Discovery of novel quinoline papain-like protease inhibitors for COVID-19 through topology constrained molecular generative model

Yongzhi Lu[1#], Ting Ran[1#], Qi Yang[1#], Guihua Zhang[1#], Jiayi Chen[1,2], Peiqi Zhou[1], Wenqi Li[1,3], Minyuan Xu[1], Jielin Tang[1], Minxian Dai[1], Jinpeng Zhong[1], Hua Chen[1], Pan He[1], Anqi Zhou[1], Bao Xue[1], Jiyun Zhang[1,2], Kunzhong Wu[1,4], Xinyu Wu[1,5], Miru Tang[1], Xinwen Chen[1*], Hongming Chen[1*] and Jinsai Shang[1,6*]

Affiliations

[1]State Key Laboratory of Respiratory Disease, The First Affiliated Hospital of Guangzhou Medical University, School of Basic Medical Sciences, Guangzhou National Laboratory, Guangzhou Medical University, Guangzhou 511436, China

[2]School of Pharmaceutical Sciences, Sun Yat-Sen University, Guangzhou 510006, China

[3]College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China

[4]State Key Laboratory of Medicinal Chemical Biology, College of Life Sciences, Nankai University, Tianjin 300071, China

[5]State Key Laboratory of Virology, Wuhan Institute of Virology, Chinese Academy of Sciences, Wuhan, 430071, China

[6]Lead contact

[#] These authors contributed equally to this work

* Correspondence: Xinwen Chen; Hongming Chen; Jinsai Shang

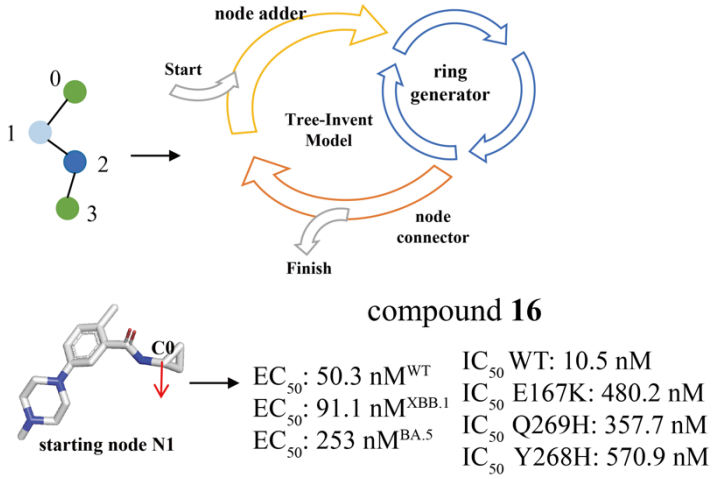
Email: chen_xinwen@gzlab.ac.cn; chen_hongming@gzlab.ac.cn; shang_jinsai@gzlab.ac.cn

Abstract

The rapid emergence of drug-resistant SARS-CoV-2 variants poses a persistent challenge to current antiviral strategies. Mutations in key viral targets, including RNA-dependent RNA polymerase (RdRp), main protease ($3CL^{pro}$), and papain-like protease (PL^{pro}), have been shown to markedly reduce the efficacy of approved therapeutics, highlighting the urgent need for next-generation antivirals capable of overcoming resistance. Here, we report the discovery of a novel class of PL^{pro} inhibitors through an AIDD strategy based on a topology-constrained molecular generative model (Tree-Invent) integrated with structure-guided optimization. Scaffold hopping from a previously reported lead enabled the identification of a quinoline-based chemical series with substantially improved metabolic stability and antiviral potency. Structure-guided optimization yielded compound GZNL-2016, which exhibited potent enzymatic inhibition of PL^{pro} (IC_{50} = 10.5 nM), robust antiviral activity against multiple SARS-CoV-2 variants, including Omicron BA.5 and XBB.1, and favorable pharmacokinetic properties following oral administration. Notably, GZNL-2016 retained substantial inhibitory activity against the clinically relevant drug-resistant mutant PL^{pro} E167K (IC_{50} = 480.2 nM; K_i = 439.3 nM), in contrast to previously reported inhibitors that exhibit markedly reduced potency. In a SARS-CoV-2 infection mouse model, oral administration of GZNL-2016 significantly reduced pulmonary viral titers, demonstrating in vivo antiviral efficacy. Collectively, this study establishes an AI-enabled strategy for rapid antiviral discovery and identifies GZNL-2016 as a promising lead compound to address the threat of coronavirus infections caused by drug-resistant mutant SARS-CoV-2 variants.

Keywords: SARS-CoV-2, Papain-Like Protease, AIDD, DUBs, Drug Design.

Graphical abstract



We leveraged an AI generative model to discover a novel PL^{pro} inhibitor with excellent liver stability, low CYP, hERG inhibition and reasonable oral PK properties. At the same time, the compound **16** exhibits high efficacy for the resistance mutation E167K, which resulted in severe resistance to the previously reported inhibitor Jun12682 and PF-07957472.

Introduction

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) is the infectious virus responsible for coronavirus disease 2019 (COVID-19). It was initially discovered in China in December 2019 and has rapidly evolved into a global epidemic with a significant cost to human society. Over the past five years, scientists worldwide have been exploring various approaches to treat infected individuals and reduce the severity of the disease. To date, several COVID-19 vaccines have been developed and granted emergency authorization for market distribution[1]. All these vaccines target the spike protein to prevent the virus from entering the host cells. However, the high mutation rate of the spike protein, especially within the receptor-binding domain (RBD), diminishes the long-term protective effectiveness of the vaccines[2]. This underscores the need to develop alternative therapeutic modalities, particularly small molecule antiviral drugs targeting more conserved viral proteins.

Currently, the FDA has approved three small-molecule drugs, including remdesivir, molnupiravir and nirmatrelvir[3]. They all target relatively conserved proteins across different SARS-CoV-2 variants. The first two drugs are viral RNA-dependent RNA polymerase (RdRp) inhibitors, while the third one is an inhibitor of the viral main protease (M^{pro}), also known as 3CL protease ($3CL^{pro}$). Remdesivir was the first drug authorized for emergency use, but its administration is limited to intravenous infusion for hospitalized patients with severe symptoms[4]. Its clinical efficacy has been found to be sub-optimal, as demonstrated by several clinical trials[5]. Similar to remdesivir, molnupiravir is a nucleotide analogue and functions as a mutagen, inhibiting SARS-CoV-2 replication by increasing the viral mutation rate[6]. However, there are concerns that molnupiravir may also induce similar mutations in the host[7]. Nirmatrelvir, combined with a cytochrome P450 3A (CYP3A) inhibitor (*i.e.,* ritonavir), is an oral medication and suitable for non-hospitalized patients[8]. Ensiltrelvir, a M^{pro} inhibitor approved in Japan[9], has demonstrated high efficacy in clinical trials. However, it is also a potent CYP3A inhibitor, which may lead to severe adverse drug-drug interactions with other medications[10]. Therefore, new antivirals with alternative mechanisms are still urgently needed.

The papain-like protease (PL^{pro}) is one of the two viral proteases responsible for cleaving the polyproteins encoded by the SARS-COV-2 genome[11]. Additionally, it can remove the ISG-15 and ubiquitin modifications on host proteins[12]. This dual enzymatic function allows it to both promote virus replication and suppress the host's immune response. Moreover, PL^{pro} is a highly conserved protein across SARS-CoV-2 variants[13], making it a promising drug target for anti-COVID-19 therapy. The catalytic site of the protein is situated within a small buried pocket, adjacent to an induced substrate-binding cavity by the reshaping of the BL2 loop, a flexible beta turn/loop region (Gly266–Gly271)[14].

GRL0617 was originally discovered as a non-covalent anti-SARS-CoV PL^{pro} inhibitor binding at the induced substrate-binding cavity and later validated to have inhibitory potency against SARS-COV-2 PL^{pro} at the micromolar level[14]. Analogs of **GRL0617** were developed with moderate enzymatic and cellular activities, such as **Jun9-84-3**[15], **XR8-23**[16], **Jun11273**[17] and covalent inhibitors **CP7**[18]. Recently, a few compounds[19] (such as **Jun12862, PF-07957472,**

GZNL-P4) were reported to have in vivo efficacy in SARS-CoV-2 infected mice, along with oral bio-availability. **GZNL-P4** is a potent compound identified in our previous study, but structural optimization is needed to improve its low liver stability. The development of PL^{pro} inhibitors at the clinical stage remains rare. To date, only one PL^{pro} inhibitor has been reported to have entered clinical trials in China, and its structure has yet to be disclosed[20]. Therefore, the development of PL^{pro} inhibitors remains necessary to explore their potential as a COVID-19 therapy.

