ACS BIO & MED CHEM Au
AN OPEN ACCESS JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/biomedchemau

Article

Efficient Strategy to Design Protease Inhibitors: Application to Enterovirus 71 2A Protease

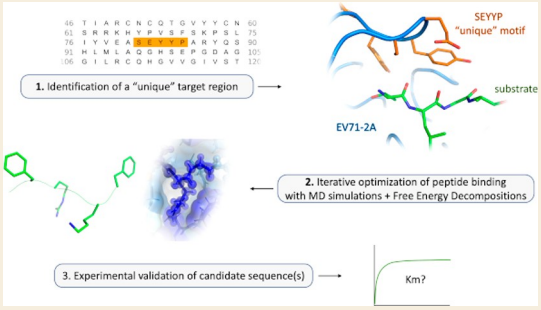
Ting Chen, Cédric Grauffel, Wei-Zen Yang, Yi-Ping Chen, Hanna S. Yuan,* and Carmay Lim*

Cite This: *ACS Bio Med Chem Au* 2022, 2, 437−449

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: One strategy to counter viruses that persistently cause outbreaks is to design molecules that can specifically inhibit an essential multifunctional viral protease. Herein, we present such a strategy using well-established methods to first identify a region present only in viral (but *not* human) proteases and find peptides that can bind specifically to this "unique" region by maximizing the protease−peptide binding free energy iteratively using single-point mutations starting with the substrate peptide. We applied this strategy to discover pseudosubstrate peptide inhibitors for the multifunctional 2A protease of enterovirus 71 (EV71), a key causative pathogen for hand-foot-and-mouth disease affecting young children, along with coxsackievirus A16. Four peptide candidates predicted to bind EV71 2A protease more tightly than the natural substrate were experimentally validated and found to inhibit protease activity. Furthermore, the crystal structure of the best pseudosubstrate peptide bound to the EV71 2A protease was determined to provide a molecular basis for the observed inhibition. Since the 2A proteases of EV71 and coxsackievirus A16 share nearly identical sequences and structures, our pseudosubstrate peptide inhibitor may prove useful in inhibiting the two key pathogens of hand-foot-and-mouth disease.



KEYWORDS: *Peptide-based drug design, pseudosubstrate, MM/PBSA, binding free energy calculations, free energy decomposition, EV71 2A protease*

INTRODUCTION

The current COVID-19 pandemic caused by SARS-CoV-2 underscores the need for effective strategies to counter viruses that persistently cause outbreaks. One such example that still poses a global health threat is enterovirus 71 (EV71), a nonenveloped, positive-sense, single-stranded RNA virus belonging to the *Picornaviridae* family, which includes 10 enterovirus species (e.g., coxsackieviruses A and B, echoviruses, enteroviruses, polioviruses) and 3 rhinovirus species.[1] EV71 is one of the most pathogenic enteroviruses, causing epidemics in the past two decades,[2] in particular sporadic outbreaks of hand-foot-and-mouth disease (HFMD).[1] EV71 infections may lead to severe neurological/cardiopulmonary complications such as aseptic meningitis, acute flaccid paralysis, encephalitis, myocarditis, pulmonary edema, and hemorrhage, resulting in death.[1,3,4] Although formalin-inactivated EV71 vaccines are available, they cannot prevent/treat HFMD, as coxsackievirus-A16 (CVA16) also causes HFMD.[5−8] No anti-EV71 drug is available;[9] hence, treatment for EV71-infected patients is mainly palliative.[2] Consequently, there is a great need to develop drugs to treat EV71 infections, which can be life-threatening in children.

