A rational design strategy and validation for protease-resistant fusion-inhibitor antiviral peptides

Kailu Yang[1,2,3,4,5], Chuchu Wang[1,2,3,4,5,6], Francesco Topi[7], Serena Muratcioglu[8], Ravi Kant[7,9,10], Tomas Strandin[7], Yih Tyng Bong[7], Lauri Kareinen[9,11], Sanna Maki[7], Saber H. Saber[12], Tobias Binder[7,14], Tarja Sironen[7,9], Subu Subramanian[8], K. Ian White[1,2,3,4,5], Richard A. Pfuetzner[1,2,3,4,5], Luis Esquivies[1,2,3,4,5], Olli Vapalahti[7,9,13], Merja Joensuu[14], John Kuriyan[8,15], Jussi Hepojoki[7], Giuseppe Balistreri[7,9,*], Axel T. Brunger[1,2,3,4,5,*]

[1] Department of Molecular and Cellular Physiology, Stanford University, Stanford, United States.
[2] Department of Neurology and Neurological Sciences, Stanford University, Stanford, United States.
[3] Department of Structural Biology, Stanford University, Stanford, United States.
[4] Department of Photon Science, Stanford University, Stanford, United States.
[5] Howard Hughes Medical Institute, Stanford University, Stanford, United States.
[6] Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore.
[7] Department of Virology, Faculty of Medicine, University of Helsinki, Helsinki, Finland.
[8] Department of Biochemistry, Vanderbilt University School of Medicine, Nashville, TN, United States.
[9] Department of Veterinary Biosciences, Faculty of Veterinary Medicine, University of Helsinki, Helsinki, Finland.
[10] Department of Tropical Parasitology, Institute of Maritime and Tropical Medicine, Medical University of Gdansk, 81-519 Gdynia, Poland.
[11] Finnish Food Authority, Helsinki, Finland.
[12] Queensland Brain Institute, University of Queensland, Brisbane, QLD, Australia.
[13] HUS Diagnostic Center, Clinical Microbiology, Helsinki University Hospital, University of Helsinki, 00029 Helsinki, Finland.
[14] Australian Institute for Bioengineering and Nanotechnology, University of Queensland, Brisbane, QLD, Australia.
[15] Department of Chemistry, Vanderbilt University, Nashville, TN, United States.

Emails: Kailu Yang, kailu@stanford.edu; Chuchu Wang, ccw0820@stanford.edu; Francesco Topi: francesco.topi@helsinki.fi; Serena Muratcioglu, serena.muratcioglu@vanderbilt.edu; Ravi Kant, ravi.kant@helsinki.fi; Tomas Strandin: tomas.strandin@helsinki.fi; Yih Tyng Bong: rita.bong@helsinki.fi; Lauri Kareinen, lauri.kareinen@helsinki.fi; Sanna Maki, sanna.z.maki@helsinki.fi; Saber H. Saber, s.abdelkader@uq.edu.au; Tobias Binder: tobias.binder@fau.de; Tarja Sironen, tarja.sironen@helsinki.fi; Subu Subramanian, subramaniansk@gmail.com; Kristopher White, kiwhite@stanford.edu; Richard A. Pfuetzner, rpfuetzn@stanford.edu; Luis Esquivies, lesquivi@stanford.edu; Olli Vapalahti, olli.vapalahti@helsinki.fi; Merja Joensuu, m.joensuu@uq.edu.au; John Kuriyan, john.kuriyan@vanderbilt.edu; Jussi Hepojoki, jussi.hepojoki@helsinki.fi; Giuseppe Balistreri, giuseppe.balistreri@helsinki.fi; Axel T. Brunger, brunger@stanford.edu

*Corresponding authors: Axel T. Brunger, 318 Campus Drive, Room E300, Stanford, CA 94305, United States, Phone: +1-650-736-1809, Email: brunger@stanford.edu; Giuseppe Balistreri, Room E353, Haartmaninkatu 3, 00290, Helsinki, Finland, Phone: +358-050-314-9167, Email: giuseppe.balistreri@helsinki.fi

Abstract

Peptide-based fusion inhibitors are promising pharmaceuticals in the fight against enveloped viruses relying on membrane fusion for host infection. However, peptide therapeutic applications have long been hindered by their poor stability *in vivo*. Here, we discovered that peptide inhibitors with the wildtype sequence of the heptad repeat 2 (HR2) domain of the SARS-CoV-2 spike protein are efficiently cleaved by Transmembrane Protease, Serine 2 (TMPRSS2), a key protease involved in the SARS-CoV-2 virus-cell fusion pathway. We then identified the corresponding cleavage sites and designed three protease-resistant peptides using ranking based on deep mutational scanning and natural occurrence. The three candidates all exhibit inhibitory activity in a cell-cell fusion assay. A high-resolution cryo-EM structure of the top candidate, HR2-NHN, bound to its HR1 target reveals the molecular basis for its potent activity. The top candidate of the cell-based screening assay significantly improved efficacy relative to the wildtype peptide when administered 12 h before infection in both an authentic virus-cell infection assay and a mouse assay. More broadly, our results suggest that the design strategies for protease-resistant peptides could be applied to a broad spectrum of other enveloped viruses and pave the way for the development of safe, prophylactic antivirals that can be administered before exposure.

Introduction

In enveloped viruses such as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), membrane fusion is necessary for viral entry into the host cell[1,2]. Importantly, a conformational change within the spike protein (S) drives membrane fusion[3,4], and peptide fragments of S derived from its heptad repeat 2 (HR2) motif bind its conserved heptad repeat 1 (HR1) motif[5], effectively inhibiting fusion[6,7]. Indeed, HR2-based peptides thus have great potential as antivirals since they can be easily administered, for example, intranasally[8] when an individual is expected to be at high risk of virus exposure. Furthermore, peptides without chemical modification are expected to exhibit minimal adverse effects, including liver toxicity, compared with traditional small-molecule drugs[9]. However, a major concern regarding the therapeutic use of these peptides is poor *in vivo* stability. While chemical modifications such as lipidation[10,11] and hydrocarbon stapling[12,13] are among the common ways to minimize degradation, such modifications involve costly synthesis. Alternatively, improved peptide stability could be achieved by mutating protease cleavage sites likely to be encountered upon administration to the airways.

We sought to improve the stability of unmodified inhibitory HR2 peptides for SARS-CoV-2 *in vivo*. We discovered that HR2-based inhibitor peptides are efficiently cleaved by Transmembrane Protease, Serine 2 (TMPRSS2), a key protease that cleaves S at the S2' site, allowing S to transition to the post-fusion state. Based on this discovery, we hypothesized that the *in vivo* stability of HR2 peptides could be improved by mutating TMPRSS2 cleavage sites, since the peptides would likely encounter TMPRSS2-expressing host cells upon administration. After identifying the cleavage sites by mass spectrometry, we designed three triple mutant HR2 peptides resistant not only to TMPRSS2 cleavage but also to cleavage by the entire trypsin-like protease family, using rankings from both deep mutational scanning and natural occurrence. All three protease-resistant design candidates exhibit inhibitory activity in a cell-cell fusion assay. We determined the high-resolution cryo-EM structure of the top candidate with the highest inhibitory activity (referred to as HR2-NHN) bound to HR1; this structure explains the effect of a key mutation, R1185H, on binding. Finally, tests in an authentic SARS-CoV-2 cell-infection assay and a mouse assay both validate HR2-NHN as an equally potent and more stable inhibitor compared to previously studied wildtype HR2 peptides.

Our study thus offers compelling new opportunities for prophylaxis or for ameliorating symptoms if administered soon after infection with SARS-CoV-2, which continues to evolve and spread, with two major variants (XFG and NB.1.8.1) currently under monitoring by the WHO. Variants of SARS-CoV-2 continue to challenge presently available therapies and pose a substantial health threat, including Long COVID[14]. Although these variants appear to have limited immune escape from the latest mRNA booster vaccines, an urgent need remains for orthogonal antiviral drugs that target relatively conserved regions of the virus. Such a strategy would lead to drugs that would remain effective against new variants, avoid large-scale immunization, be easily administered to prevent infection in advance or alleviate symptoms after infection, and have minimal side effects. More broadly, given widespread expression of trypsin-like proteases at mucosal and epithelial barriers[15-17] and universal presence of basic residues within HR2-equivalent regions of many type I enveloped viruses (**Fig. S1**)[18-32], our results inspire re-design of protease-resistant versions against a broad spectrum of other deadly viruses including Ebola virus[19], Nipah virus[24], respiratory syncytial virus (RSV)[20], and human immunodeficiency virus (HIV) [18].

Results

Widely used wildtype HR2 peptides are highly susceptible to TMPRSS2 cleavage, which abolishes their inhibitory activity

HR2 peptides inhibit SARS-CoV-2 infection, presumably by binding to the pre-hairpin intermediate of S following cleavage of S at the S1/S2 and the S2′ sites mainly by furin and TMPRSS2[6,33-35]. Although TMPRSS2 cleavage of the HR2 domain of the full-length S protein is not reported in any of its conformational states, we speculated that TMPRSS2 might cleave individual HR2 inhibitor peptides, considering that they are monomeric[6], more accessible, and less structured than the HR2 domain within the context of full-length S. Therefore, we assessed the cleavage of three previously studied (**Fig. 1a**)[6,36] fluorescently labeled HR2 peptides (42G_FL, longHR2_FL, shortHR2_FL) by TMPRSS2 using clear-native polyacrylamide gel electrophoresis (CN-PAGE) (**Fig. 1b**). All HR2 peptides tested were completely cleaved by TMPRSS2 after 1 h incubation at 37 °C, and cleavage was blocked by the serine protease inhibitor phenylmethylsulfonyl fluoride (PMSF) (**Fig. 1b**). Mass spectrometry-based proteomics revealed three fragments with two definitive cleavage sites at R1185 and K1191 and another possible cleavage site at K1181 (**Figs. S2 and S3**). We then asked if TMPRSS2-cleaved HR2 peptides could still inhibit SARS-CoV-2 fusion, using a previously established cell-cell fusion assay[6,37]. In this assay, ACE2 and S-expressing HEK cells were mixed, and content mixing was assessed with α-complementation of *E. coli* β-galactosidase. At variance to our previous studies, we developed a modified protocol to prepare solubilized peptides from the lyophilized state to achieve higher purity and more reliable results (see **Materials and Methods**). We found that after TMPRSS2 treatment, longHR2_42 loses inhibitory activity against the fusion mediated by the S protein of SARS-CoV-2 (**Fig. 1c**).

a

```
42G_FL:       1157KNHT SPDVD LGGDIS GINAS VVNIQ KEIDR LNEVA KNLNE SLIDL Q1201C_fluorescein
longHR2_FL:   1157KNHT SPDVD LG DIS GINAS VVNIQ KEIDR LNEVA KNLNE SLIDL Q1201C_fluorescein
shortHR2_FL:                1168DIS GINAS VVNIQ KEIDR LNEVA KNLNE SLIDL Q1201C_fluorescein
longHR2_42:          1162PDVD LG DIS GINAS VVNIQ KEIDR LNEVA KNLNE SLIDL QEL1203
shortHR2:                   1168DIS GINAS VVNIQ KEIDR LNEVA KNLNE SLIDL QEL1203
controlHR2:                 1168DIS GINAS VVNIQ KEIDR YNEVA KNWNE SLIDL QEL1203
HR2-HN:              1162PDVD LG DIS GINAS VVNIQ KEIDH LNEVA NNLNE SLIDL QEL1203
HR2-QHN:             1162PDVD LG DIS GINAS VVNIQ QEIDH LNEVA NNLNE SLIDL QEL1203
HR2-QHN_FL:   1157KNHT SPDVD LG DIS GINAS VVNIQ QEIDH LNEVA NNLNE SLIDL Q1201C_fluorescein
HR2-NNS:             1162PDVD LG DIS GINAS VVNIQ NEIDN LNEVA SNLNE SLIDL QEL1203
HR2-NNS_FL:   1157KNHT SPDVD LG DIS GINAS VVNIQ NEIDN LNEVA SNLNE SLIDL Q1201C_fluorescein
HR2-NHN:             1162PDVD LG DIS GINAS VVNIQ NEIDH LNEVA NNLNE SLIDL QEL1203
HR2-NHN_FL:   1157KNHT SPDVD LG DIS GINAS VVNIQ NEIDH LNEVA NNLNE SLIDL Q1201C_fluorescein
```

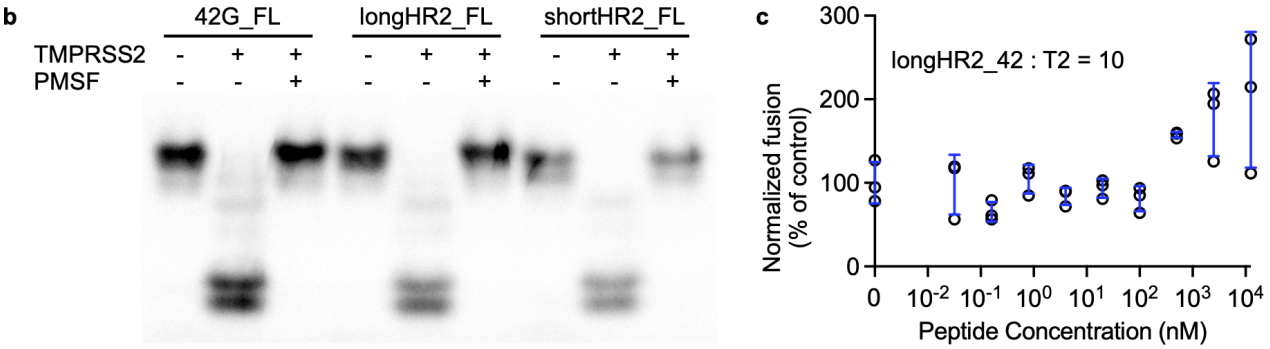


Fig. 1. TMPRSS2 cleaves HR2 peptides and abolishes their inhibitory activity.
(**a**) Primary amino acid sequences of HR2 peptides reported previously[6,36] or in this study. The suffix "FL" indicates peptide labeling with fluorescein. For all other peptides without labeling, any difference from the wildtype sequence is indicated by an underscore and bold font.
(**b**) TMPRSS2 cleaves HR2 peptides as assessed by CN-PAGE fluorescence imaging.
(**c**) Following TMPRSS2 cleavage, longHR2_42 does not inhibit fusion mediated by S, as assessed by a cell-cell fusion assay. The molar ratio between peptide and TMPRSS2 (T2) is marked in the panel. The raw data points are plotted as black circles, while the error bars (SD) are plotted in blue.

Design of protease-resistant mutants based on deep mutational scanning and natural occurrence

Given these proteolysis results, we sought to design protease-resistant HR2 peptides. However, increasing peptide stability by mutagenesis potentially risks loss of antiviral efficacy, and the search space of possible mutations at three sites is large ($20^3 = 8000$ sequences, including wildtype). To address these challenges, we used a two-pronged strategy to design stabilizing mutations at all three TMPRSS2-cleavage sites of HR2. First, we performed deep mutational scanning of the basic residues at the cleavage sites (K1181, R1185, and K1191) to rank all possible HR2 mutants by their ability to bind HR1 *in vitro* (**Fig. 2a and b**). In this approach, a GFP-fused HR1 peptide is used as a bait to fish out the stronger binders from a library of *E. coli* cells that expose a mutant HR2 library at their surface; this mutant library includes all possible single mutations in the three basic residues of the HR2 peptide. After fluorescence-assisted cell sorting (FACS), the bacteria that display the strongest fluorescence—i.e., the strongest binding to HR1-GFP—are subjected to deep sequencing to rank the most highly enriched mutations. Second, we surveyed the SARS-CoV-2 spike variants database[38,39] to rank naturally occurring mutations of the three basic residues in the context of the full SARS-CoV-2 S protein (K1181, R1185, and K1191) (**Fig. 2c**). Among the ranked lists of possible mutations to non-basic residues (**Fig. 2, b and c**) we excluded mutations to hydrophobic residues to avoid potential solubility issues, and mutations to cysteine to avoid internal disulfide bonds or more extensive oxidative oligomerization. The deep mutational scanning data (**Fig. 2b**) suggest K1181N, R1185N, and

K1191S (referred to as HR2-NNS, **Fig. 1a**) as a design candidate. Note that the introduction of most mutations at these positions was beneficial for HR1 binding; the majority of the HR2 variants bound more tightly than the wildtype peptide. On the other hand, the natural occurrence (**Fig. 2c**) suggests K1181N, R1185H, and K1191N (referred to as HR2-NHN, **Fig. 1a**) as a design candidate. In addition, we selected K1181Q, R1185H, and K1191N (referred to as HR2-QHN; **Fig. 1a**) as potential candidates (excluding the naturally occurring basic arginine and hydrophobic isoleucine at the K1181 position).

We prepared all three peptides (quality control data shown in **Figs. S4 to S6**) and verified that they are indeed resistant to TMPRSS2 cleavage (**Fig. 3a**).