The high mutation frequency of coronaviruses has given rise to a severe and increasingly pressing challenge of drug resistance. During the COVID-19 pandemic, vaccines and antibody drugs played a crucial role; however, later-emerging Omicron subvariants (e.g., BA.1, BA.2, BA.5, BQ.1.1, XBB, XBB.1.5, etc.) exhibited high or complete neutralization escape against nearly all authorized monoclonal antibodies, leading to their successive withdrawal from clinical use [21]. Similarly, mutations in Mpro (e.g., Q192R, H172Y, H172Y/Q189E, L50F/E166V, and L50F/E166A/L167V) have been demonstrated to reduce susceptibility to nirmatrelvir and ritonavir, with the decrease ranging from moderate (25–99-fold) to high (> 100-fold) [21-22]. *In vitro* studies further indicate that multiple mutations in RdRp (including E802D, V166A, N198S, S759A, V792I, C799F/R, etc.) can decrease sensitivity to remdesivir, with the magnitude of reduction varying from several-fold to several hundred-fold [23]. Several key mutations in PL^{pro} have been confirmed to significantly impair the efficacy of existing inhibitors, which constitutes a core challenge for the development of PL^{pro} inhibitors. Tan et al. reported that residues E167, Y268, and Q269 are mutation hotspots associated with resistance to existing PL^{pro} inhibitors [24]. Especially, the mutations of E167K and Y268N in PL^{pro} dramatically decrease the inhibitory activity of Jun2682 and PF-02957472 with the Ki from 26.51 nM and 6.85 nM to greater than 10 μ M and 8.036 μM, respectively [24]. Although no PL^{pro} inhibitor has reached clinical approval to date, drug resistance remains a significant concern. Consequently, developing PL^{pro} inhibitors that can overcome drug resistance is critical for future pandemics.

Artificial intelligence (AI) technology has shown great potential to accelerate drug discovery[25]. Among these models, deep molecular generation models have garnered particular attention due to their ability to autonomously produce novel molecular structures. Early molecular generation models primarily involved random sampling from the learned continuous molecular chemical space to generate a large number of molecular structures[26]. However, due to the vastness of chemical space, the probability of discovering active molecules through sampling these models is relatively low. In recent years, there has been an increasing focus on achieving goal-directed molecular design through the development of conditional molecular generation models. Key approaches include altering the form of model input[27], utilizing conditional deep neural networks [28], applying transfer learning[29], and employing reinforcement learning[30], etc. Some of the latest methods directly decode the learned three-dimensional structural features of protein pockets into molecular chemical structures[31]. Although numerous methods have been developed, only a few of them have been reported with application in real-world scenarios.

In earlier research, we developed Tree-Invent[32], a graph-based molecular generation model with topology constraints. This model is versatile across different drug design scenarios and excels

in achieving precise control over structural topology during molecular generation. In this model, a molecular graph is represented as a topological tree comprising ring nodes, non-ring nodes, and edges connecting the nodes. Molecule generation is achieved by carrying out operations of node addition, ring generation, and node connection in an autoregressive manner. Reinforcement learning (RL)[33] is also integrated within the Tree-Invent model to facilitate the exploration of chemical space toward desired properties. We applied this method to conduct scaffold hopping exercise and discovered a novel quinolone chemical series that exhibits nanomolar potency in enzymatic assay and excellent anti-viral potency in cellular experiments. Our best compound exhibits nanomolar level inhibition activity to PL^{pro} E167K and demonstrates reasonable in vitro and in vivo oral PK properties. The significant selectivity over normal human cells and weak inhibition on the CYP and hERG proteins indicate its promising safety profile for antiviral treatments.

Results

Scaffold hopping through topology constrained molecular generation with Tree-Invent

So far, exploration of the structure-activity relationship (SAR) based on the structure of **GRL-0617** has resulted in several potent PL^{pro} inhibitors (Fig. S1A). In current study, we utilized the Tree-Invent model (Fig. S2) to design novel PL^{pro} inhibitors, using the crystal structure of PL^{pro} enzymatic domain in complex with known inhibitor GRL0617 (PDB ID: 7JRN), which exhibits potency at micromolar range (Fig. S1B), where the naphthalene ring serves as the head group, is located at the BL2 groove sandwiched between the aromatic residue Y269 and the hydrophobic residues P247 and P248. The aniline group in the tail group is located at the P3 sub-pocket and extends into the P2 sub-pocket near L162 through a methyl group. The linker group is located at the P4 sub-pocket, where the hydrogen bonds with D164 and Q269 are formed by the amide group. Accordingly, the structure of **GRL0617** can be divided into three parts, namely the head, linker and tail part, corresponding to three sub-pockets in the substrate binding site (Fig. S1C) [14].

In our previous work, we identified a potent compound, **GZNL-P4**, which features a tricyclic dihydroacenaphthylene (DHAN) scaffold. The co-crystal structure of **GZNL-P4** with PL^{pro} (PDB ID: 8YX2) reveals that its tail group extends into the solvent-exposed region at the P3 sub-pocket. The positively charged piperazine nitrogen in this group can form salt bridge with E167, an interaction absent in the **GRL0617** complex structure (Fig. 1A). However, **GZNL-P4** has poor liver metabolic stability (rat liver microsome stability $t_{1/2}$ 7.95 min), which is believed to be associated with the head DHAN group, as observed in the metabolism of **GRL0617** [34], so finding other head groups is needed to address the stability issue.

Herein, we employed Tree-Invent model to explore the chemical space of head group. The structure of **GZNL-P4** was transformed to a topology tree with four nodes connected by single bonds (Fig. 1B). To focus on the generation of the head group, node N1 was predefined as the substructure comprising the linker and tail part of **GZNL-P4** and kept unchanged in the process of molecule generation. Node N2 was designated as a changeable node grown from node A to replace the head group of **GZNL-P4**. Analysis of the binding modes of **GZNL-P4** and **GRL0617**

179 indicates that the aromatic ring system of the head group is important for binding through π-π
180 stacking interaction with Y268 and hydrophobic interaction with P247 and P248 [14]. To mimic
181 this binding mode, node N2 was confined to contain 5,6- or 6,6-fused bicyclic systems with at
182 least one aromatic ring. Furthermore, two nodes, *i.e.,* N3 and N4, were additionally appended to
183 node N2 as R-group decorations. However, their generation was non-essential and entirely
184 contingent upon the model's decision regarding node addition. With these topological constraints,
185 Tree-Invent was implemented with RL to generate molecules (Fig. 1C), and two rounds of RL
186 were conducted for the generation of 5,6- and 6,6-fused bicyclic molecules. The crystal structure
187 of **GRL0617** was chosen as the protein model for carrying out docking during the generative
188 process and the docking score was transformed into reward score during the RL process, aiming
189 to generate molecules with high affinity.

190 Through 1000 epochs of RL, the learning curves of the Tree-Invent model converge (Fig.
191 S3A), suggesting the model has learned how to generate high-score molecules with 5,6- and 6,6-
192 fused bicyclic systems. Combining the generation sets of the two RL runs, 2817 unique molecules
193 were obtained. Among them, 1270 molecules have 5,6-fused ring structure and 1547 molecules
194 have 6,6-fused ring structure (Fig. S3B). Moreover, the substructures corresponding to node N1 in
195 these molecules are kept the same as in **GZNL-P4**. Half of the molecules have a naked ring
196 structure in the head group, while the others have no more than two substitutes on the ring structure
197 as expected (Fig. S3C). This demonstrates that the Tree-Invent model is capable of generating
198 molecules satisfying the topological constraints. It was also found that the naphthalene group of
199 **GRL0617** can be reproduced which proves the capability of this method in identifying active
200 molecules. Subsequently, all generated molecules were redocked to the binding pocket using the
201 Glide docking program under the standard precision mode (Glide-SP)[35], resulting to 367
202 molecules with similar binding mode to **GZNL-P4** and **GRL0617** and docking scores lower than
203 -7.0 kcal/mol, a cutoff to distinguish suitable molecules from all generated molecules (Fig. S3D).
204 Moreover, these remaining molecules are all compliant with the Lipinski rules[36]. Chemical space
205 analysis based on the PCA analysis [37] using the Morgan fingerprint [38] reveals that these
206 molecules can be grouped into several categories, and each cluster has some molecules with good
207 docking scores (Fig. 1D). The chemical space of molecules filtered based on docking scores was
208 displayed using tSNE plot [39] (Fig. 1D).

209 **Identification of novel scaffold of PL^{pro} inhibitors**

210 Considering synthesis feasibility, nine molecules were chosen for synthesis. The synthetic
211 route was illustrated as Scheme 1 in the supplemental text. The structures of the synthesized
212 compounds are shown in Fig. 2A. Their docking poses can be found in Figure S4. Then, they were
213 tested using a fluorescence-based biochemical assay[40] to evaluate their inhibitory activities on
214 PL^{pro} (Fig. 2B). Compounds **1** and **2**, containing a quinoline ring in the head group, are the most
215 active compounds. Their aromatic nitrogen atoms in the quinoline ring are situated at the para-
216 position to the linker substituent. Chlorination on the quinoline ring in compound **1** shows a minor
217 impact on the activity, while compound **2** is suggested to have better binding affinity (Fig. 2C).
218 The binding affinity of compound **2** with PL^{pro} measured by isothermal titration calorimetry (ITC)

219 and biolayer interferometry (BLI) are 285 nM and 187 nM (Fig. S5A, S5C), respectively.
220 Compound **6** with benzothiophene exhibits somewhat lower activity than compound **2**. Other
221 compounds having 5,6 ring systems such as indole, benzopyrazole and 1H-pyrrolopyridine have
222 much worse activity, regardless of whether the head group is a five-membered or six-membered
223 ring. Among them, compound **8** with azaindole is the weakest. N-methyl substitution on the indole
224 ring of compound **9** improves the activity, probably because it makes additional van der Waal
225 interaction with the residue P248. Compound **5** with furopyridine has a special substitution
226 position on the head group, indicating that the direction of substitution on the head group is very
227 important for activity. Through the AI-driven R-group generation process, we successfully
228 identified several PL^{pro} inhibitors with new scaffolds having reasonable potency. The liver
229 metabolic stability of compound **2** was measured, and to our delight, its half-life in liver microsome
230 has been largely improved compared to **GZNL-P4** (Fig. 5E).

