[14] GANT61 (an inhibitor of GLI1 and GLI2-induced transcription), deptropine (an antihistamine), and ebastine (an antihistamine) inhibited the enzymatic activity of cathepsin L (Figure 4a and 4b). Only GANT61 inhibited enzymatic activity of cathepsin B (Figure 4c and 4d).

Inhibition of function of NPC1 protein—The NPC1 protein has been reported as an intracellular receptor for EBOV.[15,16,19] Significant reduction of EBOV entry and infection were observed in the NPC1-deficient cells and mouse models.[15,20] Ebastine increased cholesterol accumulation in cells determined by the filipin staining assay, which indicated a functional impairment of NPC1 protein; whereas, the other eight evaluated hits did not impair NPC1 protein function (Figure 5a).

Enlargement of lysosome size—Lysosomes in cells are enlarged after treatment with certain compounds that damage lysosome functions, resulting in accumulation of lipids and other macromolecules.[21] The enlarged lysosomes are often observed in the patient cells with lysosomal storage diseases caused by mutations in lysosomal proteins and lipid accumulation.[22] EBOV entry is significantly reduced after the lysosomal functions are impaired by compounds. All of the nine evaluated hits increased LysoTracker dye staining in cells, indicating an enlargement in lysosome size (Figure 5b).

Cheminformatics Analysis

Assay Liabilities—First, using substructural pattern matchers implemented in ZINC15,[23] the 14 experimentally confirmed hits were found to be free of chemical aggregation liabilities.[24] Furthermore, we followed a recent ACS editorial[25] and this journal requirements (http://pubs.acs.org/paragonplus/submission/jmcmar/jmcmar_authguide.pdf) for the evaluation of pan-assay interference compounds (PAINS)[26]. Although several recent studies,[27,28,29] demonstrated that PAINS alerts are not reliable indicators of assay interference propensity, we have performed this analysis and found that none of the hits possessed any PAINS alerts. In addition, as the assay employed herein relied on a beta-lactamase reporter system, all 14 hits were also checked for potential beta-lactamase inhibition trends using PubChem Promiscuity.[30] No heightened beta-lactamase assay activity trends were observed, indicating that these hits are not assay artifacts (Supplementary File 14).

Chemical Similarity to Training Set Compounds—Hierarchical clustering analysis revealed that majority of the hits are structurally dissimilar from each other, aside from afimoxifene and endoxifen (Supplementary Figure 3). Clustering thus indicates the hits discovered in this study access a wide range of chemical space across several unique chemotypes. The structural similarity based on the Tanimoto coefficient (Tc) of the 14 hits were then compared with compounds in the antiviral training sets (Supplementary Table 2) in order to assess the uniqueness of hits. In addition to being highly structurally similar to each other, afimoxifene and endoxifen both have Tc above 0.90 to tamoxifen, which was a

previously reported anti-Ebola inhibitor.[31] Likewise, tetrandrine is highly structurally similar (Tc=0.97) to cepharanthine, a training set active compound. The most potent hit in this study, vindesine, had a Tc of 0.96 to vinblastine, which was the most potent hit in the original screen.[31] These hits, while not entirely unique from a chemical perspective, illustrate that the developed QSAR models are robust and that the experimental assays are reproducible. The remaining 10 hits were considerably dissimilar from any training set compounds (Tc = 0.63–0.89), thereby constituting novel anti-Ebola chemotypes as compared to the training set compounds.

Comparison to Previously Reported EBOV Inhibitors

The potencies and structures of the 14 hits identified in this study were compared to a compiled set of 60 previously published compounds with either *in vitro* or *in vivo* anti-Ebola activity.[2,7,32,33] The full list of previously known published compounds and their potencies can be found in the Supplementary File 15.

The most potent hit identified from virtual screen was vindesine (0.272 μM), a vinca alkaloid microtubule inhibitor. Previously, other vinca alkaloids were reported as sub-micromolar inhibitors of the EBOV *in vitro*. These vinca alkaloids, vinblastine (0.048 μM), vinorelbine (0.066 μM), and vincristine (0.141 μM),[7] are highly structurally similar to vindesine (Table 2). Colchicine and nocodazole, microtubule inhibitors that are structurally distinct from the vinca alkaloids, were also previously reported as sub-micromolar inhibitors. The identification of vindesine as one of the most potent hits identified to date highlights the robustness of the developed QSAR models, as well as the efficacy of this class of compounds and compounds with the same associated MOA as viable anti-Ebola compounds.

The second most potent hit identified from the virtual screen was BIX-01294 (0.97 μM). This compound is among most potent reported anti-Ebola compounds. Moreover, BIX-01294 is structurally dissimilar from other previously reported compounds (Table 2) and has a unique primary MOA (G9a histone methyltransferase inhibition).[34] In contrast to vindesine, the identification of BIX-01294 demonstrates the ability to QSAR-based virtual screening to retrieve structurally novel chemotypes.

The next most potent series of hits includes afimoxifene (1.36 μM), tetrandrine (2.16 μM), NVP-ADW742 (3.05 μM), and endoxifen (3.05 μM). Afimoxifene and endoxifen are metabolites of tamoxifen (Table 2), which was previously reported as a sub-micromolar Ebola inhibitor [7]. Likewise, tetrandrine is structurally similar to cepharanthine (Table 2),[7] as both are isolated from the same plant genus. NVP-ADW742, on the other hand, is structurally dissimilar from any previously reported compound (Table 2). However, additional tyrosine kinase inhibitors have shown efficacy against the EBOV with a range of potencies *in vitro*, such as sunitinib (1.91 μM) and nilotinib (24.3 μM).[7]

The remaining hits, *i.e.*, ZINC91973695, deptropine, GANT 61, ZINC67869167, Hh-Ag1.5, cediranib, ebastine, osanetant, have potencies ranging from 6.09 μM – 9.65 μM (Table 1). Each of these hits is structurally unique with respect to previously published compounds

(Table 2). In addition to being structurally novel among EBOV inhibitors, ZINC91973695 and ZINC67869167 have no previously reported bioactivities.

DISCUSSION

The power of virtual screening is its ability to quickly process millions of compounds and prioritize a small set of highly confident predictions for experimental confirmation. This approach not only saves time and cost as compared to the experimental high throughput screening, but also may lead to the evaluation of additional approved drugs that could be missed in the physical compound screening library. A combination of virtual screening with experimental confirmation is especially useful for challenging assays due to high biosafety requirements, limited patient samples, expensive reagents, or difficult formats (small animal or 3D cell culture). In this study, we prioritized 102 compounds from an *in silico* library of ~17 million compounds for testing in the EBOV entry assay using QSAR modeling and virtual screening. Fourteen of these hits were experimentally confirmed, including 5 selected hits against live-EBOV infections, and their anti-Ebola mechanisms of action were determined using.

The EBOV entry process has been extensively studied. Viral envelope glycoproteins attach to the surface of host cell, and the virus enters through micropinocytosis and endocytosis. Although a cell surface receptor and a few other components are still not clear, several key host factors including cathepsin B/L in the endosome[14] and Niemann Pick C1 protein (NPC1) in the lysosome have been reported as regulators of EBOV entry.[15,16] The chemical biology and anti-Ebola MOAs of the 14 experimentally validated hits were evaluated for interactions with both host and viral targets.

In addition to discovering compounds with unique scaffolds, we also uncovered the anti-Ebola MOAs of these compounds. We have found that the antihistamines ebastine and deptropine inhibited Ebola entry through negatively regulated lysosome function and blocking cathepsin L activity. We also found that osanetant, an anti-psychotic, induces the enlargement of lysosomes and impairs lysosomal function. Additionally, BIX-01294 showed sub-micromolar activity to inhibit EBOV entry. Our LysoTracker dye staining data indicated that BIX-01294 may block EBOV entry through a blockade of lysosome function in host cells. BIX-01294 is a G9a histone methyltransferase (HMTase) inhibitor.[34] HMTases have not been implicated in EBOV entry or replication. Additional chemical biology experiments to test the importance of HMTases in EBOV entry should be performed. Two hedgehog-signaling pathway modulators, Hh-Ag1.5 and GANT61,[35] showed moderate anti-Ebola activity. Our results revealed multiple mechanisms of action involved in the inhibition of EBOV entry by GANT61. GANT61 caused enlargement of lysosomes and inhibited both cathepsin L and cathepsin B, which are known to impair EBOV entry. Two hits from the QSAR-based screen, ZINC91973695 and ZINC67869167, have no previously reported bioactivities. Results of our chemical biology evaluations showed that both compounds induced enlargement of lysosomes, which may implicate the blockage of lysosomal function as a mechanism of action for these two compounds with novel anti-Ebola scaffolds.

