Triterpenoids manipulate a broad range of virus-host fusion via wrapping the HR2 domain prevalent in viral envelopes

Longlong Si[1], Kun Meng[1], Zhenyu Tian[1], Ziwei Zhang[1], Veronica Soloveva[2], Jiaqi Sun[1], Haiwei Li[1], Ge Fu[1], Qing Xia[1], Sulong Xiao[1], Lihe Zhang[1], Demin Zhou[1]*

[1]State Key Laboratory of Natural and Biomimetic Drugs, School of Pharmaceutical Sciences, Peking University, 38 Xueyuan Road, Beijing 100191, China

[2]U.S. Army Medical Research Institute of Infectious Diseases, Frederick, Maryland, MD 21702, USA

Address correspondence to:

Professor Demin Zhou

School of Pharmaceutical Sciences, Peking University

#38 Xueyuan Road, Beijing 100191, China

Tel.: 86-10-8280-5857; Fax: 86-10-8280-5857

Email: deminzhou@bjmu.edu.cn

Key Words: Ebola, influenza virus, HIV, envelope protein, membrane fusion inhibitor, HR2, pentacyclic triterpenoids.

 doi:10.20944/preprints201801.0177.v1

ABSTRACT

Recent years have witnessed a breakthrough in identification of a trimer-of-hairpins motif within viral envelopes that triggers a broad range of virus-host fusion. Identifying a domain capable of controlling virus-host fusion remains a challenge due to sequence diversity, heavy glycan shielding and multiple conformations. Here, we report that HR2, a prevalent heptad repeat sequence comprising an alpha-helical coil anchored in viral membranes, is an accessible site to triterpenes, a class of widely distributed natural products. Triterpenes and their derivatives inhibit the entry of Ebola, HIV, and influenza A viruses with distinct structure-activity relationships. Specifically, triterpenoid probes, upon activation by ultraviolet light, capture the viral envelope via crosslinking the HR2 coil. Profiling the Ebola HR2 sequence using amino acid substitution, surface plasmon resonance (SPR) and nuclear magnetic resonance (NMR) spectroscopy disclosed six constitutive residues that are accessible to triterpenoids, leading to wrapping of the hydrophobic helix by triterpenoids and blocking of the HR1-HR2 interaction, which is critical in the trimer-of-hairpins formation. This finding was also observed in the envelopes of HIV and influenza A viruses and might potentially extend to a broader variety of viruses. Our findings might translate into a shared mechanism that host utilize natural product triterpenoids to antagonize membrane fusion of respective viruses, complementing the current repertoire of antiviral agents.

 doi:10.20944/preprints201801.0177.v1

INTRODUCTION

Virus-host fusion is executed by viral envelopes via constituent fusion proteins during the life cycle of enveloped viruses[1,2]. Viral fusion proteins usually consist of two subunits: a host membrane-binding subunit and a viral membrane-anchoring subunit[3]. Recent work has identified three distinct classes of viral fusion proteins, which belong to evolutionarily distant viruses and exhibit limited sequence homology, along with at least four different mechanisms by which viral fusion-competent conformational changes are triggered. In addition, viral fusion proteins contain different types of fusion peptides and vary in their reliance on accessory proteins, further increasing the diversity of fusion proteins[3]. Despite this staggering diversity, all characterized viral fusion proteins are found to mediate host-virus membrane fusion by transitioning from a variety of pre-fusion states to a membrane-embedded homotrimeric prehairpin intermediate and then to a post-fusion trimer-of-hairpins conformation, which brings the viral and host membranes into close proximity and thereby facilitates their union[3] (**Fig. 1a** and **Supplementary Video 1**). Clearly, the dynamic process of forming the trimer-of-hairpins represents a potential target for therapeutic intervention against viral infections[4]. However, identifying an exposed region as a targetable site remains a challenge due to heavy glycan shielding, sequence diversity and multiple conformations.

Triterpenes, a class of secondary plant metabolites characterized by hydrophobic pentacyclic scaffolds, are found in a wide variety of herbal medicines with the capability of preventing various pathogen infections[5]. Cell-based studies have

 doi:10.20944/preprints201801.0177.v1

confirmed that oleanolic acid (OA) and betulinic acid (BA), a lupane-type triterpene, show inhibitory activity against HIV infection; the BA analogue bevirimat (PA-457) is in the process of clinical trials[6-10]. Very recently, analogues/derivatives of OA and echinocystic acid (EA), an OA-type triterpene, were found to exert a variety of inhibitory activities, strongly related to their structures, against influenza virus and HCV infection[11-13]. Further investigations suggested that these triterpenoid leads might bind to viral fusion proteins, including hemagglutinin (HA2) of influenza, E2 of HCV and GP41 of HIV, thereby disrupting viral entry into the host cells[11,13,14]. Whether the observed antiviral aspects of triterpenoids extend to other viruses and translate into a shared mechanism that can chemically antagonize the fusion of viruses with cellular membranes remains a topic of ongoing investigation.

Here, we report triterpenoids as a class of entry inhibitors of Ebola virus, expanding the list of viruses for which triterpenoids act as entry inhibitors. More importantly, we identified HR2, the prevalent heptad repeat sequence comprising an alpha-helical coil anchored in viral membranes[15,16], as a site accessible to triterpenoids that antagonize Ebola virus-cell fusion. Furthermore, as a critical element of the trimer-of-hairpins, HR2 was identified as a common target in the triterpenoid-mediated inhibition on viral entry of influenza virus and HIV, potentially extending to an increasingly diverse array of viruses. This finding provides mechanistic insights into the triterpenoid-HR2 complex as a switch in the modulation of syncytium formation, expanding both our understanding of the evolutionary function of triterpenoids in nature and the current repertoire of antiviral agents.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 January 2018 doi:10.20944/preprints201801.0177.v1

RESULTS

Identification of triterpenoids as inhibitors by EBOV entry assay

Inspired by the key role of triterpenoids in inhibiting the entry of HIV, HCV and influenza virus and their strong structure-activity relationship, we speculated on whether the entry of Ebola viruses could also be inhibited by triterpenoids. To alleviate safety concerns, Ebola pseudoparticles (EBOVpp), i.e., Ebola glycoprotein (GP) assembled onto retroviral core particles, were used to screen a triterpenoid library consisting of hundreds of chemicals. Vesicular stomatitis virus (VSV) GP and WSN influenza envelope hemagglutinin (HA)/neuraminidase (NA) protein pseudoparticles were tested in parallel to exclude toxic and non-specific effects[11,13]. We identified a panel of triterpenoid compounds that inhibited the EBOVpp to different extents, which were strongly related to their structures, but did not inhibit VSV and WSN (**Fig. 1b**, **Supplementary Fig. 1a** and **b**). Specifically, **Y11** and **Y13**, analogues of EA and OA, respectively, conjugated with acetylated galactose, displayed remarkable potencies: the IC_{50} value for **Y11** was approximately 50 nM for all five EBOV species tested in A549 and 293T cells (**Fig. 1b**, **Supplementary Fig. 1a** and **Table 1**). Furthermore, analyses of the strong structure-activity relationship revealed that a 3-keto group in the triterpenoid backbone and an acetyl substitution in the conjugated galactose are required for maintaining high anti-EBOV activity (**Supplementary Fig. 1a**). In addition, we found that **Y11** inhibited the propagation of native EBOV (Makona) in HeLa cells (multiplicity of infection (MOI) = 1.5) and human foreskin fibroblast (HFF) cells (MOI = 20): the EC_{50} values were 0.46 and 0.99 μM, and the EC_{90} values were

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 January 2018 doi:10.20944/preprints201801.0177.v1

7.57 and 16.1 μM, respectively (**Supplementary Fig. 1c**); these values were comparable to those of the previously reported EBOV inhibitor 3.47, which was used as a control[17,18].

Identification of EBOV GP as the target by time-of-addition assays

As the viral envelope was the only difference between EBOVpp and the VSV and WSN pseudoviruses[19], we determined whether EBOV GP was the potential target of the triterpenoid lead **Y11** by a series of time-of-addition experiments[11,13,19], including pretreatment of the viruses, pretreatment of the cells, and pre-attachment, post-attachment and post-entry assays conducted in parallel, as previously reported (**Supplementary Fig. 2a**). We found that the pretreatment of EBOVpp rather than of cells with **Y11** led to significant viral entry inhibition (**Supplementary Fig. 2b**). This result is markedly different from that of E-64d, which inhibits EBOV entry into cells by targeting the host cysteine protease. Furthermore, the subsequent addition of **Y11** to the cultured virus/host complexes at a low temperature, allowing virion attachment but not membrane fusion due to the energy requirements[11], had a greater inhibitory effect than the other treatments (**Supplementary Fig. 2b**). This effect suggested the possibility that EBOV GP is the target and that virion fusion is the primary targeting stage.

Identification of the envelope subunit HR2 as the target site of triterpenoid probes

Next, to directly capture the target site of the triterpenoid lead compounds, **Y18**, a triterpenoid probe containing dual-functional chemical moieties, namely, a diazirine[20] and an alkyne[21] substitution at 28-COOH, was synthesized and tested with significant

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 January 2018 doi:10.20944/preprints201801.0177.v1

activity detected (**Fig. 1b**, **Supplementary Figs. 1b**, **2** and **Table 1**). The diazirine and alkyne moieties were introduced for photocrosslinking of the target proteins, followed by click conjugation with biotin to capture and enrich the crosslinked protein (**Fig. 2a**). Incubation of **Y18** followed by ultraviolet activation with the pseudovirus-producing cell lysates specifically crosslinked EBOV GP, which was precipitated by an azido-biotin resin in western blot analysis, but not VSVg or IAV HA (**Fig. 2a** and **b**). This specific crosslinking was prevented by the competitive compounds **Y0** and **Y11**, as demonstrated in a parallel experiment, but not by the inactive analogue **Y12**, confirming the interaction of GP with **Y18**. The **Y18**/GP photocrosslinked complex was subjected to high-resolution mass spectrometry, and a 573.41819 increase in molecular weight, corresponding to photoactivated **Y18**, was detected in the peptide 623-IDQIIHDFVDK-633, which is known as HR2 anchoring in the viral membrane[22,23] (**Fig. 2c** and **Supplementary Fig. 3a**). The crosslinking site within the peptide was localized to I627, a constitutive residue of HR2, by the MALDI-MS analysis of tryptic digests. This result suggested that the EBOV envelope subunit HR2 is the target domain, with I627 as an accessible residue.