231 To better understand the binding mode of compound **2**, we resolved its co-crystal structure in
232 complex with PL^{pro} (Fig. 3A). The residues surrounding the binding pocket maintain nearly
233 identical conformations to those observed in the co-crystal structure of **GRL0617**, and there is a
234 good alignment between the binding poses of the two compounds (Fig. 3E). The binding mode
235 analysis shows that the quinoline ring of compound **2** is overlapped with the head group of
236 **GRL0617**. As expected, the quinoline ring makes π-π stacking interaction with Y268, and extends
237 the interaction with P247 and P248. The nitrogen atom on the quinoline ring is exposed to solvent,
238 making no obvious binding with protein residues. The cyclopropyl group binds at the bottom of
239 the pocket surrounding Y264 and T301, and the amide linker group makes hydrogen bonds with
240 D164, Y268 and Q269. The tail group, i.e., piperazine-substituted benzyl group, is exposed to
241 solvent, and the positively charged piperazine nitrogen in this group can form a salt bridge with
242 E167, an interaction absent in the GRL0617 complex structure, while the benzyl group extends
243 into the sub-pocket near L162 through a methyl group. This binding mode is consistent with that
244 of **GZNL-P4**, which provides clear guidance for structural optimization.

245 **Structural optimization of compound 2**

246 To further improve the potency of compound **2**, we explored substituents at the R^2 position
247 to optimize its interaction with E167. We selected commercially available amine-containing
248 reagents to replace the N-methylpiperazine group of compound **2** and designed a virtual library of
249 2347 compounds containing positively charged amine groups (Fig. 2D). Molecular docking using
250 the co-crystal structure of compound **2** as docking model was then carried out, compounds whose
251 docking pose is similar to the binding pose of compound **2** and simultaneously maintains the salt
252 bridge with E167 in an appropriate distance (the cutoff is set to 4.0 Å) were selected. At last, 29
253 top-ranked compounds (by Glide docking score) were selected to synthesize (Scheme 2 in the
254 supplemental text) and their structures are listed in Table 1. They can be divided into two types in
255 terms of the ring count in the R^2 group.

256 Following SAR analysis reveals that the R^2 group greatly influences the inhibitory activity of
257 PL^{pro} (Table 1 and Fig. S6). Several compounds show superior activity to compound **2**, and three
258 compounds demonstrate IC_{50} values lower than 0.015 μM. Among type I molecules with a single

ring in the R^2 group, compound **10** with piperazine substitute has worse potency than compound **2** having a methylpiperazine group, which means that the activity is sensitive to the position of methyl substituent on the piperazine ring. Compound **2** with 4-methylpiperazine exhibits much higher activity than compounds **14** and **15** which contain the substitutes of 2- and 3-methylpiperazine. The compounds with hydroxyalkyl groups substituted on the 4-position of piperazine have slightly worse activity than compound **2**. However, substituting with dimethylaminoalkyl groups, as seen in compound **12**, negatively impacts activity. N-sulfonylation of piperazine significantly reduces activity, likely due to the disruption of the salt bridge formed between piperazine and the acid group in E167. It is worthwhile noting that the bridged piperazine significantly enhances activity. Compound **16** (also named as GZNL-2016) is the most potent in this series with IC_{50} of 10.5 nM. Compounds bearing similar cycloamine groups, including the ones with spiral and fused rings, also demonstrate reasonable activity, although they don't show improvement compared to compound **2**. These groups can mimic the salt bridge interaction of piperazine. Nevertheless, the introduction of seven-member-ring results in a reduction of activity, which may be due to the sub-optimal salt bridge interaction of the secondary amine in the enlarged ring. Compared to compound **2** with a Kd of 285 nM (Fig. S5A), compound **16** exhibits higher binding affinity with PL^{pro}. The measured Kd of compound **16** by isothermal titration calorimetry (ITC) and biolayer interferometry (BLI) are 68.6 nM and 78.9 nM (Fig. S5B, S5D), respectively.

Type II molecules contain the R^2 substituents involving two cycloamine groups linked by a single bond. Interestingly, the majority of compounds exhibited comparable activity to compound **2**, except for compounds **35** and **36**. The reduced activity of compound **35** is attributed to the N-methyl-piperazine substituent, which may form a sub-optimal salt bridge with E167. The weak activity observed for compound **36** indicates that piperidine is worse than piperazine for the substitution on the benzene ring, as evidenced by the decrease in the activity of compound **20**. On the other hand, compounds featuring azetidines substituted with pentacyclic/hexacyclic amines demonstrated the highest activity within the sub-group. Notably, compounds **30** (also named as GZNL-2030) and **31** exhibit potency comparable to compound **16**, making them also promising candidates for further profiling. Corresponding to the activity, the binding affinity of compound **30** with PL^{pro} measured by biolayer interferometry (BLI) is 447 nM (Fig. S5E). Most compounds among **10-36** can improve the thermal stability of PL^{pro} with a delta Tm larger than 3 ℃ (Fig. S5F). These compounds with larger delta Tm exhibit better potency for PL^{pro} (Fig. S5G). Correspondingly, compound **16** exhibits the largest delta Tm (Fig. S5F), which is consistent with its most potent inhibitory activity against PL^{pro} (Table 1).

The binding poses of the most potent compounds from the crystal structures (compounds **16**, **30**) and docking (compound **31**) are consistent with that of compound **2**, which is characterized by the sandwiched conformation of the quinoline ring, the hydrogen bonds with D164, Y268 and Q269, and the salt-bridge with E167 (Fig. 3A-D). Q269 formed a hydrogen bond with compound **2**, but a π-π stacking interaction with compound **16** (Fig. 3B). In the binding pose of compound **16**, the hydrogen bond network composed of D164, E167 and compound **16** mediated by two water molecules W1 and W2 may lower energy status, which may explain the superior performance of compound **16** to other compounds.

300 **The enzymatic inhibition of compound 16 on the PL^{pro} mutations**

301 Mutations at residues E167 and Y268 of PL^{pro} lead to a dramatic decrease in sensitivity to the
302 existing PL^{pro} inhibitors Jun12682 and PF-07957472, with a nearly 1000-fold increase in Ki [24].
303 To assess the susceptibility of compound **16** to distinct PLpro mutations, we evaluated its
304 inhibitory activity against a panel of PLpro mutants, including E167K, E167S, Q269H, Y268N,
305 Y268H, and the double mutant E167G/Q269H (Fig. 4A, S7). The results demonstrated that
306 compound **16** essentially abolished its inhibitory activity against Y268N, while its potency was
307 also significantly diminished against E167S and E167G/Q269H double mutant with IC_{50} of 4.421
308 μ M and 6.883 μM, respectively. Encouragingly, compound **16** retained relatively high inhibitory
309 activity against E167K, Q269H, and Y268H, with IC_{50} values of 480.2 nM, 357.7 nM, and 570.9
310 nM, respectively. For compound **16**, the calculated Ki for PL^{pro} E167K using the Cheng-Prusoff
311 equation [41] was 439.3 nM. In stark contrast, both Jun12682 and PF-07957472 exhibited
312 dramatically declined inhibition activity with Ki around 10 μM against E167K [24]. The binding
313 affinity Kd to PL^{pro} E167K determined using ITC was 256 nM, 1.33 μM, and 1.02 μM, respectively
314 (Fig. 4B-E). This may account for the reason why Compound **16** still retains favorable inhibitory
315 activity against E167K.

316 **Antiviral activity, liver stability and PK profiling of compound 16**

317 The antiviral activities of several potent compounds were measured using a cell-based assay.
318 The antiviral potency with 50% effective concentration (EC_{50}) of compound **2** against wild-type
319 SARS-CoV-2 and two epidemic variants Omicron BA.5 and XBB.1 in infected Vero E6 cells are
320 0.4954 ± 0.1486 μM, 0.6199 ± 0.1061 μM and 0.6159 ± 0.0345 μM (Fig. 5A-C). At the same time,
321 compound **2** shows rather low toxicity to the normal human cell line HEK293T (CC_{50} > 200 μM)
322 (Fig. 5D). Structural optimization on compound **2** resulted in three more potent compounds,
323 compound **16**, **30** and **31**, which have enhanced anti-viral activity and a large enough safety
324 window (CC_{50}/EC_{50} for all variants: >379, >292, and >348 folds for compounds **16**, **30**, and **31**,
325 respectively). Notably, compound **16** has the best antiviral activity compared to compounds **30**
326 and **31** on all three tested virus variants (Fig. 5A-C). And, the variant of Omic-BA.5 is the least
327 sensitive to all three PL^{pro} inhibitors.

328 Compared to the previously reported compound **GZNL-P4** (human and rat liver microsome
329 stability $t_{1/2}$ of 3.8 min and 7.95 min, respectively[19b]), the liver stability of the Quinoline series
330 significantly improved the metabolism stability in both rat and human liver microsomes (Fig. 5E).
331 In rat liver microsomal stability assay, compound **16** displayed a prolonged half-life ($t_{1/2}$ = 120.12
332 min) and a low CLint value (20.77 mL/min/kg), which were 2.66-fold, 2.01-fold, 2.86-fold longer
333 in $t_{1/2}$ and 2.66-fold, 2.01-fold, 2.86-fold lower in CLint than those of compound **2** ($t_{1/2}$ = 45.20
334 min, CLint = 55.20 mL/min/kg), compound **30** ($t_{1/2}$ = 59.70 min, CLint = 41.80 mL/min/kg), and
335 compound **31** ($t_{1/2}$ = 41.98 min, CLint = 59.40 mL/min/kg), respectively. In human liver
336 microsomal stability assay, compound **16** maintained favorable metabolic performance compared
337 with compounds **2** and **31,** but slightly worse than compound **30**.