The enterovirus genome encodes four structural capsid proteins (VP1−VP4) and seven nonstructural proteins ($2A^{pro}$, 2B, 2C, 3A, 3B/VPg, $3C^{pro}$, and $3D^{pol}$).[10] The life cycle of all enteroviruses including EV71 involves the following steps (Figure 1):

(i) *entry* of EV71 into host cells by binding to cell surface receptors, resulting in receptor-mediated endocytosis,
(ii) *release* of the viral RNA genome via a pore in the endosomal membrane into the cytoplasm,
(iii) *translation* of the viral RNA by the host protein synthesis machinery into a single large polyprotein,
(iv) *cleavage* of the polyprotein by virus-encoded proteases to produce capsid proteins (VP0, VP1 and VP3), which recognize cell surface receptors and encapsulate the viral genome, and replication proteins ($2A^{pro}$−2C and 3A−$3D^{pol}$) that are crucial for viral RNA replication and translation,
(v) *replication* of the viral genome by the viral RNA-dependent RNA polymerase enzyme $3D^{pol}$,

Received: January 5, 2022
Revised: April 7, 2022
Accepted: April 7, 2022
Published: April 19, 2022

https://doi.org/10.1021/acsbiomedchemau.2c00001
ACS Bio Med Chem Au 2022, 2, 437−449

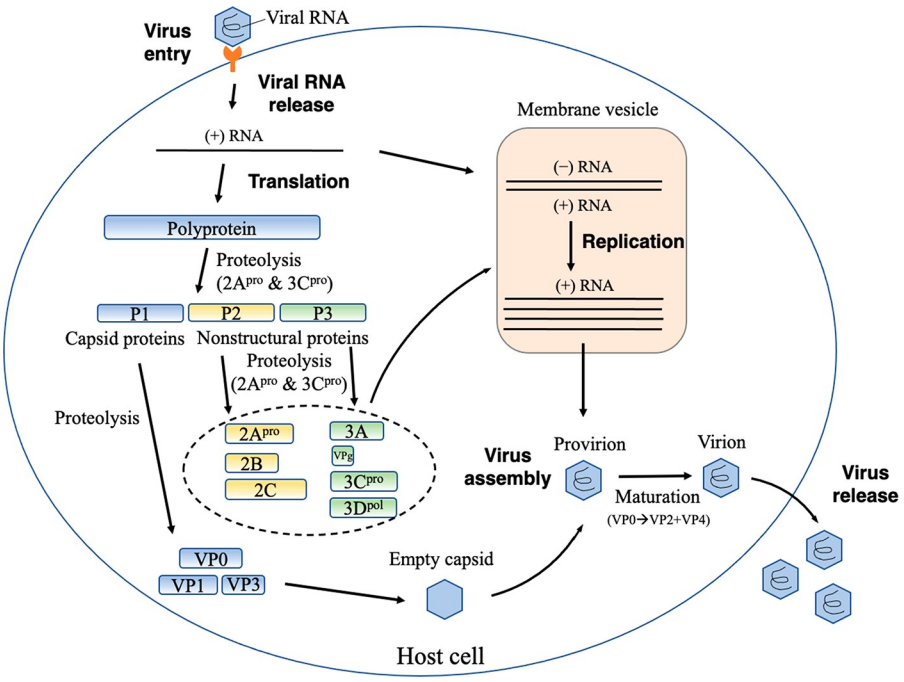


Figure 1. Virus genome structure and life cycle. Viral capsid protein (blue hexagon) binds to a cell surface receptor (orange) and enters the host cell via endocytosis. After uncoating, the viral genome RNA is released into the cytoplasm where it serves as a template for (i) translation into a single large polyprotein (blue oblong, left) and (ii) viral replication in membranous vesicles (right). Virus-encoded proteases ($2A^{pro}$ and $3C^{pro}$) cleave the polyprotein to constituent capsid proteins (VP0, VP1, VP3) and replication proteins ($2A^{pro}$−2C and 3A−$3D^{pol}$). Viral RNA-dependent RNA polymerase enzyme ($3D^{pol}$) catalyzes first the synthesis of a (−) RNA, which serves as a template for synthesis of a (+) RNA, which in turn serves as a template for translation or (−) RNA synthesis or are encapsidated by viral capsid proteins to form provirions. Cleavage of VP0 into VP2 and VP4 yields infectious, mature virions that are subsequently released from the host cell.