Verification of the specific affinity between HR2 and triterpenoid leads by SPR

To confirm the interaction between triterpenoids and HR2 identified by chemical probes and mass spectrometry, we then tested whether the triterpenoids had high affinity to HR2 using a surface plasmon resonance (SPR) assay. The N-trimer (N39) and C-terminal GP2 peptides (**Supplementary Fig. 3a**), representing HR1 and HR2, respectively, as previously reported[24], were immobilized separately on a CM5 chip with

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 January 2018 doi:10.20944/preprints201801.0177.v1

the tested compounds flowing across their surface. We found that **Y11** and **Y18** bound to HR2, exhibiting a strong dose-dependent response, with respective K_D values of 5.4 and 16.8 μM; no specific binding to HR1 was found in the parallel experiment (**Fig. 2d** and **Supplementary Fig. 3b**). The inactive analogue **Y12** and the unrelated compound E-64d, the EBOV entry inhibitor targeting the host cysteine protease[18,25], did not bind to either immobilized peptide. In addition, the interaction between HR1 and HR2 was significantly reduced in the presence of **Y11**: the K_D value increased from 1.0 μM to 23.4 μM; however, this interaction was not reduced by either **Y12** or E-64d (**Fig. 2e** and **Supplementary Fig. 3c**). This result confirmed the specific affinity between HR2 and the triterpenoid leads, which resulted in a significant reduction in HR1-HR2 interaction.

NMR characterization of HR2 residues interacting with the triterpenoid leads

We then elucidated the global spatial interactions between HR2 and the triterpenoid leads by nuclear magnetic resonance (NMR)[26,27] (**Fig. 3** and **Supplementary Fig. 4**). The addition of **Y11** to an HR2 peptide solution was found to induce chemical shift changes for the acylamino groups of residues I623, D624, Q625, I626, D627, D629 and F630 but not H628, indicating that H628 is irrelevant for triterpenoid binding (**Fig. 3a**). Specifically, **Y11** was revealed by NMR TOCSY to contact I627 and F630, as the signals of I627 β-H and F630 α-H disappeared upon the addition of **Y11** (**Fig. 3b**). Furthermore, the NOE signals between Q625 β-H and **Y11** 25-H and between I627 α-H and **Y11** Gal acetyl-H (**Fig. 3c** and **Supplementary Fig. 4**), determined by NMR ROESY, demonstrated the proximity of these groups and, more

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 January 2018 doi:10.20944/preprints201801.0177.v1

importantly, demonstrated the orientation of the triterpenoid **Y11** across the HR2 core surface to be head-to-head and tail-to-tail (**Fig. 3d** and **Supplementary Fig. 5**). Further elucidation of the NOE signal network between **Y11** and HR2, mainly including (7.22, 0.26), (7.06, 0.17), (7.06, 1.93), (1.15 0.31), (0.48, 1.43), and (0.42, 2.52) (**Fig. 3c** and **Supplementary Fig. 4**), revealed more details of their spatial interactions.

The interactions of residues of the HR2 core with **Y11**, elucidated by the NMR data, were clearly reflected by the docking calculations[11,22,28,29] showing that most hydrophobic residues were shielded by the triterpenoid hydrophobic scaffold (**Fig. 3d** and **Supplementary Fig. 5**). More importantly, the docking calculations revealed the mild discrepancy between the conformations of **Y18** and **Y11** in wrapping the HR2 coil: the replacement of the spatial Gal moiety (**Y11**) with a linear group (**Y18**) brought the triterpenoid 3=O closer to K622, leading to hydrogen bond formation between 3=O of **Y18** and the amido group of K622, while Q625 was more distant from the triterpenoid **Y18** due to the lack of spatial Gal moiety (**Supplementary Fig. 5**). This difference might explain why K622 was critical for **Y18** while Q625 appeared to be irrelevant (**Table 1**), further verifying their head-to-head and tail-to-tail orientations and the global residue interactions. This was also supported by the observation that I627 in HR2 is photocrosslinked by the diazirine moiety in **Y18** (**Fig. 2c**), reflecting their proximity (**Supplementary Fig. 5**).

Verification of HR2 residues that interact with triterpenoid leads by biological mapping

We then used amino acid substitution mapping[30] to biologically verify which

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 January 2018 doi:10.20944/preprints201801.0177.v1

residues of HR2 were potentially interacting with the triterpenoids (**Table 1** and **Supplementary Table 2**). We first performed alanine screening to evaluate the effect of each individual substitution on virus infectivity. The mutations of residues K622, D624, Q625, I627, H628, D629 and F630 resulted in the maintenance of 5 - 80% of parental infectivity, while less than 0.1% infectivity was detected for the mutations at residues I623 and I626 (**Table 1**). This result reflected the criticality of each constitutive residue of HR2 for maintaining infectivity, and the latter two residues were excluded from further exploration. Then, the potency of **Y11** against each mutant virus was tested to determine the contribution of each residue. We found that K622A, D624A, Q625A and D629A reduced the potency of **Y11** by 2- to 4-fold, reflecting their moderate contributions to the binding of HR2 to the triterpenoid lead. Impressively, F630A and I627A reduced the potency of **Y11** by approximately 7- and 10-fold (**Table 1**), respectively, revealing their significant contributions to binding the triterpenoid lead. In contrast, H628A appeared to have no effect on the potency of **Y11**, similar to the mapping of all residues for interaction with compound E-64d. This result indicated that H628, unlike other residues, does not bind the triterpenoid, potentially due to its distant location. Similar results were observed for **Y18**, except that K622A decreased the potency of **Y18** 10-fold, while Q625A did not affect the potency of **Y18** (**Table 1**), suggesting that K622 was critical while Q625 was irrelevant for **Y18** binding. This finding is consistent with the docking calculations showing that a hydrogen bond formed between 3=O of **Y18** and the amino group of K622, while Q625 was more distant from **Y18**. These data further confirmed the molecular structure model of the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 January 2018 doi:10.20944/preprints201801.0177.v1

interactions between the triterpenoid lead and HR2.

Identification of HR2 as a targeting domain in the triterpenoid-mediated inhibition of HIV entry

In light of the role of BA and its derivative **Y19** as entry inhibitors of HIV, **Y20**, a triterpenoid probe containing a diazirine moiety, was synthesized and tested for activity with its parent **Y19** as a control compound[31] (**Fig. 4a** and **b**). Using the same approach as in the exploration of EBOV GP, the incubation of **Y20** with the HIV envelope protein followed by MS analysis of the ultraviolet-activated mixture revealed a molecular weight increase of 772.5598, corresponding to the photoactivated **Y20**, in the peptide 623-WNNMTWMEWDREINNYTSLIHSLIEESQNQQEK-655, which contains HR2 in the HIV GP41 protein[32-34] (**Fig. 4c** and **Supplementary Fig. 6a**). The crosslinking site within the peptide was localized to M626 according to the MALDI-MS analysis of the tryptic digests. The SPR assays indicated that **Y20** binds tightly to HIV-HR2 but not to HIV-HR1[14,35], both were separately immobilized on a CM5 chip with the tested compounds flowing across its surface (**Supplementary Fig. 6a**, **b** and **c**). Also similar to the findings in the EBOV study, the addition of **Y20** to HIV-HR1/HIV-HR2 mixtures significantly reduced their affinity (**Supplementary Fig. 6d**). The docking simulations also indicated multiple hydrophobic interactions between the hydrophobic backbones of **Y19** and **Y20** and the residues that constitute HR2, namely, W628, W631, I635, Y638, I642, L645, and L646 (Protein Data Bank: 4TVP[32]) (**Fig. 4d**). The -COOH group of **Y19** or **Y20** lies in the polar pocket generated by D624, N625 and T627, and the 3-OH group of **Y19** or **Y20** lies near the polar residues S649

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 January 2018 doi:10.20944/preprints201801.0177.v1

and Q650. In the above binding model, the diazirine group of **Y20** is in close proximity to M626, the residue that is crosslinked by **Y20**, which was consistent with the photocrosslinking results (**Fig. 4c**), further confirming the binding orientation. We thus concluded that the HR2 in the HIV envelope protein GP41 is a target for the triterpenoid leads.

Identification of HR2 as a targeting site in the triterpenoid-mediated inhibition of influenza virus entry

Given the significant inhibitory effect of OA and its derivative **Y3** on influenza virus entry, **Y21**, a triterpenoid probe containing a diazirine group on the galactose moiety, was synthesized and tested for activity with its parent **Y3** as a control compound[11] (**Fig. 5a and b**). Using the same approach as in the exploration of EBOV GP and HIV GP, the incubation of **Y21** with HA protein followed by MS analysis of the ultraviolet-activated mixture revealed a molecular weight increase of 825.5027, corresponding to photoactivated **Y21**, in the peptide 471-NNAKEIGNGCFEFYHK-486, which is known as HR2 in HA protein[36] (influenza A/California/04/2009 (H1N1)) (**Fig. 5c** and **Supplementary Fig. 7a**). The crosslinking site was localized to one of the residues between N471 to I476 based on the MALDI-MS analysis of the tryptic digests. The interaction between the influenza virus HR2 and the triterpenoid **Y3** was further verified by mapping experiments (**Supplementary Fig. 7b**), which were consistent with the docking calculation that clearly revealed the interactions of the hydrophobic backbones of the lead compounds **Y3** and **Y21** with HR2 (Protein Data Bank: 3LZG[37]), including residues V465, K466, Q468, L469, and N472 (**Fig. 5d**). More importantly, a

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 January 2018 doi:10.20944/preprints201801.0177.v1

hydrogen bond was formed between the diazirine group of **Y21** and N472, the potential crosslinking site within HR2, which was consistent with the photocrosslinking observations (**Fig. 5c**). Therefore, we concluded that HR2 is the direct target in the triterpenoid-mediated inhibition of influenza virus entry.

DISCUSSION

Here, we report the addition of Ebola virus to the list of viruses for which triterpenoids act as antagonists against viral entry; more importantly, we report the prominent HR2 as a common domain accessible to triterpenoid antagonists. HR2 is a prevalent domain in viral membrane-anchoring subunits, like HR1, and is characterized by a heptad repeat sequence designated as ***abcdefg***[38,39] (**Fig. 1a** and **Supplementary Fig. 8a**). As exceptional sequences, both HR1 and HR2 share dual hydrophobic residues at positions *a* and *d* and self-fold into alpha-helical coils that interact with each other via hydrophobic regions at the protein surfaces, oligomerizing to form super-helical assemblies[2]. This phenomenon has been observed in an increasing variety of viruses, including HIV[39], influenza[40] and Ebola[15] viruses, in which the N-terminal HR1 domain of the viral membrane-anchoring subunit forms a parallel, trimeric coiled-coil onto which the C-terminal HR2 region coils, forming a trimer-of-hairpins and ultimately bringing the virus and host membranes into close proximity for fusion (**Fig. 1a**, **Supplementary Fig. 8** and **Video 1**). The tight van der Waals forces between the coils make the formation of the coiled-coil thermodynamically favorable, rendering the identification of chemical antagonists to block the formation of the trimeric α-helical coiled-coil highly challenging. Thus far, only long complementary peptide sequences,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 January 2018 doi:10.20944/preprints201801.0177.v1

such as T20 peptide, have been developed for the inhibition of HIV fusion[41]. Our finding that HR2 is accessible to triterpenoids implied that it might be an exposable domain in viral fusion proteins during virus-host fusion; this possibility was further supported by the EBOV HR2 peptide-mediated induction of antibodies capable of inhibiting EBOV entry, including all five EBOV subtypes (**Supplementary Fig. 9**). Given the prevalence of the HR2 motif in a variety of viral envelopes, the triterpenoid-mediated inhibition of viral entry via HR2 might be potentially extended to an increasing variety of viruses.