338 *In vivo* PK studies were performed in mice via oral gavage at a single dose of 20 mg/kg, and
339 the key PK parameters were quantified and compared among compounds **16**, **30**, and **31** (Fig. 5F).
340 Compound **16** exhibited significant superiority in systemic exposure, as evidenced by its area
341 under the concentration-time curve from time 0 to the last measurable concentration (AUC_{0-t}) was
342 1862.128 h·ng/mL. This value was 19.6% and 52.8% higher than that of compound **30** (AUC_{0-t} =
343 1556.770 h·ng/mL) and **31** (AUC_{0-t} = 1219.430 h·ng/mL), respectively, indicating significantly
344 enhanced oral bioavailability. For the elimination half-life ($T_{1/2}$), compound **16** exhibited a
345 moderate value of 0.757 h, which was comparable to those of compound **30** ($T_{1/2}$ = 0.699 h) and
346 **31** ($T_{1/2}$ = 0.930 h). The time to reach peak plasma concentration (Tmax) of compound **16** was
347 0.667 h, slightly longer than that of compound **30** (Tmax = 0.417 h) and **31** (Tmax = 0.333 h).
348 However, it still ensured rapid oral absorption and subsequent onset of pharmacological effects.
349 The peak plasma concentration (Cmax) of compound **16** was 750.218 ng/mL, which was lower
350 than that of compound **30** (Cmax = 1095.919 ng/mL) and **31** (Cmax = 1245.280 ng/mL) but still
351 maintained a sufficiently high effective plasma concentration level to elicit the desired therapeutic
352 effects.
353 In general, compound **16** has the best inhibition activity against the enzyme and virus as well
354 as excellent liver metabolic stability. Based on these favorable properties, compound **16** was
355 selected as the lead compound for further profiling. The CYP and hERG inhibition at 10 μM and
356 30 μM for compound **16** were summarized in Figure 5F, the results indicate this compound has a
357 low risk of inhibiting selected CYP isoforms and hERG channel. All of these favorable
358 characteristics indicate that compound **16** is a promising lead compound suitable for further
359 optimization.

360 ***In vivo* antiviral efficacy of compound 16**

361 To assess the *in vivo* anti-viral activity of compound **16**, we treated the model mice infected
362 with SARS-CoV-2 EG.5 by oral administration (Fig. 6A). K18-hACE2 transgenic mice aged 8
363 weeks were used as our mouse model. Thirty-two female hACE2 transgenic mice were divided
364 into four groups with eight mice in each group to evaluate the efficacy of mock, vehicle, positive
365 control PF-07321332 of 500 milligrams per kilogram (mpk), and compound **16** of 25 mpk in the
366 therapeutic treatment. The weight loss plot shows that the weight loss is less than 5% for all groups
367 (Fig. 6B). The lung live viral titers for the group treated with PF-07321332 at 500 mpk and
368 compound **16** at 25 mpk significantly decreased at 3 days post-infection (Fig. 6C).

369 **Conclusion and Discussion**

370 The continuous mutation of SARS-CoV-2 enables the virus to escape neutralization by
371 antibody therapies[21]. Furthermore, resistance to small-molecule drugs is emerging due to
372 mutations in key viral targets such as RdRp, 3CLpro, and PLpro, posing a major challenge to
373 pandemic control. The mutation of $3CL^{pro}$ at E166V dramatically decreased the susceptibility to
374 Nirmatrelvir (IC_{50} increased 100-fold) [42]. Gandhi et al. reported a De novo emergence of a
375 remdesivir resistance mutation during treatment [23a]. Although no PL^{pro} inhibitors are clinically
376 available, a number of PL^{pro} inhibitors have been reported, such as Jun12682, PF-07957472 [19a,

[19c] and GZNL-P36 [19b]. In addition, drug-resistant mutations of PL^{pro} have also been documented, among which E167 and Y268 are two key drug-resistant mutation sites. In particular, E167G, E167A, E167S, E167K and Y268N confer strong drug resistance to Jun12682 and PF-07957472. Specifically, the Ki value of PF-07957472 increases by more than 130-fold against E167K and 969-fold against Y268N; the Ki value of Jun12682 increases by more than 130-fold against E167K and 385-fold against Y268N. Thus, the drug development targeting the E167 and Y268 mutations of PL^{pro} will help address the potential drug resistance that may arise in the future.

In the current study, we leveraged an AI generative model to discover a novel series of PL^{pro} inhibitors targeting the SARS-CoV-2 virus. Using the generative model, Tree-Invent, we successfully identified a novel scaffold of PL^{pro} inhibitors exhibiting comparable activity to a known reference compound **GZNL-P4**, though it has much improved liver metabolic stability. Further optimization work on this series results in a lead compound **16** with an improved enzyme potency and cellular-level antiviral activity. This compound demonstrates excellent liver stability, low CYP and hERG inhibition and it also demonstrates reasonable oral PK properties. At the same time, unlike the previously reported PL^{pro} inhibitors Jun12682 and PF-07957472 [24], compound **16** exhibits strong activity against E167K and can effectively overcome the drug resistance caused by the E167K mutation. We believe that this study offers valuable insights for the future development of PL^{pro} inhibitors as a potential treatment for COVID-19 disease, especially provides a scaffold compound and points the way forward for overcoming E167K-mediated resistance.

Methods and Materials

Detailed descriptions of the in vitro pharmacology studies, *in vivo* pharmacology studies, transcriptomics studies, X-ray crystallography, computational study, and synthetic methods can be found in Supplementary Information.

Data analysis

The data in the figures represent mean ± SD. All data were analyzed using GraphPad Prism 9.0 software. Statistical comparison between different groups was performed using the corresponding statistical analysis labeled in the figure legends combining several experiments. The binding mode of ligands with protein was illustrated by PyMOL2.6.0a0 (open source).

Data Availability

Crystal structures generated during the current study are available in the Protein Data Bank (PDB) under accession codes 8ZSE (PL^{pro} bound to compound **2**, here renamed as GZNL-2002), 21MS (PL^{pro} bound to compound **16**, here renamed as GZNL-2016), and 21MU (PL^{pro} bound to compound **30**, here renamed as GZNL-2030). The supplementary note for synthetic compounds can be found in Supplementary Information.

Acknowledgments

This work is supported by the fundings: Overseas Experts Supporting Programs under National Research Platform (WGZJ22-001); National Natural Science Foundation of China (82170473); The Major Program of Guangzhou National Laboratory (GZNL2024A01005,

GZNL2023A01008); Guangdong Natural Science Foundation (2021QN020451, 2021CX020227, 2024A1515011589); Basic and Applied Basic Research Projects of Guangzhou Basic Research Program (2023A04J0161). We thank the staff at Guangzhou Laboratory Core Facility and Animal Center. We also thank the staff at BL10U2/BL02U1/BL17B1/BL18U1/BL19U1 beamlines at Shanghai Synchrotron Radiation Facility (SSRF) of the National Facility for Protein Science in Shanghai (NFPS), Shanghai Advanced Research Institute, Chinese Academy of Sciences, for providing technical support in X-ray diffraction data collection and analysis.

Author Contributions

J.S, Y.L., Q.Y., W.L., M.D., J.T., J.C., J.Z., K.W., and X.W. performed the cellular and biochemical assays. Y.L. and J.S. performed crystallography. G.Z., P.Z., H.C., P.H., and T.R. performed the medicinal chemistry. T.R., M.X., M.T., and C.H. performed the computational studies. Y.L., W.L., M.D., J.C., J.Z., K.W., and X.W. performed cloning, protein purification and biophysical analysis. Y.L., Q.Y., T.R., G.Z., X.C., C.H., and J.S. conceived the experiments and wrote the manuscript with input from all authors.

Conflict of Interest

The authors declare no competing interests.

References

[1] M. Li, H. Wang, L. Tian, Z. Pang, Q. Yang, T. Huang, J. Fan, L. Song, Y. Tong, H. Fan, *Signal transduction and targeted therapy* **2022**, *7*, 146.

[2] W. T. Harvey, A. M. Carabelli, B. Jackson, R. K. Gupta, E. C. Thomson, E. M. Harrison, C. Ludden, R. Reeve, A. Rambaut, S. J. Peacock, D. L. Robertson, *Nature reviews. Microbiology* **2021**, *19*, 409-424.

[3] G. Li, R. Hilgenfeld, R. Whitley, E. De Clercq, *Nature reviews. Drug discovery* **2023**, *22*, 449-475.

[4] aA. Amstutz, B. Speich, F. Mentré, C. S. Rueegg, D. Belhadi, L. Assoumou, C. Burdet, S. Murthy, L. E. Dodd, Y. Wang, K. A. O. Tikkinen, F. Ader, M. Hites, M. Bouscambert, M. A. Trabaud, M. Fralick, T. C. Lee, R. Pinto, A. Barratt-Due, F. Lund-Johansen, F. Müller, O. P. O. Nevalainen, B. Cao, T. Bonnett, A. Griessbach, A. Taji Heravi, C. Schönenberger, P. Janiaud, L. Werlen, S. Aghlmandi, S. Schandelmaier, Y. Yazdanpanah, D. Costagliola, I. C. Olsen, M. Briel, *The Lancet. Respiratory medicine* **2023**, *11*, 453-464; bR. L. Gottlieb, C. E. Vaca, R. Paredes, J. Mera, B. J. Webb, G. Perez, G. Oguchi, P. Ryan, B. U. Nielsen, M. Brown, A. Hidalgo, Y. Sachdeva, S. Mittal, O. Osiyemi, J. Skarbinski, K. Juneja, R. H. Hyland, A. Osinusi, S. Chen, G. Camus, M. Abdelghany, S. Davies, N. Behenna-Renton, F. Duff, F. M. Marty, M. J. Katz, A. A. Ginde, S. M. Brown, J. T. Schiffer, J. A. Hill, *The New England journal of medicine* **2022**, *386*, 305-315.