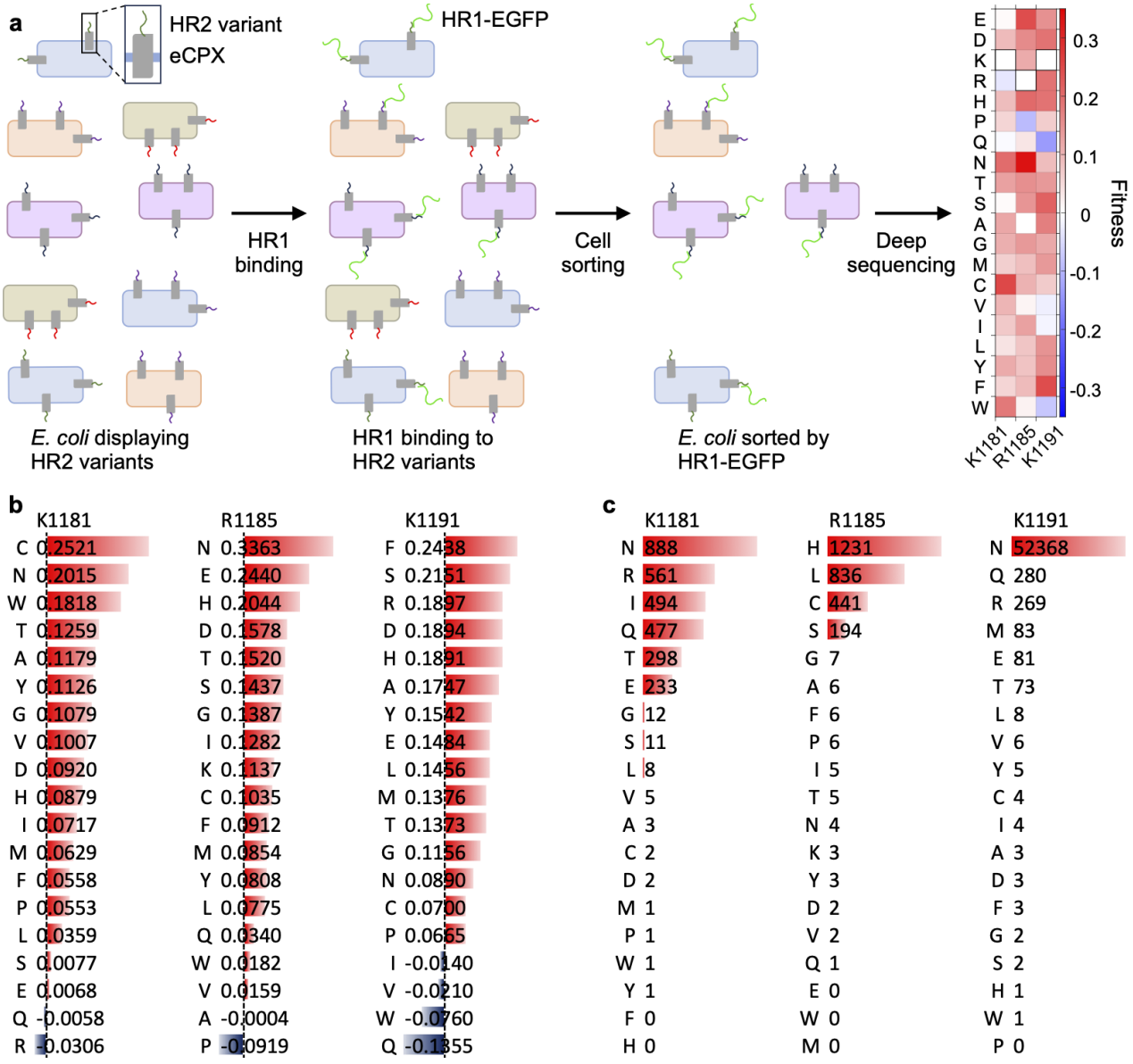


Fig. 2. Ranking of mutations based on deep mutational scanning and natural occurrence. (**a**) Schematic representation of the high-throughput platform screening for HR1 binding. A library of *E. coli* cells displaying HR2 variants is mixed with GFP-labelled HR1 peptides to allow binding between HR1 and HR2. The cells are ranked based on GFP fluorescence and subjected to deep sequencing. Enrichment is calculated as the ratio of the number of sequencing reads for each variant in the sorted and unsorted populations (see the Methods section for details). All

enrichment values are $\log_{10}$-transformed and normalized relative to the wildtype sequence, which has a value of 0. The effect of single amino acid substitutions of the three basic residues (K1181, R1185, and K1191) on HR1 binding is shown. Red and blue colors indicate that the specific mutant binds better or worse, respectively, than the wildtype sequence.
(**b**) Enrichment rank for the three basic residues susceptible to TMPRSS2 cleavage of every single amino acid substitution ranked by relative enrichment, with the most beneficial mutations shown first.
(**c**) All possible mutations for the three basic residues susceptible to TMPRSS2 cleavage ranked by natural occurrence. The number denotes the number of observations of each mutation in the GISAID-derived SARS-CoV-2 spike variants database[38,39].

Screening of HR2 mutants by a cell-cell fusion assay

To assess the efficacy of the three designed peptides, we first tested their capacity to inhibit cell-cell fusion in the cell-cell fusion assay[6,37] in the presence of TMPRSS2. We found the HR2-NHN peptide to have the highest inhibitory activity, with the other two peptides close behind (**Fig. 3, b to d**).

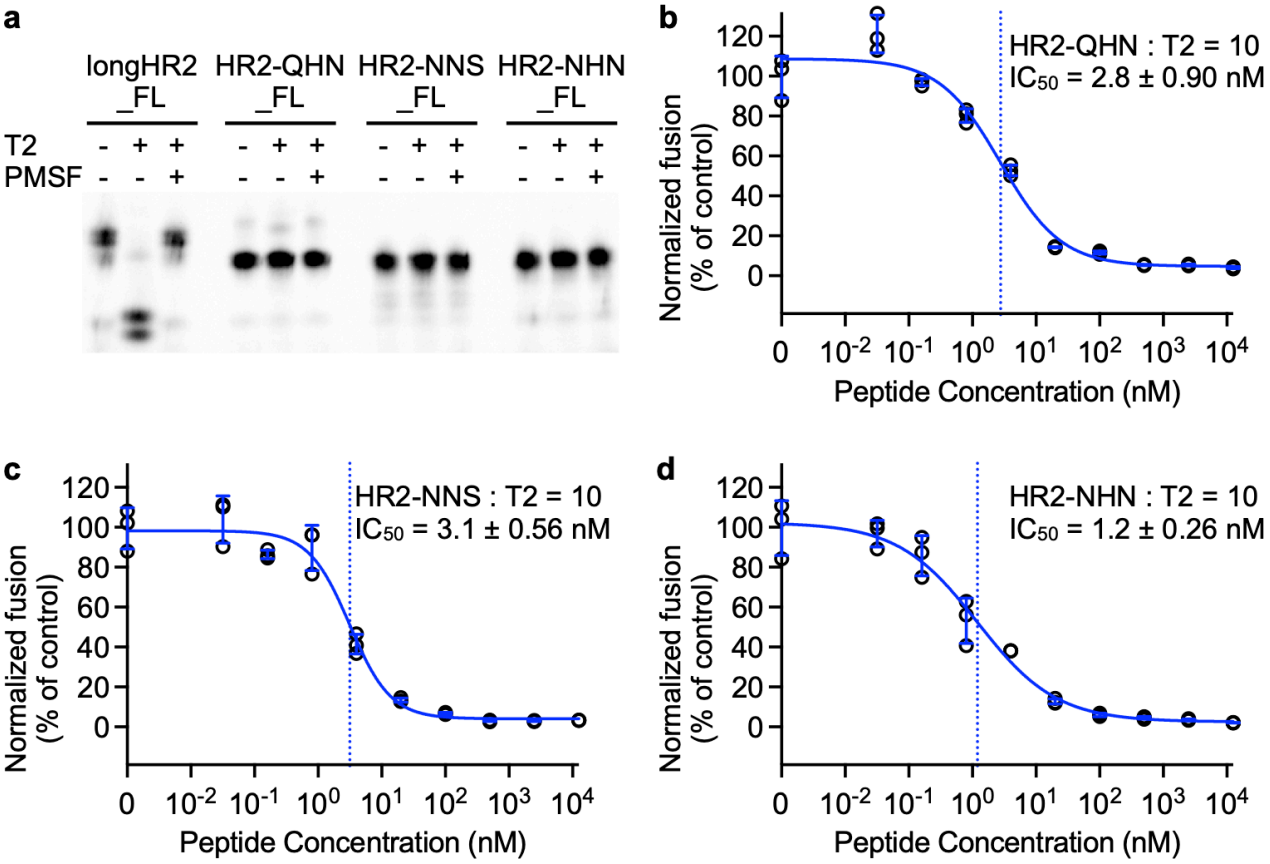


Fig. 3. Cleavage assay and screening of the designed peptides.
(**a**) The wildtype peptide (longHR2_FL) and the three newly designed mutant peptides (HR2-QHN_FL, HR2-NNS_FL, and HR2-NHN_FL, see **Fig. 1a** for their sequences) are labeled, subjected to TMPRSS2 (T2) cleavage, and then analyzed by CN-PAGE fluorescence imaging (see Methods section for details).
(**b–d**) Inhibitory activities of HR2-QHN (**b**), HR2-NNS (**c**), and HR2-NHN (**d**) in the cell-cell fusion assay. The molar ratio between peptide and TMPRSS2 (T2) is marked in the panel. The raw data points are plotted as black circles, while the error bars (SD), fitted curves, and vertical dashed lines at IC_{50} are plotted in blue.

The R1185H mutation disrupts a salt bridge but drives formation of an additional hydrogen bond between HR1 and HR2

Given the finding that the designed HR2 peptides bind HR1 so tightly, we sought to determine the molecular basis of these interactions, particularly because the basic R1185 of HR2 was previously shown to form a salt bridge with the acidic D936 of HR1. To do so, we performed single-particle cryo-EM of HR1 bound to the best-performing HR2-NHN using a previously reported molecular scaffolding method [40] (**Fig. S7, Table S1**). The scaffolding method links the trimeric N termini of four HR1 fragments to four trimeric C termini of the Dps4 dodecamer from Nostoc punctiforme, and then complexes with HR2 peptides, enabling structure determination of the small HR1—HR2 complex by single-particle cryo-EM. While the R1185H mutation disrupts the salt bridge between R1185 of HR2 and D936 of HR1, the histidine side chain forms a compensatory hydrogen bond with S940 of HR1 (**Fig. 4**). In contrast, K1181 and K1191 do not show well-ordered sidechain density, consistent with the fact that they are surface-exposed and do not interact with HR1. After mutating these two lysine residues to asparagine, no new interaction with HR1 is formed, although the sidechain densities of the two asparagine residues are well-defined.

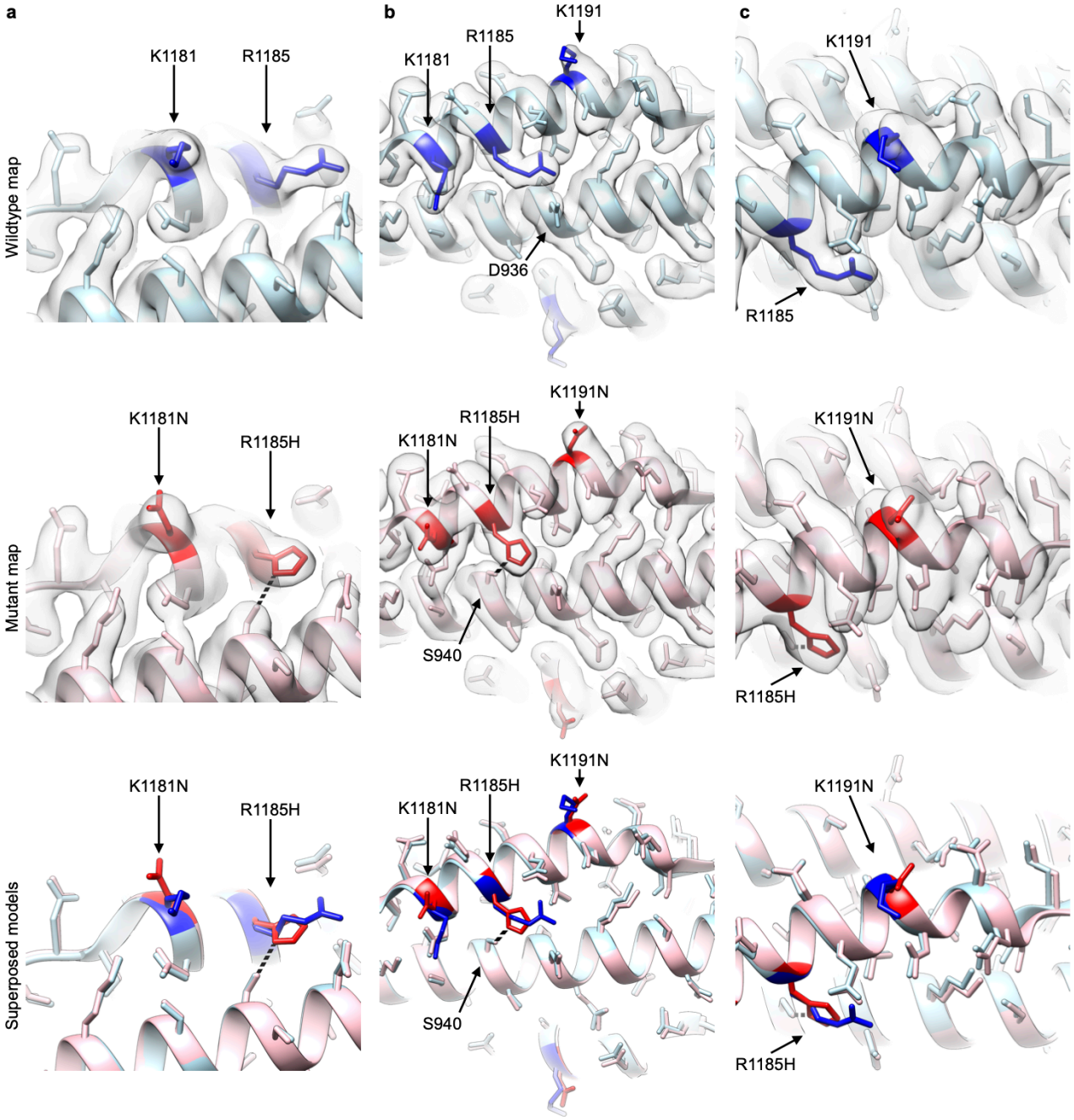


Fig. 4. **Structural insights into how the three mutations affect interaction with HR1.** Comparison of cryo-EM structures of the wildtype longHR2_42 and HR2-NHN bound to HR1 in three different views. Panel (**a**) is focused on K1181N, panel (**b**) is focused on the entire helix of HR2, and panel (**c**) is focused on K1191N, with R1185H visible nearby in all three panels. The wildtype cryo-EM structure (light blue) and map are shown in the top row, the HR2-NHN (pink) structure and map are displayed in the middle row, and the superposed wildtype and HR2-NHN models without maps are displayed in the bottom row. The mutated residues are colored blue for the wildtype structure and red for the mutant structures. The hydrogen bonds between R1185 and S940 are shown as black dashed lines in the middle and bottom rows for panels (**a–c**).

The TMPRSS2-resistant HR2-NHN maintains inhibitory activity longer than the wildtype longHR2_42 in an authentic SARS-CoV-2–cell infection assay

Protease-resistant peptides are expected to have a longer bioavailability when proteases such as TMPRSS2 or cathepsins are present at the cell surface or in the endosomal compartment of cells

in vitro or *in vivo*. To test if the antiviral activity of the best-performing TMPRSS2-resistant peptide HR2-NHN persisted longer than the original peptide longHR2_42, we preincubated Calu-3 cells, a lung-derived human line endogenously expressing ACE2 and TMPRSS2[41], with a serial dilution of the respective peptides for 72 h or 0 h (*i.e.*, just before infection) followed by infection for 18 h with a recombinant SARS-CoV-2 Wuhan (Wuh) strain expressing the mCherry red fluorescent protein[42] for detection by high-content fluorescence microscopy (**Fig. 5**). When administered just before infection (0 h), both peptides inhibited infection in a dose dependent manner and with similar IC_{50}s (**Fig. 5, a-b**). Pre-incubation with the cells for 72 h prior to infection resulted in a decrease in the antiviral activity of the original longHR2_42 peptide, the IC_{50} of which shifted from 2.6 nM to 8.9 nM, but not for the protease-resistant HR2-NHN, which retained its antiviral activity (IC_{50} 1.9 nM) (**Fig. 5, c-d**). Statistical analysis of the area under the curve (AUC) confirmed that this difference was significant (**Fig. S8, a-b**). Both peptides also inhibited the Omicron variant Xbb1.5[43] , but with lower efficacy than the Wuh strain (**Fig. 5, e-f, and S8 c**). The decreased IC_{50}s against Omicron were expected, as we have previously shown that mutations in the HR1 region of the Omicron Spike protein impair binding of longHR2_42[36].