(vi) assembly of the nascent viral RNA and the capsid proteins into provirions that are converted into infectious, *mature* virions upon cleavage of VP0 into VP2 and VP4,

(vii) *release* of *mature* virions from host cell.

Each of the above critical steps in the EV71 life cycle can be targeted by anti-EV71 agents.[2] Indeed, several EV71inhibitors have been identified and tested in clinical trials, but they have all failed due to lack of efficacy and/or toxicity issues.[11]

Herein, we chose to target the EV71-$2A^{pro}$ because it plays multiple essential roles in (i) EV71 polyprotein processing and thus virus replication, (ii) inhibition of cap-dependent host mRNA translation to facilitate cap-*independent* viral RNA translation, and (iii) evasion of innate immunity, the first line of host defense against invading pathogens. The EV71-$2A^{pro}$ is a Cys protease that performs the first cleavage of the polyprotein at the junction between its N-terminus and the VP1 C-terminus, hence it is a precursor of replicative proteins.[8] In addition to cleaving the viral polyprotein, EV71-$2A^{pro}$ also cleaves several *host* proteins that are virus sensors or innate immune regulators in order to optimize virus spread and suppress antiviral cellular responses: It cleaves host elongation factors, eIF4GI/II, and affect antiviral protein translations without affecting viral RNA translation.[12,13] EV71-$2A^{pro}$ can also degrade antiviral signaling molecules and interferon-α/β receptor 1, thus antagonizing the antiviral actions of type I interferon.[14,15] Furthermore, high-resolution crystal structures of EV71-$2A^{pro}$ free and in complex with substrate are available. Despite its importance for EV71 infection, EV71-$2A^{pro}$ remains a relatively unexplored target with only two known weak inhibitors:[16] A human rhinovirus-$2A^{pro}$ peptide inhibitor (LVLQTM)[17] and a synthetic derivative of furoquinoline alkaloid (called CW-33)[18] reduced EV71 replication at a high concentration of 200 μM and an $EC_{50} \sim 200$ μM, respectively. Hence, more potent inhibitors specific to EV71-$2A^{pro}$ are warranted.

Our aim is to present an efficient in silico strategy for (i) finding a "unique" drug target region in an essential viral protease and (ii) designing peptides that can bind more tightly than the native substrate to the viral enzyme, which therefore cannot efficiently cleave the native substrate. We provide proof-of-concept for our strategy using EV71-$2A^{pro}$. Note that several strategies[19,20] are already available for modifying peptides so that they are cell permeable or stable in plasma and hence we do not focus on these aspects in this work. Below, we describe how an appropriate virus region for drug design was identified by comparing the sequences and structures of EV71-$2A^{pro}$ and other virus or human proteins. Next, we present an efficient and comprehensive protocol using molecular dynamics (MD) simulations followed by free energy decomposition for identifying peptides that can bind more tightly than the native substrate to EV71-$2A^{pro}$. We experimentally confirm that the predicted peptides could indeed compete with the natural substrate in binding to EV71-$2A^{pro}$ and inhibit protease activity. We further solved the crystal structure of EV71-$2A^{pro}$ bound to the best pseudosubstrate peptide to provide further insight into the mode of action/interactions. In summary, our strategy for choosing an appropriate drug target region and designing peptides that can compete with the substrate has led to the identification of a specific EV71-$2A^{pro}$ inhibitor. By design, the pseudosubstrate peptide found may also simultaneously inhibit other 2A proteases whose structures are nearly identical to that of EV71-

https://doi.org/10.1021/acsbiomedchemau.2c00001
ACS Bio Med Chem Au 2022, 2, 437−449

2A^{pro} such as CVA16-2A^{pro}, the other major causative agent for HFMD.