[5] M. Yasir, C. R. Lankala, P. Kalyankar, A. Ishak, M. Mekhail, C. Sestacovschi, E. Kima, *Cureus* **2023**, *15*, e43060.

[6] F. Kabinger, C. Stiller, J. Schmitzová, C. Dienemann, G. Kokic, H. S. Hillen, C. Höbartner, P. Cramer, *Nature structural & molecular biology* **2021**, *28*, 740-746.
[7] M. D. Waters, S. Warren, C. Hughes, P. Lewis, F. Zhang, *Environmental and molecular mutagenesis* **2022**, *63*, 37-63.
[8] D. R. Owen, C. M. N. Allerton, A. S. Anderson, L. Aschenbrenner, M. Avery, S. Berritt, B. Boras, R. D. Cardin, A. Carlo, K. J. Coffman, A. Dantonio, L. Di, H. Eng, R. Ferre, K. S. Gajiwala, S. A. Gibson, S. E. Greasley, B. L. Hurst, E. P. Kadar, A. S. Kalgutkar, J. C. Lee, J. Lee, W. Liu, S. W. Mason, S. Noell, J. J. Novak, R. S. Obach, K. Ogilvie, N. C. Patel, M. Pettersson, D. K. Rai, M. R. Reese, M. F. Sammons, J. G. Sathish, R. S. P. Singh, C. M. Steppan, A. E. Stewart, J. B. Tuttle, L. Updyke, P. R. Verhoest, L. Wei, Q. Yang, Y. Zhu, *Science (New York, N.Y.)* **2021**, *374*, 1586-1593.
[9] Y. Unoh, S. Uehara, K. Nakahara, H. Nobori, Y. Yamatsu, S. Yamamoto, Y. Maruyama, Y. Taoda, K. Kasamatsu, T. Suto, K. Kouki, A. Nakahashi, S. Kawashima, T. Sanaki, S. Toba, K. Uemura, T. Mizutare, S. Ando, M. Sasaki, Y. Orba, H. Sawa, A. Sato, T. Sato, T. Kato, Y. Tachibana, *Journal of medicinal chemistry* **2022**, *65*, 6499-6512.
[10] R. Shimizu, T. Matsuzaki, R. Oka, T. Sonoyama, T. Fukuhara, A. Kuwata, Y. Matsuo, R. Kubota, *Journal of clinical pharmacology* **2023**, *63*, 918-927.
[11] B. H. Harcourt, D. Jukneliene, A. Kanjanahaluethai, J. Bechill, K. M. Severson, C. M. Smith, P. A. Rota, S. C. Baker, *Journal of virology* **2004**, *78*, 13600-13612.
[12] D. Shin, R. Mukherjee, D. Grewe, D. Bojkova, K. Baek, A. Bhattacharya, L. Schulz, M. Widera, A. R. Mehdipour, G. Tascher, P. P. Geurink, A. Wilhelm, G. J. van der Heden van Noort, H. Ovaa, S. Müller, K. P. Knobeloch, K. Rajalingam, B. A. Schulman, J. Cinatl, G. Hummer, S. Ciesek, I. Dikic, *Nature* **2020**, *587*, 657-662.
[13] S. Kandwal, D. Fayne, *Virology* **2023**, *581*, 97-115.
[14] J. Osipiuk, S. A. Azizi, S. Dvorkin, M. Endres, R. Jedrzejczak, K. A. Jones, S. Kang, R. S. Kathayat, Y. Kim, V. G. Lisnyak, S. L. Maki, V. Nicolaescu, C. A. Taylor, C. Tesar, Y. A. Zhang, Z. Zhou, G. Randall, K. Michalska, S. A. Snyder, B. C. Dickinson, A. Joachimiak, *Nature communications* **2021**, *12*, 743.
[15] C. Ma, M. D. Sacco, Z. Xia, G. Lambrinidis, J. A. Townsend, Y. Hu, X. Meng, T. Szeto, M. Ba, X. Zhang, M. Gongora, F. Zhang, M. T. Marty, Y. Xiang, A. Kolocouris, Y. Chen, J. Wang, *ACS central science* **2021**, *7*, 1245-1260.
[16] Z. Shen, K. Ratia, L. Cooper, D. Kong, H. Lee, Y. Kwon, Y. Li, S. Alqarni, F. Huang, O. Dubrovskyi, L. Rong, G. R. J. Thatcher, R. Xiong, *Journal of medicinal chemistry* **2022**, *65*, 2940-2955.
[17] P. Jadhav, B. Huang, J. Osipiuk, X. Zhang, H. Tan, C. Tesar, M. Endres, R. Jedrzejczak, B. Tan, X. Deng, A. Joachimiak, J. Cai, J. Wang, *European journal of medicinal chemistry* **2024**, *264*, 116011.
[18] B. C. Sanders, S. Pokhrel, A. D. Labbe, Mathews, II, C. J. Cooper, R. B. Davidson, G. Phillips, K. L. Weiss, Q. Zhang, H. O'Neill, M. Kaur, J. G. Schmidt, W. Reichard, S. Surendranathan, J. Parvathareddy, L. Phillips, C. Rainville, D. E. Sterner, D. Kumaran, B. Andi, G. Babnigg, N. W. Moriarty, P. D. Adams, A. Joachimiak, B. L. Hurst, S. Kumar, T. R. Butt, C. B. Jonsson, L. Ferrins, S. Wakatsuki, S. Galanie, M. S. Head, J. M. Parks, *Nature communications* **2023**, *14*, 1733.
[19] aR. M. Garnsey MR, Nguyen LT, Cardin R, Tillotson J, Mashalidis E, Yu A, Aschenbrenner L, Balesano A, Behzadi A, Boras B, Chang JS, Eng H, Ephron A, Foley T, Ford KK, Frick JM, Gibson S, Hao L, Hurst B, Kalgutkar AS, Korczynska M,

Lengyel-Zhand Z, Gao L, Meredith HR, Patel NC, Polivkova J, Rai D, Rose CR, Rothan H, Sakata SK, Vargo TR, Qi W, Wu H, Liu Y, Yurgelonis I, Zhang J, Zhu Y, Zhang L, Lee AA. , *Sci. Adv.* **2024**, *10*, eado4288; bY. Lu, Q. Yang, T. Ran, G. Zhang, W. Li, P. Zhou, J. Tang, M. Dai, J. Zhong, H. Chen, P. He, A. Zhou, B. Xue, J. Chen, J. Zhang, S. Yang, K. Wu, X. Wu, M. Tang, W. K. Zhang, D. Guo, X. Chen, H. Chen, J. Shang, *Nature communications* **2024**, *15*, 10169; cB. Tan, X. Zhang, A. Ansari, P. Jadhav, H. Tan, K. Li, A. Chopra, A. Ford, X. Chi, F. X. Ruiz, E. Arnold, X. Deng, J. Wang, *Science (New York, N.Y.)* **2024**, *383*, 1434-1440.

[20] W. He, L. Liu, B. Li, L. Xu, H. Li, J. Zheng, S. Wan, C. Wu, T. Wang, M. Wang, China, **2022**.

[21] M. Vukovikj, A. Melidou, P. Nannapaneni, T. Normark, A. Kraus, E. K. Broberg, *Euro Surveill* **2025**, *30*.

[22] E. Heilmann, F. Costacurta, S. A. Moghadasi, C. Ye, M. Pavan, D. Bassani, A. Volland, C. Ascher, A. K. H. Weiss, D. Bante, R. S. Harris, S. Moro, B. Rupp, L. Martinez-Sobrido, D. von Laer, *Sci Transl Med* **2023**, *15*, eabq7360.

[23] aS. Gandhi, J. Klein, A. J. Robertson, M. A. Peña-Hernández, M. J. Lin, P. Roychoudhury, P. Lu, J. Fournier, D. Ferguson, S. A. K. Mohamed Bakhash, M. Catherine Muenker, A. Srivathsan, E. A. Wunder, Jr., N. Kerantzas, W. Wang, B. Lindenbach, A. Pyle, C. B. Wilen, O. Ogbuagu, A. L. Greninger, A. Iwasaki, W. L. Schulz, A. I. Ko, *Nature communications* **2022**, *13*, 1547; bL. J. Stevens, A. J. Pruijssers, H. W. Lee, C. J. Gordon, E. P. Tchesnokov, J. Gribble, A. S. George, T. M. Hughes, X. Lu, J. Li, J. K. Perry, D. P. Porter, T. Cihlar, T. P. Sheahan, R. S. Baric, M. Gotte, M. R. Denison, *Sci Transl Med* **2022**, *14*, eabo0718.

[24] H. Tan, Q. Zhang, K. Georgiou, S. Zhang, K. Li, G. Lambrinidis, A. Kolocouris, X. Deng, J. Wang, *Nature communications* **2025**, *16*, 4548.

[25] V. Chenthamarakshan, S. C. Hoffman, C. D. Owen, P. Lukacik, C. Strain-Damerell, D. Fearon, T. R. Malla, A. Tumber, C. J. Schofield, H. M. E. Duyvesteyn, W. Dejnirattisai, L. Carrique, T. S. Walter, G. R. Screaton, T. Matviiuk, A. Mojsilovic, J. Crain, M. A. Walsh, D. I. Stuart, P. Das, *Science advances* **2023**, *9*, eadg7865.