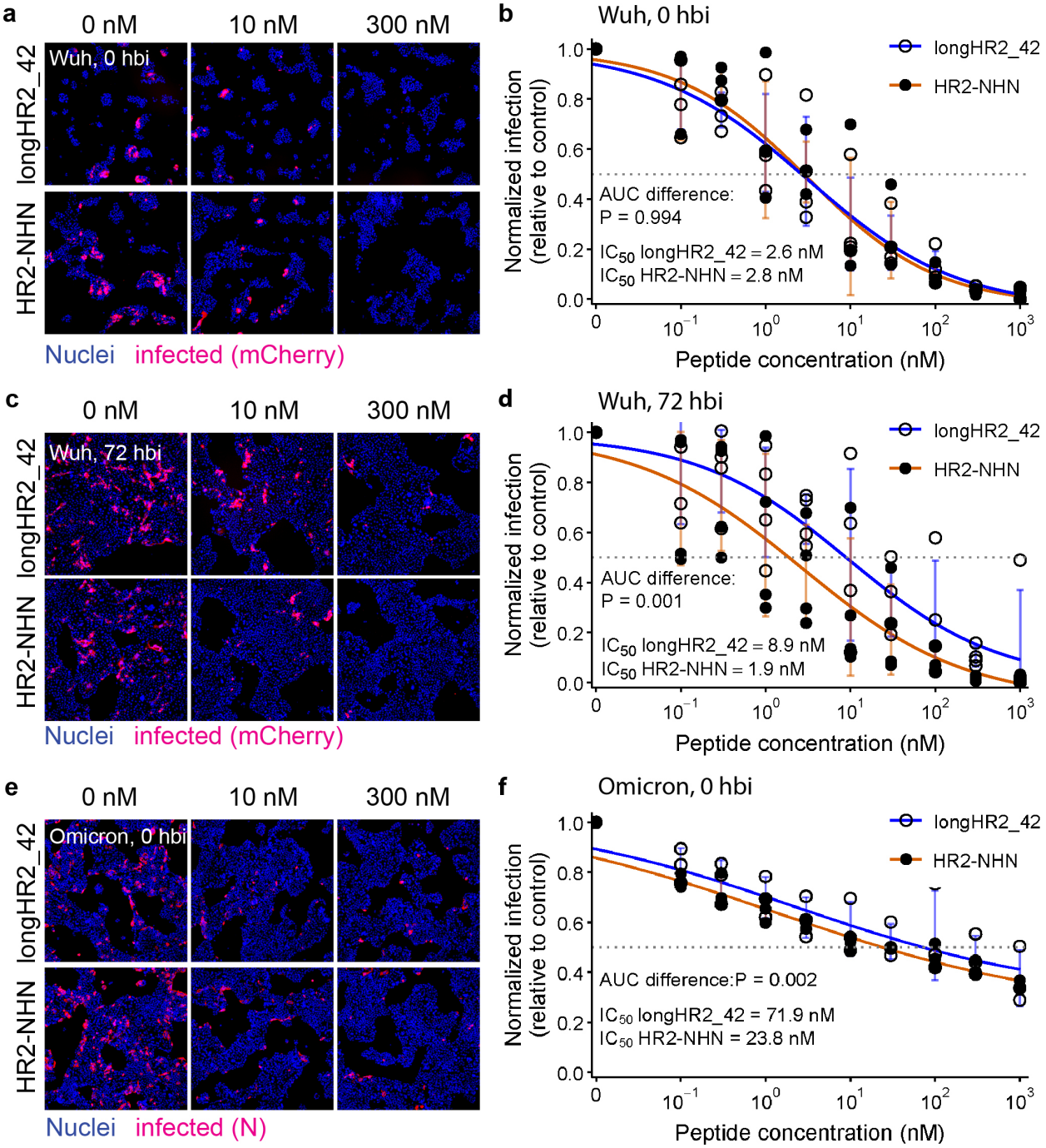


Fig. 5. The TMPRSS2-resistant HR2-NHN maintains inhibitory activity longer than wildtype longHR2_42 in an authentic SARS-CoV-2 cell infection assay.
Representative fluorescence images (**a, c, e**) and quantification after automated image analysis (**b, d, f**) of Calu-3 cells treated with indicated doses of longHR2_42 or HR2-NHN at 0 hours (**a-b and e-f**) or 72 hours (**c-d**) before infection (hbi) for 18 hours with SARS-CoV-2-mCherry Wuhan strain (**a-d**) or Omicron Xbb1.5 (**e-f**). The estimated IC_{50} of the two peptides after different preincubation times and the statistical significance (*P*) of the difference between the Area Under the Curve (AUC) for each peptide are indicated in the graphs (see also **Fig. S8**). Individual points represent independent repetitions; error bars indicate SD. Solid lines show the fitted dose-response curves. The horizontal dotted line denotes normalized infection of 0.5. Blue curves, longHR2_42; orange curves, HR2-NHN. IC_{50}, curve fitting, respective AUCs, and statistical analysis were performed in R using the bootstrap resampling method for dose and time responses.

Supporting the results obtained in Calu-3 cells, longer bioavailability for HR2-NHN compared to longHR2_42 was also observed in A549 cells stably over-expressing human ACE2 and TMPRSS2 (named A549-AT)[44] (**Fig. S8**). Over 50% of the inhibitory activity of the protein-resistant peptide HR2-NHN against SARS-CoV-2 Wuh was retained following pre-incubation for 12 h, while, in marked contrast, the inhibitory activity of the wildtype longHR2_42 is completely lost for the same period of pre-incubation (**Fig. S8, a and b**). With a pre-incubation period of 24 h, both peptides lose all inhibitory activity, probably because the peptides are either endocytosed into the cells after 24 h or degraded by multiple host proteases in A549-AT cells.

The TMPRSS2-resistant HR2-NHN maintains antiviral activity longer than the wildtype longHR2_42 in SARS-CoV-2-infected mice

To test if HR2-NHN also had improved bioavailability *in vivo* compared to longHR2_42, we measured their activity in wildtype BALB/c female mice with the beta variant of SARS-CoV-2, which infects wildtype mice and causes lung infection within 24 h[45].

First, to determine the optimal peptide concentration to use *in vivo* before a direct comparison of longHR2_42 and HR2-NHN, we tested in mice (4 mice per treatment group) two concentrations of longHR2_42, 1 mg/kg and 0.1 mg/kg and administered the peptide just before infection (0 hbi, 1 mg/kg and 0.1 mg/kg), 3 hours before infection (3 hbi, 1 mg/kg), or 3 hours post infection (3hpi, 1 mg/kg) (**Fig. 6a**). The peptides (or PBS control) and the virus (10^5 pfu/mouse) were administered intranasally as 20µl intranasal inoculations. In this first test, a second inoculation of the peptide was performed at 24 h post infection (hpi) (**Fig. 6a**), and after anesthesia at 48 hpi, mice were culled and RNA extracted for qRT-PCR analysis (**Fig. 6a**, end).

When the longHR2_42 was administered intranasally at 1 mg/kg at 0 hbi (*i.e.*, few seconds before intranasal virus inoculation), infection in the lungs was strongly inhibited as determined by qRT-PCR of RNA extracted from the lungs at 48 hpi (**Fig. 6b**, 0 hbi, 1 mg/kg). This inhibition was dose-dependent, and at a dose of 0.1 mg/kg the inhibitory effect was no longer significant (**Fig. 6b**, 0 hbi, 0.1 mg/kg). Significant inhibitory activity was observed at a concentration of 1 mg/kg when the longHR2_42 peptide was administered prophylactically 3 hbi (**Fig. 6b**, 3 hbi, 1 mg/kg). To assess whether longHR2_42 would have antiviral activity when administered after virus infection (i.e., therapeutic activity), the peptide was administered at 3 h post-virus inoculation. Although to a lesser extent compared to prophylactic administration, significant antiviral activity was still observed (**Fig. 6b**, 3 hpi).

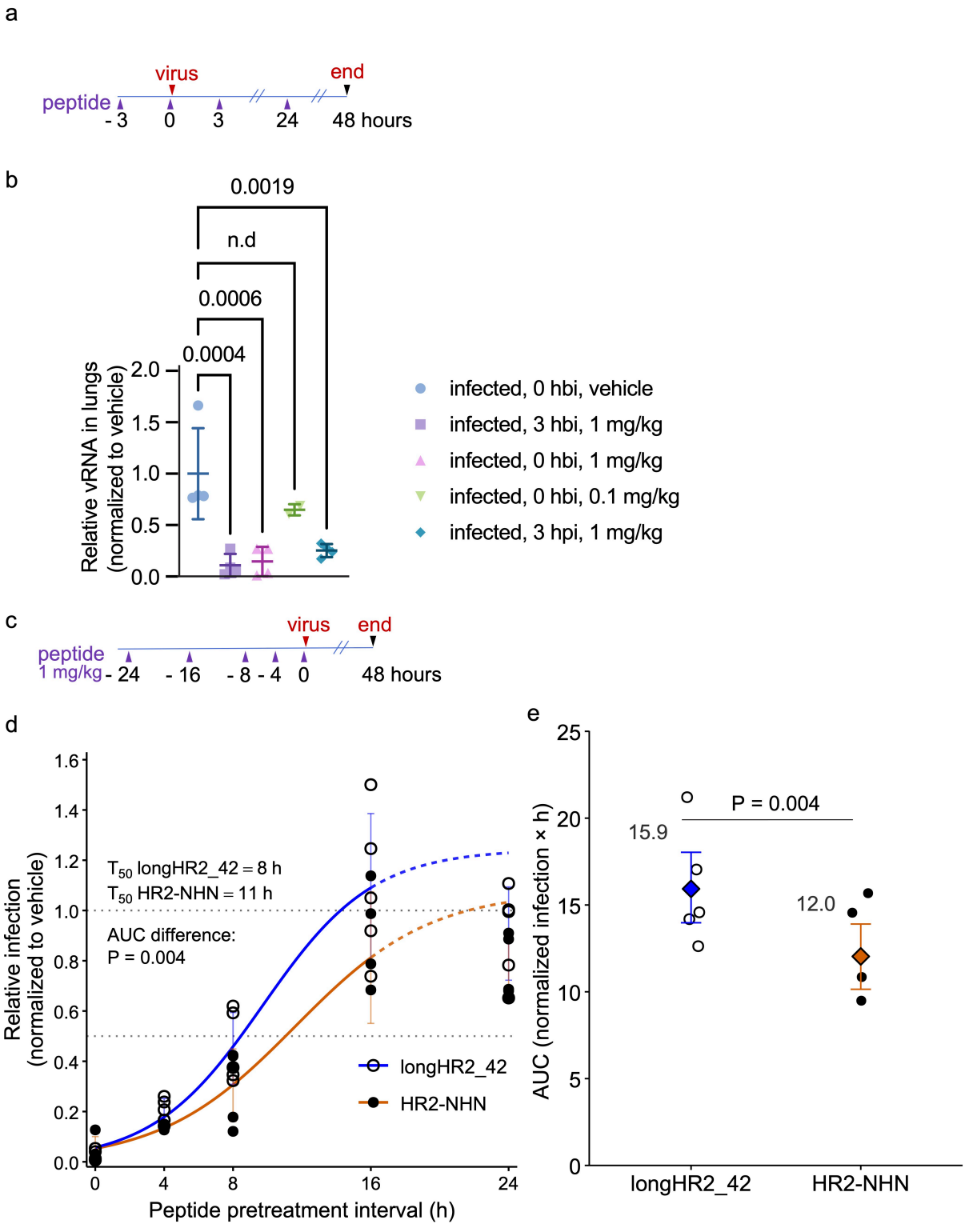


Figure 6. HR2-NHN peptide maintains antiviral activity longer than the wildtype longHR2_42 in SARS-CoV-2 infected mice.

(**a**) Workflow of *in vivo* experiment identifying a protective dosing regimen for longHR2_42. (**b**) Relative lung viral RNA (vRNA), normalized to the mean of the vehicle-treated group (defined as

1.0), for mice treated with vehicle or longHR2_42 at the indicated pretreatment or post-infection time and doses. Individual points represent individual mice; the group mean and SD are indicated for each treatment; hbi, hours before infection; hpi, hours post-infection. The P values indicating the statistical significance of each sample relative to vehicle-control treated mice ii indicated in the graph. Statistical analysis was performed with Ordinary one-way Anova Dunnett's multiple comparison test. For the mouse treatment "0 h 0.1 mg/kg", two of the four mice died during anesthesia prior to infection; thus, statistical analysis was not performed for these mice (n.d.). (**c**) Workflow of an *in vivo* experiment to determine the duration of antiviral activity of the peptides. (**d**) Re-infection assay with viruses extracted from lungs of mice pretreated with peptides at indicated times used to infect Vero-E6 cells for 18 hours. Individual points represent infections with viruses extracted from individual mice. Error bars indicate SD. Infection was normalized to the vehicle-treated controls (defined as 1.0). Solid lines show the fitted time-response curves over 0–16 h; dashed lines indicate extrapolation beyond the fitted interval to 24 h. Horizontal dotted lines denote normalized infection values of 1.0 and 0.5. Estimated T_{50} values and the exact P value for the 0–24 h AUC comparison are indicated. n = 5 mice per peptide at each time point. Blue, longHR2_42; orange, HR2-NHN. **(e)** Area under the curve AUC was calculated from the observed mean normalized infection values from each time point in (**d**). Diamonds indicate the group AUC estimates and error bars indicate bootstrap 95% confidence intervals. The exact bootstrap P value for the comparison between longHR2_42 and HR2-NHN is shown. Blue, longHR2_42; orange, HR2-NHN.

As a first comparison between the original and the TMPRSS2-resistant peptide, mice were pretreated at 0, 12, and 24 hours before infection at a concentration of 1 mg/kg, as a single dose at the indicated times. The second dose at 24 hpi was not administered. The extent of lung infection was determined by qRT-PCR at 48 hpi as described above (**Fig. S9a**). Immunohistochemistry of mouse lungs from each treatment group was used to confirm the PCR results (**Fig. S9b**). Confirming the *in vitro* results, the antiviral activity of HR2-NHN was significantly higher than longHR2_42 when the peptides were administered 12 h before infection, as assessed by qRT-PCR (**Fig. S9a**, -12 hbi) and confirmed by a qualitative reduction in N protein staining (**Fig. S9b**, -12 hbi). However, when administered at a single dose 24 hbi, both peptides had lost antiviral activity (**Fig. S9b**, -24 hbi).

To estimate the half-life of the two peptides' antiviral activity, a more extensive time course (0, -4, -8, -16, -24 hbi) was performed in mice (**Fig. 6c**). To increase assay sensitivity, we reduced the peptide concentration to 0.5 mg/kg. While quantification of viral RNA by qRT-PCR is a sensitive method, it does not distinguish between intracellular viral genomes (or RNA transcripts) and extracellular infectious viruses, which are relevant to the spread of infection in the respiratory tract and to host-to-host transmission. To assess the extent of residual infectious viruses in the lungs of mice infected and treated with either vehicle control or peptides, at 48 hpi, mice were sacrificed, and lung homogenates were used to re-infect highly susceptible Vero-E6 cells. As described above for the cell culture assays, at 18 hpi, infection rates were quantified by immunofluorescence with antibodies raised against the viral N protein, followed by automated fluorescence imaging and image analysis. Under these conditions, the apparent $T_{1/2}$ of the original peptide longHR2-42 was approximately 8.45 h, while that of the protease-resistant HR2-NHN was 11h (**Fig. 6d**). Statistical analysis of the AUC confirmed that these differences were statistically significant (**Fig. 6e**). If administered more than 16 h prior to infection, none of the peptides retained inhibitory effects (**Fig. 6d** and **Fig. S9c**).

Discussion

Peptide therapeutics are notorious for their poor stability *in vivo*, and efforts to improve their stability and bioavailability are thus critical. Here, we describe a strategy to achieve this in the context of proof-of-concept peptide therapeutics targeting SARS-CoV-2. We observed that wildtype HR2-based inhibitors can be efficiently cleaved by TMPRSS2 (**Fig. 1b**), a key protease in the infection process. We therefore developed TMPRSS2-resistant peptides with vastly improved protease resistance and longer-lived potency through a strategy drawing on deep mutation scanning (**Fig. 2, a and b**) and natural occurrence (**Fig. 2c**), rationalized binding through structural studies with cryo-EM (**Fig. 4 and Fig. S7**) and tested our designs using multiple *in vitro* and *in vivo* approaches (**Figs. 3, 5, and 6**). The most efficacious protease-resistant peptide, HR2-NHN, is an HR2-based inhibitor with improved bioavailability that resists TMPRSS2 cleavage and outperforms its wildtype predecessor (longHR2_42) when administered 12 hours before infection in a mouse assay (**Fig. 4**).

We used two ranking schemes, deep mutational scanning and surveying of natural occurrence (**Fig. 3, b and c**), for the design of three TMPRSS2-cleavage resistant peptides, HR2-QHN, HR2-NNS, and HR2-NHN (**Fig. 1a**). All three designs exhibit inhibitory activity (**Fig. 3**). The peptide HR2-NHN is the top candidate from the natural occurrence approach, and HR2-NNS is the top candidate from the deep mutagenesis approach. The only difference between these two top designs is the choice of asparagine or histidine at position 1185, and the histidine mutant exhibits higher inhibitory activity. Combining different ranking schemes increases the exploration of possible mutants. In this context, we note the features of deep mutational scanning. First, it focuses on the binding between HR1 and HR2 and is not biased by other restraints that likely dominate some mutations in natural occurrence. For example, some frequently occurring mutations may be advantageous for prefusion conformations of the spike protein, or simply for improved viral genome stability, rather than for tighter binding between HR1 and HR2. Second, it theoretically works for any virus with a similar entry mechanism, even with limited sequencing data for a novel or rare virus.

Furthermore, although the peptides were designed to resist cleavage by TMPRSS2 specifically, we mutated all three basic residues. Therefore, the resultant HR2-NHN peptide is likely resistant to the entire trypsin-like protease family, which requires a basic residue at the P1 position for cleavage[46]. This could generally enhance bioavailability and help explain the contrast between HR2-NHN and wild-type longHR2_42 when preincubated with cells at different times before infection and when delivered intranasally in a mouse model of SARS-CoV-2 infection, even with a single, relatively low dose of 1 mg/kg.

Our *in vivo* results indicate that intranasal delivery of monomeric, chemically unmodified fusion inhibitory peptides has both prophylactic and therapeutic potential. The described approach to confer protease resistance through rational mutagenesis extended the half-life of the respective peptide's antiviral activity. This approach can be applied to fusion-inhibitor peptides raised against different viruses, such as other coronaviruses, influenza virus, and respiratory syncytial virus.

A recent study investigating the effects of PEGylation and lipidation on the bioavailability of SARS-CoV-2 inhibitor peptides estimated that unmodified peptides lose potency after approximately 8 h of administration, a finding that correlated with decreased plasma concentrations in treated mice[47]. Our study is consistent with this estimate and demonstrates that the mutagenesis strategy presented here can improve peptide stability *in vivo* without the need for chemical modifications, a significant advantage given the simplicity of manufacturing these monomeric peptides for potential therapeutic use.