A few analogs of previously reported Ebola entry inhibitors or compounds with the same of mechanisms of action were also identified. The most potent hit from our screen (and one of the most potent reported EBOV inhibitors) was vindesine, a vinca alkaloid microtubule inhibitor. Indeed, the vinca alkaloid microtubule inhibitors vinblastine, vincristine, and vinorelbine were also potent hits in the original screen.[31] Likewise, though afimoxifene and endoxifen are novel hits, Selective Estrogen Receptor Modulators (SERMs) have been shown in several studies to have anti-Ebola activity.[31,36,37] The same is true for the receptor tyrosine kinase (RTK) inhibitors cediranib and NVP-ADW742, as sunitinib has been previously reported to have anti-Ebola activity.[31] Last, tetrandrine, a calcium-ion channel blocker, was reported[38] as potent anti-Ebola inhibitor during the course of our study. Thus, these results demonstrate the ability of our QSAR models to reliably retrieve compounds with anti-Ebola activities and confirm the reproducibility of the VLP-assay.

CONCLUSIONS

Our study is the first case of QSAR-driven experimental discovery of novel anti-Ebola agents with limited host cell toxicity. Robust and predictive QSAR models for both anti-Ebola activity and host cytotoxicity were developed and used for virtual screening of ~17 million compounds in order to identify Ebola inhibitors with high therapeutic windows (selectivity indices). Ultimately, 102 VS hits were tested in both Ebola VLP and ATP content cytotoxicity assays; 14 of these hits had $IC_{50} < 10$ μM and $SI > 10$, which is comparable to the measured potencies of several previously reported compounds. The two most potent hits in the screen were vindesine, a vinca alkaloid microtubule inhibitor, and BIX-01294, an HMTase inhibitor (Table 2). In a live-EBOV assay, vindesine had an IC_{50} of 0.34 μM. Several of the hits were SERMs and RTKs, which have MOAs known to be related to anti-Ebola activity. We investigated the previously uncharacterized MOAs for anti-Ebola activity of several hits, including both host factors and direct Ebola VLP interactions. Two compounds, ZINC91973695 and ZINC67869167, represent novel chemotypes and can be considered as leads for future anti-Ebola chemical optimization.

In addition to the identification of these compounds, this study demonstrates that FDA-approved drugs, such as vindesine, and compounds that have not yet passed clinical trials for their primary indications, like cediranib, can be repurposed as antivirals. Such compounds are of particular interest, as they may have the potential, pending additional pre-clinical evaluation, to be granted orphan drug status in the United States for EBOV disease. The integrated computational and experimental strategy employed in this study represents an advancement for the rapid discovery of Ebola therapeutics.

EXPERIMENTAL SECTION

Computational

Data Collection, Curation, and Classification

Antiviral Data: Prior to this work, the Ebola VLP was prepared at the Icahn School of Medicine at Mount Sinai, and VLP-based qHTS screening campaigns were performed at the National Center for Advancing Translational Sciences (NCATS) at the National Institutes of

Health (NIH).[31] The results of 4 qHTS screening campaigns (2 primary and 2 confirmatory) were extracted from PubChem (AIDs 1117318, 1117313, 1117312, and 1117308).[39,40] These data are available in Supplementary Files 1–4.

Each of the four screens has three readouts, including a blue, green, and ratio (blue/green) channel. The blue channel analyzes the efficacy of the compound at inhibiting VLP entry activity in the host cell. The green channel indicates the healthy and viable cells that loaded with CCF2-AM. The ratio channel screen measures the ratio of blue/green spectra. The beta-lactamase in the VLP hydrolyzes the CCF2-AM dye used in the assay to give a blue fluorescence spectrum. An effective inhibitor will prohibit the beta-lactamase in the VLP from hydrolyzing CCF2-AM, resulting in reduction of the intensity of the blue fluorescence spectrum. A low blue emission spectrum indicates that the compound is inhibitory, while a high green emission spectrum reflects the absence of host cytotoxicity. A simplified schema of the assay is depicted in Figure 6.

In total, 3121 compounds were tested. These data were then curated according to our well-established protocols.[41–43] Briefly, mixtures, inorganics, and organometallics were removed. Additionally, replicate compounds were identified and sets were removed if screening results conflicted; if the results were concordant, then one representative compound was selected. After curation, 3104 unique compounds remained.

Cytotoxicity Data: Host cell cytotoxicity data for a subset of compounds tested for anti-EBOV activity were obtained from the researchers at the NCATS. Compounds were tested for host cytotoxicity potential in HeLa and HEK cell lines. In total, 171 unique compounds were tested in HeLa cell line, and 156 unique compounds were tested in HEK cell line. All 156 compounds tested in the HEK cell line were also tested in the HeLa cell line. Data curation was performed as above, and one organometallic was removed, leaving 170 and 155 compounds for consideration from the HeLa and HEK cell lines, respectively. These data are available in Supplementary Files 5–6.

Determination of Antiviral Activity: Only compounds with dose-response curve classes[44] of 1.1, 1.2, 2.1, 2.2, and 4 were considered for potential inclusion into the QSAR model training set. In order to comprehensively characterize the results of the screens, two separate protocols were used to classify "active" and "inactive" compounds for subsequent QSAR modeling. In the first protocol (P1), a compound was classified as "active" if and only if the compound had an $AC_{50} < 10$ μM and Maximum Inhibition $\geq 70\%$ in both a primary and confirmatory screen. Similarly, an "inactive" compound had an $AC_{50} \geq 10$ μM and Maximum Inhibition $< 70\%$ in both a primary and confirmatory screen.

In the second protocol (P2), an "activity" score was calculated for each compound, *j*, according to the following equation

$$Activity\ score(j) = 50 \times \left[\frac{\left(\max\left(AC_{50}\right) - AC_{50}(j)\right)}{\left(\max\left(AC_{50}\right) - \min\left(AC_{50}\right)\right)}\right] + 50 \times \left[1 - \frac{(\max(\text{efficacy}) - \text{efficacy}(j))}{(\max(\text{efficacy}) - \min(\text{efficacy}))}\right]$$

where *activity score(j)* is the relative activity of a specific compound; max(AC_{50}) and min(AC_{50}) are the maximum and minimum AC_{50} in the screen, respectively, and AC_{50}(j) is the AC_{50} of a specific compound; max(efficacy) and min(efficacy) are the maximum and minimum efficacies in the screen, respectively, and efficacy(j) is the efficacy of a specific compound. If a compound had an activity score ≥70 in either primary <u>or</u> confirmatory screen, the compound was classified as "active". Similarly, if a compound had an activity score < 70 in either a primary <u>or</u> confirmatory screen, the compound was classified as "inactive".

Determination of Cytotoxicity: For both the HeLa and HEK cell lines, a compound was considered "toxic" if the associated pAC_{50} > 4.0 (AC_{50} < 100 μM); whereas, a compound was considered "non-toxic" if the pAC_{50} ≤4.0 (AC_{50} ≥100 μM) or the curve class was 4, indicating no response. Only compounds with dose-response curve classes[44] of 1.1, 1.2, 2.1, 2.2, and 4 were considered.

Antiviral Training Set Balancing: In both protocols (P1 and P2), the data were imbalanced towards the inactive class. Thus, in order to balance the active and inactive classes in a 1:1 ratio, the inactive class was down-sampled.[45] Fifty percent of the corresponding inactives with the highest Tanimoto similarity[46], *i.e.*, the most similar inactives to the compounds from active class based on MACCS keys fingerprint,[47] were chosen, and the remaining 50% of the corresponding inactives were randomly selected. Important to note that all rationally chosen inactives had different nearest neighbors among actives.[48] For P1 and P2, a total of 166 compounds (83 active and 83 inactive) and 1224 compounds (612 active and 612 inactive) formed the respective training sets. These compounds are available in Supplementary Files 7–8.

Cytotoxicity Training Set Balancing: The "toxic" and "non-toxic" classes were relatively balanced; thus, no down-sampling of the larger class was required. For HeLa and HEK cell lines, a total of 170 compounds (90 toxic and 80 non-toxic) and 155 compounds (83 toxic and 72 non-toxic) formed the respective training sets. These compounds are available in Supplementary Files 9–10.