The HR2 motif is usually considered an undruggable target due to its helicity; chemical ligands with high affinity and specificity have not been previously reported[4,18,33,35,42,43]. Triterpenoids, which provide a pentacyclic backbone for hydrophobic interactions with the six non-polar residues of HR2, as determined by amino acid substitution, SPR and NMR spectroscopy, can cross-wrap HR2 in a putative head-to-head and tail-to-tail orientation. In particular, the hydrophobic regions of the coil surfaces composed by the residues *a* and *d*, which provide the thermodynamic driving force for the coiled coils and allow multiple α-helical coils to wrap around each other, were shielded (**Figs. 3d**, **4d** and **5d**, **Supplementary Fig. 8** and **Video 2**). Additionally, the high conservation of HR2 residues among different EBOV strains (**Supplementary Fig. 3a**), as demonstrated by the impaired infectivity phenotype of all HR2 mutants (**Table 1**), and the much lower glycan shielding than at other parts of the viral membrane make HR2 an ideal target for disrupting the HR1-HR2 interaction[23,44,45]. Despite the evolutionary distance among pathogenic viruses,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 January 2018 doi:10.20944/preprints201801.0177.v1

such as the devastating Ebola, HIV and influenza A virus in this study, the structural homology of the trimer-of-hairpins, based on experimentally determined three-dimensional structures, may translate into a shared mechanism that hosts can triterpenoids to antagonize membrane fusion in a variety of viruses. Overall, the combined factors of helix wrapping, hydrophobicity and the terminal hydrogen bond (**Fig. 5d** and **Supplementary Fig. 5**) give rise to a distinct HR2-triterpenoid molecular architecture.

All organisms must survive their corresponding viruses and therefore have essential antiviral systems. As plants employ natural defensive chemistry against pathogens[46], the broad antiviral aspects of triterpenes imply their role as a class of defense chemicals, likely owing to the accessibility of the hydrophobic HR2 to the pentacyclic hydrophobic backbone. Although the evolution of viruses remains controversial[47-51], the high degree of structural similarity, including the HR1-HR2 interaction conserved among distinct viral fusion proteins[3] and the similar trimer-of-hairpins causing the virus-host membrane fusion, suggests that HR2-containing viral fusion proteins have a common origin[52,53] and that triterpenes are their native antagonists. The differences in the amino acid sequence of HR2 in evolutionarily distant viruses would shape viral phylogenies and trigger the evolution of chemical defense as an effective counterstrategy. The scaffold variations in the lead triterpenoids suggest that triterpenes might have evolved to carry out functions specific to the viruses they encounter (**Supplementary Fig. 10**). As demonstrated in this study, OA-based triterpenoids have a high propensity to bind 462-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 January 2018 doi:10.20944/preprints201801.0177.v1

$Y_{(a)}E_{(b)}K_{(c)}V_{(d)}K_{(e)}S_{(f)}Q_{(g)}L_{(a)}K_{(b)}N_{(c)}N_{(d)}A_{(e)}$-473, the HR2 sequence shared by influenza viruses; BA-based triterpenoids tend to bind 627-$T_{(g)}W_{(a)}M_{(b)}E_{(c)}W_{(d)}D_{(e)}R_{(f)}E_{(g)}I_{(a)}N_{(b)}N_{(c)}Y_{(d)}T_{(e)}S_{(f)}L_{(g)}I_{(a)}H_{(b)}S_{(c)}LI_{(d)}E_{(e)}ES_{(a)}Q_{(b)}$-650, an HR2 variant shared by HIV; and EA-based triterpenoids tend to bind 622-$K_{(g)}I_{(a)}D_{(b)}Q_{(c)}I_{(d)}I_{(e)}H_{(f)}D_{(g)}F_{(a)}$-630, an HR2 variant shared by EBOV (**Supplementary Fig. 8**). Clearly, despite the lack of sequence conservation, the residues at the same positions in HR2 have similar characteristics, giving rise to similar spatial structures, which provide the structural basis by which triterpenoids act as broad-spectrum viral fusion antagonists.

The triterpenoid wrapping of HR2 has significant implications for the development of antiviral drugs. Unlike neutralizing antibodies and peptide fusion inhibitors that target long epitopes with multiple conformations, triterpenoids recognize relatively smaller domains and cross the whole surface of the HR2 coil, and thus, they can be developed as powerful fusion inhibitors. Currently, a series of triterpenoid leads, including BA-, OA- and EA-based chemicals targeting HIV, HCV, influenza and Ebola, are being explored. Notably, bevirimat (PA-457), a BA-based HIV inhibitor, has been developed in clinical trials as a maturation inhibitor and therefore deserves to be reconsidered as an anti-fusion antagonist for further development. Triterpenes, mainly consisting of the lupane, oleanane, and ursane types[11,13], are secondary plant metabolites found in various plants, constituting up to 30% of dry weight in a few species[54]. This provides a rich source of chemical diversity for a stock of deterrents to Ebola, HIV and influenza virus and may potentially extend to a wider variety of viruses.

 doi:10.20944/preprints201801.0177.v1

Overall, the broad antiviral properties of triterpenoids and the identification of the prevalent HR2 coil as an exposable target pave the way for modulations of virus-host fusion that are of general significance, emerging as a new frontier in drug discovery and development.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 January 2018 doi:10.20944/preprints201801.0177.v1

METHODS

Cell lines. Human embryonic kidney (HEK) 293T cells, Madin-Darby canine kidney (MDCK) cells, human lung adenocarcinoma A549 cells, HeLa cells, CEM 4 cells, and human foreskin fibroblast (HFF) cells were grown in Dulbecco's Modified Eagle's Medium (DMEM) (Gibco BRL, Inc., Gaithersburg, MD, USA) supplemented with 10% fetal bovine serum (FBS) (PAA Laboratories, Linz, Austria) and cultured at 37 °C with 5% CO_2.

Reagents. Rabbit anti-GP (Ebola/Zaire) polyclonal antibody was obtained from Immune Technology Corp. (New York, USA). Rabbit anti-HA (H1N1) monoclonal antibody was obtained from Sino Biological Inc. (Beijing, China). Mouse anti-VSVG monoclonal antibody, horseradish peroxidase (HRP)-conjugated anti-rabbit IgG and anti-mouse IgG were obtained from Sigma-Aldrich (St. Louis, MO, USA).

E-64d (TCI Development Co., Ltd., Shanghai, China) and compound 3.47 were > 98% pure and were used as reference compounds.

All tested compounds were dissolved in DMSO to a stock concentration of 10 mM. The purity of all evaluated biological compounds was > 95%, as determined by HPLC (Agilent 1260) with a detection wavelength of 215 nm.

Plasmids. Plasmids expressing wild-type EBOV GP were obtained by inserting GP cDNA from the corresponding Ebola virus subtype (Zaire, strain Mayinga 1976, GenBank accession number U23187.1; Sudan, strain Gulu, GenBank accession number YP_138523.1; Bundibugyo, strain Uganda 2007, GenBank accession number ACI28624.1; Ivory Coast, strain Cote d'Ivoire 1994, GenBank accession number

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 January 2018 doi:10.20944/preprints201801.0177.v1

ACI28632.1; and Reston, strain Siena 1992, GenBank accession number AAC54891.1) between the KpnI and XbaI sites of the PCMV3 mammalian expression vector (Sino Biological Inc., Beijing, China). Plasmids expressing the wild-type HIV-1 envelope protein were obtained by inserting HIV-1 gp160 cDNA (strain NDK) between the KpnI and XbaI sites of the PCMV3 mammalian expression vector. The HA and NA genes were amplified from influenza A virus strain A/WSN/33(H1N1) cDNA by PCR and then inserted between the KpnI and EcoRI or XhoI sites, respectively, of the pcDNA4/TO vector (Invitrogen, Carlsbad, CA). Other plasmids included pCMV-VSVG (Addgene), pNL 4-3 (NIH AIDS Reagent Program), and pAdvantage (Promega). The EBOV GP and influenza virus HA mutants were constructed with a Site-Directed Mutagenesis Kit (Agilent Technologies) and confirmed by sequencing (BGI, Beijing). All plasmids used for transfection were amplified using a Maxiprep Kit (Promega) according to the manufacturer's instructions.

Pseudotype virus preparation. The detailed pseudotype viral propagation protocol has been previously described[55]. Briefly, 5×10^5 HEK293T cells were seeded into 6-well plates 24 h prior to transfection. Then, HEK293T cells were cotransfected with 1.0 μg of pNL4-3 plasmid, 12 ng of pCMV3-EBOV GP and 0.3 μg of pAdvantage with the transfection reagent MegaTran 1.0 (Origene) according to the manufacturer's instructions to produce Ebola virus GP pseudotyped HIV virions (HIV/EBOV GP). HEK293T cells were cotransfected with 1.0 μg of pNL4-3 plasmid, 0.5 μg of pcDNA4/TO-HA, 0.5 μg of pcDNA4/TO-NA and 0.3 μg of pAdvantage to produce pseudotyped influenza A virus (IAVpp). HEK293T cells were cotransfected with 1.0 μg

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 January 2018 doi:10.20944/preprints201801.0177.v1

of pNL4-3 plasmid, 0.7 μg of pCMV-VSVG and 0.3 μg of pAdvantage to produce VSVG pseudotyped HIV virions (VSVpp). HEK293T cells were cotransfected with 1.0 μg of pNL4-3 plasmid, 0.7 μg of pCMV3-HIV Env and 0.3 μg of pAdvantage to produce HIV pseudotyped virions (HIVpp). The supernatants containing the pseudotyped viruses were collected twice (at 48 h and 72 h after transfection), combined, clarified by the removal of floating cells and cell debris with low-speed centrifugation, and filtered through a 0.45-μm pore-size filter (Nalgene). The culture supernatants containing pseudotyped viruses were either used immediately or flash frozen in aliquots and stored at -80 ℃ until use. Each aliquot was thawed only once for use in a single round of replication.

Chemistry. The detailed experimental procedures and characterization data for all tested compounds are presented in the Supplementary Information.