We note that this is the first time that our unmodified peptides with N-terminal extension (the wildtype longHR2_42 and the mutant HR2-NHN) were studied in mice. When the peptides were administered simultaneously with the virus, or even three hours prior to infection, we observed complete blockage of infection with a single low dose at 1 mg/kg, compared to 4–20 mg/kg from a recent study using a dimer of shortHR2 with linked PEG and cholesterol[47,48]. Several factors may have contributed to the significant difference in potency. First, the N-terminal extension of the peptide likely plays a pivotal role in competing with the HR2 domain of S for binding to the HR1 domain. Second, although the linked PEG and cholesterol moieties offer several advantages, including increased solubility, improved stability, and enhanced membrane binding to increase local concentration, they also increase the overall mass of the inhibitor, potentially affecting efficacy. Third, we used wildtype mice and the beta variant of the virus, and infection is milder than in mice transgenically expressing human ACE2 and thus easier to inhibit. Nevertheless, these mutations presented here could also be incorporated into other HR2-based designs.

Considering its superior stability, we envision that the new HR2-NHN inhibitor from this study could become a better prototype than the wildtype peptide for further tests and improvements to further enhance the efficacy of HR2-NHN. In particular, applying chemical modifications such as lipidation and hydrocarbon stapling might further improve the stability and potency of HR2NH[10-13]. Finally, we propose that our strategy could be readily performed for other enveloped viruses that usually contain basic residues in their HR2-equivalent regions (**Fig. S1**)[18-32], including Ebola virus[19], Nipah virus[24], RSV[20], and HIV[18].

Materials and Methods

Purification of TMPRSS2

The purification of the secreted, C-terminally His-tagged TMPRSS2 ectodomain (residues 109-492) was performed following a previously established protocol[35]. Briefly, the plasmid was obtained from Addgene (plasmid no. 176412). Expression was carried out in Sf9 insect cells using a baculovirus system. The culture supernatant, containing the secreted protein, was harvested 4–5 days post-infection when cell viability reached 55–60%. After clarification by centrifugation, the supernatant pH was adjusted to 7.4, and the protein was captured via batch incubation with Ni-NTA resin. The resin was then collected, transferred to a gravity-flow column, washed with three column volumes of ice-cold PBS, and the protein was eluted using PBS containing 500 mM imidazole. For zymogen activation, the eluate was concentrated and dialyzed against an activation buffer (25 mM Tris pH 8.0, 75 mM NaCl, 2 mM $CaCl_2$) for 6 hours at room temperature. The final purification step involved size-exclusion chromatography on a Superdex 75 column equilibrated in 50 mM Tris pH 7.5, 250 mM NaCl. Fractions corresponding to the main protein peak were verified by SDS-PAGE with and without DTT, pooled, and concentrated with a 10 kDa MWCO concentrator (EMD Millipore cat. no. UFC901008). The SEC profile and SDS-PAGE (**Fig. S10**) were essentially the same as the protocol we followed[35]. For long-term storage, the purified active enzyme was prepared in a buffer containing 25% glycerol (50 mM Tris pH 7.5, 250 mM NaCl, 25% glycerol), flash-frozen in liquid nitrogen, and stored at -80°C.

Fluorescent labeling of peptides

The HR2 peptides for fluorescent labeling were prepared in-house as described before[6]. Briefly, HR2 peptides ranging from either residues 1157–1201 (longHR2_FL), or residues 1168–1201 (shortHR2_FL), or residues 1157–1201 with the glycine insertion (42G_FL)[36], of the Wuhan strain SARS-CoV-2 S protein were cloned into the pETDuet-1 plasmid with N-terminal hexa-histidine and SUMO tags and a C-terminal cysteine for maleimide labeling. The peptides were

recombinantly expressed in *E. coli* BL21(DE3) at 37 °C overnight using autoinducing LB medium[49]. Cells were harvested, resuspended in lysis buffer (50 mM Tris pH 8, 300 mM NaCl, 20 mM imidazole, 0.5 mM TCEP), and lysed by sonication. Lysate was clarified by centrifugation at 40,000 RPM for 30 min. Supernatant was bound to Nickel-NTA resin for 45 min at 4℃ with stirring. After washing with the lysis buffer, the bound protein was eluted with the lysis buffer supplemented with 300 mM imidazole. The eluted protein was then subjected to SUMO protease cleavage, heat treatment at 95 °C for 10 min, and SEC using a Superdex 30 HiLoad 16/60 column equilibrated with 20 mM sodium phosphate, pH 7.4, 150 mM NaCl, and 0.5 mM TCEP. The purified peptides were concentrated to ~1 mM with a 2 kDa MWCO concentrator (Sartorius cat no. VN02H91), flash frozen in liquid nitrogen, and stored in a −80 °C freezer for future use.

The maleimide labeling of the HR2 peptides was performed by mixing 100 μM peptide and 2.5 mM fluorescein-5-maleimide (Thermo Fisher, cat. no. 62245) in degassed buffer (50 mM sodium phosphate, pH 6.8, 150 mM NaCl, 0.05 mM TCEP), followed by overnight incubation at 4 °C. Following labeling, free dyes were removed by SEC using a Superdex 75 HR 10/300 GL column in PBS buffer (137 mM NaCl, 2.7 mM KCl, and 10 mM phosphate, pH 7.4). Labeled peptides were then further purified by high-performance liquid chromatography (HPLC) with a C18 column in a gradient of water and acetonitrile supplemented with 0.1% trifluoroacetic acid, followed by lyophilization and a final round of SEC using a Superdex 75 HR 10/300 GL column equilibrated with PBS buffer to remove any residual trifluoroacetic acid. The purified peptides were confirmed by mass spectrometry, concentrated to ~50 μM, protected from light, and stored at -80°C.

TMPRSS2 cleavage of the fluorescently labeled peptides

Concentrations used for the cleavage were 1.5 μM for the fluorescently labeled peptides (42G_FL, longHR2_FL, shortHR2_FL), 0.3 μM for TMPRSS2, and 5 mM for PMSF, respectively. After mixing on ice, the cleavage was allowed to proceed for 1 h at 37 °C.

CN-PAGE and in-gel fluorescence imaging

The cleaved samples were mixed with 4X native loading buffer (120 mM Tris pH 6.8, 20% glycerol, 0.02% bromophenol), and each sample was loaded onto an AnykD precast polyacrylamide gel (Bio-Rad, cat. no. 4569036). The gel was run in 25 mM Tris, 192 mM glycine, pH 8.3, at 70 V for 4 hours at 4°C. Protein LoBind tubes (Eppendorf, cat. no. 022431081) minimized potential adsorption of protein, and samples were protected from light throughout the entire process. The gel was then immediately imaged using an iBright 1500 imager (Invitrogen, excitation wavelength: 494 nm, emission wavelength: 512 nm).

Preparation of the HR2 peptide library

To create the library for saturation mutagenesis of the HR2 peptide (shortHR2), the residues in the wildtype sequence were replaced by each of the 19 amino acids, one at a time. The 5’ and 3’ flanking sequences ‘GGTCAAAGCGGTCAAcatatg’ and ‘tctagaCGCATTAGTCCGATG’, respectively, were appended onto each coding DNA sequence. All these sequences were purchased as an oligonucleotide pool generated by massively parallel synthesis (Twist Bioscience, San Francisco, CA). The oligonucleotide sequences were amplified and gel purified. The 5’ end of the forward primer contained the sequence CATATG, the recognition motif for the NdeI restriction enzyme, while the 5’ end of the reverse primer contained the sequence TCTAGA, the recognition motif for the XbaI restriction enzyme. Both NdeI and XbaI recognition sites were introduced into the pBAD33 backbone containing eCPX. The oligo pool and the pBAD33 plasmid were then cut by NdeI/XbaI and ligated (using T4 DNA ligase) in the ligase buffer to

circularize the linear product. We transformed the ligated DNA into NEB® 5-alpha Electrocompetent *E. coli*. Library coverage was maintained at ~1000-fold per variant and assessed by colony counts of transformants. We isolated and saved mini-prepped plasmid to be used in the surface-display assay.

Purification of HR1 for the surface display assay

HR1 complementary DNA (cDNA) was inserted between a His6-SUMO sequence and the GFP gene and cloned into a pSMT3 vector. The SUMO-HR1-GFP fusion was expressed in BL21(DE3) cells by overnight induction with 0.5 mM isopropyl β-D-1-thiogalactopyranoside at 18 °C. Cells were pelleted and then resuspended in a buffer containing 50 mM Tris, pH 8.0, 300 mM NaCl, 10 mM imidazole, 2 mM β-mercaptoethanol, and a mixture of protease inhibitors. Cells were lysed by French press, and the lysates were clarified by centrifugation at 35,000 × g for 1 h. The SUMO-HR1-GFP fusion was purified using a HisTrap Fast Flow column (Sigma). Eluted fractions were then dialyzed against a buffer containing 20 mM Tris-HCl (pH 8.0), 150 mM NaCl, and 1 mM dithiothreitol at 4 °C overnight. After dialysis, the sample was concentrated and gel-filtered in the same buffer on a HiLoad 16/60 Superdex 200 (GE Healthcare) column. Peak fractions that contain SUMO-HR1-GFP fusion were pooled and concentrated to 4 mg/mL through centrifugation with a 10 kDa MWCO concentrator (EMD Millipore cat. no. UFC901008).

Surface display of the HR2 peptide library

We used 100 ng DNA (shortHR2 peptide library) to transform 100 µL electrocompetent *E. coli* MC1061 cells. Details for the peptide expression can be found in Shah et al. After induction of peptide expression on the N-terminus of the surface display scaffold eCPX[50], *E. coli* cells were spun down and resuspended in a buffer containing 50 mM HEPES (pH 7.5), 150 mM NaCl, 0.2% bovine serum albumin, and incubated with 1 µM SUMO-HR1-GFP on ice for an hour. The labeled and washed cells were then resuspended and diluted 5-fold in the same buffer for sorting. Cells were sorted on a Sony SH800 cell sorter. The distribution of GFP fluorescence (HR1 binding) for the whole cell population was measured, and a sorting gate was set to collect the highest 25–35% of this distribution. For each sample, 1,000,000 cells were collected into a 15 mL conical tube containing 3 mL of LB on ice. For each library, an analogous tube was prepared with approximately 1,000,000 unsorted cells diluted in LB. Details of the DNA isolation and deep sequencing can be found in Shah et al[51].

Analysis of deep sequencing data

A custom shell script utilizing grep was used to count the number of occurrences of each variant in the raw sequencing file. Enrichment ranks were calculated following a previously described method[52]. Briefly, for each HR2 variant *i*, the enrichment rank (E^i) was computed based on its relative abundance in the GFP+ sorted population compared to the unsorted population and normalized to the corresponding ratio for the wildtype (WT) HR2 sequence. The enrichment rank was calculated using the following formula:

$$E^i = log_{10}\left(\frac{C^i_{GFP+}}{C^i_{unsorted}}\right) - log_{10}\left(\frac{C^{WT}_{GFP+}}{C^{WT}_{unsorted}}\right)$$

where C^i and C^{WT} refer to the counts of the HR2 variant *i* and the wildtype HR2, respectively, in either the GFP+ or unsorted populations. Using this approach, an enrichment rank of +1 or −1 corresponds to a 10-fold increase or decrease, respectively, in the abundance of a variant relative to wildtype in the GFP+ population compared to the unsorted population.

Peptide synthesis and preparation

All the non-fluorescently labeled peptides listed in **Fig. 1a** were synthesized by GenScript USA Inc. HPLC and liquid chromatography–mass spectrometry profiles for the three new designs

(HR2-QHN, HR2-NNS, and HR2-NHN) are shown in **Figs. S3 to S5** (provided by the manufacturer).

We developed a new protocol for dissolving HR2 peptides with the following goals: (A) minimize the contact of peptides with other substances, such as DMSO and concentrator membranes, that may introduce contaminants that interfere with concentration calculations by measuring absorbance at 205 nm; (B) minimize the leftover salts and other substances from the peptide powder, such as trifluoroacetic acid, that may be harmful to cells and viruses; keep the protocol simple enough that it can be easily reproduced in most biology laboratories. A 5-mL 1800-Da-MWCO desalting column (Thermo Scientific, cat. no. 43426) was prepared according to the manufacturer's manual and equilibrated with 10 mL of 20 mM HEPES-Na pH 7.5. 2 mg of the lyophilized peptide powder from GenScript, was resuspended in 0.5mL of ddH2O. To clarify the peptide suspension, 0.1 M NaOH was added dropwise until it turned clear. The clarified peptide solution was immediately applied to the pre-equilibrated column and allowed to enter the column bed. 1 mL aliquots of 20 mM HEPES-Na pH 7.5 were applied to the column sequentially, and fractions were collected in 1.5-mL Eppendorf Protein LoBind tubes (Eppendorf, cat. no. 022431081). The peptide concentration for each fraction was quantified with a Nanodrop spectrometer by measuring absorbance at 205 nm. Usually, the peptide should be in fraction #2 with a concentration of 0.5-0.8 mg/mL, or around 200 μM.

Detailed comparisons for preparing peptides in solution using different protocols are rarely published, so we only compared our new protocol using a desalting column with our previously published protocol using a concentrator[6]. Using our cell-cell fusion assay, we found that the IC_{50} barely changes for the well-behaved longHR2_42 (**Fig. S11**). However, the IC_{50} for the shortHR2 decreased from 263 nM to 27 nM (**Fig. S11**), indicating that the new protocol improves the sample purity either substantially, or only removes trace amount of contaminants that may have substantial impact on concentration measurement by absorption at 205 nm[53]. The most commonly used way of protein concentration measurement by absorption at 280 nm is not applicable since the HR2 peptides do not have any aromatic residues, while accurate quantification by mass of peptide powder requires an overwhelmingly large amount than needed in our screening assay, plus the weight of peptide powder inevitably contains that from salts during lyophilization.

HEK cell-cell fusion assay

We optimized the cell-cell fusion assay based on the α-complementation of *E. coli* β-galactosidase for comparing the inhibitory activity of different peptides with higher throughput. Suspension culture Expi293F cells (Thermo Fisher, cat. no. A14527), a clonal derivative of Human Embryonic Kidney (HEK) 293 cells, were grown to a density of $1 \sim 2 \times 10^6$ cells/mL in FreeStyle 293 expression medium (Thermo Fisher, cat. no. 12338026). The cells were then pelleted, resuspended in medium to a density of 1×10^6 cells/mL, and allowed to recover at 37 °C for 30 min. One group of cells was then co-transfected using polyethyleneimine (PEI, Sigma) (125 μg PEI/mL cells) with Wuhan strain full-length SARS-CoV-2 S protein construct (12.5 μg DNA/mL cells) and the α-fragment of *E. coli* β-galactosidase construct (12.5 μg DNA/mL cells) to generate the S protein-expressing cells. Since HEK cells do not express TMPRSS2, it is likely that S is cleaved by other host cell proteases that render S fusogenic upon binding to its receptor ACE2. Using the same amount of PEI, the other group of cells was co-transfected with the full-length ACE2 (12.5 μg DNA/mL cells) construct and the ω-fragment of *E. coli* β-galactosidase construct (12.5 μg DNA/mL cells) to generate the ACE2 receptor-expressing cells. Note that the endogenous ACE2 level is very low in HEK cells, requiring ectopic expression for efficient fusion. As a negative control, two additional groups of cells were transfected with either the α-fragment or the ω-fragment of *E. coli* β-galactosidase construct alone. After incubation of the

cells in flasks at 37 °C for 24 h, the cells were pelleted. The S-expressing cells were resuspended in FreeStyle 293 expression medium supplemented with different concentrations of peptide or peptide-TMPRSS2 mixture (50 μL, 2×10^6 cells/mL), respectively. The ACE2-expressing cells and negative control cells were resuspended in 50 μL fresh FreeStyle 293 expression medium (pH 7.4) to be 2×10^6 cells/mL. S-expressing, and ACE2 cells or α-fragment and ω-fragment cells were then mixed in a 96-well plate (Greiner Bio-One) to initiate cell-cell fusion at 37 °C for 2 h. Fusion was arrested by adding 100 μL β-galactosidase substrate from the Gal-Screen reporter system (Invitrogen). The mixture was incubated at 37 °C in the dark for 1 h before luminescence was recorded using a Tecan Infinite M1000.

Cryo-EM structure determination

The cryo-EM structure of the HR2-NHN bound to HR1 was determined following a molecular scaffolding method described previously[40]. Briefly, the scaffolded HR1HR2 complex was generated by co-expressing the scaffolded HR1 and Small Ubiquitin-like Modifier (SUMO)-tagged HR2-NHN in *E. coli* BL21(DE3) using auto-inducing lysogeny broth (LB) medium[49], followed by nickel affinity chromatography and size-exclusion chromatography (SEC) with a Superose 6 Increase 10/300 GL column in 25 mM Hepes-Na, pH 7.4, 150 mM NaCl, 0.5 mM EDTA, 0.5 mM Tris(2-carboxyethyl)phosphine (TCEP). The sample was concentrated to 20 μM, supplemented with 0.05% Nonidet P-40, and plunge frozen on a Quantifoil 2/1 holey carbon grid using a Vitrobot Mark IV (Thermo Fisher). A total of 20,031 movies were recorded using a Titan Krios electron microscope (Thermo Fisher) equipped with a K3 camera (Gatan) using the EPU automation software (Thermo Fisher Scientific), at a nominal magnification of 130,000× and a pixel size of 0.325 Å. Each movie contained 40 frames with a total electron dose of 52 $e^-/Å^2$. The data were processed using a combination of MotionCor2 [54], Gctf[55], EMAN2 [56], cryoSPARC[57], and RELION[58], as described previously[40]. More details for data collection and processing are summarized in **Fig. S7 and Table S1.**

For model building, the PDB 8czi was used as the template. The mutations of HR2 were performed in Coot[59] and then refined by the automated structure refinement protocol in Rosetta[60]. The structure was then subjected to real-space refinement (minimization_global, local_grid_search, adp) in PHENIX[61]. Coot[59] was used for further fitting of sidechains and manual inspection.