Modelability Index (MODI): The MODelability Index (MODI) estimates the likelihood of obtaining predictive QSAR models for a binary data set of compounds.[49] MODI is defined as a weighted ratio of the number of nearest-neighbor pairs of compounds in descriptor space with the same activity class versus the total number of pairs. MODI threshold of 0.65 was previously found to separate the modelable from non-modelable data sets.[49] MODI was calculated for all antiviral and cytotoxicity datasets prior to QSAR modeling in the present study as described earlier.[49]

Computational Methods

QSAR Model Generation and Validation—Three separate packages, Chembench,[50,51] HiT QSAR,[52] and GUSAR,[53] were employed for consensus classification modeling of both antiviral activity (P1 and P2) and host cytotoxicity (HeLa and HEK). QSAR models built on Chembench used Dragon 6.0 descriptors[54] and the random forest[55] machine-learning

algorithm. For models built with HiT QSAR, Simplex Representation of Molecular Structure (SiRMS) descriptors[56] and random forest (RF) were used. GUSAR models utilized a combination of Multilevel Neighborhoods of Atoms (MNA) and Quantitative Neighborhoods of Atoms (QNA) descriptors[57] and a radial-basis function with self-consistent regression (RBF-SCR) as the machine-learning algorithm.[53] We have followed best practices of QSAR modeling developed earlier by our group. All models were rigorously validated using five-fold external cross validation.[9] Y-randomization was performed for all models.[9] Models were statistically evaluated according to, sensitivity (SE), specificity (SP), correct classification rate (CCR), positive predictive value (PPV), and negative predictive value (NPV). These statistical metrics are calculated by the equations 1–5, respectively.

$$SE = \frac{TP}{TP + FN} \quad (1)$$

$$SP = \frac{TN}{TN + FP} \quad (2)$$

$$CCR = \frac{SE + SP}{2} \quad (3)$$

$$PPV = \frac{TP}{TP + FP} \quad (4)$$

$$NPV = \frac{TN}{TN + FN} \quad (5)$$

Here, TP and TN represent the number of true positives (correct classifications of actives), and true negatives (correct classifications of inactives), respectively; whereas, FP and FN represent the number of false positives (incorrect classifications of actives) and false negatives (incorrect classifications of inactives), respectively.

Virtual Screening: Two *in silico* libraries, the ZINC drug-like library[23] and previously untested drugs and experimental compounds from the NCATS Chemical Genomics Center Pharmaceutical Collection (NPC), totaling ~17 million compounds after curation (see above), were virtually screened using the developed QSAR models of antiviral activity and host cytotoxicity. A model was deemed acceptable for virtual screening if and only if the CCR, SE, SP, PPV, and NPV were all above 0.60, and no associated Y-randomized model

had a CCR above 0.60. An applicability domain (AD) was used for all models. Consensus prediction was utilized, meaning that for a compound to be considered a virtual screening "hit", it must be within the AD of each model and be predicted as "active" and "non-toxic" in *all* developed QSAR models of antiviral activity and host cytotoxicity, respectively (Figure 2). Once virtual screening "hits" were experimentally validated, the Sequential Agglomerative Hierarchical Nonoverlapping (SAHN) method implemented in the ISIDA/Cluster program[58] was used to probe the uniqueness of hit chemotypes and to identify the most structurally similar compounds in the training set and in the published literature.

Experimental Methods

Ebola VLPs containing a beta-lactamase-fused VP40 and GP were prepared in Dr. García-Sastre's lab, as previously described.[6] LiveBLAzer FRET–B/G Loading Kit and CCF2-AM, Dulbecco's modified Eagle's medium (DMEM), and Opti-MEM reduced serum medium were purchased from Life Technologies (Carlsbad, CA, USA). An ATP content cell viability assay kit was purchased from Promega (Madison, WI, USA). 1536-well polystyrene plates were purchased from Greiner Bio-One (Monroe, NC, USA). Compounds were purchased from Sigma-Aldrich (St. Louis, MO, USA), Santa Cruz (Dallas, TX, USA), ChemBridge Corporation (San Diego, CA, USA), Enamine Ltd (Kiev, Ukraine), Maybridge Chemical Company (Altrincham, United Kingdom), Vitas-M Laboratory (Champaign, IL, USA), Ambinter (Orléans, France) and AKos Consulting & Solutions Deutschland GmbH (Steinen-Schlächtenhaus, Germany) at the highest available purity. All of the compounds were dissolved as a 10 mM stock solution in dimethyl sulfoxide (DMSO) and diluted in DMSO at a 1:3 dilution to generate eleven concentrations in 384-well plates, followed by reformatting into one 1536-well compound plate for high throughput screening.

Materials—All commercially available reagents, compounds, and solvents were purchased and used without further purification. Column chromatography on silica gel was performed on RediSep column using the Teledyne Isco CombiFlash Rf system. Preparative purification was performed on a Waters semi-preparative HPLC. The column used was a Phenomenex Luna C18 (5 micron, 30 × 75 mm) at a flow rate of 45 mL/min. The mobile phase consisted of acetonitrile and water (each containing 0.1% trifluoroacetic acid). A gradient of 10% to 50% acetonitrile over 8 minutes was used during the purification. Fraction collection was triggered by UV detection (220 nm).

[1]H spectra were recorded using an INOVA 400 MHz spectrometer (Varian). Samples were analyzed on an Agilent 1200 series LC/MS using a Zorbax Eclipse XDB-C18 reverse phase (5 micron, 4.6 × 150 mm) column and a flow rate of 1.1 mL/min. The mobile phase was a mixture of acetonitrile and H_2O each containing 0.05% trifluoroacetic acid. LC Method A: a gradient of 4% to 100% acetonitrile over 7 minutes was used during analytical analysis. LC Method B: a gradient of 4% to 100% acetonitrile over 3 minutes was used during analytical analysis. High resolution mass spectrometry was recorded on Agilent 6210 Time-of-Flight LC/MS system. Spectra of compounds are listed in Supplementary File 18.

Cell culture methods—HeLa and HEK293 cells were purchased from the American Type Culture Collection (ATCC, Manassas, VA, USA). The cells were cultured in DMEM

supplemented with 10% fetal bovine serum (FBS, GE healthcare, Piscataway, NJ, USA) and 100 U/mL of penicillin and 100 μg/mL of streptomycin (Life Technologies, Carlsbad, CA, USA) at 37 °C in a humidified atmosphere with 5% CO2. Cells were passaged at 90% confluency.

Ebola VLP beta-lactamase assay for HTS in 1536-well plates—A chemical biology screening campaign was performed. Ebola VLP assay was conducted as previously described.[7] Briefly, HeLa cells were seeded at 750 cells/well in 3 μL of assay medium (DMEM+10% FBS) in 1536-well assay plates. Compounds were prepared in a 1536-well compound plate, and 23 nL of each compound was transferred into 1536-well assay plate using an NX-TR pintool station (WAKO Scientific Solutions, San Diego, CA, USA). After 1 h incubation at 37 °C with 5% CO_2, 1 μL/well of VLP solution was added to the assay plates using a BioRapTR FRD dispenser. The plates were then spinoculated, followed by incubation at 37 °C with 5% CO_2 for 4.5 h. 1 μL CCF2-AM beta-lactamase substrate was added in to each well, and the plates were incubated for 2 h at room temperature. The assay was detected at dual fluorescence intensities (Ex1= 405±20, Em1= 460±20, and Ex2= 405±20, Em2= 530±20 nm) using EnVision plate reader (PerkinElmer, Boston, MA, USA).

Cell viability assay with the ATP content assay kit—The cell viability assay was performed as previously described.[7] Briefly, HeLa and HEK293 cells were plated at 750 cells/well in 3 μL in 1536-well assay plates, followed by the addition of tested compounds at 23 nL/well. After a 4.5 h incubation at 37 °C and 5% CO_2, cell viability was measured by adding 3 μL of ATP content assay mixture to each well. Luminescence values were obtained using a ViewLux plate reader (PerkinElmer, Boston, MA, USA).

Ebola live virus assays—Vero E6 cells were plated in the 96-well plate (black with optical bottom). Briefly, serial dilutions of 5 drugs (diluted in DMEM 2% FBS starting at 10 μM) and DMSO as control, were added to the wells, and incubated for 1 h at 37 °C with 5% CO_2. The cells were infected with EBOV/Mayinga-eGFP at a MOI of 0.1 TCID50/cell. The assay was run in triplicate at a biosafety level-4 (BSL-4) facility. The fluorescence was read 72 h after infection using a BioTek Synergy HT.

Filipin staining and LysoTracker-red staining—The assays were performed as previously described.[59] Fibroblast cells were plated at 1,000 cells/well in 4 μL of assay medium (DMEM + 10% FBS) in 1536-well assay plates and incubated overnight at 37 °C and 5% CO_2. Compounds were added to the assay plate at 23 nL/well. After 24 h incubation at 37 °C and 5% CO_2, 2 μl/well of 50 ng/ml filipin or 0.5 μM LysoTracker Red DND-99 was added to the plate. After 1 hr. incubation at 37 °C and 5% CO_2, the plates were washed twice. The fluorescence intensities were then read with a fluorescence plate reader (GE Healthcare, Chicago, Illinois, USA). U18666A [3-β-(2-[diethylamino]ethoxy)-androst-5-en-17-one, monohydrochloride] was used as the positive control.[60]

Cathepsin B/L assay—Cathepsin B/L assays were performed as previously described.[8] Briefly, recombinant 5 ng of cathepsin B, or cathepsin L were added into each well in 384-well plate. Indicated drugs were added into the recombinant enzymes, followed by initiation of the reaction by addition of fluorescent substrate. The activity measurements were done

using Tecan plate reader (Tecan US, Inc., Morrisville, NC, USA). Cathepsin L inhibitor and ED64 were used as positive controls.[8]

Thermal shift binding assay with Ebola VLP—The thermal shift binding assay was performed as previously described.[17] Ebola VLPs were pre-incubated with indicated drugs for 10 min at room temperature. The mixture was then subsequently heated at 49 °C for 3 min, followed by centrifugation at 13, 000 × g at 4 °C for 20 min. The supernatant was collected and denatured by heating at 75 °C for 10 min in the presence of SDS loading buffer (Life Technologies, Carlsbad, California, United States). The samples were separated by SDS-PAGE gel electrophoresis and detected by anti-beta-lactamase antibodies (Life Technologies, Carlsbad, California, United States).