Infection assay using pseudotyped viruses. Compounds were tested using entry assays for HIV/EBOV GP, HIVpp, IAVpp and VSVpp as previously described[11,13]. Infections were performed in 96-well plates by adding diluted HIV/EBOV GP, HIVpp, IAVpp or VSVpp to 5×10^3 293T or A549 cells per well in the presence and absence of the test compounds. The mixtures were then incubated for 72 h at 37 ℃ to screen the compound libraries. Luciferase activity, which reflected the number of pseudoparticles in the host cells, was measured 3 days after infection using the Bright-Glo reagent (Promega) and an FB15 luminometer (Berthold Detection Systems) according to the manufacturer's protocol. Test compounds were serially diluted to a final concentration of 1 μM in 1% dimethyl sulfoxide (DMSO), and the maximum

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 January 2018 doi:10.20944/preprints201801.0177.v1

activity (100% of the control) and background were derived from the control wells containing DMSO alone or from uninfected wells, respectively. The individual signals in each of the compound test wells were then divided by the average signals of the control values (wells lacking inhibitor) after subtracting the background, and the results were multiplied by 100 to determine the percent activity. The corresponding inhibition values were calculated by subtracting this value from 100. The specificity of the compounds for inhibiting HIV/EBOV GP was determined by evaluating the inhibition of IAVpp and VSVpp infection in parallel. Each sample was tested in duplicate, and all the experiments were repeated at least three times.

Native Ebola virus infection inhibition assay under BSL-4 conditions. HeLa cells and HFF cells were seeded into 96-well plates and exposed to native EBOV (Makona). HeLa cells were incubated with a serially diluted lead compound (10, 3.33, 1.11, 0.37, 0.12, 0.04, 0.01, 0.005 and 0 μM) or 1% DMSO 2 h prior to the addition of the virus (MOI = 1.5). The viruses were added to the HFF cells at an MOI of 20, as measured in HeLa cells. The infection was stopped after 48 h by fixing the cells with a formalin solution. Immunostaining was performed with an Alexa Fluor 488-conjugated anti-GP antibody to detect the infected cells. Images were acquired with the PE Opera confocal platform using a 10× objective and analyzed using Acapella software. Genedata software was used to calculate the percent infection from the GP staining data, and this value was then converted into percent inhibition. The number of nuclei per well was used to determine the percentage of viable cells (compared to that in the infected but untreated controls). The experiment was repeated twice to generate replicate 1 (rep1)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 January 2018 doi:10.20944/preprints201801.0177.v1

and replicate 2 (rep2). Compound 3.47 was used as a positive control[17].

Time-of-addition experiment. The "time-of-addition" experiment was designed as previously described[11,13,19] to determine the mechanism of action of the antiviral compounds, and the procedure is shown schematically in Supplementary Fig. 2. Briefly, A549 cells were seeded into 96-well plates at a density of 5×10^3 cells per well 24 h prior to infection and incubated at 37 ℃ with 5% CO_2. Then, the A549 cells were infected with HIV/EBOV GP and treated with test compounds (0.5 μM **Y11**, 2 μM **Y18**, or 2 μM E-64d) at the various stages of viral entry. Duplicate wells were used for each time point, and control infected cultures were treated with only a drug vehicle (DMSO). At 72 h after infection, the cells were assayed, and the infectivity was measured as described above.

Photoaffinity labeling and click chemistry. Photo-activation and click chemistry were performed using a previously described method with some modifications[17,56,57]. Briefly, cell lysates containing the packaged pseudotyped viruses were incubated with 50 μM **Y0**, **Y11**, or **Y12** or with DMSO for 30 min at room temperature. The lysates were then incubated with 5 μM **Y18** for an additional 30 min and exposed to ultraviolet light (365 nm) for 10 min on ice. Azido-immobilized agarose and 10 mM L-ascorbic acid were then added sequentially. The azido-immobilized agarose was prepared by incubating excess biotin-azide (Invitrogen) with streptavidin (SA)-immobilized agarose (Invitrogen). The cycloaddition reaction (click chemistry) was initiated by adding 1 mM $CuSO_4$, and the samples were incubated overnight at 4 ℃. The agarose was washed, and the captured target proteins were analyzed by western blotting.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 January 2018 doi:10.20944/preprints201801.0177.v1

Photoaffinity labeling and mass spectrometry measurements. Purified EBOV (Zaire, strain Mayinga 1976) GP protein was solubilized in ultrapure water to a final concentration of 1 mg/mL. The protein was then incubated with 5 μM **Y18** or 1% DMSO for an additional 30 min and exposed to ultraviolet light (365 nm) for 10 min on ice. The probe-labeled protein was analyzed by peptide mapping. After an overnight trypsin digest at 37 ℃, the reaction was quenched by the addition of trifluoroacetic acid (TFA). The tryptic peptides were re-suspended and separated by C18 RP-HPLC (Easy-nLC II, Thermo Fisher Scientific, CA) with in-line UV detection. The fractions were collected, dried, and then re-suspended in TFA. Each fraction was further separated on an analytical C18 column. The peptides were identified by MALDI (MS and MS-MS) analysis. The photoaffinity labeling of HIV GP41 with **Y20** and of influenza virus HA with **Y21** and subsequent MS measurements were performed using the same procedure.

Production and purification of the recombinant peptide eboIZN39IQ (N39). The design of the recombinant peptide eboIZN39IQ (N39) has been previously described[24]. The optimized gene encoding eboIZN39IQ (N39) was synthesized by BGI (Beijing, China) and cloned into the pET21a (+) vector between the NdeI and XhoI sites to generate pET21-N39. The optimized gene sequence was

CACCACCATCATCATCATATTGAAGGCCGCGGCCACATGGATATCAAGAA

AGAAATTGAGGCGATCAAGAAAGAGCAGGAAGCGATCAAGAAGAAGATC

GAGGCGATCGAGAAAGAACTGCGTCAACTGGCAAACGAAACCACCCAAG

CACTGCAACTGTTTCTGCGCGCAACCACCGAACTGCGTACCTTTAGCATCC

TGAACCGCAAAGCGATCGATTTTCTGCTGCAGCGCATGAAGCAGATCGAA

 doi:10.20944/preprints201801.0177.v1

GACAAGATCGAAGAAATTGAGAGCAAGCAGAAGAAGATCGAGAACGAG ATCGCGCGTATCAAAAAACTGATCGGCGAACGTTACTAAA. A 6×His tag was added at the N-terminus for affinity purification, followed by the addition of a Factor Xa cleavage site (IEGR). BL21 (DE3) was used as the host strain for eboIZN39IQ (N39) expression. The strain hosting pET21-N39 was inoculated into 2 mL of Luria-Bertani (LB) medium containing 100 μg/mL ampicillin and grown at 37 ℃ while shaking at 220 rpm. The overnight *E. coli* cultures were then diluted to an optical density (OD) of 0.2 in LB medium and incubated at 37 ℃. When the culture ODs reached approximately 1.0 (OD_{600}), isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM. After overnight induction at 30 ℃, the cells were harvested by centrifugation and re-suspended in His-Bind buffer (20 mM Tris-HCl (pH 8.0), 250 mM NaCl, 5 mM imidazole). The proteins were extracted by passing the cells twice through a microfluidizer at 1200 bar with cooling. The supernatant containing the recombinant peptide eboIZN39IQ (N39) was collected by centrifugation and mixed with Ni-NTA His-Bind resin (Novagen) for 2 h at 4 ℃. The unbound proteins were then removed by washing the resin with 10 volumes of wash buffer (20 mM Tris-HCl (pH 8.0), 250 mM NaCl, 20 mM imidazole). The resin-enriched eboIZN39IQ (N39) peptide was eluted with the same buffer, except that the concentration of imidazole was 500 mM. The purified peptide was dialyzed into 5% acetic acid, further purified by RP-HPLC on a C18 column (GE Healthcare) and then lyophilized. The peptide was then digested with Factor Xa protease (NEB) according to the manufacturer's protocol. The digested peptide was dialyzed into 5% acetic acid, purified by HPLC and then

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 January 2018 doi:10.20944/preprints201801.0177.v1

lyophilized again. The final peptide sequence was **GHMDIKKEIEAIKKEQEAIKKKIEAIEKE***LRQLANETTQALQLFLRATTELRTF SILNRKAIDFLLQR***MKQIEDKIEEIESKQKKIENEIARIKKLIGERY**, where IZ_m and IQ are shown in bold and the Ebola virus N-trimer in italics.

Design and syntheses of eboC24 (C24) peptide. The eboC24 (C24) peptide was designed and chemically synthesized by Abace Biology (Beijing, China) according to a previously described method with some modifications. All peptides with > 95% purity were lyophilized, and their molecular weights were verified by LC/MS. The resulting eboC24 (C24) peptide sequence was Lys (N3)-IEPHDWTKNITDKIDQIIHDFVDK-NH_2, and its molecular weight was 3074.45 Da.

SPR. The interactions between the peptides and the compounds were analyzed with the Biacore T200 system (GE Healthcare, Uppsala, Sweden) at 25 °C. The recombinant eboIZN39IQ (N39) peptide was immobilized on a sensor chip (CM5) with an Amine Coupling Kit (GE Healthcare, Buckinghamshire, UK). The final levels of immobilized eboIZN39IQ (N39) were typically approximately 3000 response units (RU). The eboC24 (C24) peptide was conjugated to DIBO-biotin (Invitrogen) by click chemistry and then immobilized on an SA sensor chip, with final levels of approximately 600 RU. Various concentrations of the compounds were subsequently injected as analytes, and PBS-P (10 mM phosphate buffer containing 2.7 mM KCl, 137 mM NaCl, and 0.05% Surfactant P20, pH 4.5) was used as the running buffer. For the binding studies, appropriate concentrations of the analytes were added to the running buffer at a flow rate of 30 μL/min, a contact time of 120 s and a dissociation time of 60 s, and the chip

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 January 2018 doi:10.20944/preprints201801.0177.v1

platforms were washed with running buffer and 50% DMSO. The data were analyzed with Biacore evaluation software (T200 Version 1.0), and the curve was fitted with a 1:1 binding model.

The eboIZN39IQ (N39) peptide was immobilized on a CM5 sensor chip at approximately 3000 RU to explore whether **Y11**-eboC24 binding had any effect on the interactions between the eboC24 and eboIZN39IQ (N39) peptides in the SPR experiments. EboC24 concentrations of 25, 12.5, 6.25, 3.12, 1.56, 0.78, 0.39 and 0 μM were passed through the chip at a flow rate of 10 μL/min, a contact time of 60 s and a dissociation time of 600 s in the absence or presence of **Y11** or other compounds. The N39 peptide-SA chip was regenerated with glycine (pH 2.5). The data were analyzed with Biacore evaluation software (T200 Version 1.0), and the curve was fitted with a 1:1 binding model.