Animals

A total of 30 female BALB/c mice (Envigo, Indianapolis, IN, USA) were transferred to the University of Helsinki biosafety level-3 (BSL-3) facility and acclimatized to individually ventilated biocontainment cages (ISOcage; Scanbur, Karl Sloanestran, Denmark) for seven days with ad libitum water and food (rodent pellets). For subsequent experimental infection, the mice were placed under isoflurane anesthesia and inoculated intranasally with 20 μl of virus dilution or DMEM (non-infected control). The animals were held in an upright position for a few seconds to allow the liquid to flow downwards in the nasal cavity. Mice did not lose weight when infected with the SARS-CoV-2 beta variant. Their well-being was further monitored carefully for signs of illness (e.g., changes in posture or behavior, rough coat, apathy, ataxia). Euthanasia was performed under terminal isoflurane anesthesia with cervical dislocation. Experimental procedures were approved by the Animal Experimental Board of Finland (license number ESAVI/32448/2023).

Cells

A549 stably expressing human *ACE2* and *TMPRSS2* fused to GFP (A549-AT)[44] were grown in Dulbecco's modified Eagle's medium (DMEM), 2 mM glutamine, 10 % heat-inactivated fetal bovine serum (FBS), and 1x penicillin/streptomycin antibiotic mix.

Virus isolation, propagation, and sequencing

All experiments with wildtype or mutant SARS-CoV-2 were performed in BSL3 facilities of the University of Helsinki with appropriate institutional permits. Viruses isolated from consensual COVID-19 patients (permits HUS/32/2018 § 16) were propagated once in VeroE6-TMPRSS2 cells[42], titrated by plaque assay, and stored at -80°C in DMEM, 2% FBS, and 1x Penicyllin/Streptomycin. All SARS-CoV-2 viruses were sequenced by next-generation sequencing at the University of Helsinki. The SARS-CoV-2 Beta variant virus used in mice experiments was isolated from infected patient nasopharyngeal samples as described[42] and amplified using VeroE6-TMPRSS2 [42]. Once propagated, their genomic sequence was confirmed using an Illumina platform available at the Department of Virology, University of Helsinki [42]. The Beta variant sequence has been previously described in detail [45] and deposited in the NCBI GenBank database under accession number MW717678. For virus production, at three days post inoculation, the collected medium from the infected cells was centrifuged twice at 4500xg for 10 min at 4°C, and the cleared supernatant was aliquoted in cryotubes and stored at -80 °C in the BSL3 facility. The virus was propagated in Minimum Essential Medium (MEM) containing 2% FBS, 20 mM HEPES, pH 7.2, 2 mM glutamine, and 1x penicillin/streptomycin antibiotic mix. Virus titrations were performed by standard plaque assay in VeroE6-TMPRSS2 cells as previously described[42]. Infected cells were maintained in incubators at 37 °C and 5% CO_2 in the BSL-3 facility at Helsinki University Hospital.

Virus infection and peptide treatments in cell cultures

Human bronchial adenocarcinoma-derived lung Calu-3 cells (ATCC, Cat. no. HTB-55; male) and A549-ACE2-TMPRSS2-GFP (A549-AT)[62] cells were seeded in DMEM containing 10% (A549-AT) or 20% (Calu-3) Dulbecco's Modified Eagle's High Glucose, pyruvate Medium (DMEM, Gibco, Cat. no. 11995065), 1x non-essential amino acids (NEAA, Gibco, Cat. no. 11140050), 10% heat-inactivated fetal bovine serum (FBS; Gibco, Cat. no. 26140079), and 100U/ml penicillin and streptomycin (P/S; Gibco, Cat. no. 15140122). A549-AT cells were seeded at 15,000 cells per well and Calu-3 cells were seeded at 22,000 or 12,000 cells per well in 96-well imaging plates (catalogue number 6005182; PerkinElmer) 24 h (for time 0 hbi experiments) or 96 h (for the 72 hbi experiments) before infection with SARS-CoV-2 mCherry[63] (a recombinant virus based on the original Wuhan strain) or the Omicron Xbb1.5 variant. The peptides were diluted in the same infection medium and added at the indicated times. The amount of virus used to infect the cells was adjusted to obtain 15- 20% infected cells in vehicle-treated controls, as determined by high-content fluorescence imaging at 18 hpi and automated image analysis after fixation with 4 % paraformaldehyde (in PBS) for 20 min at room temperature.

High-content fluorescence imaging

Fixed cells were washed three times with Dulbecco-modified PBS containing 0.2% BSA (DPBS/BSA), permeabilized with 0.1% Triton X-100 in DPBS/BSA, and nuclear DNA stained with Hoechst DNA dye (ThermoFisher Scientific, catalog number H3570). Automated fluorescence imaging was performed using a Molecular Devices Image-Xpress Nano high-content epifluorescence microscope equipped with a 10× objective and a 4.7-megapixel CMOS camera (pixel size: 0.332 µm). Image analysis was performed with CellProfiler-4 software (www.cellprofiler.org). Nuclei were automatically detected using the Otsu algorithm included in the software. To automatically identify infected cells, an area surrounding each nucleus (5-pixel expansion of the nuclear area) was used to estimate the fluorescence intensity of the virally

expressed mCherry fluorescent protein, using an intensity threshold such that <0.01% of 'positive cells' were detected in noninfected wells.

Infections and peptide treatments in mice

Eight-week-old female BALB/c mice were anesthetized using isoflurane and intranasally inoculated (n = 4 or 5 per group) with 2×10^5 PFU of the SARS-CoV-2 Beta variant in 20 µl of DMEM[45]. The peptides solubilized as described above in 20 mM HEPES pH 7.2 at a concentration of 1 mg/ml (of 0.1 mg/ml when indicated) were applied as a 20 µl drop/mouse at indicated times. The virus and peptide drops were applied to the openings of the animal's nostrils, and the liquid was naturally inhaled during anesthesia. At 48 hpi, animals were euthanized under terminal isoflurane anesthesia with cervical dislocation and dissected immediately after sacrifice. The right lungs were dissected, collected, and frozen at -80°C before processing for PCR and analysis of viral RNAs, or for homogenization and virus extraction, as described in the next paragraph; left lungs were fixed in 10% buffered formalin for 48 h and stored in 70% ethanol for histological and immunohistochemical examinations.

Lung homogenization and virus reinfection in VeroE6 cells

A portion of the right lung from each mouse was washed once with ice cold PBS, weighted, and 100 mg homogenized by mixing with 100 mg of glass beads (1.0 mm diameter, Sigma-Aldrich, catalogue number Z250473) in 500 µl ice-cold MEM containing 2% FBS, 20 mM HEPES pH 7.2, 2 mM glutamine, and 1x penicillin/streptomycin antibiotic mix. Samples were vortexed three times at full speed for 30 seconds (total vortexing time 1.5 minutes). After centrifugation at 400xg for 10 min, 25 µl of the supernatant from each sample was used to infect Vero-E6 cells grown on 96-well imaging plates in a total volume of 100 µl. At 18 hpi cells were fixed in 4% PFA for 20 mi at room temperature and processed for immunofluorescence staining of viral protein N, automated imaging and image analysis as described above.

RNA Isolation and qRT-PCR

RNA was extracted from lung samples using Trizol (Thermo Scientific) according to the manufacturer's instructions. Isolated RNA was subjected to one-step RT-qPCR analysis as described using primers specific for the viral genome encoding for the subgenomic E[64] gene with TaqMan fast virus 1-step master mix (ThermoFisher Scientific, catalogue number 4444432) using AriaMx instrumentation (Agilent, Santa Clara, CA, USA). The actin RT-qPCR is described in[65].

Primer and probe sequences used in the RT-qPCR.
Subgenomic E Forward cgatctcttgtagatctgttctc[64]
Probe acactagccatccttactgcgcttcg[64]
Reverse atattgcagcagtacgcacaca[64]

Beta-actin Forward actgccgcatcctcttcct[65]
Probe cctggagaagagctatgagctgcctgatg[65]
Reverse tcgttgccaatggtgatgac[65]

Immunohistochemistry

The left lungs of the sacrificed mice were trimmed for histological examination and paraffin-wax embedded. The heads were sawn longitudinally in the midline using a diamond saw (Exakt 300; Exakt, Oklahoma, OK, USA), then decalcified and processed as previously described[45]. Consecutive sections (3–5 µm) were prepared from lungs and subjected to immunohistochemistry for the detection of SARS-CoV-2 antigen expression using a rabbit polyclonal anti-SARS-CoV

NP antibody that cross-reacts with NP of SARS-CoV-2 (Rockland Immunochemicals, Limerick, USA, cat no. 200-402-A50).

Statistics and data analysis

Data from the human embryonic kidney (HEK) cell-cell membrane fusion assay from three independent biological replicates, determined for each concentration of inhibitor, are used. For the HEK cell-cell membrane fusion assay, normalized fusion was calculated as $(\text{Luminescence}_{(+\text{inhibitor})} - \text{Luminescence}_{(\alpha\&\omega)})/(\text{Luminescence}_{(+\text{PBS})} - \text{Luminescence}_{(\alpha\&\omega)})$, where "+inhibitor" or "+PBS" refers to adding inhibitor or PBS to the mixture of the cells expressing the α-fragment of *E. coli* β-galactosidase and S, and the cells expressing the ω-fragment of *E. coli* β-galactosidase and ACE2, and "α&ω" refers to the mixture of the cells expressing the α-fragment only and the cells expressing the ω-fragment only.

After the normalization, the arithmetic means of the three replicates were used to fit the inhibition curves to obtain estimates of the IC_{50} values; the estimates were obtained by nonlinear regression of inhibitor concentration vs. response in GraphPad Prism version 9.1.0 for macOS (GraphPad Software, San Diego, CA, https://www.graphpad.com). The fitted model is $Y = \text{Bottom} + (\text{Top} - \text{Bottom})/(1 + (IC_{50}/X)^{\text{HillSlope}})$, where Y is the extent of inhibition, X is the inhibitor concentration, Bottom and Top are the minimal and maximal inhibition. The SE of the IC_{50} estimation was calculated using OriginPro 9.1 (OriginLab Corporation).

For the peptide dose responses *in vitro* in Calu-3 cells (**Fig. 5 and S8**), and for the estimation of bioavailability in mice (**Fig. 6 and S9**), all statistical analyses were performed in R. The R code used is available in the GitHub repository (available at the link: https://github.com/bongrita/SARSCoV_Fusion_inhibitor_peptide.git).

For the *in vitro* dose-response experiments (**Fig. 5**), normalized infections were modeled as a function of peptide concentration using four-parameter log-logistic regression (4PL) (LL.4, drc package of R)[66]. The IC_{50} values were defined as the concentration at which normalized infection reached a value of 0.5 and were estimated using the "ED()" function of R with 95% confidence intervals calculated by the delta method.

IC_{50} comparisons for longHR2_42 and HR2-NHN were performed using joint dose-response models and the R function "compParm()" ratio test[66]. To test if the difference between the fitted curves was statistically significant, area under the fitted curve (AUC) was calculated over 0.1–1000 nM peptide concentrations by trapezoidal integration and compared between the two peptides using non-parametric bootstrap resampling with 1000 iteration[67]. AUC was recalculated for each iteration, with 95% confidence intervals derived from the bootstrap percentile distribution and two-sided P values calculated from the bootstrap tail probabilities.

For the *in vivo* mouse re-infection time-course experiment, infection measurements from vehicle- or peptide-treated mice were normalized to the mean infection level of the vehicle-treated control group, which was defined as 1.0, and data are presented as mean ± SD. Overall infection across the 0–24 h pretreatment interval was summarized by trapezoidal AUC calculated from the observed mean normalized infection values at 0, 4, 8, 16 and 24 h, and differences between peptides were assessed by bootstrap resampling of mice within each time point with 1,000 iterations[66]. To estimate the duration of antiviral activity (T_{50}), a logistic time-response curve was fitted separately for each peptide using data from 0–16 h; the 24-h time point was excluded from T_{50} estimation because the decrease between 16 h and 24 h violated the monotonicity assumption of the logistic model. T_{50} was defined as the pretreatment interval at which fitted normalized

infection reached 0.5, corresponding to 50% of the vehicle-treated infection level. Differences in the AUC for each peptide were assessed by bootstrap resampling within each time point with 1000 iterations[66]. For all analyses, statistical significance was defined as $P < 0.05$. Statistical analysis of the mouse experiment described in Fig. 6b was performed with Ordinary one-way Anova Dunnett's multiple comparison test using GraphPad Prism version 9.1.0 for macOS (GraphPad Software, San Diego, CA, https://www.graphpad.com). For the mouse treatment "0h 0.1 mg/kg", two of the mice dies during anesthesia prior infection thus statistical analysis was not performed with the remaining 2 mice.

Figure preparation

The figures of PDB structures and maps were made using UCSF Chimera[68]. The data fitting of all inhibition assays was performed and plotted using GraphPad Prism version 9.1.0 for macOS (GraphPad Software, San Diego, CA, https://www.graphpad.com) or R. Figure panels were assembled with Adobe Illustrator and Image-J (Fiji).

Data availability

The EM maps and corresponding structures reported here have been deposited in EMDB (Electron Microscopy Data Bank) and PDB with the following accession IDs: EMD-70569 and PDB 9okn. The R code used for statistical analysis, generation of graphs, and calculation of AUCs and IC50s is available on GitHub (accessible at the link: https://github.com/bongrita/SARSCoV_Fusion_inhibitor_peptide.git).

Acknowledgments

We thank Bing Chen for kindly providing plasmids and protocols for the cell-cell membrane fusion assay, Rebekah C Gullberg, Casey Thompson, Yuhao Li, Sean Whelan, and Judith Frydman for stimulating discussions, the Sarafan ChEM-H High-Throughput Screening Knowledge Center for providing the Tecan microplate reader, Robert Sullivan (University of Queensland) for the help in the histology of mice lungs, the Finnish Institute for Molecular Medicine (FIMM) High Content Imaging unit services were used for imaging. We also thank the Stanford Cryo-Electron Microscopy Center (cEMc) and the Stanford-SLAC Cryo-EM Center (S^2C^2) for support. This article is subject to HHMI's Open Access to Publications policy. HHMI laboratory heads have previously granted a nonexclusive CC BY 4.0 license to the public and a sublicensable license to HHMI in their research articles. Pursuant to those licenses, the accepted manuscript of this article can be made freely available under a CC BY 4.0 license immediately upon publication.

Author contributions:

Conceptualization: KY, JK, GB, and ATB
Molecular cloning, protein expression, purification, and labeling: KY, KIW, RAP, and LE
TMPRSS2 cleavage test: KY
Deep mutational scanning: SM, SS, and JK
Design of mutant peptides: KY
Cell-cell fusion assay: CW
Structure determination: KY
Virus-Cell and -mouse infection assays: FT, RK, LK, SM, SHS, TStr, TB, YTB, TSir, JH, OV, MJ, and GB
Supervision: JK, GB, and ATB
Writing: KY, CW, SM, RK, KIW, JK, GB, and ATB

Conflict of interests

All authors declare they have no conflict of interest.

References

1 White, J. M., Ward, A. E., Odongo, L. & Tamm, L. K. Viral Membrane Fusion: A Dance Between Proteins and Lipids. *Annu Rev Virol* **10**, 139-161, doi:10.1146/annurev-virology-111821-093413 (2023).

2 Harrison, S. C. Viral membrane fusion. *Nat Struct Mol Biol* **15**, 690-698, doi:10.1038/nsmb.1456 (2008).

3 Harrison, S. C. Viral membrane fusion. *Virology* **479-480**, 498-507, doi:10.1016/j.virol.2015.03.043 (2015).

4 Xing, L. *et al.* Early fusion intermediate of ACE2-using coronavirus spike acting as an antiviral target. *Cell* **188**, 1297-1314 e1224, doi:10.1016/j.cell.2025.01.012 (2025).

5 Abbasian, M. H. *et al.* Global landscape of SARS-CoV-2 mutations and conserved regions. *J Transl Med* **21**, 152, doi:10.1186/s12967-023-03996-w (2023).

6 Yang, K. *et al.* Nanomolar inhibition of SARS-CoV-2 infection by an unmodified peptide targeting the prehairpin intermediate of the spike protein. *Proc Natl Acad Sci U S A* **119**, e2210990119, doi:10.1073/pnas.2210990119 (2022).

7 Xia, S. *et al.* Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion. *Cell Res* **30**, 343-355, doi:10.1038/s41422-020-0305-x (2020).

8 de Vries, R. D. *et al.* Intranasal fusion inhibitory lipopeptide prevents direct-contact SARS-CoV-2 transmission in ferrets. *Science* **371**, 1379-1382, doi:10.1126/science.abf4896 (2021).

9 Rossino, G. *et al.* Peptides as Therapeutic Agents: Challenges and Opportunities in the Green Transition Era. *Molecules* **28**, doi:10.3390/molecules28207165 (2023).

10 Ingallinella, P. *et al.* Addition of a cholesterol group to an HIV-1 peptide fusion inhibitor dramatically increases its antiviral potency. *Proc Natl Acad Sci U S A* **106**, 5801-5806, doi:10.1073/pnas.0901007106 (2009).

11 Wang, Y. *et al.* The efficacy and safety of inhaled peptide YKYY017 for COVID-19 patients with mild illness: a phase 2 randomized controlled trial. *Nat Commun* **16**, 7272, doi:10.1038/s41467-025-62214-x (2025).

12 Bird, G. H. *et al.* Hydrocarbon double-stapling remedies the proteolytic instability of a lengthy peptide therapeutic. *Proc Natl Acad Sci U S A* **107**, 14093-14098, doi:10.1073/pnas.1002713107 (2010).