Data analysis and statistics

Half maximal inhibitory concentration (IC_{50}) values of compound activity data were calculated using Prism software (GraphPad Software, Inc. San Diego, CA). All values were expressed as the mean ± SEM (n ≥3).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

Abbreviations used

AD	Applicability Domain
Bla	Beta-Lactamase
BSL	Biosafety level
CCR	Correct Classification Rate
EBOV	Ebola Virus
FDA	Food and Drug Administration
FN	False Negatives
FP	False Positives
FRET	Fluorescence Resonance Energy Transfer

MLT	Machine Learning Technique
MNA	Multilevel Neighborhoods of Atoms
MOA	Mechanism of Action
MODI	Modelability Index
NCATS	National Center for Advancing Translational Sciences
NIH	National Institutes of Health
NPC1	Niemann Pick C1
NPV	Negative Predictive Value
PAINS	Pan-Assay INterference compoundS
PPV	Positive Predictive Value
QNA	Quantitative Neighborhoods of Atoms
RF	Random Forest QSAR, Quantitative Structure-Activity Relationship
RBF	SCR, Radial-Basis Function with Self-Consistent Regression
RTK	Receptor Tyrosine Kinase
SAHN	Sequential Agglomerative Hierarchical Nonoverlapping
SE	Sensitivity
SERMs	Selective Estrogen Receptor Modulators
SI	Selectivity Index
SiRMS	Simplex Representation of Molecular Structure
SP	Specificity
Tc	Tanimoto coefficient
TN	True Negatives
TP	True Positives
VLP	Virus-like Particle
VS	Virtual Screening

REFERENCES

(1). Green A Ebola Outbreak in the DR Congo. The Lancet. 2017.

(2). Warren TK; Jordan R; Lo MK; Ray AS; Mackman RL; Soloveva V; Siegel D; Perron M; Bannister R; Hui HC; Larson N; Strickley R; Wells J; Stuthman KS; Van Tongeren SA; Garza NL; Donnelly G; Shurtleff AC; Retterer CJ; Gharaibeh D; Zamani R; Kenny T; Eaton BP; Grimes E; Welch LS; Gomba L; Wilhelmsen CL; Nichols DK; Nuss JE; Nagle ER; Kugelman JR; Palacios

G; Doerffler E; Neville S; Carra E; Clarke MO; Zhang L; Lew W; Ross B; Wang Q; Chun K; Wolfe L; Babusis D; Park Y; Stray KM; Trancheva I; Feng JY; Barauskas O; Xu Y; Wong P; Braun MR; Flint M; McMullan LK; Chen S-S; Fearns R; Swaminathan S; Mayers DL; Spiropoulou CF; Lee WA; Nichol ST; Cihlar T; Bavari S Therapeutic Efficacy of the Small Molecule GS-5734 against Ebola Virus in Rhesus Monkeys. Nature 2016, 531 (7594), 381–385. [PubMed: 26934220]

(3). Smither SJ; Eastaugh LS; Steward JA; Nelson M; Lenk RP; Lever MS Post-Exposure Efficacy of Oral T-705 (Favipiravir) against Inhalational Ebola Virus Infection in a Mouse Model. Antiviral Res. 2014, 104, 153–155. [PubMed: 24462697]

(4). McMullan LK; Flint M; Dyall J; Albariño C; Olinger GG; Foster S; Sethna P; Hensley LE; Nichol ST; Lanier ER; Spiropoulou CF The Lipid Moiety of Brincidofovir Is Required for in Vitro Antiviral Activity against Ebola Virus. Antiviral Res. 2016, 125, 71–78. [PubMed: 26526586]

(5). Burd EM Ebola Virus: A Clear and Present Danger. J. Clin. Microbiol 2015, 53 (1), 4–8. [PubMed: 25392362]

(6). Tscherne DM; Manicassamy B; García-Sastre A An Enzymatic Virus-like Particle Assay for Sensitive Detection of Virus Entry. J. Virol. Methods 2010, 163 (2), 336–343. [PubMed: 19879300]

(7). Kouznetsova J; Sun W; Martínez-Romero C; Tawa G; Shinn P; Chen CZ; Schimmer A; Sanderson P; McKew JC; Zheng W; García-Sastre A Identification of 53 Compounds That Block Ebola Virus-like Particle Entry via a Repurposing Screen of Approved Drugs. Emerg. Microbes Infect 2014, 3 (12), e84. [PubMed: 26038505]

(8). Sun W; He S; Martínez-Romero C; Kouznetsova J; Tawa G; Xu M; Shinn P; Fisher EG; Long Y; Motabar O; Yang S; Sanderson PE; Williamson PR; García-Sastre A; Qiu X; Zheng W Synergistic Drug Combination Effectively Blocks Ebola Virus Infection. Antiviral Res. 2017, 137, 165–172. [PubMed: 27890675]

(9). Tropsha A Best Practices for QSAR Model Development, Validation, and Exploitation. Mol. Inform 2010, 29, 476–488. [PubMed: 27463326]

(10). Ekins S; Freundlich JS; Clark AM; Anantpadma M; Davey RA; Madrid P Machine Learning Models Identify Molecules Active against the Ebola Virus in Vitro. F1000Research 2015, 4.

(11). Ekins S; Lingerfelt MA; Comer JE; Freiberg AN; Mirsalis JC; O'Loughlin K; Harutyunyan A; McFarlane C; Green CE; Madrid PB Efficacy of Tilorone Dihydrochloride against Ebola Virus Infection. Antimicrob. Agents Chemother. 2017, AAC.01711–17.

(12). Lipinski CA; Lombardo F; Dominy BW; Feeney PJ Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. Adv. Drug Deliv. Rev 2001, 46 (1–3), 3–26. [PubMed: 11259830]

(13). Southall Noel T., Jadhav Ajit, Huang Ruili, Nguyen Trung, and Y. W. Enabling the Large-Scale Analysis of Quantitative High-Throughput. In Handbook of Drug Screening, Second Edition; 2009; pp 442–463.

(14). Chandran K; Sullivan NJ; Felbor U; Whelan SP; Cunningham JM Endosomal Proteolysis of the Ebola Virus Glycoprotein Is Necessary for Infection. Science 2005, 308 (5728), 1643–1645. [PubMed: 15831716]

(15). Côté M; Misasi J; Ren T; Bruchez A; Lee K; Filone CM; Hensley L; Li Q; Ory D; Chandran K; Cunningham J Small Molecule Inhibitors Reveal Niemann-Pick C1 Is Essential for Ebola Virus Infection. Nature 2011, 477 (7364), 344–348. [PubMed: 21866101]

(16). Carette JE; Raaben M; Wong AC; Herbert AS; Obernosterer G; Mulherkar N; Kuehne AI; Kranzusch PJ; Griffin AM; Ruthel G; Cin PD; Dye JM; Whelan SP; Chandran K; Brummelkamp TR Ebola Virus Entry Requires the Cholesterol Transporter Niemann–Pick C1. Nature 2011, 477 (7364), 340–343. [PubMed: 21866103]

(17). Molina DM; Jafari R; Ignatushchenko M; Seki T; Larsson EA; Dan C; Sreekumar L; Cao Y; Nordlund P Monitoring Drug Target Engagement in Cells and Tissues Using the Cellular Thermal Shift Assay. Science (80-.). 2013, 341 (6141).