Mapping the HR2 domain using amino acid substitutions. We performed mutagenesis on the HR2 domain and tested the activity of the lead compounds against the EBOV mutants to identify which HR2 residues interacted with the lead compounds. Briefly, each HR2 residue (residues 610-636) was separately replaced with alanine (A) by site-directed mutagenesis (Agilent Technologies) and confirmed by sequencing (BGI, Beijing, China). The resulting plasmids were used to package the EBOV mutants. The infectivity of the EBOV mutants and the antiviral activity against them were evaluated as described above. The HR2 domain from influenza virus was mapped in the same manner by using amino acid substitutions.

NMR. The HR2 peptide samples used for NMR spectroscopy were prepared by

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 January 2018 doi:10.20944/preprints201801.0177.v1

dissolving an appropriate amount of peptide in 0.5 mL of solution (H_2O:D_2O = 9:1) to obtain a peptide concentration of 1.0 mM. The peptide-**Y11** complex was prepared by the same procedure by mixing the two components at a ligand:peptide ratio of 1:1. The NMR spectra were recorded on a Bruker DRX-600 spectrometer equipped with a cryoprobe. One-dimensional (1D) NMR spectra were recorded in Fourier mode with quadrupole detection. The TOCSY and NOESY experiments were performed in phase-sensitive mode using quadrupole detection in ω1 with a time-dependent phase increase in the initial pulse. The data block sizes were 2048 addresses in t_2 and 512 equidistant t1 values. Prior to the Fourier transformation, the time domain data matrices were multiplied by shifting the sine functions in both dimensions. The TOCSY experiments were performed at 298 K with a mixing time of 200 ms. The NOESY experiments were performed at 298 K with a mixing time of 300 ms. The NMR spectra were analyzed using MestReNova software.

Docking simulation. The **Y11** and **Y18** compounds were docked to the EBOV GP protein by AutoDock 4.2. The structural template of EBOV GP (Protein Data Bank: 5JQ3) was obtained from the RCSB Protein Data Bank (http://www.rcsb.org/pdb/home/home.do). Grid points covering EBOV HR2 were generated with AutoDock Tools (ADT) to perform blind docking experiments. The box size was 80 × 80 × 80 points with a standard spacing of 0.375 Å. Docking simulations for the compounds were performed using the Lamarckian genetic algorithm (GA) and a protocol with 50 GA runs, an initial population of 300 randomly placed individuals, a maximum number of 25 million energy evaluations, a mutation rate of 0.02, and a

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 January 2018 doi:10.20944/preprints201801.0177.v1

crossover rate of 0.80. The resulting conformations, which differed by less than 2.0 Å in the positional root-mean-square deviation (RMSD), were clustered together. The default values were used for all other parameters. All relevant torsion angles were treated as rotatable during the docking process to allow searching of the conformational space. After AutoGrid and AutoDock were run, the possible positions of the compounds on EBOV GP were obtained. The same procedure was conducted for the docking simulations of **Y19** and **Y20** with HIV-1 GP160 (PDB: 4TVP) and of **Y3** and **Y21** with influenza virus HA (PDB: 3LZG).

Production and purification of a polyclonal antibody against the HR2 peptide (KIDQIIHDF). A polyclonal antibody was generated by immunizing New Zealand rabbits with the synthesized HR2 peptide-mariculture keyhole limpet hemocyanin (mcKLH) conjugate by a previously described method with some modifications[58]. The maleimide-activated mcKLH carrier protein (Pierce) was reconstituted by adding dH_2O to Hypo-vial™ to yield a 10 mg/mL solution. A mixture of 2 mg of peptide and 2 mg of activated mcKLH was used, and the mcKLH formed a suspension, not a solution, that was typically opaque to whitish-blue in color. The suspension was not solubilized by vortexing or heating, as these treatments would cause mcKLH to further precipitate from solution. Next, 2 mg of sulfhydryl-containing hapten was dissolved in 200 μL of dH_2O. The peptide and activated mcKLH solutions were then immediately mixed, and the mixture was allowed to react for 2 h at room temperature. The conjugate was purified by dialysis against PBS buffer. For the first immunization, the purified peptide-mcKLH conjugate (100 μg) in PBS (1 mL) was mixed with an equal volume of Freund's

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 January 2018 doi:10.20944/preprints201801.0177.v1

complete adjuvant (Bio Basic Inc.) to form a stable emulsion. One rabbit was subcutaneously injected at 2 to 4 different sites. Three booster injections were administered at 20-day intervals using 100 μg of the purified peptide-mcKLH conjugate with incomplete Freund's adjuvant. Blood samples (0.5-1 mL) were collected prior to each injection, and bleeding (30 mL) was performed 10 days after the last booster to evaluate the immune response. Polyclonal antibodies (IgGs) were purified from rabbit serum by affinity chromatography on a 5-mL HiTrap Protein A column (GE Healthcare) according to the manufacturer's instructions. Binding was performed in 0.02 M sodium phosphate (pH 7.0), and elution was performed in 0.1 M citric acid (pH 3.0). The eluted IgGs were collected, immediately neutralized to physiological pH by adding 1 M Tris-HCl buffer (pH 9.0), and then concentrated to 1 mg/mL by using 30 kDa molecular weight cut-off concentrators. Next, an N-hydroxysuccinimide (NHS)-activated Sepharose column (GE Healthcare) was used for affinity purification of the HR2-specific IgGs from the total rabbit antibodies. Concentrated HR2 peptide (20 mg) in coupling buffer (0.2 M $NaHCO_3$ and 0.5 M NaCl, pH 8.3) was applied to the NHS-activated column and allowed to bind for 2 h at room temperature. Excess uncoupled, activated NHS groups were deactivated by sequential washing with two buffers: 0.5 M ethanolamine and 0.5 M NaCl (pH 8.3) followed by 0.1 M sodium acetate and 0.5 M NaCl (pH 4.0). The IgGs purified from rabbit serum using Protein A were passed through the HR2 column. Unbound IgGs were removed by wash buffer (0.05 M sodium phosphate and 0.15 M NaCl, pH 7.0) to obtain pure HR2-specific IgGs, which were eluted with 3-5 column volumes of 0.1 M citric acid buffer and then neutralized with 1

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 January 2018 doi:10.20944/preprints201801.0177.v1

M Tris-HCl buffer (pH 9.0). Purified HR2-specific IgG samples were buffer-exchanged with 0.05 M sodium phosphate buffer (pH 7.0) and then concentrated to 1 mg/mL. All animal experiments were performed in accordance with the guidelines of the Institutional Animal Care and Use Committee of Peking University.

Detection of the HR2 peptide (KIDQIIHDF)-specific polyclonal antibody. An indirect ELISA was employed to analyze sera and purified IgG, as previously described[58,59]. Briefly, the HR2 peptide antigen was diluted in carbonate coating buffer, and then 100 μL of this solution was added to the wells of an ELISA microplate and incubated overnight at 4 °C. After the plate was washed 3 times with PBS containing 0.05% Tween 20 to remove unbound antigen, 150 μL of blocking buffer (3% bovine serum albumin (BSA) in wash buffer) was applied to all wells for 1 h at 37 °C. After removing the blocking buffer, 100 μL of rabbit sera or IgGs diluted in 1% blocking buffer was added to each well, and the plate was incubated for an additional 1 h at room temperature. After 3 washes, HRP-conjugated goat anti-rabbit IgG was diluted (1:5000) in 1% blocking buffer, added to each well (100 μL per well), and allowed to bind to the captured rabbit IgG for 1 h at room temperature. After 5 additional washes, the bound conjugate was detected with 50 μL of 3,3',5,5'-tetramethylbenzidine (TMB) (Sigma) substrate, and the reaction was stopped after 8 min by the addition of 50 μL of 1 M H_2SO_4. The spectroscopic absorbance of each well was measured by an automated plate reader at a wavelength of 450 nm.

 doi:10.20944/preprints201801.0177.v1

Acknowledgments.

This work was supported by the National Natural Science Foundation of China (grant nos. 21572015, 81530090, 81373271 and 81361168002) and the National Basic Research Program of China (grant no. 2016YFA0501500). We thank Dr. James M. Cunningham (Brigham and Women's Hospital, Harvard Medical School) for assistance with the anti-native Ebola virus screening work.

Author contributions

D.Z. and L.S. conceived the study, designed experiments, interpreted data and wrote the manuscript. L.Z. revised the manuscript. L.S. performed most of the experiments and interpreted data. K.M., Z.T., Z.Z., V.S., J.S., H.L., G.F., Q.X., and S.X. provided assistance on part of the experiments. All authors commented on the paper and approved its publication.

Competing financial interests

The authors declare no competing financial interests.

Additional information The genome sequences have been deposited in GenBank under the accession numbers U23187.1 (Zaire EBOV GP), YP_138523.1 (Sudan EBOV GP), ACI28624.1 (Bundibugyo EBOV GP), ACI28632.1 (Ivory Coast EBOV GP), AAC54891.1 (Reston EBOV GP), and M27323.1 (HIV-1, group M, subtype D, strain NDK). Supplementary information and chemical compound information are available online at http://www.nature.com/naturechemicalbiology/. Reprints and permission information is available at http://www.nature.com/reprints/index.html. Correspondence and requests for materials should be addressed to D.Z.

(deminzhou@bjmu.edu.cn).

References

1 Vigant, F., Santos, N. C. & Lee, B. Broad-spectrum antivirals against viral fusion. *Nat Rev Microbiol* **13**, 426-437 (2015).

2 Harrison, S. C. Viral membrane fusion. *Nat Struct Mol Biol* **15**, 690-698 (2008).

3 White, J. M., Delos, S. E., Brecher, M. & Schornberg, K. Structures and mechanisms of viral membrane fusion proteins: multiple variations on a common theme. *Crit Rev Biochem Mol Biol* **43**, 189-219 (2008).

4 Eckert, D. M. & Kim, P. S. Mechanisms of viral membrane fusion and its inhibition. *Annu Rev Biochem* **70**, 777-810 (2001).

5 Domingo, V., Arteaga, J. F., Quilez del Moral, J. F. & Barrero, A. F. Unusually cyclized triterpenes: occurrence, biosynthesis and chemical synthesis. *Nat Prod Rep* **26**, 115-134 (2009).

6 Freed, E. O. HIV-1 assembly, release and maturation. *Nat Rev Microbiol* **13**, 484-496 (2015).

7 Smith, P. F. *et al.* Phase I and II study of the safety, virologic effect, and pharmacokinetics/pharmacodynamics of single-dose 3-o-(3',3'-dimethylsuccinyl)betulinic acid (bevirimat) against human immunodeficiency virus infection. *Antimicrob Agents Chemother* **51**, 3574-3581 (2007).

8 Li, F. *et al.* PA-457: a potent HIV inhibitor that disrupts core condensation by targeting a late step in Gag processing. *Proc Natl Acad Sci U S A* **100**, 13555-

13560 (2003).