METHODS

Sequence and Structure Analyses

EV71-2A^{pro} sequence analysis was performed using BLASTp (https://blast.ncbi.nlm.nih.gov/Blast.cgi), and its conserved domain was identified by CDD.[21] Multiple sequence alignment was performed using CLUSTALW.[22] Structures in the Protein Data Bank (PDB)[23] that are similar to the EV71-2A^{pro} structure (PDB 4fvd) were identified using TM-align in the COFACTOR server.[24,25] Pairwise structural alignment was carried out using flexible structure alignment by chaining aligned fragment pairs allowing twists (FATCAT).[26]

Molecular Dynamics Simulations

System Setup. The starting point for the MD simulations is the 1.66 Å crystal structure of the EV71-2A^{pro} C110A mutant bound to a partial substrate, $T_3L_2G_1K_{1'}$ (PDB 4fvd). Residues at P4 and P2′ were added using MODELLER 9.17.[27] Mutations of the peptide residues were performed using SCWRL4.[28] Two protonation schemes were employed for the catalytic dyad: neutral C110/H21 and charged C110$^-$/H21$^+$. The protonation states of the other ionizable side chains at pH 7 were determined using PROPKA3:[29] all the other His were found to be neutral, Arg/Lys residues were positively charged, whereas Asp/Glu residues were negatively charged. Missing hydrogen atoms were added using the HBUILD module in the CHARMM program[30] with the CHARMM36 force field.[31] The all-hydrogen protein−peptide complex was inserted at the center of a cubic TIP3P[32] water box of edge length 60 Å. Water molecules falling within 2.8 Å of any protein heavy atom were removed. To neutralize the system, no more than three water molecules located >4.5 Å from the complex were randomly replaced by Na^+ or Cl^- counterions.

Simulation Protocol. MD simulations were performed at physiological pH at a temperature of 298.15 K and 1 atm pressure using NAMD2.12.[33] All bonds to hydrogen atoms were constrained by the SHAKE[34] algorithm. Long-range electrostatic forces were treated using the particle mesh Ewald method[35] with a grid spacing of 1 Å and a nonbond cutoff of 12 Å. The nonbonded interactions were updated every 1 fs. First, the solvated protein−peptide complex was minimized using 10 000 steps of conjugated gradient. Next, it was subjected to three 100 ps rounds of equilibration during which the restraints on the backbone and side chain atoms were progressively removed. The first equilibration employed a time step of 1 fs, whereas the remaining two equilibration rounds used a time step of 2 fs. Eight replicates were generated and subjected to 2 ns production dynamics, yielding a total of 16 ns for each EV71-2A^{pro}−peptide complex. To monitor the stability of each EV71-2A^{pro}−peptide complex, we computed the root-mean-square-deviation (RMSD) of the backbone atoms from those in the starting structure. Since the RMSDs during the second half of each production run were quite stable, frames were saved every picosecond after 1 ns, resulting in 8000 conformations for computing ΔG^{bind}_{sln}, the binding free energy of a peptide to EV71-2A^{pro}.

Binding Free Energy Calculations

As EV71-2A^{pro} underwent negligible structural change upon peptide binding with a backbone RMSD of only 0.24 Å between the 4fvd complex structure and 1.9 Å apo structure (PDB 4fvb), the ΔG^{bind}_{sln} was derived from a set of conformations extracted from the eight 2 ns trajectories of each EV71-2A^{pro}−peptide complex using the Molecular Mechanics Poisson−Boltzmann Surface Area (MM-PBSA) method.[36−41] Compared to alchemical methods (e.g., free energy perturbation or thermodynamic integration), which require the sampling of many intermediate states between the bound and unbound states,[40,42] MM-PBSA, an end-point method using implicit solvation, is more efficient at computing the binding free energies of numerous peptide candidates to EV71-2A^{pro}. The ΔG^{bind}_{sln} was computed as a sum of the gas-phase binding free energy (ΔG^{bind}_{gas}) and the solvation free energy difference between bound and unbound states ($\Delta\Delta G_{solv}$) averaged over a set of MD conformations; i.e.,

$$\Delta G^{bind}_{sln} \sim \Delta G^{bind}_{gas} + \Delta\Delta G_{solv} \quad (1)$$

where

$$\Delta\Delta G_{solv} = \Delta G_{solv}{}^{complex} - \Delta G_{solv}{}^{protein} - \Delta G_{solv}{}^{peptide} \quad (2)$$