(18). Jae LT; Brummelkamp TR Emerging Intracellular Receptors for Hemorrhagic Fever Viruses. Trends Microbiol. 2015, 23 (7), 392–400. [PubMed: 26004032]

(19). Miller EH; Obernosterer G; Raaben M; Herbert AS; Deffieu MS; Krishnan A; Ndungo E; Sandesara RG; Carette JE; Kuehne AI; Ruthel G; Pfeffer SR; Dye JM; Whelan SP; Brummelkamp TR; Chandran K Ebola Virus Entry Requires the Host-Programmed Recognition of an Intracellular Receptor. EMBO J. 2012, 31 (8), 1947–1960. [PubMed: 22395071]

(20). Herbert AS; Davidson C; Kuehne AI; Bakken R; Braigen SZ; Gunn KE; Whelan SP; Brummelkamp TR; Twenhafel NA; Chandran K; Walkley SU; Dye JM Niemann-Pick C1 Is Essential for Ebolavirus Replication and Pathogenesis in Vivo. MBio 2015, 6 (3), e00565–15. [PubMed: 26015498]

(21). Jaishy B; Abel ED Lipids, Lysosomes, and Autophagy. J. Lipid Res. 2016, 57 (9), 1619–1635. [PubMed: 27330054]

(22). Vale GA Proceedings: Attractants for Controlling and Surveying Tsetse Populations. Trans. R. Soc. Trop. Med. Hyg 1974, 68 (1), 11.

(23). Sterling T; Irwin JJ ZINC 15--Ligand Discovery for Everyone. J. Chem. Inf. Model 2015, 55 (11), 2324–2337. [PubMed: 26479676]

(24). Jadhav A; Ferreira RS; Klumpp C; Mott BT; Austin CP; Inglese J; Thomas CJ; Maloney DJ; Shoichet BK; Simeonov A Quantitative Analyses of Aggregation, Autofluorescence, and Reactivity Artifacts in a Screen for Inhibitors of a Thiol Protease. J. Med. Chem 2010, 53 (1), 37–51. [PubMed: 19908840]

(25). Aldrich C; Bertozzi C; Georg GI; Kiessling L; Lindsley C; Liotta D; Merz KM; Schepartz A; Wang S The Ecstasy and Agony of Assay Interference Compounds. J. Chem. Inf. Model 2017, 57 (3), 387–390. [PubMed: 28244743]

(26). Baell JB; Holloway GA New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays. J. Med. Chem 2010, 53, 2719–2740. [PubMed: 20131845]

(27). Lagorce D; Oliveira N; Miteva MA; Villoutreix BO Pan-Assay Interference Compounds (PAINS) That May Not Be Too Painful for Chemical Biology Projects. Drug Discov. Today 2017.

(28). Capuzzi SJ; Muratov EN; Tropsha A Phantom PAINS: Problems with the Utility of Alerts for P an- A Ssay IN Terference Compound S. J. Chem. Inf. Model 2017, 57 (3), 417–427. [PubMed: 28165734]

(29). Jasial S; Hu Y; Bajorath J How Frequently Are Pan-Assay Interference Compounds Active? Large-Scale Analysis of Screening Data Reveals Diverse Activity Profiles, Low Global Hit Frequency, and Many Consistently Inactive Compounds. J. Med. Chem 2017, 60 (9), 3879–3886. [PubMed: 28421750]

(30). Canny SA; Cruz Y; Southern MR; Griffin PR PubChem Promiscuity: A Web Resource for Gathering Compound Promiscuity Data from PubChem. Bioinformatics 2012, 28 (1), 140–141. [PubMed: 22084255]

(31). Kouznetsova J; Sun W; Martínez-Romero C; Tawa G; Shinn P; Chen CZ; Schimmer A; Sanderson P; McKew JC; Zheng W; García-Sastre A Identification of 53 Compounds That Block Ebola Virus-like Particle Entry via a Repurposing Screen of Approved Drugs. Emerg. Microbes Infect 2014, 3 (12), e84. [PubMed: 26038505]

(32). Warren TK; Wells J; Panchal RG; Stuthman KS; Garza NL; Van Tongeren SA; Dong L; Retterer CJ; Eaton BP; Pegoraro G; Honnold S; Bantia S; Kotian P; Chen X; Taubenheim BR; Welch LS; Minning DM; Babu YS; Sheridan WP; Bavari S Protection against Filovirus Diseases by a Novel Broad-Spectrum Nucleoside Analogue BCX4430. Nature 2014, 508 (7496), 402–405. [PubMed: 24590073]

(33). Yermolina MV; Wang J; Caffrey M; Rong LL; Wardrop DJ Discovery, Synthesis, and Biological Evaluation of a Novel Group of Selective Inhibitors of Filoviral Entry. J. Med. Chem 2011, 54 (3), 765–781. [PubMed: 21204524]

(34). Kubicek S; O'Sullivan RJ; August EM; Hickey ER; Zhang Q; Teodoro ML; Rea S; Mechtler K; Kowalski JA; Homon CA; Kelly TA; Jenuwein T Reversal of H3K9me2 by a Small-Molecule Inhibitor for the G9a Histone Methyltransferase. Mol. Cell 2007, 25 (3), 473–481. [PubMed: 17289593]

(35). Wu F; Zhang Y; Sun B; McMahon AP; Wang Y Hedgehog Signaling: From Basic Biology to Cancer Therapy. Cell Chem. Biol 2017, 24 (3), 252–280. [PubMed: 28286127]

(36). Fan H; Du X; Zhang J; Zheng H; Lu X; Wu Q; Li H; Wang H; Shi Y; Gao G; Zhou Z; Tan D-X; Li X Selective Inhibition of Ebola Entry with Selective Estrogen Receptor Modulators by Disrupting the Endolysosomal Calcium. Sci. Rep 2017, 7, 41226. [PubMed: 28117364]

(37). Johansen LM; Brannan JM; Delos SE; Shoemaker CJ; Stossel A; Lear C; Hoffstrom BG; DeWald LE; Schornberg KL; Scully C; Lehar J; Hensley LE; White JM; Olinger GG FDA-Approved Selective Estrogen Receptor Modulators Inhibit Ebola Virus Infection. Sci. Transl. Med 2013, 5 (190), 190ra79–190ra79.

(38). Sakurai Y; Kolokoltsov AA; Chen C-C; Tidwell MW; Bauta WE; Klugbauer N; Grimm C; Wahl-Schott C; Biel M; Davey RA Two-Pore Channels Control Ebola Virus Host Cell Entry and Are Drug Targets for Disease Treatment. Science (80-.). 2015, 347 (6225), 995–998.

(39). Wang Y; Suzek T; Zhang J; Wang J; He S; Cheng T; Shoemaker BA; Gindulyte A; Bryant SH PubChem BioAssay: 2014 Update. Nucleic Acids Res. 2014, 42 (Database issue), D1075–82. [PubMed: 24198245]

(40). Kim S; Thiessen PA; Bolton EE; Chen J; Fu G; Gindulyte A; Han L; He J; He S; Shoemaker BA; Wang J; Yu B; Zhang J; Bryant SH PubChem Substance and Compound Databases. Nucleic Acids Res. 2015, 44 (D1), D1202–13. [PubMed: 26400175]

(41). Fourches D; Muratov E; Tropsha A Curation of Chemogenomics Data. Nat. Chem. Biol 2015, 11 (8), 535. [PubMed: 26196763]

(42). Fourches D; Muratov E; Tropsha A Trust, but Verify: On the Importance of Chemical Structure Curation in Cheminformatics and QSAR Modeling Research. J. Chem. Inf. Model 2010, 50 (7), 1189–1204. [PubMed: 20572635]

(43). Fourches D; Muratov E; Tropsha A Trust, but Verify II: A Practical Guide to Chemogenomics Data Curation. J. Chem. Inf. Model 2016, 56 (7), 1243–1252. [PubMed: 27280890]

(44). Wang Y; Jadhav A; Southal N; Huang R; Nguyen D-T A Grid Algorithm for High Throughput Fitting of Dose-Response Curve Data. Curr. Chem. Genomics 2010, 4, 57–66. [PubMed: 21331310]

(45). Capuzzi SJ; Politi R; Isayev O; Farag S; Tropsha A QSAR Modeling of Tox21 Challenge Stress Response and Nuclear Receptor Signaling Toxicity Assays. Front. Environ. Sci 2016, 4, 3.

(46). Tanimoto T IBM Internal Report In Armonk: IBM Corp; 1957.

(47). Accelrys. MACCS structural keys.

(48). Muratov EN; Artemenko AG; Varlamova EV; Polischuk PG; Lozitsky VP; Fedchuk AS; Lozitska RL; Gridina TL; Koroleva LS; Sil'nikov VN; Galabov AS; Makarov VA; Riabova OB; Wutzler P; Schmidtke M; Kuz'min VE Per Aspera Ad Astra: Application of Simplex QSAR Approach in Antiviral Research. Future Med. Chem 2010, 2 (7), 1205–1226. [PubMed: 21426164]

(49). Golbraikh A; Muratov E; Fourches D; Tropsha A Data Set Modelability by QSAR. J. Chem. Inf. Model 2014, 54 (1), 1–4. [PubMed: 24251851]

(50). Walker T; Grulke CM; Pozefsky D; Tropsha A Chembench: A Cheminformatics Workbench. Bioinformatics 2010, 26 (23), 3000–3001. [PubMed: 20889496]

(51). Capuzzi SJ; Kim IS-J; Lam WI; Thornton TE; Muratov EN; Pozefsky D; Tropsha A Chembench: A Publicly Accessible, Integrated Cheminformatics Portal. J. Chem. Inf. Model 2017, 57 (2), 105–108. [PubMed: 28045544]

(52). Kuz'min VE; Artemenko a G.; Muratov EN Hierarchical QSAR Technology Based on the Simplex Representation of Molecular Structure. J. Comput. Aided. Mol. Des 2008, 22 (6–7), 403–421. [PubMed: 18253701]

(53). Zakharov A; Peach ML; Sitzmann M; Nicklaus MC A New Approach to Radial Basis Function Approximation and Its Application to QSAR. J. Chem. Inf. Model 2014.