9 Greene, W. The brightening future of HIV therapeutics. *Nature Immunology* **5**, 867-871 (2004).

10 Mayaux, J. F. *et al.* Triterpene Derivatives That Block Entry of Human-Immunodeficiency-Virus Type-1 into Cells. *P Natl Acad Sci USA* **91**, 3564-3568 (1994).

11 Yu, M. R. *et al.* Discovery of Pentacyclic Triterpenoids as Potential Entry Inhibitors of Influenza Viruses. *J Med Chem* **57**, 10058-10071 (2014).

12 Xiao, S. *et al.* Pentacyclic triterpenes grafted on CD cores to interfere with influenza virus entry: A dramatic multivalent effect. *Biomaterials* **78**, 74-85 (2016).

13 Yu, F. *et al.* Development of Oleanane-Type Triterpenes as a New Class of HCV Entry Inhibitors. *J Med Chem* **56**, 4300-4319 (2013).

14 Labrosse, B., Treboute, C. & Alizon, M. Sensitivity to a nonpeptidic compound (RPR103611) blocking human immunodeficiency virus type 1 Env-mediated fusion depends on sequence and accessibility of the gp41 loop region. *J Virol* **74**, 2142-2150 (2000).

15 Malashkevich, V. N. *et al.* Core structure of the envelope glycoprotein GP2 from Ebola virus at 1.9-angstrom resolution. *P Natl Acad Sci USA* **96**, 2662-2667 (1999).

16 White, J. M. & Schornberg, K. L. A new player in the puzzle of filovirus entry. *Nat Rev Microbiol* **10**, 317-322 (2012).

17 Cote, M. *et al.* Small molecule inhibitors reveal Niemann-Pick C1 is essential for Ebola virus infection. *Nature* **477**, 344-348 (2011).

18 Picazo, E. & Giordanetto, F. Small molecule inhibitors of ebola virus infection. *Drug Discov Today* **20**, 277-286 (2015).

19 Basu, A. *et al.* Identification of a small-molecule entry inhibitor for filoviruses. *J Virol* **85**, 3106-3119 (2011).

20 Das, J. Aliphatic Diazirines as Photoaffinity Probes for Proteins: Recent Developments. *Chem Rev* **111**, 4405-4417 (2011).

21 Thirumurugan, P., Matosiuk, D. & Jozwiak, K. Click chemistry for drug development and diverse chemical-biology applications. *Chem Rev* **113**, 4905-4979 (2013).

22 Zhao, Y. *et al.* Toremifene interacts with and destabilizes the Ebola virus glycoprotein. *Nature* **535**, 169-172 (2016).

23 Lee, J. E. *et al.* Structure of the Ebola virus glycoprotein bound to an antibody from a human survivor. *Nature* **454**, 177-182 (2008).

24 Clinton, T. R. *et al.* Design and characterization of ebolavirus GP prehairpin intermediate mimics as drug targets. *Protein Sci* **24**, 446-463 (2015).

25 Chandran, K., Sullivan, N. J., Felbor, U., Whelan, S. P. & Cunningham, J. M. Endosomal proteolysis of the Ebola virus glycoprotein is necessary for infection. *Science* **308**, 1643-1645 (2005).

26 de Opakua, A. I. *et al.* Molecular mechanism of Galphai activation by non-GPCR proteins with a Galpha-Binding and Activating motif. *Nat Commun* **8**,

 doi:10.20944/preprints201801.0177.v1

15163 (2017).

27 Makley, L. N. *et al.* Pharmacological chaperone for alpha-crystallin partially restores transparency in cataract models. *Science* **350**, 674-677 (2015).

28 Morris, G. M. *et al.* AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem* **30**, 2785-2791 (2009).

29 Sanner, M. F. Python: a programming language for software integration and development. *J Mol Graph Model* **17**, 57-61 (1999).

30 Watanabe, S. *et al.* Functional importance of the coiled-coil of the Ebola virus glycoprotein. *Journal of Virology* **74**, 10194-10201 (2000).

31 Soler, F. *et al.* Betulinic acid derivatives: A new class of specific inhibitors of human immunodeficiency virus type 1 entry. *J Med Chem* **39**, 1069-1083 (1996).

32 Pancera, M. *et al.* Structure and immune recognition of trimeric pre-fusion HIV-1 Env. *Nature* **514**, 455-461 (2014).

33 Yi, H. A., Fochtman, B. C., Rizzo, R. C. & Jacobs, A. Inhibition of HIV Entry by Targeting the Envelope Transmembrane Subunit gp41. *Curr HIV Res* **14**, 283-294 (2016).

34 Ozorowski, G. *et al.* Open and closed structures reveal allostery and pliability in the HIV-1 envelope spike. *Nature* **547**, 360-363 (2017).

35 Eckert, D. M. & Kim, P. S. Design of potent inhibitors of HIV-1 entry from the gp41 N-peptide region. *Proc Natl Acad Sci U S A* **98**, 11187-11192 (2001).

36 Bullough, P. A., Hughson, F. M., Skehel, J. J. & Wiley, D. C. Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature* **371**, 37-43

(1994).

37 Xu, R. *et al.* Structural basis of preexisting immunity to the 2009 H1N1 pandemic influenza virus. *Science* **328**, 357-360 (2010).

38 Chambers, P., Pringle, C. R. & Easton, A. J. Heptad repeat sequences are located adjacent to hydrophobic regions in several types of virus fusion glycoproteins. *J Gen Virol* **71 (Pt 12)**, 3075-3080 (1990).

39 Chan, D. C., Fass, D., Berger, J. M. & Kim, P. S. Core structure of gp41 from the HIV envelope glycoprotein. *Cell* **89**, 263-273 (1997).

40 Joshi, S. B., Dutch, R. E. & Lamb, R. A. A core trimer of the paramyxovirus fusion protein: parallels to influenza virus hemagglutinin and HIV-1 gp41. *Virology* **248**, 20-34 (1998).

41 Kilby, J. M. *et al.* Potent suppression of HIV-1 replication in humans by T-20, a peptide inhibitor of gp41-mediated virus entry. *Nat Med* **4**, 1302-1307 (1998).

42 Gochin, M. & Zhou, G. Amphipathic properties of HIV-1 gp41 fusion inhibitors. *Curr Top Med Chem* **11**, 3022-3032 (2011).

43 Eckert, D. M., Malashkevich, V. N., Hong, L. H., Carr, P. A. & Kim, P. S. Inhibiting HIV-1 entry: Discovery of D-peptide inhibitors that target the gp41 coiled-coil pocket. *Cell* **99**, 103-115 (1999).

44 Sok, D. *et al.* Rapid elicitation of broadly neutralizing antibodies to HIV by immunization in cows. *Nature*, doi:10.1038/nature23301 (2017).

45 Wang, H. *et al.* Ebola Viral Glycoprotein Bound to Its Endosomal Receptor Niemann-Pick C1. *Cell* **164**, 258-268 (2016).

46 Berenbaum, M. R. The chemistry of defense: theory and practice. *Proc Natl Acad Sci U S A* **92**, 2-8 (1995).

47 Forterre, P. & Prangishvili, D. The origin of viruses. *Res Microbiol* **160**, 466-472 (2009).

48 Forterre, P. The origin of viruses and their possible roles in major evolutionary transitions. *Virus Res* **117**, 5-16 (2006).

49 Moreira, D. & Lopez-Garcia, P. Ten reasons to exclude viruses from the tree of life. *Nat Rev Microbiol* **7**, 306-311 (2009).

50 Koonin, E. V., Senkevich, T. G. & Dolja, V. V. The ancient Virus World and evolution of cells. *Biol Direct* **1**, 29 (2006).

51 Raoult, D. & Forterre, P. Redefining viruses: lessons from Mimivirus. *Nat Rev Microbiol* **6**, 315-319 (2008).

52 Pletnev, S. V. *et al.* Locations of carbohydrate sites on alphavirus glycoproteins show that E1 forms an icosahedral scaffold. *Cell* **105**, 127-136 (2001).

53 Lescar, J. *et al.* The Fusion glycoprotein shell of Semliki Forest virus: an icosahedral assembly primed for fusogenic activation at endosomal pH. *Cell* **105**, 137-148 (2001).

54 Alakurtti, S., Makela, T., Koskimies, S. & Yli-Kauhaluoma, J. Pharmacological properties of the ubiquitous natural product betulin. *Eur J Pharm Sci* **29**, 1-13 (2006).

55 Mohan, G. S. *et al.* Less is more: Ebola virus surface glycoprotein expression levels regulate virus production and infectivity. *J Virol* **89**, 1205-1217 (2015).

 doi:10.20944/preprints201801.0177.v1

56 Shi, H., Zhang, C. J., Chen, G. Y. & Yao, S. Q. Cell-based proteome profiling of potential dasatinib targets by use of affinity-based probes. *J Am Chem Soc* **134**, 3001-3014 (2012).

57 Kinoshita, T. *et al.* Binding of brassinosteroids to the extracellular domain of plant receptor kinase BRI1. *Nature* **433**, 167-171 (2005).

58 Zhang, Y. *et al.* Production and application of polyclonal and monoclonal antibodies against Spiroplasma eriocheiris. *Sci Rep* **5**, 17871 (2015).

59 Si, L. L. *et al.* Generation of influenza A viruses as live but replication-incompetent virus vaccines. *Science* **354**, 1170-1173 (2016).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 January 2018 doi:10.20944/preprints201801.0177.v1

Figures Legends

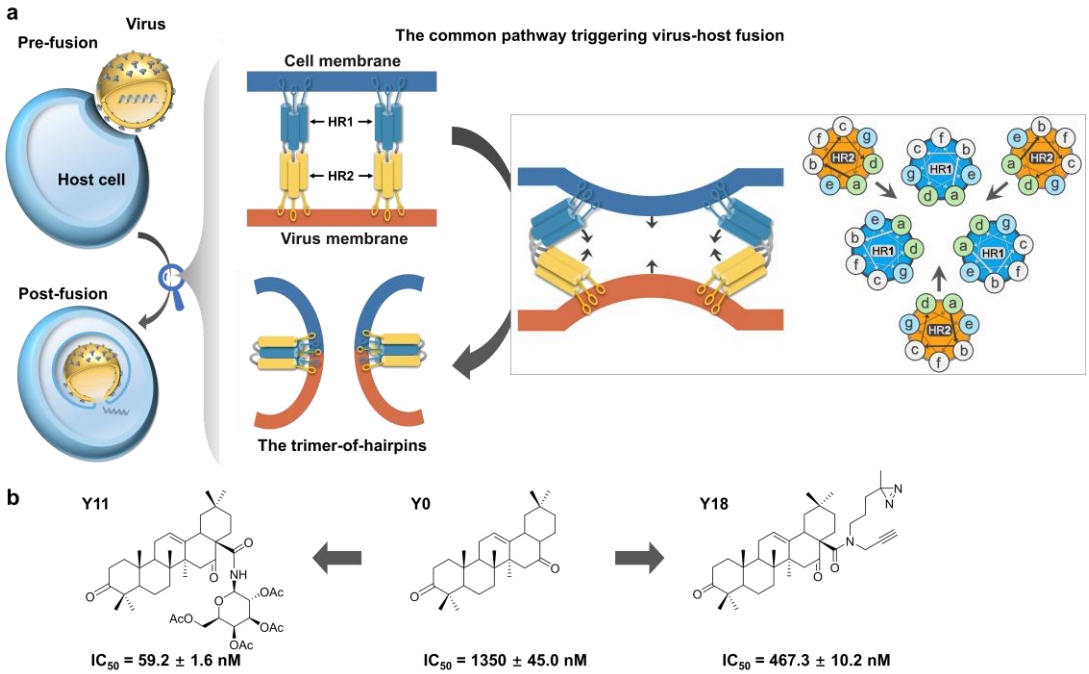


Figure 1. Schematic representation of virus-host membrane fusion via a common trimer-of-hairpins by which Ebola, HIV, influenza A and other viruses with envelopes, enter cells. (a) The pathway for converting distinct viral envelopes into the common trimer-of-hairpins conformation, which brings the viral and host membranes into close proximity for fusion. The interaction of HR2 with HR1 via hydrophobic residues, mainly at positions *a* and *d,* plays a critical role in the formation of the trimer-of-hairpins. (b) The lead compound and probe discovered among triterpenoid products that can inhibit Ebola virus-host fusion.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 January 2018 doi:10.20944/preprints201801.0177.v1