First, the 8000 conformations of each protein−peptide complex were sorted into 10 clusters based on their gas-phase electrostatic interaction energies. A representative conformation from each cluster was used to compute the binding free energy, ΔG^{bind}_{sln}, according to eq 1. The ΔG^{bind}_{gas} was approximated by the sum of the van der Waals (vdW) and electrostatic binding energies using no cutoffs. The solvation free energy (ΔG_{solv}) was approximated by the sum of electrostatic ($\Delta G_{solv}{}^{elec}$) and nonelectrostatic ($\Delta G_{solv}{}^{nonel}$) components. The $\Delta G_{solv}{}^{elec}$ contribution was estimated by finite−difference solution of the linearized Poisson−Boltzmann equation implemented in the APBS program[43] using a dielectric constant of 1 for the protein and 78.54 for the solvent with an initial grid spacing of 1.3 Å, which was successively reduced to 0.8 Å and finally to 0.3 Å. The $\Delta G_{solv}{}^{nonel}$ contribution was assumed to be proportional to the solvent accessible surface area (SASA) of the molecule; i.e., $\Delta G_{solv}{}^{nonel} \sim 0.00072 \times$ SASA, where SASA was computed using a solvent probe radius of 1.4 Å.

Free Energy Decomposition

Following previous work,[37] the contribution of each protein/peptide residue to the binding free energy was then determined from the sum of (i) the pairwise vdW and electrostatic gas-phase energies of the residue and (ii) electrostatic and nonelectrostatic solvation free energies of the residue. Since the 10 clusters contain different numbers of conformations, the per-residue contributions and ΔG^{bind}_{sln} were weighted according to the number of conformations in each cluster to yield the respective averages and standard deviations for each protein−peptide complex.

Protein Expression and Purification

The cDNA of EV71-2A^{pro} was synthesized from MDBio, Inc. (Taiwan) and cloned into *Bam*HI/*Eco*RI sites of the modified expression vector pGEX-4T-1, which expressed the recombinant protein (residues 1−150) with an N-terminal fused GST followed by a TEV protease cleavage site. The catalytically inactive mutant (EV71-2A^{pro}-C110A) was generated by using QuickChange site-directed mutagenesis kits (Stratagene) from the wild-type construct. These two plasmids were transformed respectively into *E. coli* BL21 (DE3) pLysS strain cultured in LB medium supplemented with 100 μg/mL ampicillin. Cells were grown to an optical density of 0.6 measured at a wavelength of 600 nm and induced by 0.8 mM IPTG at 18 °C for 14−16 h. The harvested cells were disrupted by using a microfluidizer (Microfluidics M-110P) in the lysis buffer containing 50 mM Tris-HCl (pH 7.4), 150 mM NaCl, and 10 mM β-mercaptoethanol. The cell lysate was centrifuged, and the supernatant was loaded onto a GSTrap 4B column (5 mL, GE HealthCare). After washing with lysis buffer, the TEV protease (1 mg) was applied to the column for on-column digestion of the GST tag overnight at 4 °C. The untagged EV71-2A^{pro} was eluted by the lysis buffer and was further purified by using a gel filtration column (HiLoad 16/60 Superdex 200, GE HealthCare) pre-equilibrated with a buffer of 50 mM Tris-HCl (pH 7.4).

EV71-2A^{pro} Inhibition Assays

The fluorogenic peptide Dabcyl-RDKITTLGKFGQDE-Edans-NH2 (50 μM) was mixed respectively with the designed peptide with concentrations ranging from 0.49 to 500 μM in a buffer of 20 mM Tris-HCl (pH 7.4) and 100 mM NaCl for 10 min at 30 °C. EV71-2A^{pro} (250 nM) was added to the peptide mixture, and the fluorescence intensity generated by cleavage of the fluorogenic peptide was continuously monitored at 500 nm with the excitation wavelength at 360 nm via a SpectraMax Paradigm Multi-Mode Microplate Reader (Molecular Devices, LLC, USA). Increasing of the

https://doi.org/10.1021/acsbiomedchemau.2c00001
ACS Bio Med Chem Au 2022, 2, 437−449

relative fluorescent signal (RFU) at the time interval 100−600 s was calculated as the initial velocity (RFU/s). The relative activity of EV71-2A^{pro} in the presence of the peptide candidate was calculated based on the reduced initial velocity as compared to that of the enzyme alone.