[26] R. Gómez-Bombarelli, J. N. Wei, D. Duvenaud, J. M. Hernández-Lobato, B. Sánchez-Lengeling, D. Sheberla, J. Aguilera-Iparraguirre, T. D. Hirzel, R. P. Adams, A. Aspuru-Guzik, *ACS central science* **2018**, *4*, 268-276.

[27] Y. Yang, S. Zheng, S. Su, C. Zhao, J. Xu, H. Chen, *Chemical science* **2020**, *11*, 8312-8322.

[28] S. Kang, K. Cho, *Journal of chemical information and modeling* **2019**, *59*, 43-52.

[29] S. Amabilino, P. Pogány, S. D. Pickett, D. V. S. Green, *Journal of chemical information and modeling* **2020**, *60*, 5699-5713.

[30] M. Olivecrona, T. Blaschke, O. Engkvist, H. Chen, *Journal of cheminformatics* **2017**, *9*, 48.

[31] O. Zhang, T. Wang, G. Weng, D. Jiang, N. Wang, X. Wang, H. Zhao, J. Wu, E. Wang, G. Chen, Y. Deng, P. Pan, Y. Kang, C. Y. Hsieh, T. Hou, *Nature computational science* **2023**, *3*, 849-859.

[32] M. Xu, H. Chen, *Journal of chemical information and modeling* **2023**, *63*, 7067-7082.

[33] Y. Matsuo, Y. LeCun, M. Sahani, D. Precup, D. Silver, M. Sugiyama, E. Uchibe, J. Morimoto, *Neural networks : the official journal of the International Neural Network Society* **2022**, *152*, 267-275.

[34] H. Cho, Y. J. Kim, J. W. Chae, M. R. Meyer, S. K. Kim, C. S. Ryu, *Frontiers in pharmacology* **2023**, *14*, 1067408.
[35] R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis, P. S. Shenkin, *Journal of medicinal chemistry* **2004**, *47*, 1739-1749.
[36] C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, *Advanced drug delivery reviews* **2001**, *46*, 3-26.
[37] M. Ringnér, *Nature biotechnology* **2008**, *26*, 303-304.
[38] H. Zhou, J. Skolnick, *The journal of physical chemistry. B* **2024**, *128*, 5363-5370.
[39] G. C. Linderman, S. Steinerberger, *SIAM journal on mathematics of data science* **2019**, *1*, 313-332.
[40] K. Ratia, S. Pegan, J. Takayama, K. Sleeman, M. Coughlin, S. Baliji, R. Chaudhuri, W. Fu, B. S. Prabhakar, M. E. Johnson, S. C. Baker, A. K. Ghosh, A. D. Mesecar, *Proceedings of the National Academy of Sciences of the United States of America* **2008**, *105*, 16119-16124.
[41] Y. Cheng, W. H. Prusoff, *Biochem Pharmacol* **1973**, *22*, 3099-3108.
[42] S. Iketani, H. Mohri, B. Culbertson, S. J. Hong, Y. Duan, M. I. Luck, M. K. Annavajhala, Y. Guo, Z. Sheng, A. C. Uhlemann, S. P. Goff, Y. Sabo, H. Yang, A. Chavez, D. D. Ho, *Nature* **2023**, *613*, 558-564.

Figures and Tables

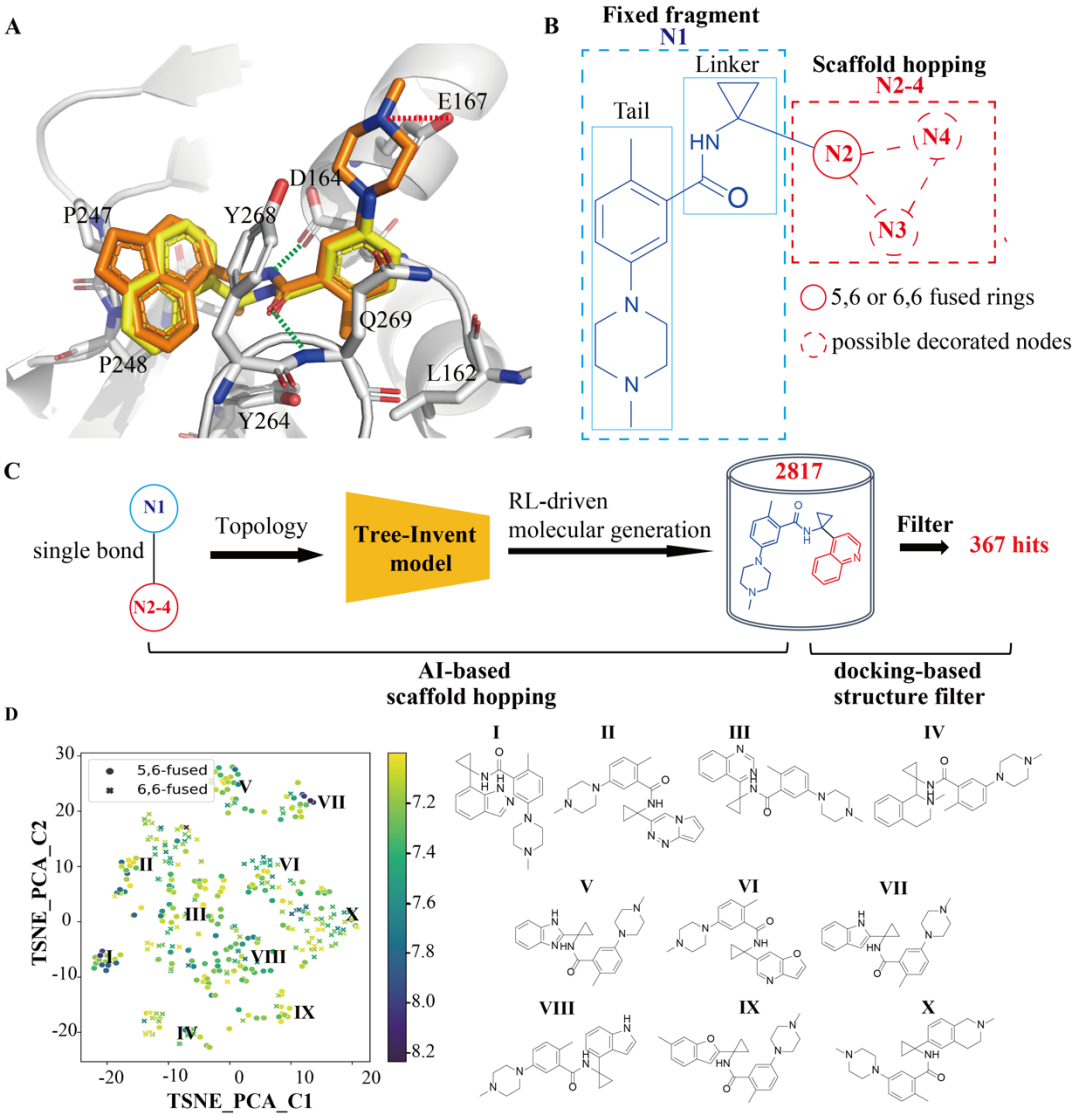


Figure 1. Scaffold hopping of the GZNL-P4 structure. A) Binding mode of **GZNL-P4**. The crystal structures of **GZNL-P4** (orange) and **GRL-0617** (yellow) are overlapped. Ligands and residues surrounding the binding pockets are shown as sticks. Protein secondary structure is shown as transparent cartoons. Green and red dashes represent hydrogen bond and salt bridge interactions. B) Transformation of **GZNL-P4** to a topological tree. The solid and dashed circles represent essential and inessential nodes for molecular generation, respectively. The dashed lines represent the bonds to the inessential nodes. C) Workflow of scaffold hopping of the **GZNL-P4** structure based on the Tree-Invent model using the RL strategy. The blue and red circles represent prefixed and changeable nodes. D) Chemical space analysis based on the PCA analysis [37] using the Morgan fingerprint [38]. tSNE plot [39] was used to display the chemical space of molecules filtered based on docking scores (represented by the color of the heatmap) and their pharmacophore model. The labels I-X indicate ten diverse structures with reasonable docking scores.