13 Bird, G. H. *et al.* A stapled lipopeptide platform for preventing and treating highly pathogenic viruses of pandemic potential. *Nat Commun* **15**, 274, doi:10.1038/s41467-023-44361-1 (2024).

14 Al-Aly, Z. & Topol, E. Solving the puzzle of Long Covid. *Science* **383**, 830-832, doi:10.1126/science.adl0867 (2024).

15 Liu, Y., Qu, H. Q., Qu, J., Tian, L. & Hakonarson, H. Expression Pattern of the SARS-CoV-2 Entry Genes ACE2 and TMPRSS2 in the Respiratory Tract. *Viruses* **12**, doi:10.3390/v12101174 (2020).

16 Baron, J. *et al.* Matriptase, HAT, and TMPRSS2 activate the hemagglutinin of H9N2 influenza A viruses. *J Virol* **87**, 1811-1820, doi:10.1128/JVI.02320-12 (2013).

17 Kalinska, M., Meyer-Hoffert, U., Kantyka, T. & Potempa, J. Kallikreins - The melting pot of activity and function. *Biochimie* **122**, 270-282, doi:10.1016/j.biochi.2015.09.023 (2016).

18 Weissenhorn, W., Dessen, A., Harrison, S. C., Skehel, J. J. & Wiley, D. C. Atomic structure of the ectodomain from HIV-1 gp41. *Nature* **387**, 426-430, doi:10.1038/387426a0 (1997).
19 Weissenhorn, W., Carfi, A., Lee, K. H., Skehel, J. J. & Wiley, D. C. Crystal structure of the Ebola virus membrane fusion subunit, GP2, from the envelope glycoprotein ectodomain. *Mol Cell* **2**, 605-616, doi:10.1016/s1097-2765(00)80159-8 (1998).
20 Zhao, X., Singh, M., Malashkevich, V. N. & Kim, P. S. Structural characterization of the human respiratory syncytial virus fusion protein core. *Proc Natl Acad Sci U S A* **97**, 14172-14177, doi:10.1073/pnas.260499197 (2000).
21 Yang, Z. N. *et al.* The crystal structure of the SIV gp41 ectodomain at 1.47 A resolution. *J Struct Biol* **126**, 131-144, doi:10.1006/jsbi.1999.4116 (1999).
22 Baker, K. A., Dutch, R. E., Lamb, R. A. & Jardetzky, T. S. Structural basis for paramyxovirus-mediated membrane fusion. *Mol Cell* **3**, 309-319, doi:10.1016/s1097-2765(00)80458-x (1999).
23 Xu, Y. *et al.* Structural basis for coronavirus-mediated membrane fusion. Crystal structure of mouse hepatitis virus spike protein fusion core. *J Biol Chem* **279**, 30514-30522, doi:10.1074/jbc.M403760200 (2004).
24 Lou, Z. *et al.* Crystal structures of Nipah and Hendra virus fusion core proteins. *FEBS J* **273**, 4538-4547, doi:10.1111/j.1742-4658.2006.05459.x (2006).
25 Liu, Y. *et al.* Structural characterization of mumps virus fusion protein core. *Biochem Biophys Res Commun* **348**, 916-922, doi:10.1016/j.bbrc.2006.07.168 (2006).
26 Igonet, S. *et al.* X-ray structure of the arenavirus glycoprotein GP2 in its postfusion hairpin conformation. *Proc Natl Acad Sci U S A* **108**, 19967-19972, doi:10.1073/pnas.1108910108 (2011).
27 Koellhoffer, J. F. *et al.* Crystal structure of the Marburg virus GP2 core domain in its postfusion conformation. *Biochemistry* **51**, 7665-7675, doi:10.1021/bi300976m (2012).
28 Koellhoffer, J. F. *et al.* Structural characterization of the glycoprotein GP2 core domain from the CAS virus, a novel arenavirus-like species. *J Mol Biol* **426**, 1452-1468, doi:10.1016/j.jmb.2013.12.009 (2014).
29 Lu, L. *et al.* Structure-based discovery of Middle East respiratory syndrome coronavirus fusion inhibitor. *Nat Commun* **5**, 3067, doi:10.1038/ncomms4067 (2014).
30 Yan, L., Meng, B., Xiang, J., Wilson, I. A. & Yang, B. Crystal structure of the post-fusion core of the Human coronavirus 229E spike protein at 1.86 A resolution. *Acta Crystallogr D Struct Biol* **74**, 841-851, doi:10.1107/S2059798318008318 (2018).
31 Zhang, X. *et al.* Crystal Structure of Refolding Fusion Core of Lassa Virus GP2 and Design of Lassa Virus Fusion Inhibitors. *Front Microbiol* **10**, 1829, doi:10.3389/fmicb.2019.01829 (2019).
32 Belot, L. *et al.* Crystal structure of Mokola virus glycoprotein in its post-fusion conformation. *PLoS Pathog* **16**, e1008383, doi:10.1371/journal.ppat.1008383 (2020).
33 Hoffmann, M., Kleine-Weber, H. & Pohlmann, S. A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. *Mol Cell* **78**, 779-784 e775, doi:10.1016/j.molcel.2020.04.022 (2020).
34 Tang, T. *et al.* Proteolytic Activation of SARS-CoV-2 Spike at the S1/S2 Boundary: Potential Role of Proteases beyond Furin. *ACS Infect Dis* **7**, 264-272, doi:10.1021/acsinfecdis.0c00701 (2021).
35 Fraser, B. J. *et al.* Structure and activity of human TMPRSS2 protease implicated in SARS-CoV-2 activation. *Nat Chem Biol* **18**, 963-971, doi:10.1038/s41589-022-01059-7 (2022).
36 Yang, K. *et al.* Structure-based design of a SARS-CoV-2 Omicron-specific inhibitor. *Proc Natl Acad Sci U S A* **120**, e2300360120, doi:10.1073/pnas.2300360120 (2023).

37 Cai, Y. *et al.* Distinct conformational states of SARS-CoV-2 spike protein. *Science* **369**, 1586-1592, doi:10.1126/science.abd4251 (2020).
38 Korber, B. *et al.* Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* **182**, 812-827 e819, doi:10.1016/j.cell.2020.06.043 (2020).
39 Korber, B. *et al. COVID-19 Viral Genome Analysis Pipeline*, <https://cov.lanl.gov/components/sequence/COV/int_sites_tbls.comp> (2020).
40 Yang, K. *et al.* Structural conservation among variants of the SARS-CoV-2 spike postfusion bundle. *Proc Natl Acad Sci U S A* **119**, e2119467119, doi:10.1073/pnas.2119467119 (2022).
41 Koch, J. *et al.* TMPRSS2 expression dictates the entry route used by SARS-CoV-2 to infect host cells. *EMBO J* **40**, e107821, doi:10.15252/embj.2021107821 (2021).
42 Cantuti-Castelvetri, L. *et al.* Neuropilin-1 facilitates SARS-CoV-2 cell entry and infectivity. *Science* **370**, 856-860, doi:10.1126/science.abd2985 (2020).
43 Ojha, R. *et al.* Dynamin independent endocytosis is an alternative cell entry mechanism for multiple animal viruses. *PLoS Pathog* **20**, e1012690, doi:10.1371/journal.ppat.1012690 (2024).
44 Kant, R. *et al.* Complete Protection from SARS-CoV-2 Lung Infection in Mice Through Combined Intranasal Delivery of PIKfyve Kinase and TMPRSS2 Protease Inhibitors. *bioRxiv*, doi:10.1101/2023.07.19.549731 (2023).
45 Kant, R. *et al.* Common Laboratory Mice Are Susceptible to Infection with the SARS-CoV-2 Beta Variant. *Viruses* **13**, doi:10.3390/v13112263 (2021).
46 Carroll, E. L. *et al.* Trypsin-Like Proteases and Their Role in Muco-Obstructive Lung Diseases. *Int J Mol Sci* **22**, doi:10.3390/ijms22115817 (2021).
47 Bi, W. *et al.* An enhanced broad-spectrum peptide inhibits Omicron variants in vivo. *Cell Rep Med* **5**, 101418, doi:10.1016/j.xcrm.2024.101418 (2024).
48 Mougari, S. *et al.* Intranasally administrated fusion-inhibitory lipopeptides block SARS-CoV-2 infection in mice and enable long-term protective immunity. *Commun Biol* **8**, 57, doi:10.1038/s42003-025-07491-4 (2025).
49 Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **41**, 207-234, doi:10.1016/j.pep.2005.01.016 (2005).
50 Rice, J. J. & Daugherty, P. S. Directed evolution of a biterminal bacterial display scaffold enhances the display of diverse peptides. *Protein Eng Des Sel* **21**, 435-442, doi:10.1093/protein/gzn020 (2008).
51 Shah, N. H. *et al.* An electrostatic selection mechanism controls sequential kinase signaling downstream of the T cell receptor. *Elife* **5**, doi:10.7554/eLife.20105 (2016).
52 Subramanian, S. *et al.* Adaptive Capacity of a DNA Polymerase Clamp-loader ATPase Complex. *Mol Biol Evol* **41**, doi:10.1093/molbev/msae013 (2024).
53 Scopes, R. K. Measurement of protein by spectrophotometry at 205 nm. *Anal Biochem* **59**, 277-282, doi:10.1016/0003-2697(74)90034-7 (1974).
54 Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331-332, doi:10.1038/nmeth.4193 (2017).
55 Zhang, K. Gctf: Real-time CTF determination and correction. *J Struct Biol* **193**, 1-12, doi:10.1016/j.jsb.2015.11.003 (2016).
56 Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol* **157**, 38-46, doi:10.1016/j.jsb.2006.05.009 (2007).
57 Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296, doi:10.1038/nmeth.4169 (2017).

58 Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* **180**, 519-530, doi:10.1016/j.jsb.2012.09.006 (2012).
59 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132, doi:10.1107/S0907444904019158 (2004).
60 Wang, R. Y. *et al.* Automated structure refinement of macromolecular assemblies from cryo-EM maps using Rosetta. *Elife* **5**, doi:10.7554/eLife.17219 (2016).
61 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221, doi:10.1107/S0907444909052925 (2010).
62 Kettunen, P. *et al.* SARS-CoV-2 Infection of Human Neurons Is TMPRSS2 Independent, Requires Endosomal Cell Entry, and Can Be Blocked by Inhibitors of Host Phosphoinositol-5 Kinase. *J Virol* **97**, e0014423, doi:10.1128/jvi.00144-23 (2023).
63 Rihn, S. J. *et al.* A plasmid DNA-launched SARS-CoV-2 reverse genetics system and coronavirus toolkit for COVID-19 research. *PLoS Biol* **19**, e3001091, doi:10.1371/journal.pbio.3001091 (2021).
64 Dagotto, G. *et al.* Comparison of Subgenomic and Total RNA in SARS-CoV-2 Challenged Rhesus Macaques. *J Virol* **95**, doi:10.1128/JVI.02370-20 (2021).
65 Zivcec, M., Safronetz, D., Haddock, E., Feldmann, H. & Ebihara, H. Validation of assays to monitor immune responses in the Syrian golden hamster (Mesocricetus auratus). *J Immunol Methods* **368**, 24-35, doi:10.1016/j.jim.2011.02.004 (2011).
66 Ritz, C., Baty, F., Streibig, J. C. & Gerhard, D. Dose-Response Analysis Using R. *PLoS One* **10**, e0146021, doi:10.1371/journal.pone.0146021 (2015).
67 Efron, B. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics* **7**, 26, doi:10.1214/aos/1176344552 (1979).
68 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612, doi:10.1002/jcc.20084 (2004).

Supplementary Materials for

A rational design strategy and validation for protease-resistant fusion-inhibitor antiviral peptides

Kailu Yang *et al.*

*Corresponding authors: Axel T. Brunger, Email: brunger@stanford.edu; Giuseppe Balistreri, Email: giuseppe.balistreri@helsinki.fi

This PDF file includes:

Figs. S1 to S11
Table S1

1ebo	TCHILGPDC**R**IEPHDWT**K**NITD**K**IDQIIHDFVD**K**	Ebola virus
1env	WMEWD**R**EINNYTSLIHSLIEESQNQQE**K**NEQELLELD**K**	Human immunodeficiency virus
1g2c	PLVFPSDEFDASISQVNE**K**INQSLAFI**RK**SDELLHNVNAG	Human respiratory syncytial virus A
1qbz	TP**K**WNNETWQEWE**RK**VDFLEENITALLEEAQIQQE**K**NMYELQ**K**L	Simian immunodeficiency virus
1svf	QILSIDPLDISQNLAAVN**K**SLSDALQHLAQSDTYLSAI	Simian virus 5
1wdf	DLSLDFE**K**LNVTLLDLTYEMN**R**IQDAI**KK**LNESYINL	Murine hepatitis virus
1wp7	DISSQISSMNQSLQQS**K**DYI**K**EAQ**R**LLDTV	Nipah henipavirus
1wp8	DISSQISSMNQSLQQS**K**DYI**K**EAQ**K**ILDTV	Hendra henipavirus
2fyz	DISTELS**K**VNASLQNTV**K**YI**K**ESNHQLQSVIV	Mumps orthorubulavirus
3mko	P**K**CWLVTNGSYLNETHFSDQIEQEADNMITEML**R**	Lymphocytic choriomeningitis mammarenavirus
4g2k	TC**K**VLGPDCSIGIEDLS**R**NISEQIDQI**KK**DEQ**K**	Lake Victoria Marburg Virus
4n21	LCNVV**K**GPMCSSDIDDFS**K**NVSDMID**K**VHEEM**KK**FYHE	California Academy of Science Virus
4njl	SIPNFGSLTQINTTLLDLTYEMLSLQQVV**K**ALNESYIDL**K**ELGN	Middle East respiratory syndrome-related coronavirus
5yl9	VPDLVVEQYNQTILNLTSEISTLEN**K**SAELNYTVQ**K**LQTLIDNINSTLVDL**K**W	Human coronavirus 229E
6jgy	P**K**CWLVSNGSYLNETHFSDDIEQQADNMITEMLQ**K**E	Lassa mammarenavirus
6tmr	PDGQILIPEMQSEQL**K**QHMDLL**K**AAVFPL**R**HPLI	Mokola virus

Fig. S1. Basic residues are common in HR2-equivalent regions of type I enveloped viruses. We surveyed the PDB for X-ray crystal structures of post-fusion bundles of fusion proteins from various type I enveloped viruses, and then listed the PDB code, the sequence of the HR2-equivalent region with basic residues highlighted, and the name of the virus, as shown above.

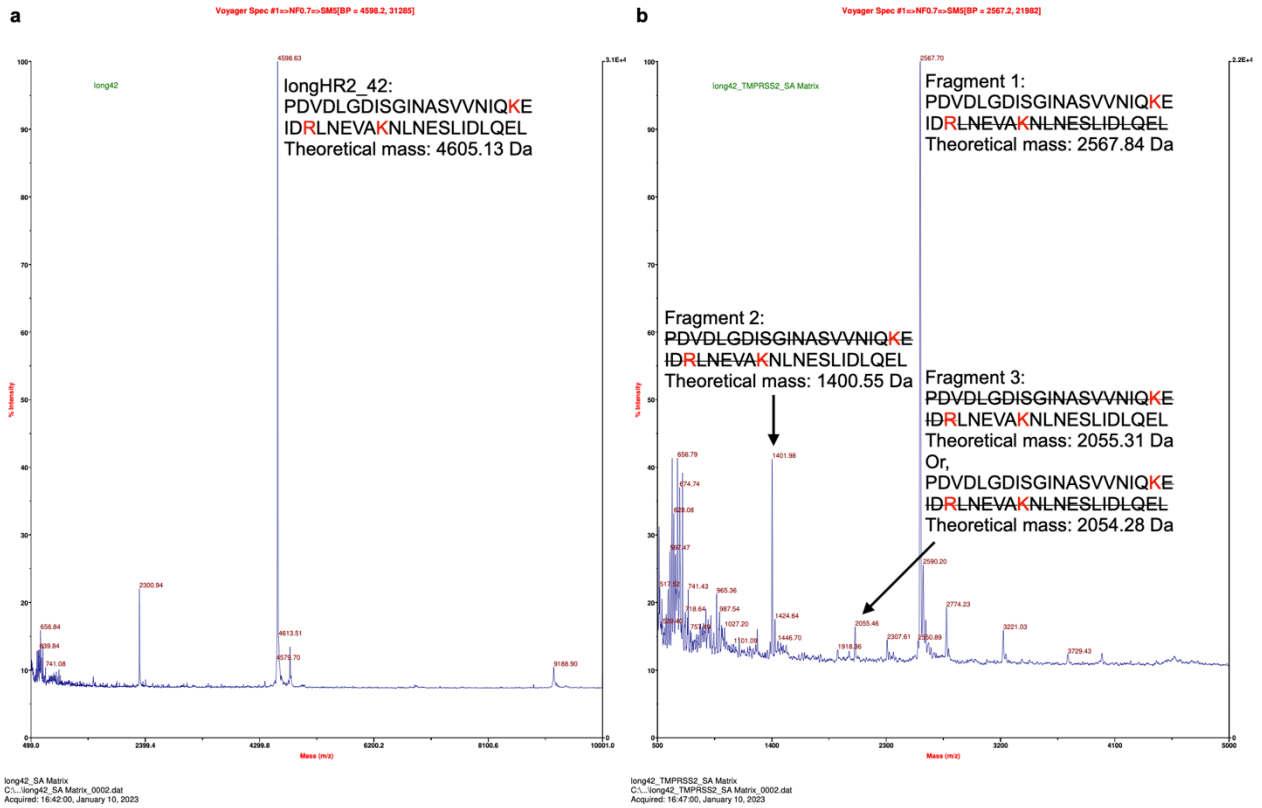


Fig. S2. Mass spectrometry of longHR2_42 shows that TMPRSS2 cleaves the peptide at R1185 and K1191, and possibly at K1181.