(54). Kode. DRAGON 7.0 https://chm.kode-solutions.net/products_dragon.php (accessed May 7, 2016).

(55). Breiman L Random Forests. Mach. Learn 2001, 45 (1), 5–32.

(56). Kuz'min VE; Artemenko AG; Lozytska RN; Fedtchouk AS; Lozitsky VP; Muratov EN; Mescheriakov AK Investigation of Anticancer Activity of Macrocyclic Schiff Bases by Means of 4D-QSAR Based on Simplex Representation of Molecular Structure. SAR QSAR Environ. Res 2005, 16 (3), 219–230. [PubMed: 15804810]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

(57). Zakharov AV; Lagunin AA; Filimonov DA; Poroikov VV. Quantitative Prediction of Antitarget Interaction Profiles for Chemical Compounds. Chem. Res. Toxicol 2012, 25 (11), 2378–2385. [PubMed: 23078046]

(58). Varnek A; Fourches D; Horvath D; Klimchuk O; Gaudin C; Vayer P; Solov'ev V; Hoonakker F; Tetko I; Marcou G ISIDA - Platform for Virtual Screening Based on Fragment and Pharmacophoric Descriptors. Curr. Comput. Aided-Drug Des 2008, 4 (3), 191–198.

(59). Xu M; Liu K; Swaroop M; Porter FD; Sidhu R; Firnkes S; Finkes S; Ory DS; Marugan JJ; Xiao J; Southall N; Pavan WJ; Davidson C; Walkley SU; Remaley AT; Baxa U; Sun W; McKew JC; Austin CP; Zheng W δ-Tocopherol Reduces Lipid Accumulation in Niemann-Pick Type C1 and Wolman Cholesterol Storage Disorders. J. Biol. Chem 2012, 287 (47), 39349–39360. [PubMed: 23035117]