Kinetic Activity Assays

The Michaelis−Menten constant K_m of the substrate peptide and designed peptides for EV71-2A^{pro} were calculated by incubation of 250 nM EV71-2A^{pro} respectively with the fluorogenic peptide Dabcyl-RDKITTLGKFGQDE-Edans-NH2 (substrate), Dabcyl-RDKIPF-RGYYGQDE-Edans-NH2 (pseudosubstrate I), Dabcyl-RDKIGF-RGKFGQDE-Edans-NH2 (pseudosubstrate II), Dabcyl-RDKIPF-MGYRGQDE-Edans-NH2 (pseudosubstrate III), or Dabcyl-RDK-ITFMGKFGQDE-Edans-NH2 (pseudosubstrate IV) in a concentration range from 0.015 to 250 μM in a buffer of 20 mM Tris-HCl (pH 7.4) and 100 mM NaCl at 30 °C for 10 min. Increasing of fluorescence intensity at the time interval 100−600 s was calculated as the initial velocity and the relative velocity were calculated at the different substrate/pseudosubstrate peptide concentrations. The fluorescence intensity data and enzymatic activity were analyzed with SoftMax Pro 7 Software (Molecular Devices, LLC.USA), and Michaelis−Menten saturation curves were fitted with GraphPad Prism 7.0.

Protein Crystallization and Crystal Structure Determination

EV71-2A^{pro}-C110A (10 mg/mL) was premixed with peptide-II at a molar ratio of 1:2 for the preparation of the EV71-2A^{pro}-C110A/pseudosubstrate II complex. Crystals of the EV71-2A^{pro}-C110A/pseudosubstrate II complex were grown by a hanging-drop vapor diffusion method at room temperature with the initial drops containing 1 μL of EV71-2A^{pro}-C110A/pseudosubstrate II complex solution and 1 μL of the reservoir solution of 0.1 M HEPES (pH 7.5), 20% 2-propanol, and 10% polyethylene glycol (PEG) 4000. Crystals appeared in 3−4 days and were transferred into a solution containing 0.1 M HEPES (pH7.5), 20% 2-propanol, 10% PEG 4000, and 10% glycerol prior to flash cooling to 100 K for data collection. X-ray diffraction data up to a resolution 2.2 Å were collected at TPS beamline 05A, National Synchrotron Radiation Research Center (NSRRC), Taiwan. Diffraction data were processed with HKL2000, and the crystal structure was determined by molecular replacement, using the structure of apo EV71-2A^{pro} (PDB 4fvb) as the searching model by Phaser in the PHENIX program suite. After initial structure refinement, four amino acids ($F_3R_2G_1K_{1'}$) were modeled into the ill-defined electron density in the peptide-binding groove of EV71-2A^{pro}. Cycles of molecular model building and refinement were performed by PHENIX and Coot. Detailed statistics for data processing and structure refinement are listed in Table 1.

RESULTS

Strategy to Identify Viral Drug Target Region

To simultaneously inhibit the 2A proteases of several subtypes of enteroviruses, and at the same time avoid targeting human proteins, we performed a sequence analysis of EV71-2A^{pro} using BLASTp. The EV71-2A^{pro} aa sequence was found to share >30% sequence identity with only viral proteins belonging to the pico_P2A superfamily, but no human proteins. Notably, this conserved domain with PFAM code pfam00947 did not appear in any human proteins. EV71-2A^{pro} shares high sequence identity with the 2A proteases of other enterovirus species such as CVA16 (97%), coxsackievirus B4 (73%), echovirus 1 (75%), poliovirus 1 (59%), enterovirus D68 (50%), and to a lesser extent with human rhinoviruses (31−36%). Moreover, the sequences flanking the catalytic triad (C110−H21−D39) are highly conserved among the enteroviruses and rhinoviruses (Figure 2).