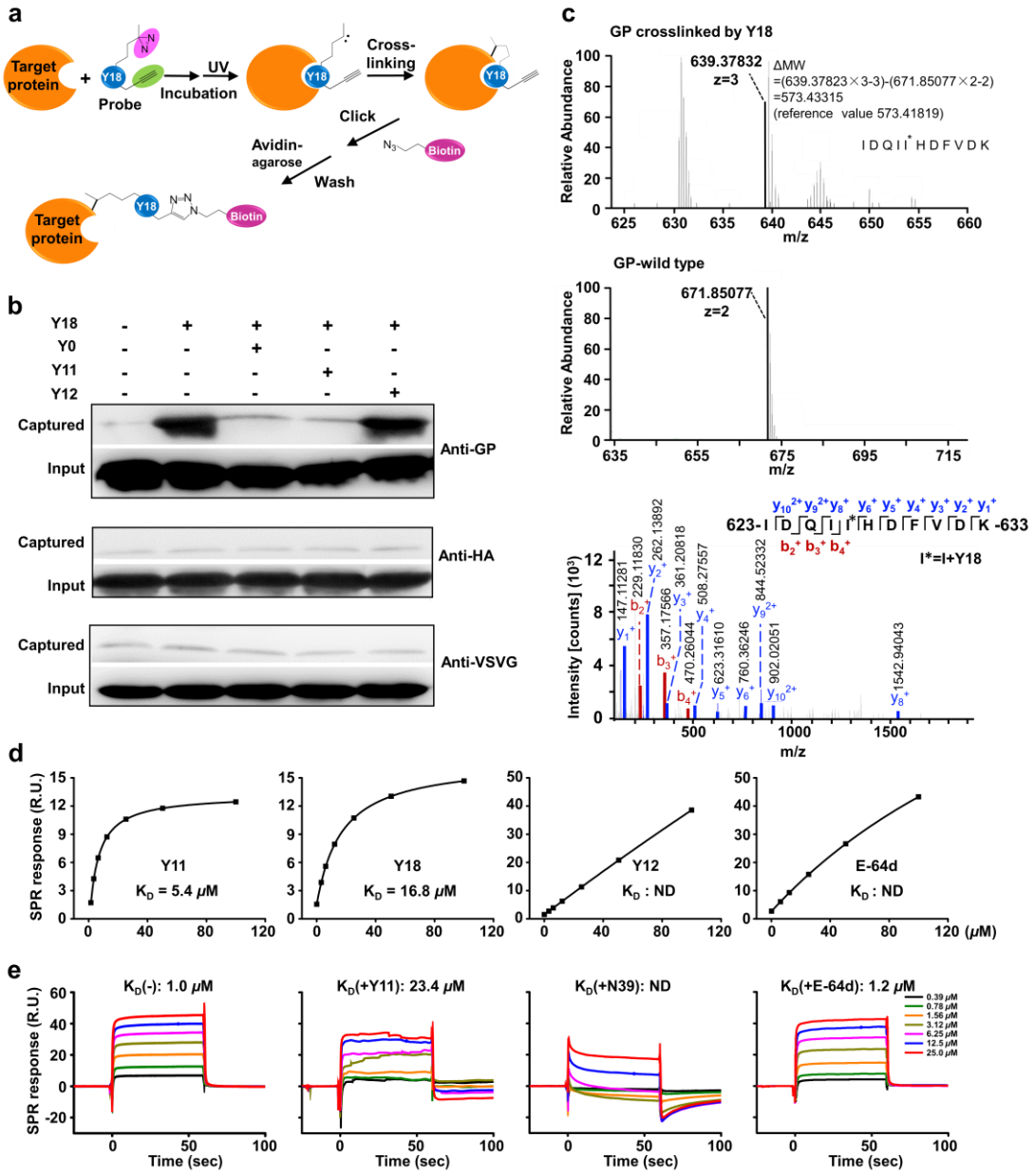
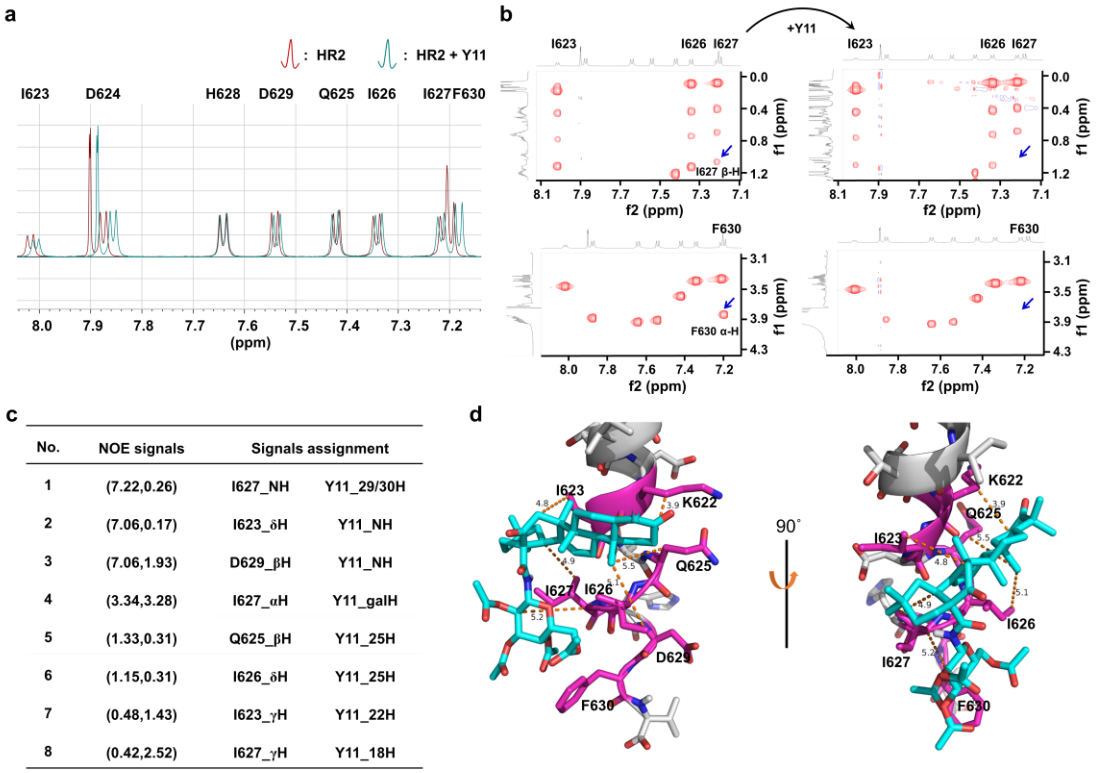


Figure 2. Identification of the EBOV envelope GP as the target of the triterpenoid leads. (**a**) Scheme representing the mechanism by which the triterpenoid probe **Y18** covalently captures the target protein via sequential photocrosslinking capture and click-mediated biotin harvest. (**b**) Western blot analyses of the crosslinked protein with **Y18** using anti-GP, anti-HA, or anti-VSVG antibodies. The packaged cell lysates, pretreated with 50 μM **Y0**, **Y11**, or **Y12** for 30 min, were incubated with 5 μM **Y18** and then exposed to ultraviolet light (365 nm) before being treated with azide-containing biotin. (**c**) The mass spectrum of the captured protein with the crosslinking site identified by peptide mapping. (**d**) SPR characterization of the affinity between the

 doi:10.20944/preprints201801.0177.v1

triterpenoid compounds, **Y11** and **Y18**, and the HR2 peptide, which was immobilized on an SA sensor chip. **Y12** and E-64d served as negative controls. (**e**) SPR characterization of the effects of the triterpenoid compounds on HR1-HR2 interaction. HR2 was allowed to flow across the HR1 chip surface in the absence or presence of the lead compound; EboIZN39IQ and E-64d served as positive and negative controls, respectively.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 January 2018 doi:10.20944/preprints201801.0177.v1



No.	NOE signals	Signals assignment	
1	(7.22,0.26)	I627_NH	Y11_29/30H
2	(7.06,0.17)	I623_δH	Y11_NH
3	(7.06,1.93)	D629_βH	Y11_NH
4	(3.34,3.28)	I627_αH	Y11_galH
5	(1.33,0.31)	Q625_βH	Y11_25H
6	(1.15,0.31)	I626_δH	Y11_25H
7	(0.48,1.43)	I623_γH	Y11_22H
8	(0.42,2.52)	I627_γH	Y11_18H

Figure 3. Characterization of the spatial interactions within the triterpenoid-HR2 complex. (**a**) Overlapped HR2 acylamino ^{1}H-NMR spectra in the absence and presence of **Y11** to differentiate the residues interacting with the triterpenoid lead. (**b**) NMR TOCSY spectra of the HR2 peptide (KIDQIIHDF) in the absence or presence of **Y11**. Upon the addition of **Y11 (**ratio of HR2:**Y11** as 1:1**)**, the signals of I627 β-H and F630 α-H from HR2 shifted. (**c**) Assignment of each of the intermolecular NOE signal from the triterpenoid-HR2 complex. (**d**) Global HR2-**Y11** complex inferred from the NOE signals plus docking simulation (Protein Data Bank: 5JQ3). The intermolecular NOEs are indicated by orange dashed lines.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 January 2018 doi:10.20944/preprints201801.0177.v1