(60). Appelqvist H; Nilsson C; Garner B; Brown AJ; Kågedal K; Ollinger K Attenuation of the Lysosomal Death Pathway by Lysosomal Cholesterol Accumulation. Am. J. Pathol 2011, 178 (2), 629–639. [PubMed: 21281795]
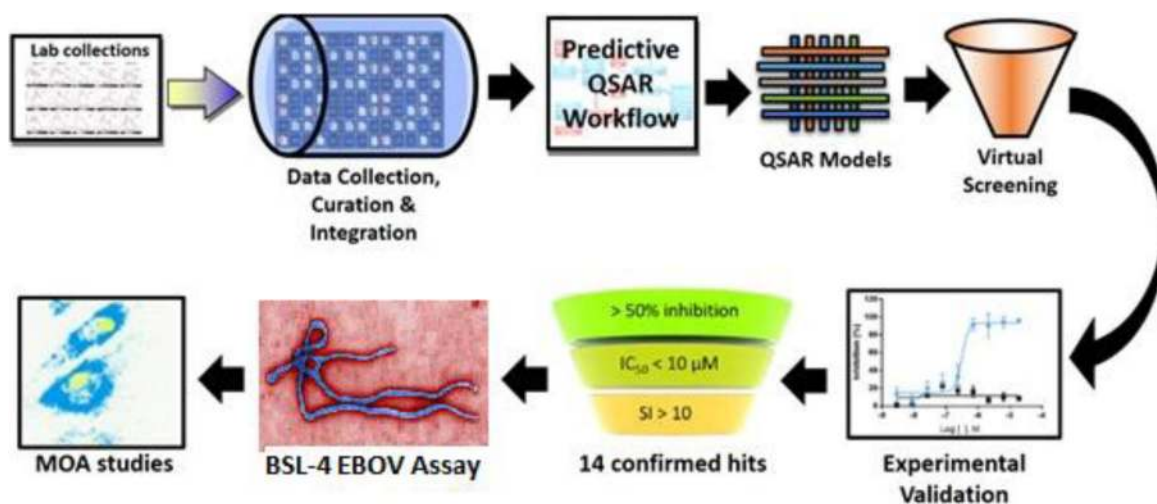
Figure 1.
Overall study design. The present study synergistically incorporates computational modeling and experimentation.
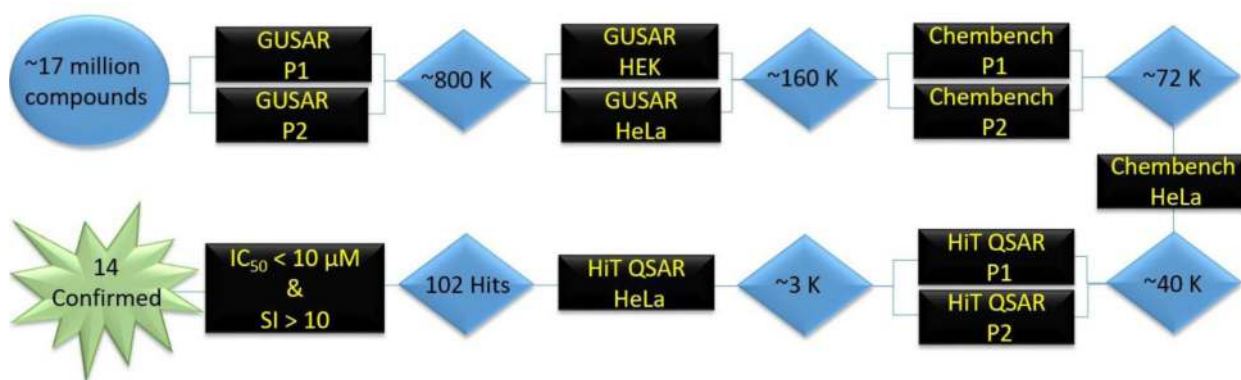
Figure 2.
Screening workflow. A virtual chemical library of ~17 million compounds was screened against a battery of antiviral (P1 and P2) and cytotoxicity (HEK and HeLa) models. Hits selected for experimental validation were predicted to be EBOV inhibitors with limited host cytotoxicity. Then computational hits were experimentally validated, then their activity was evaluated using percent inhibition, IC_{50} values, and selectivity index (SI).
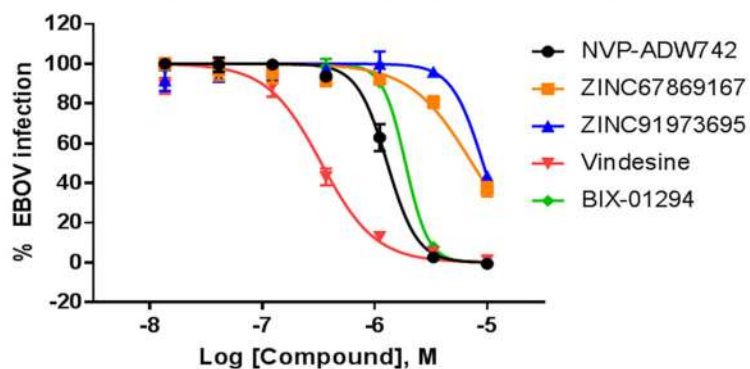
Figure 3.
Confirmatory dose-response curves against BSL-4 live-EBOV infection. Five hits were selected for screening in Vero E6 cells using a eGFP-EBOV assay. Vindesine (red) was the most potent compounds with an IC50 of 0.34 μM. The IC_{50} values of NVP-ADW742 (black), BIX-01294 (green), ZINC67869167 (orange), and ZINC91973695 (blue) were between 1 μM to 10 μM. All experiments were performed in triplicate. Data are represented as mean ± SEM.
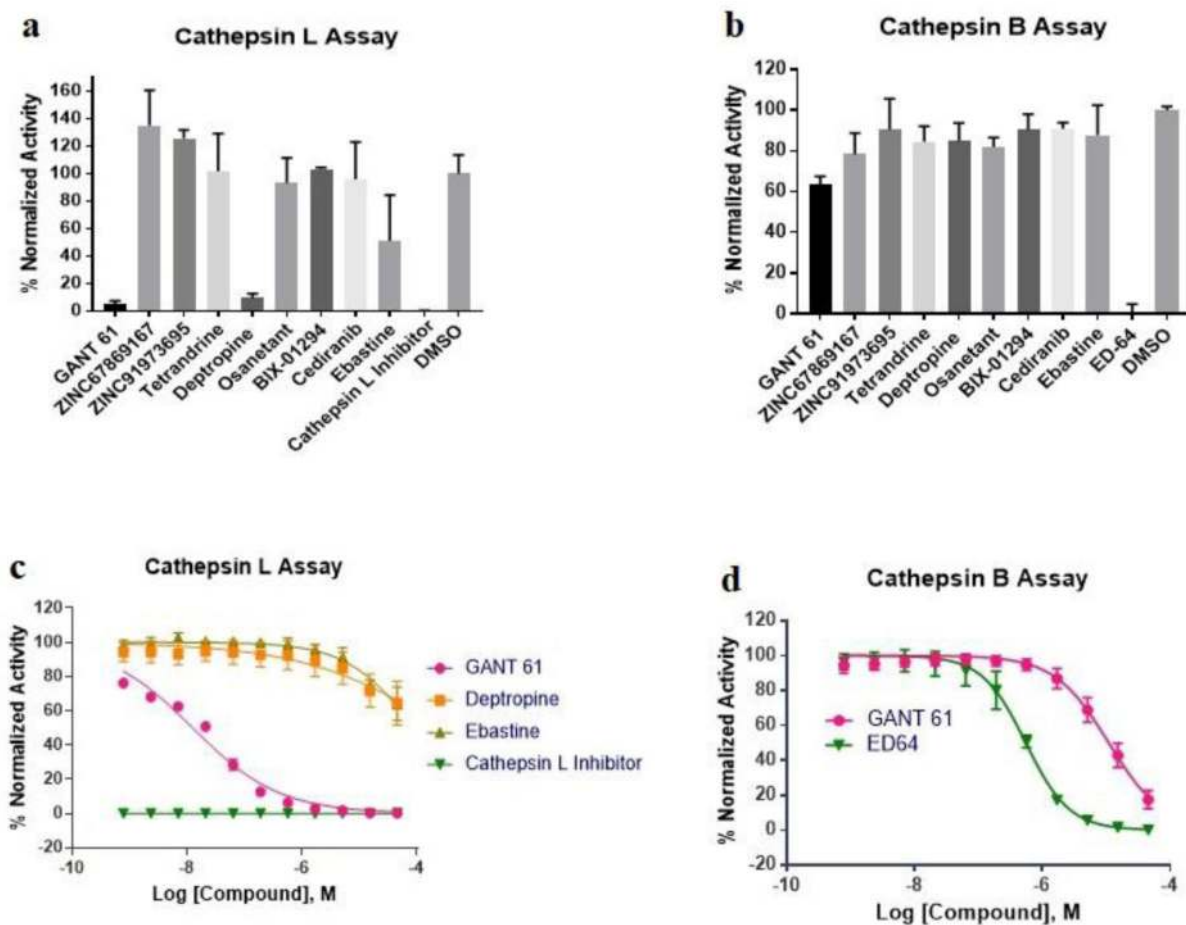
Figure 4.
Inhibition of protease activities of recombinant cathepsin L and cathepsin B by Ebola entry inhibitors. **a** and **b**. recombinant cathepsin B or cathepsin L were treated with 10 μM of GANT61, ZINC67869167, ZINC91973695, tetrandrine, deptropine, osanetant, BIX-01294, cediranib, and ebastine. **c.** Dose-response studies of GANT61, deptropine and ebastine in cathepsin L assay. **d.** Dose-response studies of GANT61 in cathepsin B assay. All experiments were performed in triplicate and data are representative of at least two independent experiments. Data are represented as mean ± SEM.
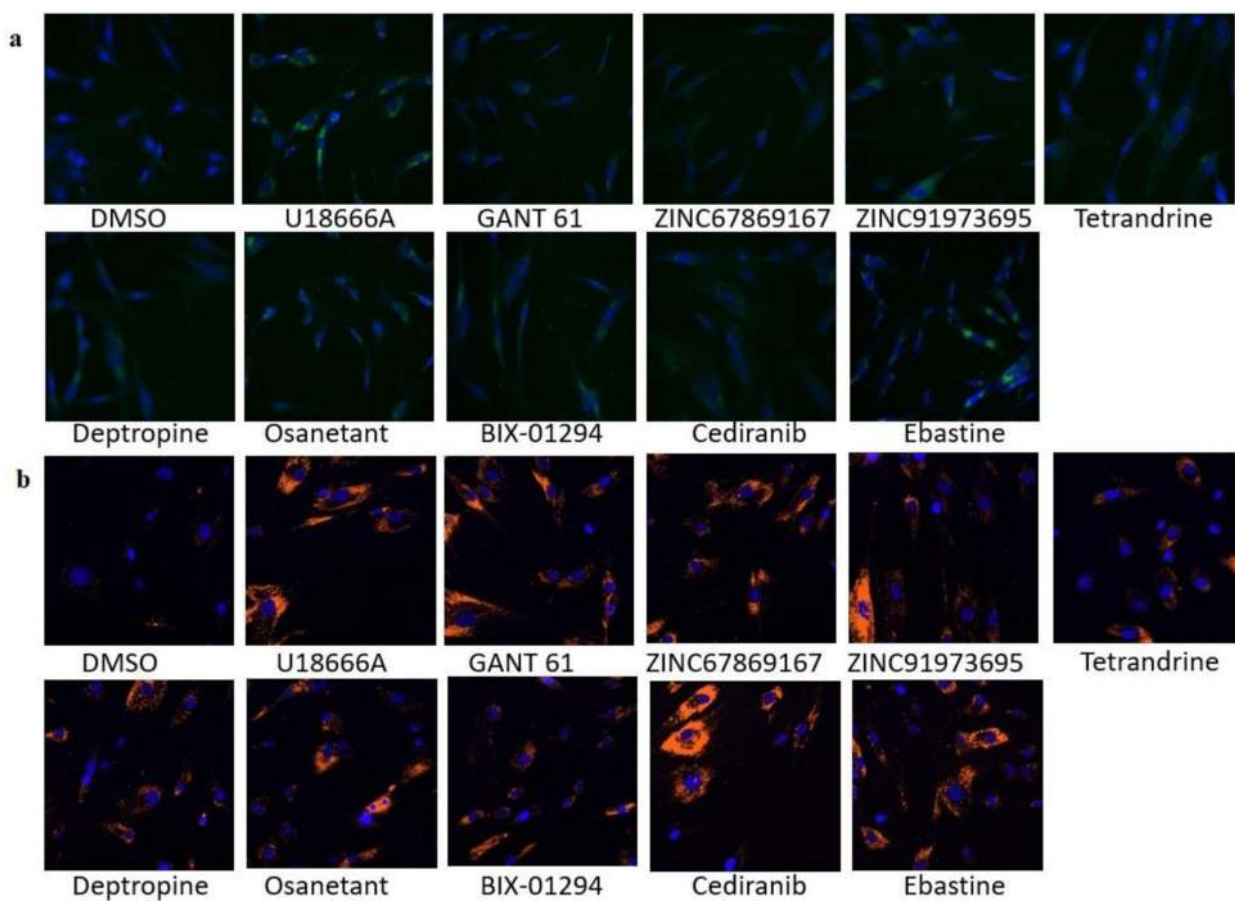
Figure 5.
Cholesterol accumulation and enlargement of lysosome induced by Ebola entry inhibitors. **a**. U18666A and ebastine increased filipin staining in fibroblasts (green: filipin; blue: nuclei). **b**. U1866A, GANT61, ZINC67869167, ZINC91973695, tetrandrine, deptropine, osanetant, BIX-01294, cediranib, and ebastine increased LysoTracker staining in fibroblasts (orange: LysoTracker red; blue: nuclei). All experiments were performed in triplicate and data are representative of at least two independent experiments. Data are represented as mean ± SEM.
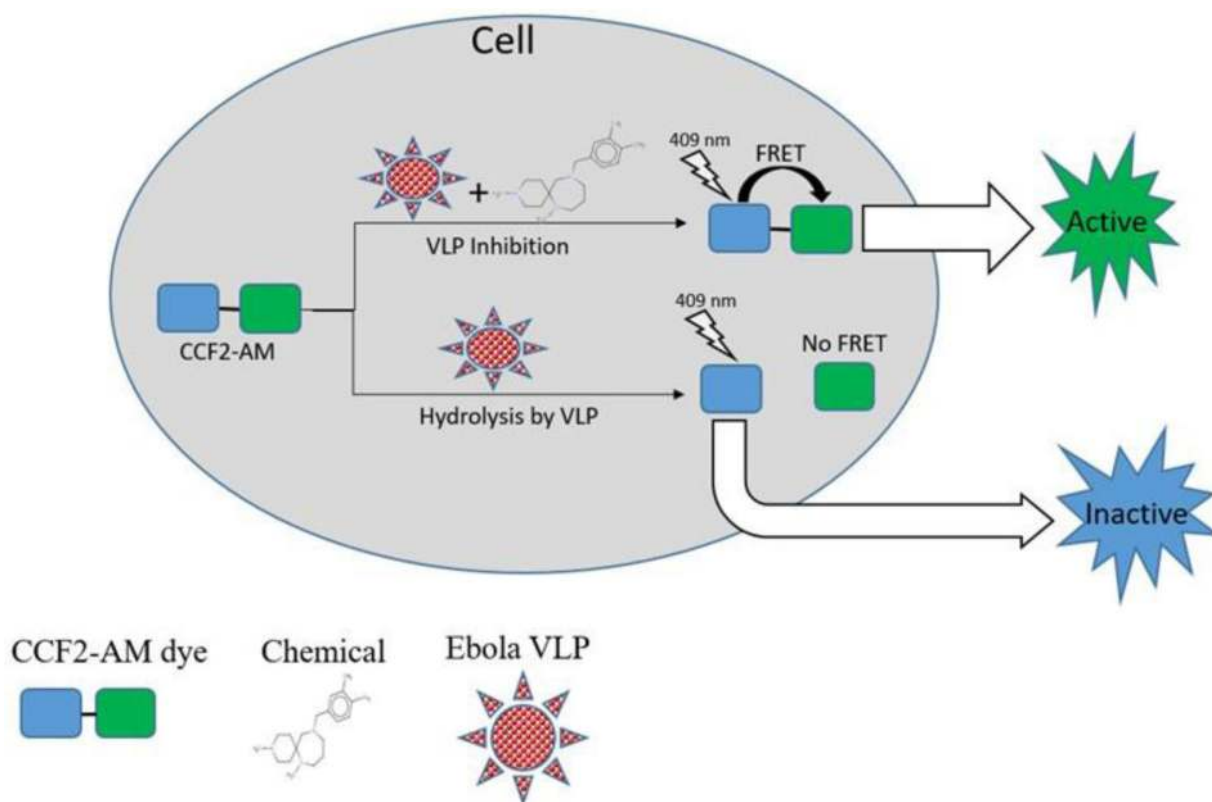
Figure 6.
Simplified schema of Ebola VLP assay. Ebola VLPs contain Ebola GP and the VP40 protein fused to a beta-lactamase (Bla) reporter. HeLa cells are loaded with the beta-lactamase substrate CCF2-AM. If the VLP enters into the cell, Bla hydrolyzes the substrate CCF2-AM, disrupting the fluorescence resonance energy transfer (FRET) in the substrate, thus causing blue fluorescence. Inhibition of the VLP by a chemical will preserve the substrate FRET, maintaining a green fluorescence. The ratio of blue/green fluorescence intensities represents the VLP activity of inside cells.
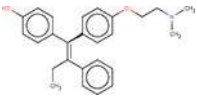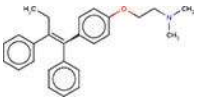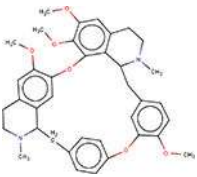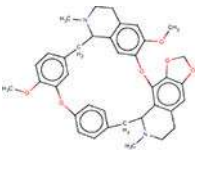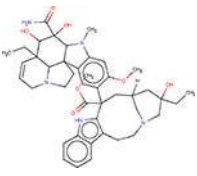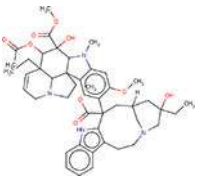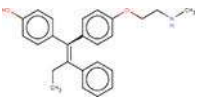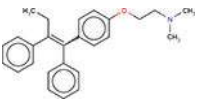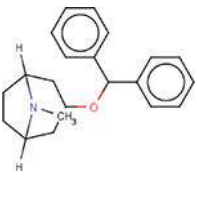
Table 1.