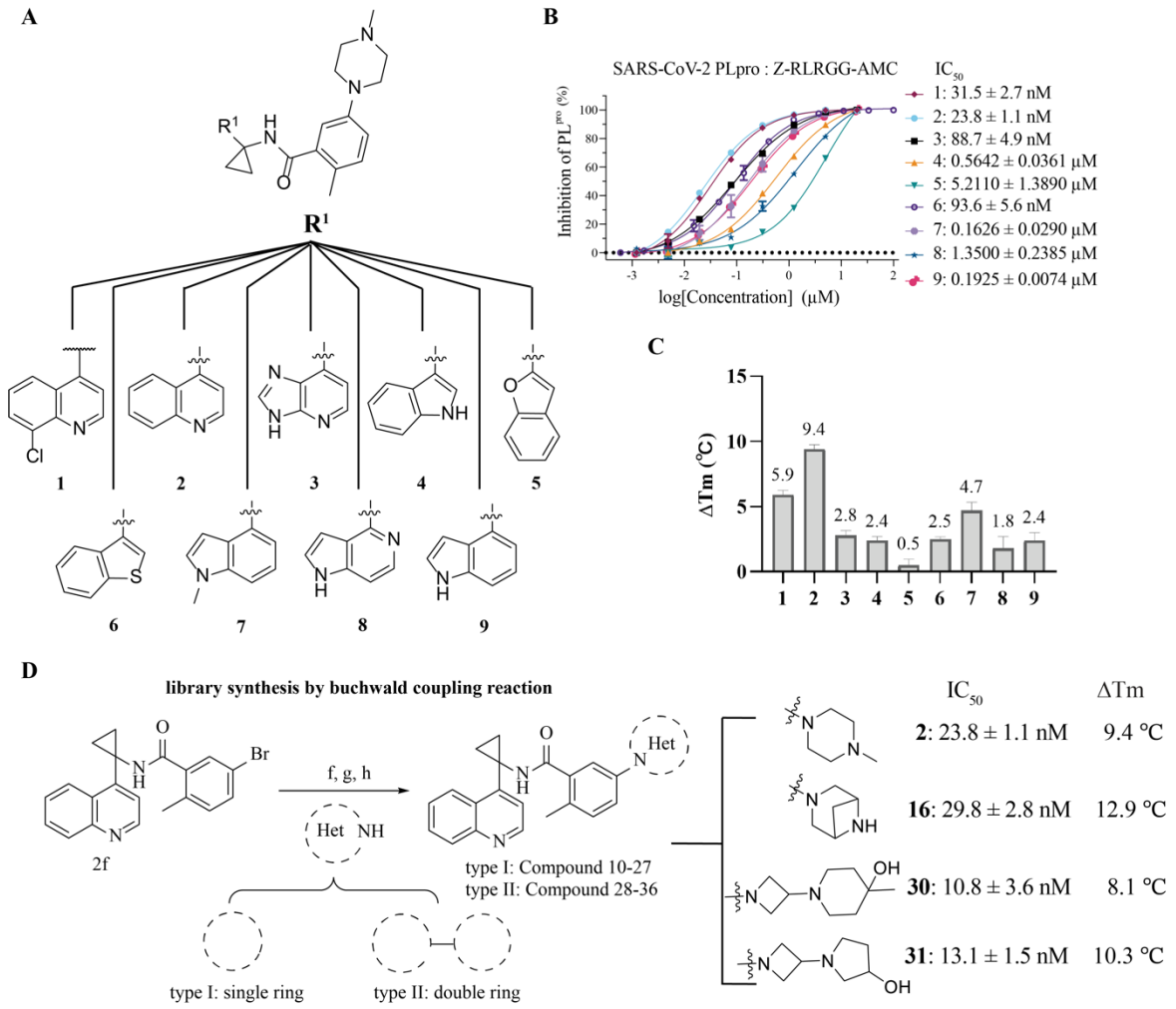


Figure 2. Lead compound identification. A) Structures of nine synthesized compounds **1-9**. R^1 groups are shown at the top. B) Dose-dependent inhibitory activity of compounds **1-9** against PL^{pro} as determined by the fluorescence-based biochemical assay. Half maximum inhibitory concentration (IC_{50}) is shown in the legend. Data are presented as mean ± s.d. of three technical replicates. C) Differential scanning fluorimetry assay analysis of SARS-CoV-2 PL^{pro} with compounds. The data are presented as the mean of three replicates. D) Scheme of library synthesis for compound **10-36**. Synthesis routes were shown in the supplemental text. The inhibition activity (IC_{50}) and the delta melt temperature (Δ Tm) of lead compounds **2**, **16**, **30**, and **31** were shown. The inhibition activity of IC_{50} is presented as mean ± s.d. of three technical replicates, but the Δ Tm is presented as the mean of three replicates.

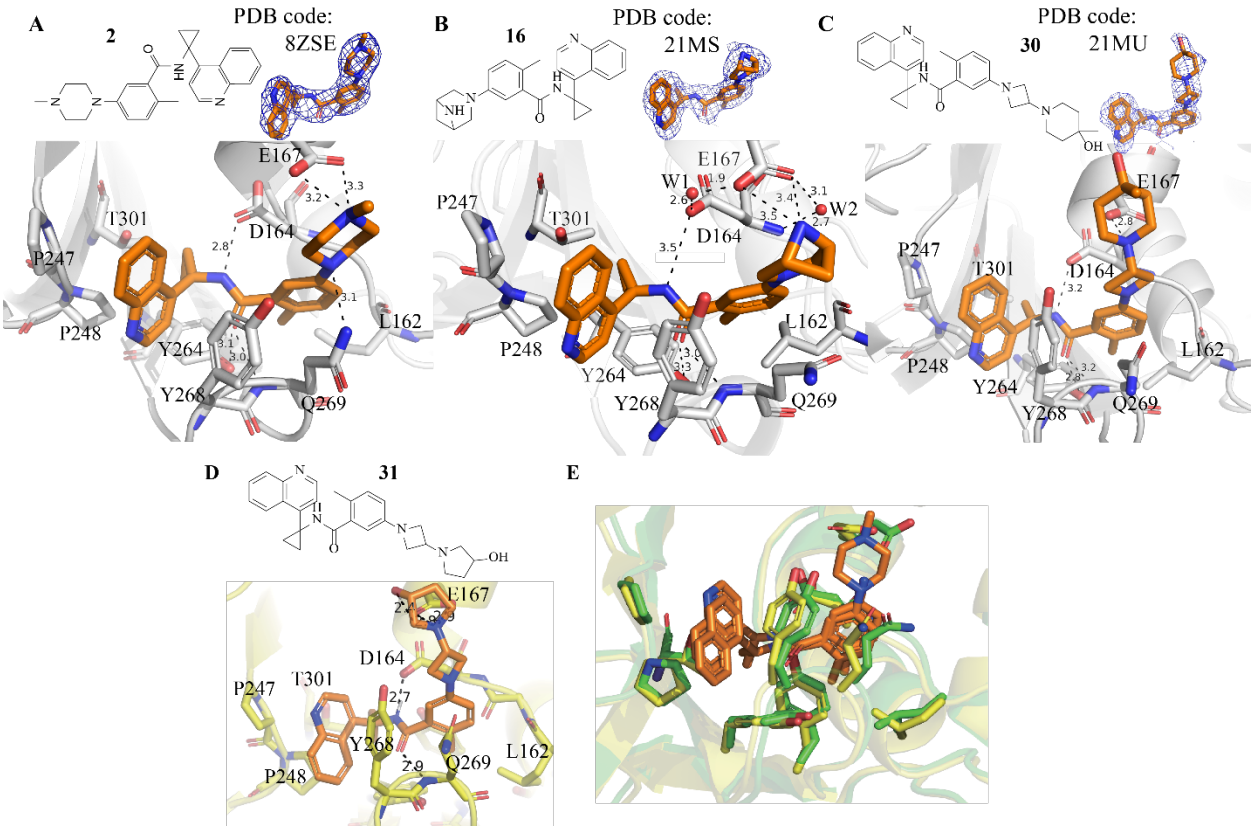
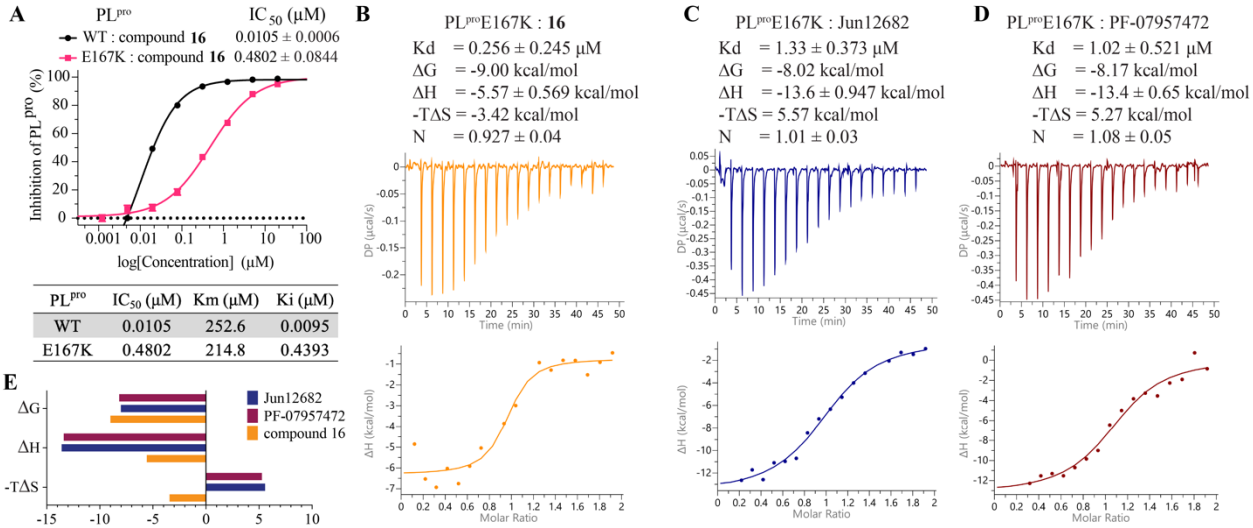
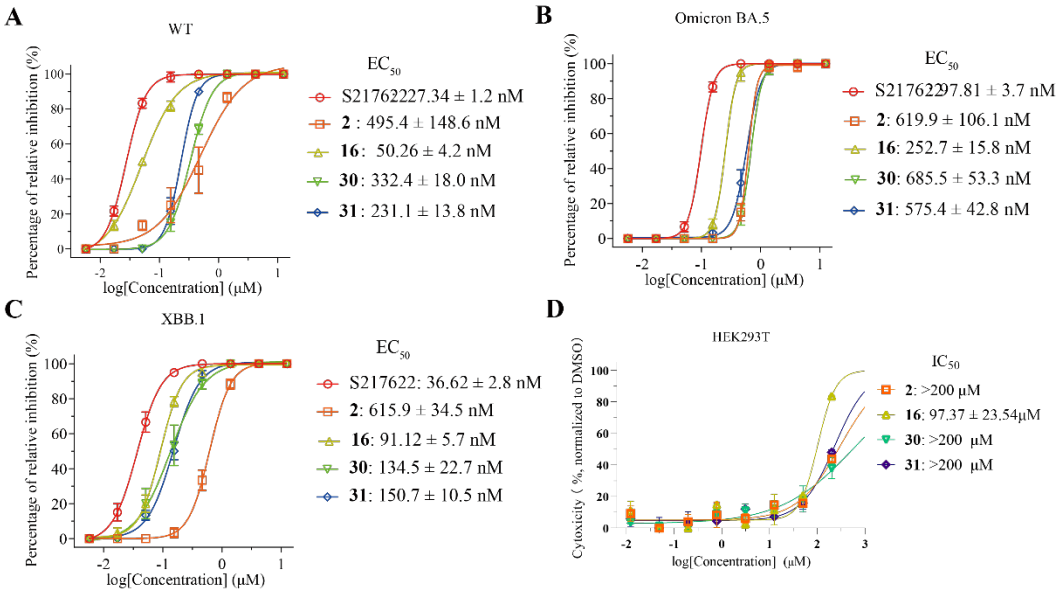