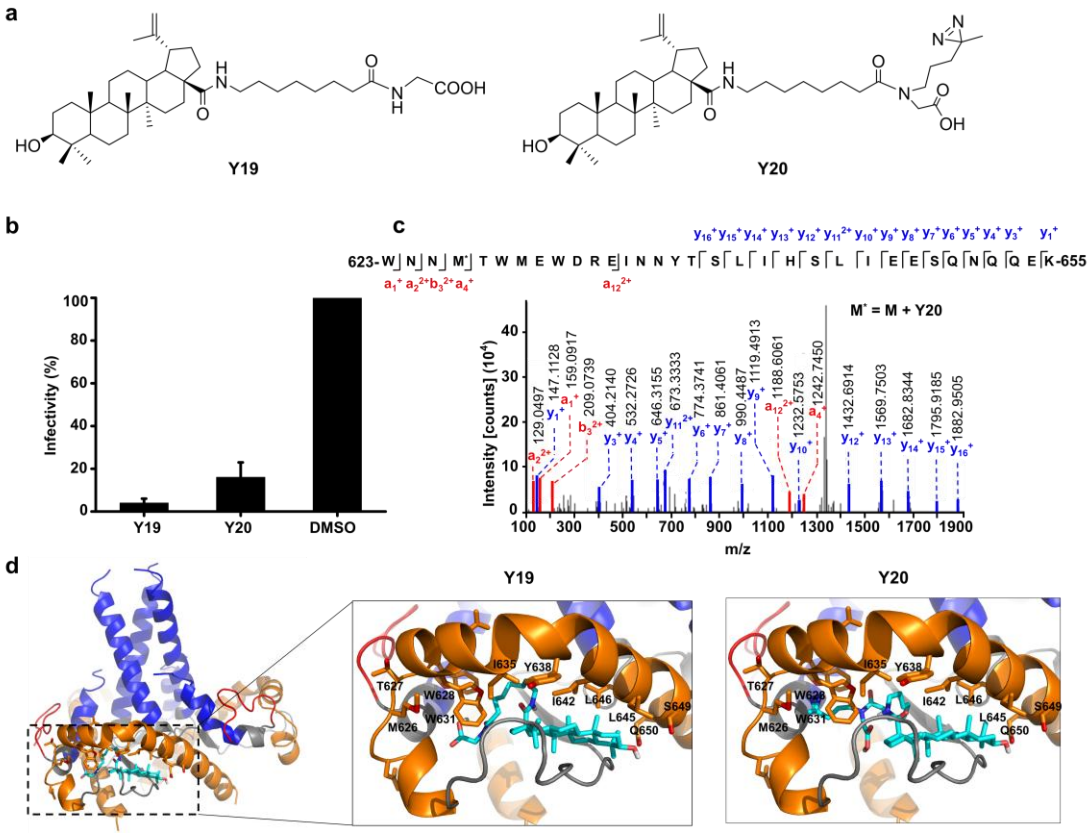


Figure 4. Identification of HIV GP41 HR2 as the target domain of the triterpenoid lead compounds. (**a**) Structures of the anti-HIV triterpenoid lead **Y19** and its photocrosslinking probe **Y20**. (**b**) Characterization of the inhibitory activity of the triterpenoids **Y19** and **Y20** on the entry of HIV-1 Env-pseudoviruses. CEM 4 was used as the host cell. (**c**) Identification of HR2 as the photocrosslinked domain by mass spectrometry. The HIV GP41 protein was photoaffinity labeled with the **Y20** probe and then analyzed by peptide mapping. A molecular weight increase of 772.5598, corresponding to photoactivated **Y20**, was observed for the peptide 623-WNNMTWMEWDREINNYTSLIHSLIEESQNQQEK-655. The crosslinking site at M626 was confirmed by the MALDI-MS analysis of tryptic digests. (**d**) Structural representations of **Y19** and **Y20** binding to the HIV GP41 HR2 domain (Protein Data Bank: 4TVP), as inferred from docking simulations. The -COOH group of the lead compounds lies in the polar pocket generated by D624, N625 and T627, and the 3-OH group of the lead compounds lies near the polar residues S649 and Q650. The diazirine group of **Y20** is in close proximity to M626, the residue that is crosslinked by **Y20**. Multiple hydrophobic interactions are observed between the lead compounds and their binding sites, namely, W628, W631, I635, Y638, I642, L645, and L646.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 January 2018 doi:10.20944/preprints201801.0177.v1

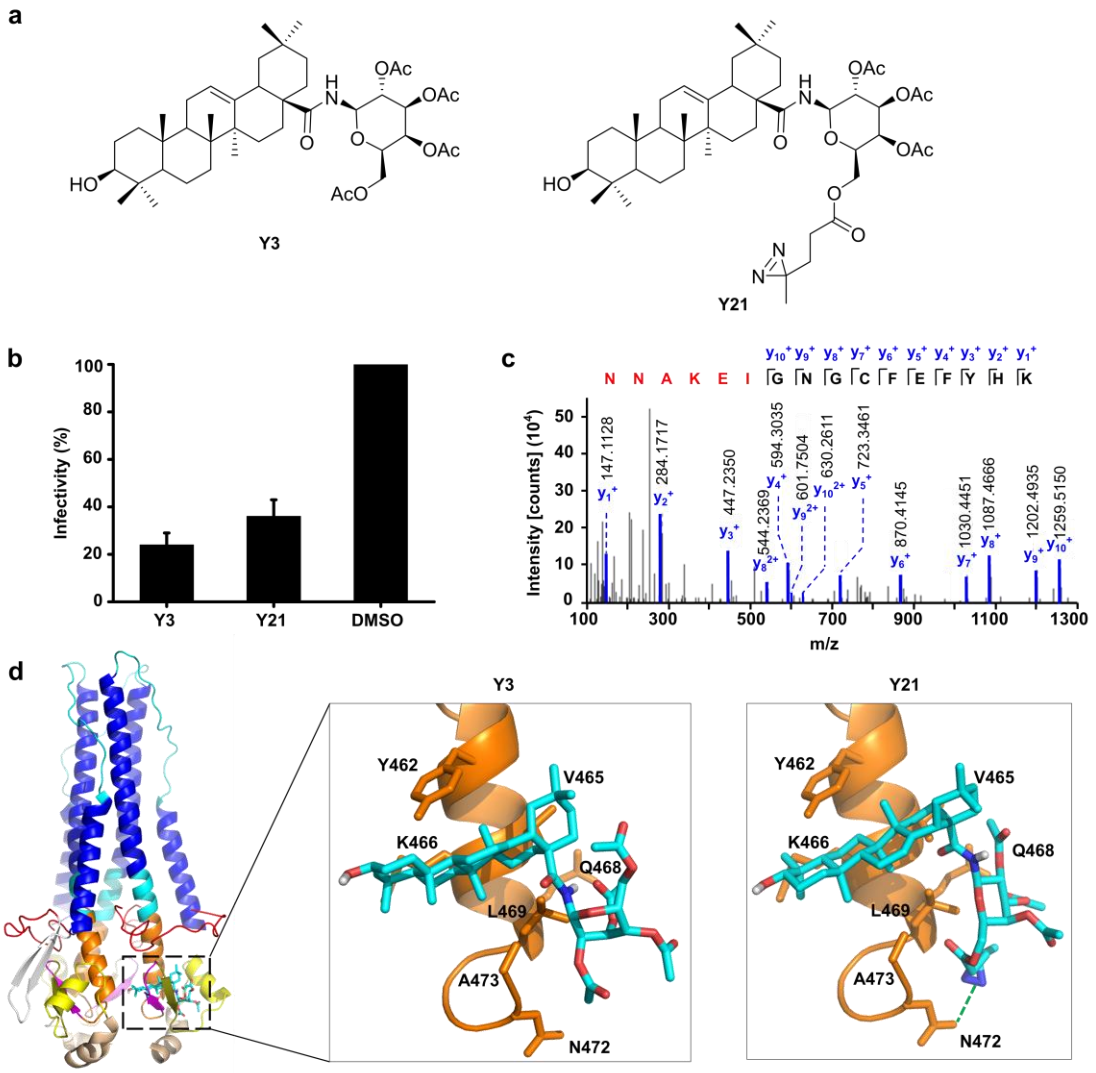


Figure 5. Identification of HR2 in influenza HA2 as the target domain for triterpenoid leads. (**a**) Structures of the anti-influenza triterpenoid lead **Y3** and its photocrosslinking probe **Y21**. (**b**) Characterization of the inhibitory activity of the triterpenoids **Y3** and **Y21** on the entry of influenza Env-pseudoviruses (IAVpp). A549 was used as the host cell. (**c**) Identification of HR2 as the photocrosslinked domain by mass spectrometry. The influenza HA protein was photoaffinity labeled with the **Y21** probe and then analyzed by peptide mapping. A molecular weight increase of 825.5027, corresponding to photoactivated **Y21**, was observed in the peptide 471-NNAKEIGNGCFEFYHK-486 (influenza A/California/04/2009 (H1N1)). The crosslinking site was within a residue between N471 to I476 based on the MALDI-MS analysis of tryptic digests. (**d**) Structural representations of **Y3** and **Y21** binding to the HR2 domain of influenza HA (Protein Data Bank: 3LZG) according to both the docking simulation and the alanine screening data. The amino acid residues that are involved in interactions with the lead compounds are labeled. The **Y21** diazirine group forms a hydrogen bond with N672, which is located near the crosslinking site.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 January 2018 doi:10.20944/preprints201801.0177.v1

Table 1. Mapping the constitutive residues of HR2 to elucidate their effects on viral infectivity and the potencies of the tested compounds

Variant	Infectivity (%)	IC_{50} (nM)		
		E-64d	Y11	Y18
K622A	19.1 ± 4.3	58.4 ± 2.3	174.9 ± 5.2	4495.0 ± 9.9
I623A	0.0 ± 0.1	ND	ND	ND
D624A	20.2 ± 1.7	70.4 ± 2.4	148.9 ± 5.3	772.1 ± 5.6
Q625A	18.8 ± 2.6	61.5 ± 4.6	179.9 ± 9.6	287.6 ± 9.6
I626A	0.1 ± 0.1	ND	ND	ND
I627A	21.3 ± 1.2	56.9 ± 5.8	549.8 ± 6.3	3751 ± 9.2
H628A	78.9 ± 5.0	75.9 ± 8.0	36.8 ± 5.3	450.7 ± 8.8
D629A	12.1 ± 3.6	85.0 ± 5.3	100.7 ± 5.7	1001.0 ± 9.9
F630A	4.6 ± 1.2	72.3 ± 7.6	428.7 ± 5.9	10199 ± 8.9
wild type	100	77.4 ± 5.6	59.2 ± 1.6	467.3 ± 10.2

Data are presented as the mean ± s.d. (n = 3). ND, not determined due to loss of infectivity upon mutation.