Experimental results for the top 14 hits. Most experimentally confirmed hits have known MOAs and therapeutic use indications.

Name	Potency, μM	Selectivity Index	Indication	MOA
Vindesine	0.272	1837	Anticancer	Microtubule Inhibitor
BIX-01294	0.966	45	Anticancer	HMTase Inhibitor
Afimoxifene	1.96	123	Anticancer	Estrogen Receptor Modulator
Tetrandrine	2.16	22	Anti-inflammation	Calcium Channel Blocker
NVP-ADW742	3.05	13	Anticancer	Tyrosine Kinase Inhibitor
Endoxifen	3.05	164	Anticancer	Estrogen Receptor Modulator
ZINC91973695	6.09	82	N/A	N/A
Deptropine	6.58	76	Antihistamine	Anticholinergic
GANT61	6.83	73	Anticancer	Hedgehog Antagonist
ZINC67869167	6.83	73	N/A	N/A
Hh-Ag1.5	7.67	65	Anticancer	Hedgehog Agonist
Cediranib	7.67	65	Anticancer	Tyrosine Kinase Inhibitor
Ebastine	9.56	51	Antihistamine	Histamine H1 Antagonist
Osanetant	9.65	52	Antipsychotic	Neurokinin 3 Receptor Antagonist

Table 2.

Structural similarity of top hits to previously published compounds. The Tanimoto coefficient (T_C) between experimentally confirmed hits and compounds in the literature was calculated using ISIDA (see Methods).
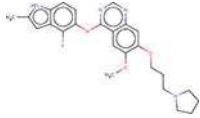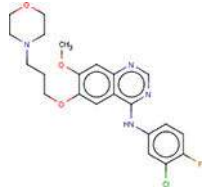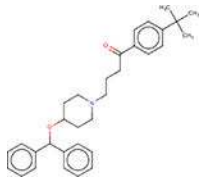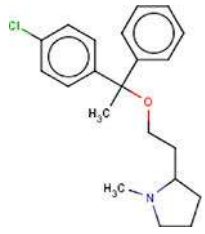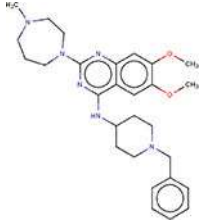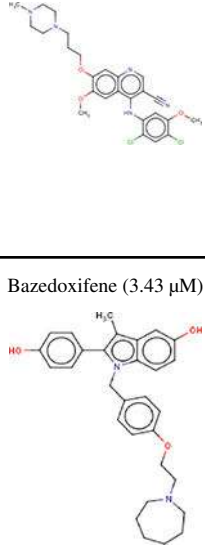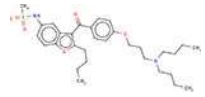
Hit (IC$_{50}$)	T$_C$	Published Compound (IC$_{50}$)
Afimoxifene (1.96 μM)	0.99	Tamoxifen (0.73 μM)
Tetrandrine (2.16 μM)	0.97	Cepharanthine (1.53 μM)
Vindesine (0.272 μM)	0.96	Vinblastine (0.048 μM)
Endoxifen (3.05 μM)	0.93	Tamoxifen (0.73 μM)
Deptropine (6.58 μM)	0.89	Benztropine (2.64 μM)
Cediranib (7.67 μM)	0.79	Gefitinib (9.68 μM)
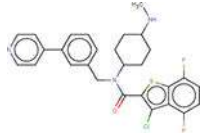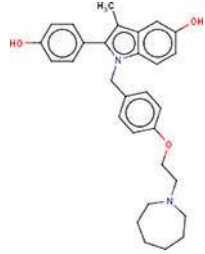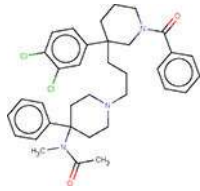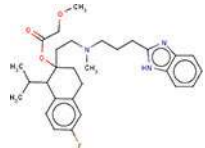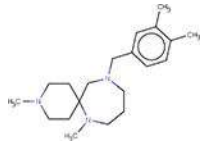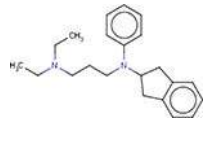
Hit (IC$_{50}$)	T$_C$	Published Compound (IC$_{50}$)
Ebastine (9.56 μM)	0.73	Clemastine (1.10 μM)
BIX-01294 (0.966 μM)	0.72	Bosutinib (3.85 μM)
NVP-ADW742 (3.05 μM)	0.72	Bazedoxifene (3.43 μM)
ZINC91973695 (6.09 μM)	0.69	Dronedarone (2.20 μM)
Hh-Ag1.5 (7.67 μM)	0.66	Bazedoxifene (3.43 μM)

Hit (IC$_{50}$)	T$_C$	Published Compound (IC$_{50}$)
Osanetant (9.65 μM)	0.71	Mibefradil (4.32 μM)
ZINC67869167 (6.83 μM)	0.71	Aprindine (7.69 μM)
GANT61 (6.83 μM)	0.47	Thioproperazine (4.32 μM)