(**a**) Mass spectrometry of longHR2_42 before TMPRSS2 cleavage. The sequence and theoretical mass of intact longHR2_42 are indicated. All basic residues (K1181, R1185, and K1191) are colored red.

(**b**) Mass spectrometry of longHR2_42 after TMPRSS2 cleavage. The sequences and theoretical masses of cleaved fragments are indicated near their corresponding peaks. Sequences not present in the fragments are shown in strikethrough format for easier comparison with the intact sequence. Fragment 1 identifies R1185 as a cleavage site. Fragment 2 identifies K1191 as a cleavage site. Fragment 3, with a measured mass of 2055.46 Da, could correspond to two possible fragments with a theoretical mass of 2055.31 Da (cleavage at R1185) and 2054.28 Da (cleavage at K1181), respectively.

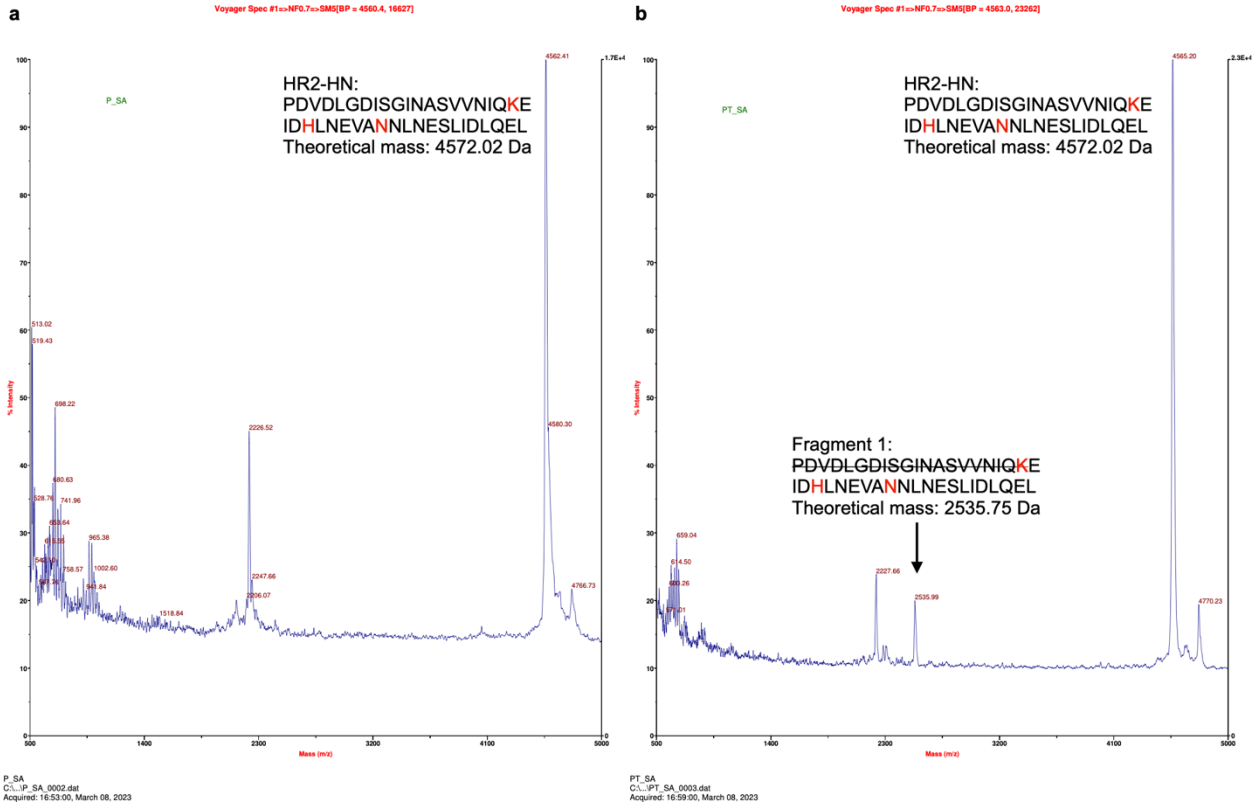


Fig. S3. Mass spectrometry of HR2-HN shows that the TMPRSS2 cleavage site at K1181 is critical.

We synthesized a mutant of HR2, named HR2-HN, with only the two confirmed cleavage sites mutated (R1185H and K1191N). We found that TMPRSS2 cleaves HR2-HN at K1181, albeit much less efficiently than it does the wildtype HR2 peptide (longHR2_42).

(**a**) Mass spectrometry of HR2-HN before TMPRSS2 cleavage. The sequence and theoretical mass of intact HR2-HN are indicated.

(**b**) Mass spectrometry of HR2-HN after TMPRSS2 cleavage. The sequences and theoretical masses of cleaved fragments are indicated near their corresponding peaks. Sequences not present in the fragments are shown in strikethrough format for easier comparison with the intact sequence.

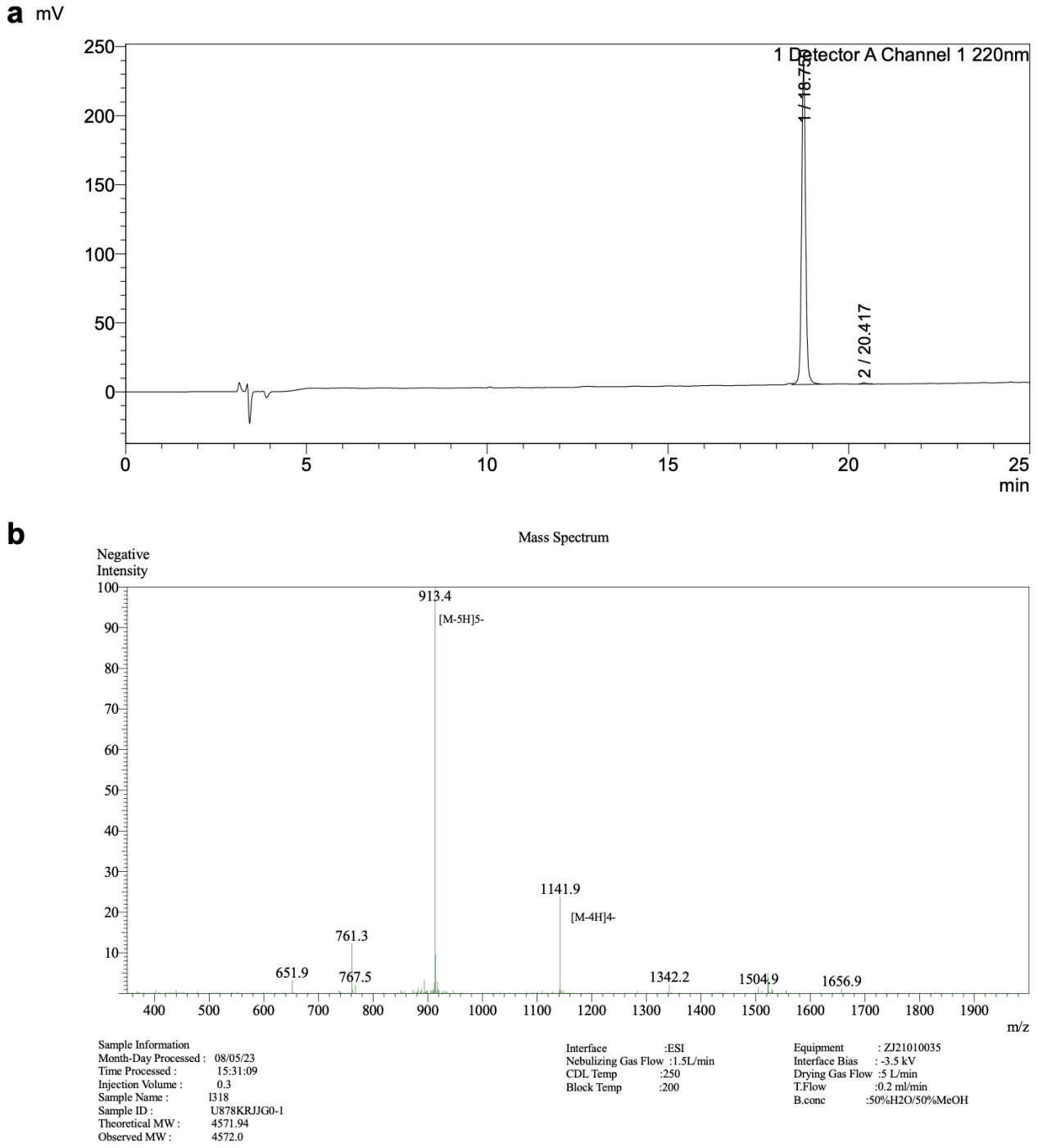


Fig. S4. HPLC (a) and LC-MS (b) profiles of HR2-QHN.

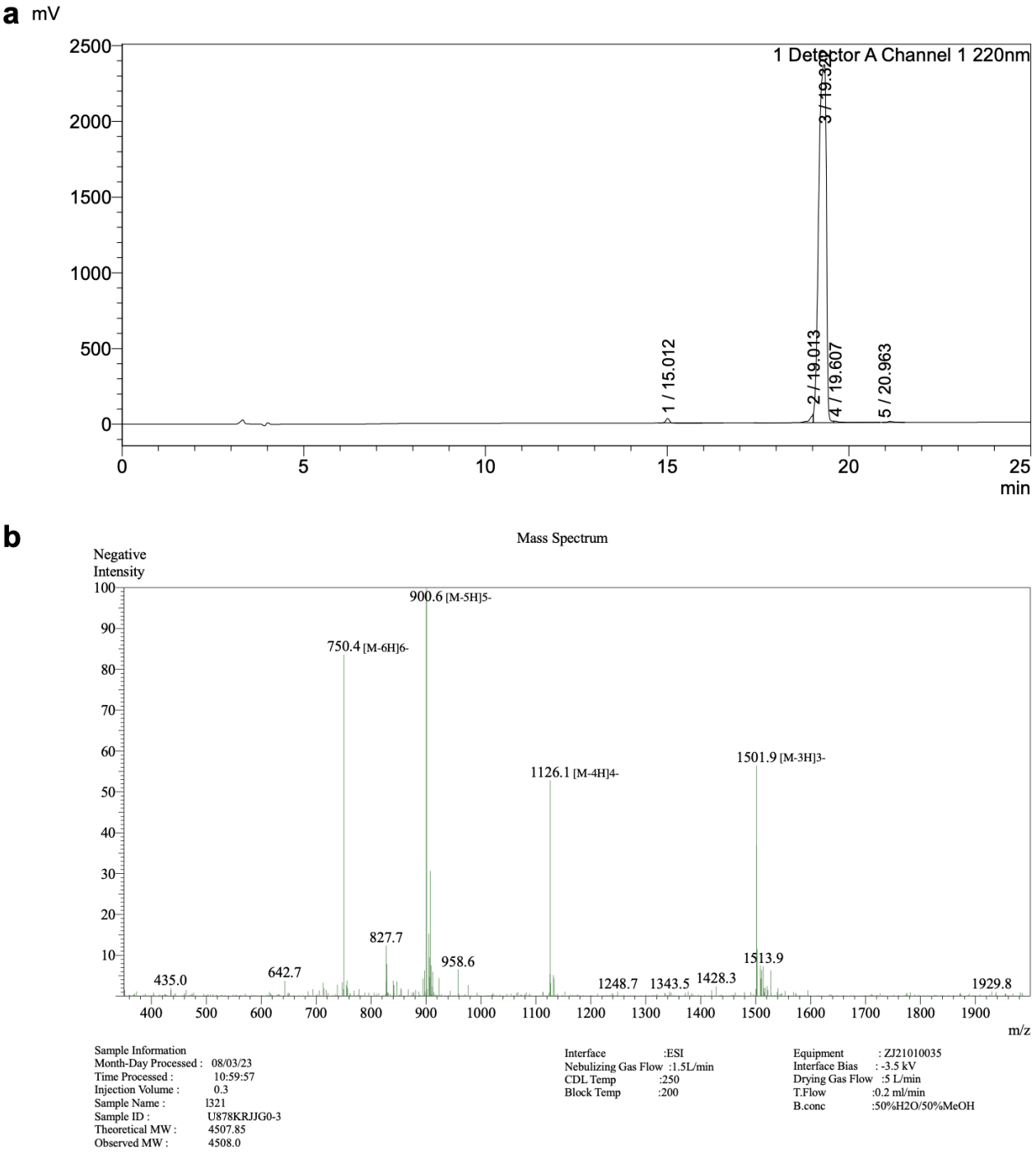


Fig. S5. HPLC (a) and LC-MS (b) profiles of HR2-NNS.

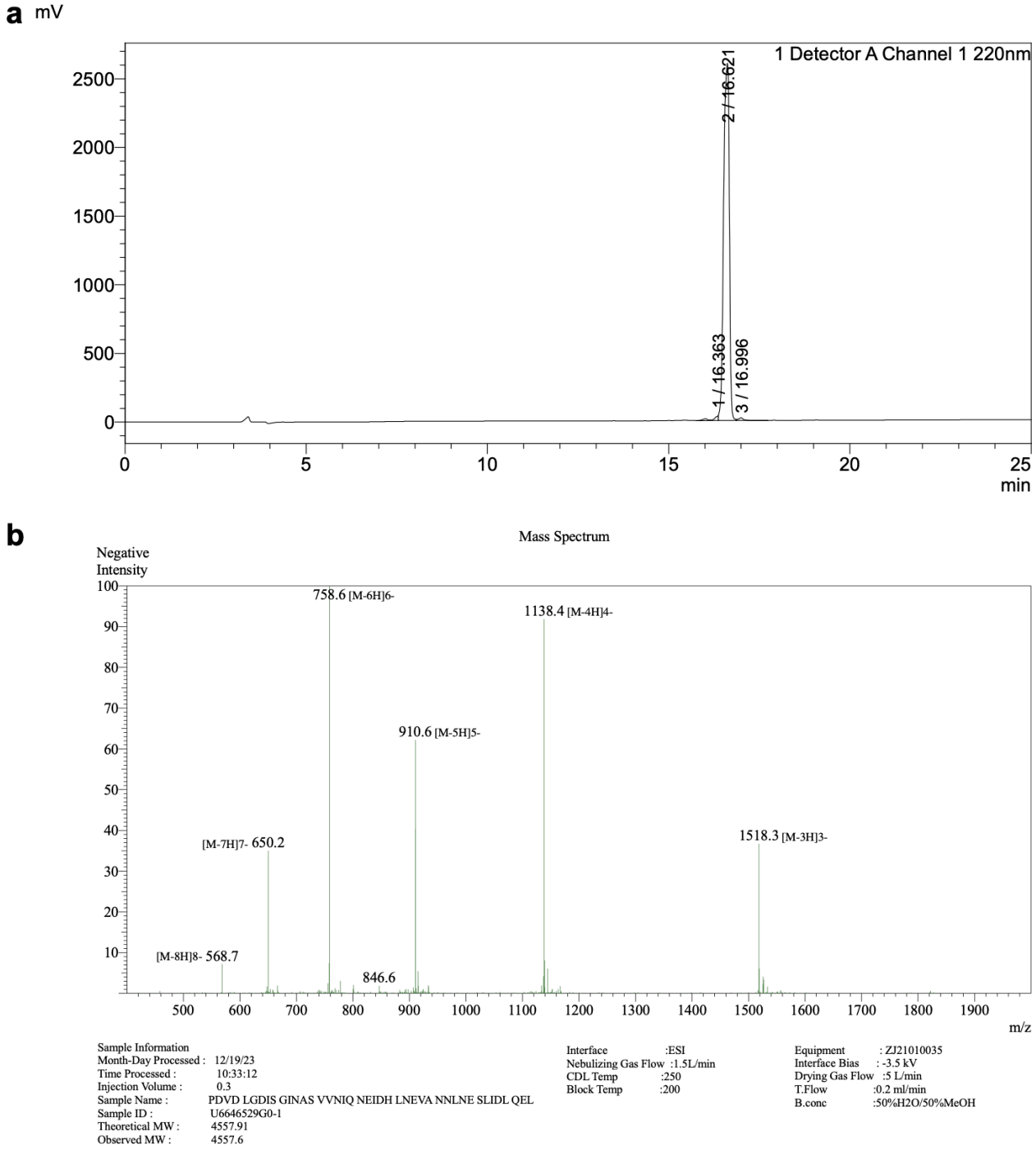


Fig. S6. HPLC (a) and LC-MS (b) profiles of HR2-NHN.