Table 1. X-ray Diffraction and Crystal Structure Refinement Statistics for EV71-2A^{pro} in Complex with Pseudosubstrate II (PDB entry: 7da6)

data collection	
space group	$C2_1$
cell dimensions	
a, *b*, *c* (Å)	85.48, 44.41, 51.37
α, β, γ (deg)	90.0, 111.57, 90.0
resolution (Å)	30.0−2.2 (2.24−2.2)[a]
total reflections	32,175
unique reflections	9,107
redundancy	3.5 (3.5)[a]
R_{merge}	0.063 (0.31)[a]
CC 1/2	0.985 (0.882)[a]
I/σ	25.03 (3.54)[a]
completeness (%)	98.9 (98.7)[a]
refinement	
resolution (Å)	22.75−2.20 (2.28−2.20)[a]
no. of reflections	9,107
R_{work}	0.1950 (0.2327)[a]
R_{free}	0.2456 (0.3267)[a]
RMSD bond lengths (Å)	0.01
RMSD bond angles (deg)	1.26
average B-factors (Å^2)	40.94
Ramachandran favored (%)	96.38
Ramachandran allowed (%)	3.62
Ramachandran disallowed (%)	0.0

[a]Statistics for the highest resolution shell are shown in parentheses.

Although the EV71-2A^{pro} does not share significant (>30%) overall sequence identity with any human protein, it may nevertheless share *structural* similarity with a human protein. Thus, to see if any human proteins have structures that are highly similar to EV71-2A^{pro}, we compared the EV71-2A^{pro} structure (PDB 4fvb) with the PDB structures in the BioLIP database of the COFACTOR server. The degree of structural similarity between the EV71-2A^{pro} structure and a PDB structure was estimated by the TM-score and the RMSD between the C$^{\alpha}$ atoms of residues that are structurally aligned by TM-align in the COFACTOR server.[24] In general, the TM-score is <0.3 for two randomly chosen unrelated protein structures and is >0.5 for two protein structures with the same fold.[24] Hence, we considered two protein structures to be similar if the TM-score is >0.5 and their pairwise RMSD is <3 Å. The results (Table 2) confirm that *no* PDB structures of human proteins are highly similar to the EV71-2A^{pro} structure.

A Common Target for Drug Design

To see if there is a common target for the design of antiviral compounds, we aligned the EV71-2A^{pro} structure (PDB 4fvb) and the top-ten analogous structures in Table 2 using FATCAT[26] to determine common features that are shared between EV71-2A^{pro} and the aligned proteins. Figure 3 (top panel) illustrates the *structural* alignment between EV71-2A^{pro} and poliovirus precursor 3cd (PDB 2ijd) structures. The results show a characteristic region (^{87}SEYYP^{91}) in the bII2-cII loop near the substrate-binding cleft of the EV71-2A^{pro} that is fully conserved in the 2A proteases of EV71, coxsackieviruses (CVA16 and CVB4), echovirus 1, and human rhinovirus 2, and partly conserved (^{89}YYP^{91}) in the 2A proteases of poliovirus 1 as well as human enterovirus D68 or rhinovirus C15 (Figure 2). Interestingly, E88 and Y89 are thought to switch the substrate-binding cleft conformation between

https://doi.org/10.1021/acsbiomedchemau.2c00001
ACS Bio Med Chem Au 2022, 2, 437−449

Figure 2. Multiple sequence alignment of 2A proteases. All sequences are extracted from the UniProt database. EV71, human enterovirus 71; CVA16, coxsackievirus A16; CVB4, coxsackievirus B4; EC1, echovirus 1; PV1, poliovirus 1; EVD68, human enterovirus D68; HRV2, human rhinovirus 2; HRVC15, human rhinovirus C15. Multiple sequence alignment was performed using CLUSTALW.[22] The catalytic triad (H21, D39, C110) and the ^{87}(S/N)(E/N/G)YYP^{91} motif are in orange, whereas the Zn^{2+}-binding cysteines are marked by an asterisk underneath. The five residues that differ between EV71-2A^{pro} and CVA16-2A^{pro} are shaded in gray.