Figure 3. Binding mode of active compounds. A-C) The binding mode of compounds **2** (A)**, 16** (B)**, 30** (C) were determined by X-ray diffraction. Compounds **2, 16, 30** were shown as orange sticks. Residues around the binding pocket are shown as gray sticks. Protein secondary structure is shown as gray cartoons. Black dashes represent hydrogen bonds and salt bridges. The contour level of ligand map was set as 1.0. D) Docking poses of compound **31**. Ligands and residues are shown as orange and yellow sticks, respectively. Black dash represents either hydrogen bond or salt bridge interactions labeled with the distance. E) Alignment of the crystal structures with compound **2** (yellow) and **GRL0617** (green, PDB code: 7JRN). Protein secondary structure is shown as cartoons and residues around the binding pocket are shown as sticks. Ligands are specifically shown as orange sticks.



PL^{pro}	IC$_{50}$ (μM)	Km (μM)	Ki (μM)
WT	0.0105	252.6	0.0095
E167K	0.4802	214.8	0.4393

Figure 4. The binding and inhibition activity of compound 16 to PL^{pro}E167K mutant. A) The inhibition activity of compound 16 for the PL^{pro}E167K mutant conferring drug resistance. The error bars are mean ± s.d based on three technical replicates. B-E) The binding affinity obtained by ITC of compound 16, Jun12682, and PF-07957472 to PL^{pro}E167K by ITC.



E

parameter	2	16	30	31
Rat Livermicrosome stability $t_{1/2}$ (min)	45.20	120.12	59.70	41.98
Rat Clint (mL/min/kg)	55.20	20.77	41.80	59.40
Human Livermicrosome stability $t_{1/2}$ (min)	30.29	39.55	47.90	35.30
Human Clint (mL/min/kg)	52.90	40.53	33.50	45.50

F

parameter	16	30	31
Mouse $T_{1/2}$ (h) PO@20mg/kg	0.757	0.699	0.930
Mouse Tmax (h) PO@20mg/kg	0.667	0.417	0.333
Mouse Cmax (ng/mL) PO@20mg/kg	750.218	1095.919	1245.280
Mouse AUC(0-t) (h * ng/mL) PO@20mg/kg	1862.128	1556.770	1219.430
inh%CYP 2C9@10μM	4.33	NT	NT
inh% CYP 2D6@10μM	37.15	NT	NT
inh%CYP 3A4@10μM (Midazolam)	56.88	NT	NT
inh%CYP 3A4@10μM (Testosteron)	37.13	NT	NT
hERG inh%@30μM	38.94	NT	NT

notes:
NT: No testing; PO: peros

Figure 5. Cellular activity and PK profiling of the compounds 2, 16, 30, and 31. A-C) Antiviral activity of compounds **2**, **16**, **30**, **31** against SARS-CoV-2 wild type (WT) (A), variants Omicron BA.5 (B), and XBB.1 (C). The EC_{50} was measured on virus infected Vero E6 cells. The error bars are mean ± s.d based on three technical replicates. D) Toxicity to normal HEK293T cell lines. Data are presented as the mean of two technical replicates. E) Table for liver stability in rat and human liver microsomes. F) Table for in vivo PK characteristics, CYP inhibition and hERG toxicity.

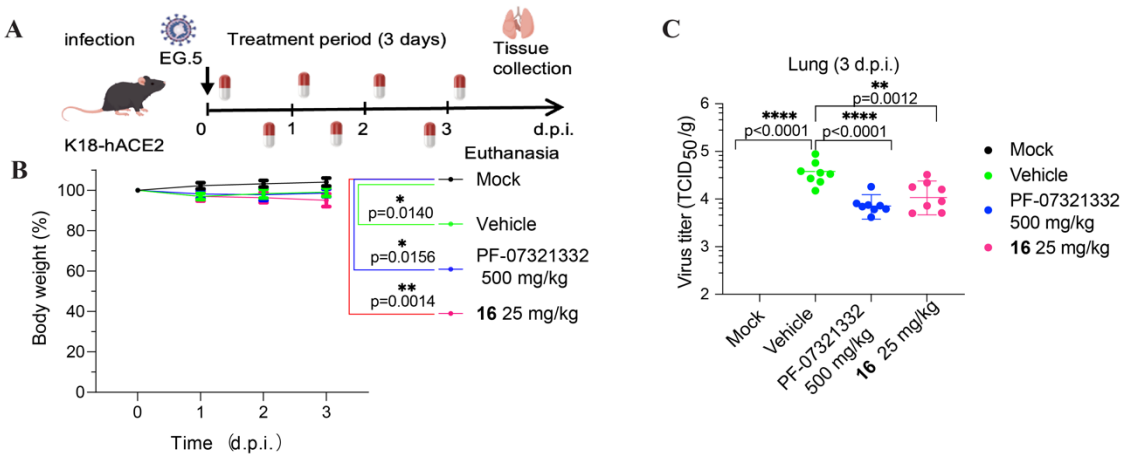
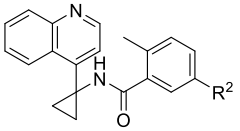
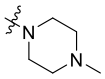
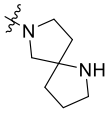
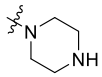
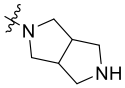
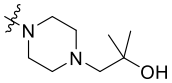
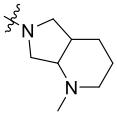
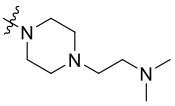
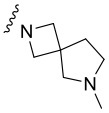
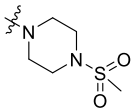
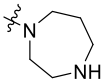
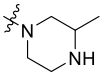
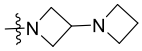
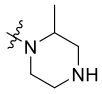
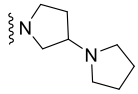
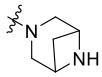
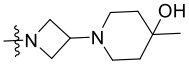
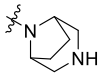
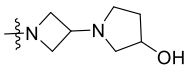
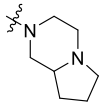
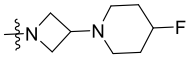
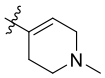
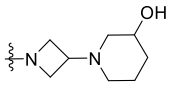
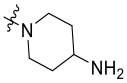
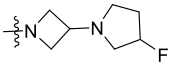
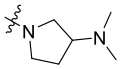
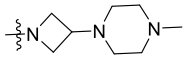
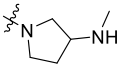
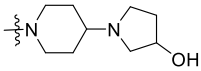


Figure 6. *In vivo* antiviral activity of PL^{pro} inhibitor compound 16. A) Experimental design for the 3-day experiment in K18-ACE2 mouse. B) Body weight loss of mice from different groups (n = 8 per group). C) Live viral titers in lungs collected at 3 d.p.i. (n = 8 per group). The error bars are mean ± s.d. Statistical differences were determined by Ordinary one-way ANOVA with Dunnett multiple comparisons test for the body weight on day 3 in (B) and the live viral titer in (C). *P< 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001; ns, not significant.

Table 1. SAR exploration of R^2 group



Compound	R^2	PL^{pro} IC_{50} (μM)	Compound	R^2	PL^{pro} IC_{50} (μM)
2		0.0238±0.0011	**23**		0.0992±0.0217
10		0.0327±0.0016	**24**		0.0563±0.0059
11		0.0288±0.0018	**25**		0.0156±0.0018
12		2.2330±0.2485	**26**		0.2861±0.0224
13		1.0300±0.1295	**27**		0.3016±0.0287
14		0.0453±0.0069	**28**		0.0376±0.0055
15		11.0700±14.6380	**29**		0.0256±0.0050
16		0.0105±0.0006	**30**		0.0108±0.0036
17		0.0316±0.0065	**31**		0.0131±0.0015
18		0.0520±0.0088	**32**		0.0103±0.0028
19		0.0284±0.0022	**33**		0.0494±0.0019
20		0.3836±0.0510	**34**		0.0519±0.0107
21		0.0483±0.0107	**35**		0.1817±0.0131
22		0.0560±0.0090	**36**		1.0540±0.1536

Data are presented as mean ± s.d. of three technical replicates.