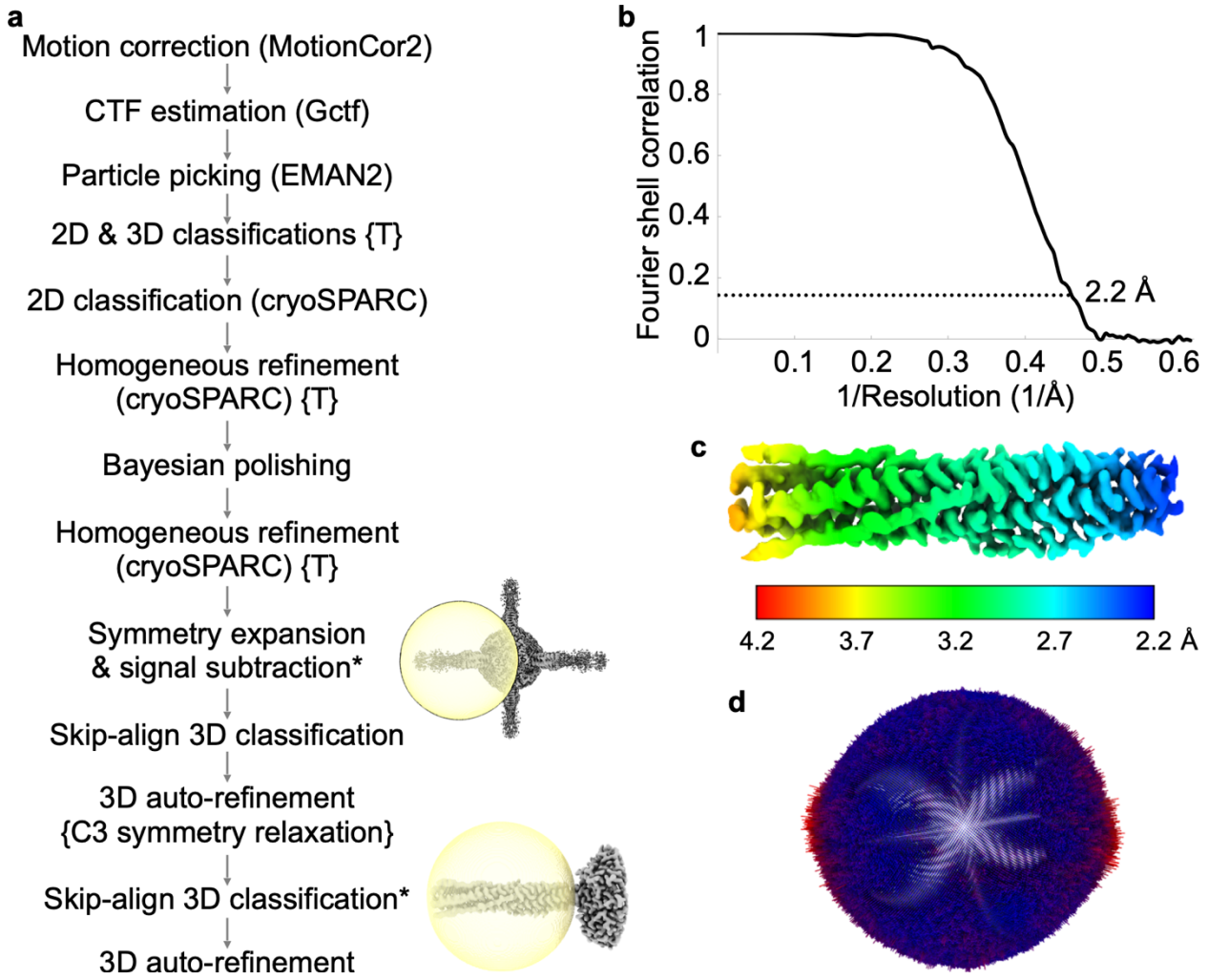


Fig. S7. Cryo-EM Structure determination of HR2-NHN bound to HR1.
(**a**) Workflow of cryo-EM data processing. RELION was used for each step unless another program is indicated in parentheses. Symmetry was imposed or relaxed as indicated in curly brackets. The steps that used a manually generated mask (rather than the default spherical mask in RELION or the default dynamic masks in cryoSPARC) are indicated with a star sign and an image showing the manually generated mask (yellow) and an average map (gray).
(**b**) Fourier shell correlations (FSC) of the final local refinements with RELION (last 3D auto-refinement). Note that the FSC calculation was performed using the default spherical mask that covers both the HR1HR2 bundle and part of the scaffold.
(**c**) Local resolution of the final reconstruction.
(**d**) Distribution of the particles' orientations in the final reconstruction depicted in the same orientation as the map in panel **c**. The length of each bar is proportional to the number of particles oriented in that direction. The bars are also colored by length, with red indicating more particles and blue fewer.

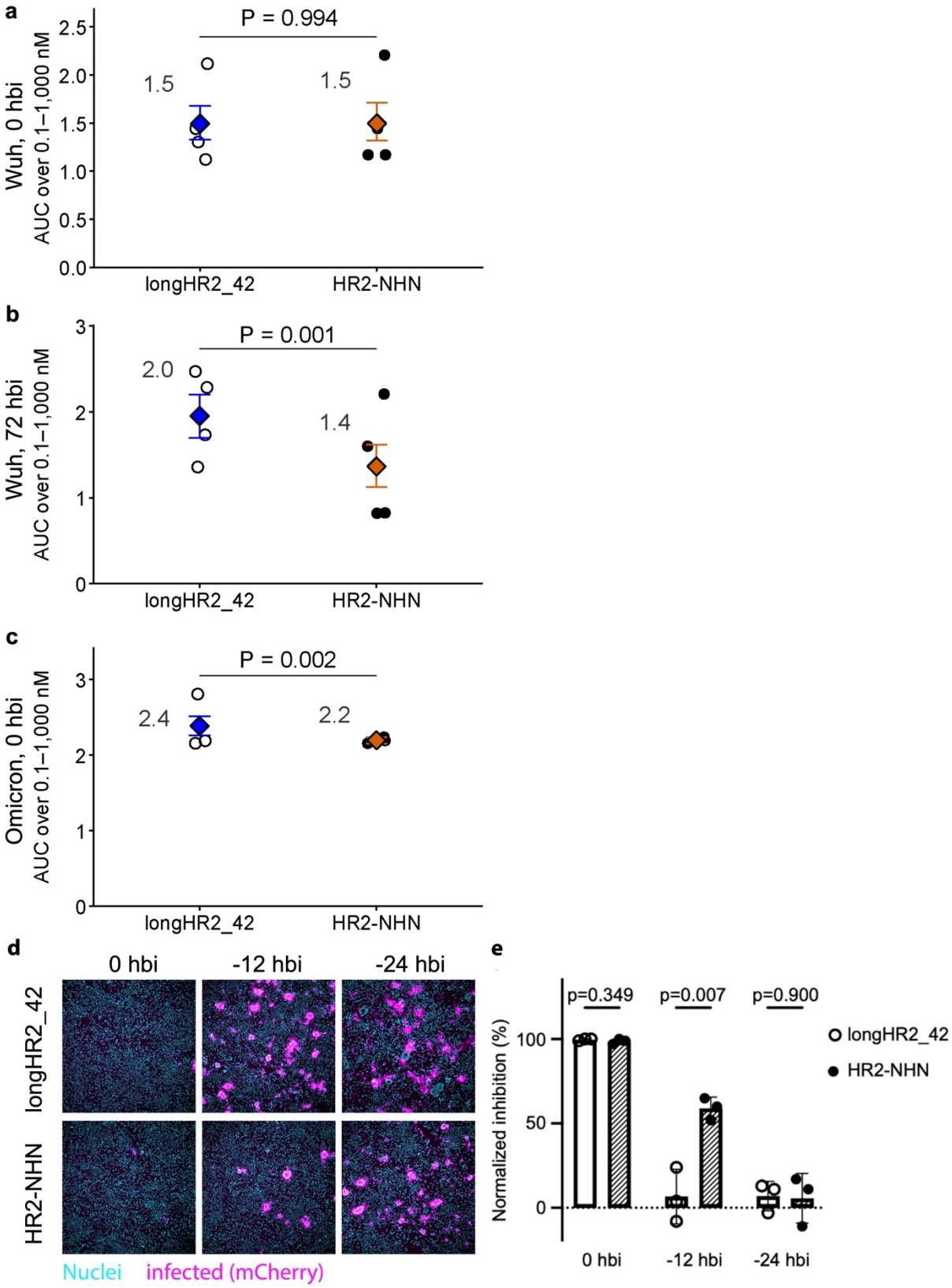


Fig. S8. Comparison between longHR2_42 and HR2-NHN inhibitory activities when preincubated with cells for different times before infection. Comparison of dose-response AUC between longHR2_42 and HR2-NHN when administered to Calu-3 cells at 0 hbi (**a**) and 72 hbi (**b**) with SARS-CoV-2-mCherry Wuhan, and at 0 hbi with Omicron Xbb1.5 (**c**). AUC was

calculated over a peptide concentration range of 0.1–1,000 nM. Individual points represent AUC values calculated from independently repeated dose-response series. Diamonds indicate the pooled AUC estimate, and error bars indicate the bootstrap 95% confidence interval. Exact bootstrap P values for the comparison between peptides are shown. Blue, longHR2_42; orange, HR2-NHN. (**d**) Example cell images and normalized inhibition (**e**) from the infection of A549-AT cells by recombinant SARS-CoV-2-mCherry with the respective peptides for 0 h, 12 h, and 24 h before infection. Nuclei are stained with Hoechst DNA dye (cyan), and infected cells are detected with mCherry red fluorescent protein (magenta) expressed by the recombinant virus. The P values indicating the statistical significance of each sample pair are indicated in the graph. Statistical analysis was performed by the unpaired two-tailed t-test.

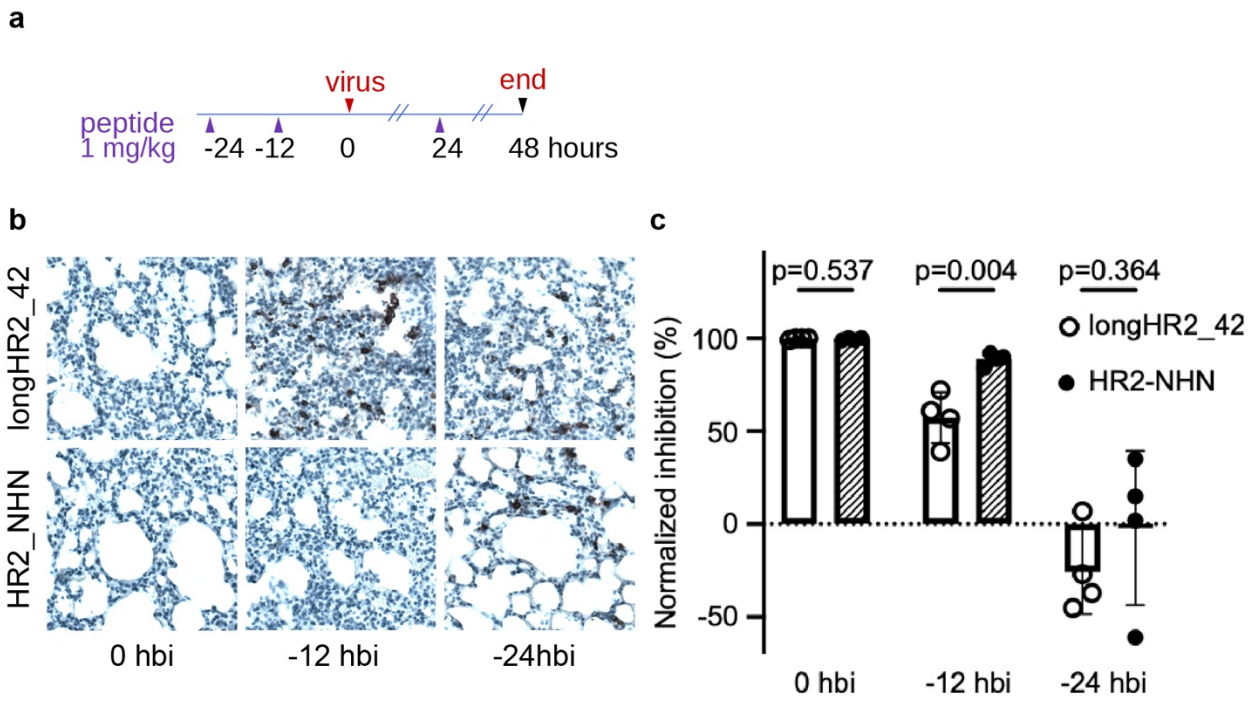


Fig. S9. The TMPRSS2-resistant HR2-NHN maintains inhibitory activity longer than wildtype longHR2_42 in an authentic SARS-CoV-2 mouse infection model.
(**a**) Workflow of *in vivo* experiment identifying the duration of antiviral activity for the peptides. (**b**) Example lung section images after immunohistochemistry to visualize viral N protein expression (dark precipitate) accumulating in bronchiolar epithelial cells and pneumocytes counterstained with hematoxylin (light blue staining) at 48 hpi, and (**c**) normalized inhibition quantified by RT-PCR from the infection of mice with SARS-CoV-2 beta. In (**b**) and (**c**), infected mice were pretreated with the respective peptides for 0 h, 12 h, and 24 h before infection. In panel (**c**), p-values are calculated by the unpaired two-tailed t-test are indicated for each time point.

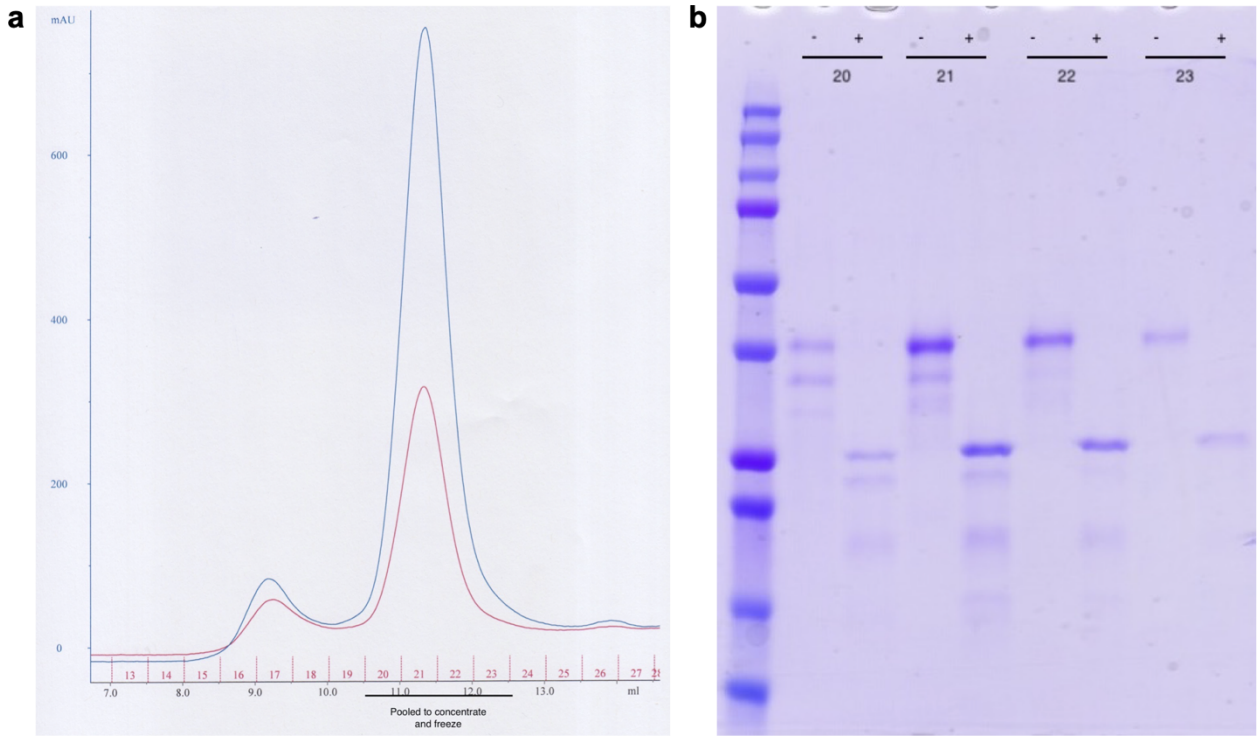


Fig. S10. Purification of TMPRSS2.

(**a**) Profile of the last SEC step of a TMPRSS2 preparation. UV absorption curves at 280 nm and 260 nm are colored blue and red, respectively.

(**b**) SDS-PAGE of the pooled fractions in the last SEC step. The fraction numbers were labeled on top of the lanes with the plus and minus signs denoting with and without DTT, respectively.

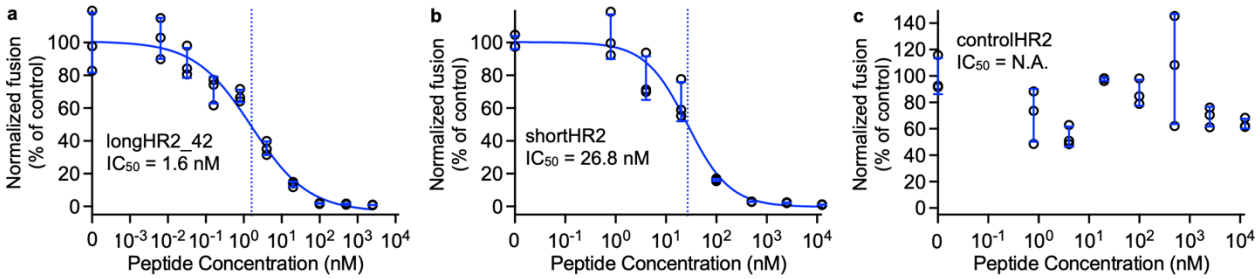


Fig. S11. Inhibitory activities of peptides prepared with our new protocol in a cell-cell fusion assay.

The raw data points are plotted as black circles, while the error bars (SD), fitted curves, and vertical dashed lines at IC_{50} are plotted in blue.

Table S1. Cryo-EM data collection, refinement, and model building

	HR1/HR2-NHN
EMDB	EMD-70569
PDB	9okn
Microscope	Titan Krios
Voltage (kV)	300
Camera	Gatan K3
Pixel size (Å)	0.65
Exposure time (s)	1.1
Number of frames per exposure	40
Total Dose ($e^-/Å^2$)	52
Number of movies	20,031
Defocus range (μm)	-2 to -0.3
Number of particles	838,766
Resolution of final global refinement (0.143 FSC, Å)	1.7
Resolution of final local refinement (0.143 FSC, Å)	2.2
Bond RMSD (Å)	0.005
Angle RMSD (°)	0.473
Molprobity score	1.31
Clashscore, all atoms	5.72
Ramachandran favored (%)	98.47
Ramachandran allowed (%)	1.53
Ramachandran outliers (%)	0
Rotamer outliers (%)	0.67
Cβ outliers (%)	0
CaBLAM outliers (%)	0