Table 2. All Structural Analogues with TM Score > 0.5 and RMSD < 3 Å Calculated by TM-Align in the COFACTOR Server[a]

rank	PDB hit	TM score[a]	RMSD[b]	identity[c]	coverage[d]	protein name
1	4mg3A	0.98	0.59	0.95	1.00	CVA16 2A protease
2	7jreA	0.92	1.53	0.50	0.99	EV68 2A protease
3	2hvA	0.91	1.59	0.35	0.99	HRV2 2A protease
4	1z8rA	0.88	1.59	0.70	0.98	CVB4 2A protease
5	2m5tA	0.85	2.05	0.33	0.99	HRV 2A protease
6	6ku7A	0.78	2.46	0.11	0.95	HRVC 3C protease
7	2ijd1	0.78	2.59	0.14	0.96	poliovirus precursor protein 3 cd
8	2in2A	0.78	2.53	0.11	0.95	HRV 3C protease
9	3zv8A	0.78	2.68	0.12	0.96	EV68 3C protease
10	1cqqA	0.78	2.58	0.14	0.95	HRV2 3C protease
11	2ea3A	0.75	2.64	0.15	0.95	*Cellulomonas bogoriensis* chymotrypsin
12	1qtfA	0.68	2.71	0.16	0.86	*Staphylococcus aureus* exfoliative toxin
13	2wv4A	0.54	2.44	0.20	0.65	HFMD 3C protease

[a]The TM-score reflects the degree of the structural alignment between the EV71 2A^{pro} structure (PDB 4fvb) and a given PDB structure; it is 1 if two proteins have identical structures. [b]The RMSD between C$^{\alpha}$ atoms of residues that are structurally aligned by TM-align. [c]The percentage sequence identity in the structurally aligned region. [d]The coverage of the structural alignment by TM-align is equal to the number of structurally aligned residues divided by the total number of EV71 2A^{pro} residues, which is 138.

"closed" and "open" states.[44] Since the active-site structures of some human proteases remain unsolved, we searched all human sequences for the presence of the ^{87}SEYYP^{91} motif using BLASTp. The results show part of this motif in only a single human protease: the ^{88}EYYP^{91} motif matches residues 763−766 of ADAMTS18, a disintegrin and metallopeptidase with thrombospondin motifs 18 (UniProt code Q8TE60). The ADAMTS18 structure predicted by AlphaFold[45] shows the ^{763}EYYP^{766} sequence located in a β-strand with very high confidence (the per-residue confidence score, pLDDT score > 90), whereas it is part of a loop in EV71-2A^{pro}. This indicates that an inhibitor interacting with the unique EV71−2A^{pro} ^{87}SEYYP^{91} loop region would unlikely interact with human host proteases.

Hence, a common attractive target for the design of antiviral compounds against the 2A proteases of EV71, coxsackievirus A16/B4, and echovirus-1 is the substrate-binding site and the nearby ^{87}SEYYP^{91} loop region (Figure 3, bottom left). The 2A proteases of EV71 and CVA16, the two major causative agents for HFMD, share not only high sequence similarity but also high structural similarity: When the backbone C$^{\alpha}$ atoms of the EV71-2A^{pro} (PDB 4fvb) and CVA16-2A^{pro} (PDB 4mg3) structures are superimposed, the RMSD is only 0.59 Å (Figure 3, bottom right). Hence, a molecule designed to inhibit EV71-2A^{pro} would also likely inhibit CVA16-2A^{pro}.

https://doi.org/10.1021/acsbiomedchemau.2c00001
ACS Bio Med Chem Au 2022, 2, 437−449

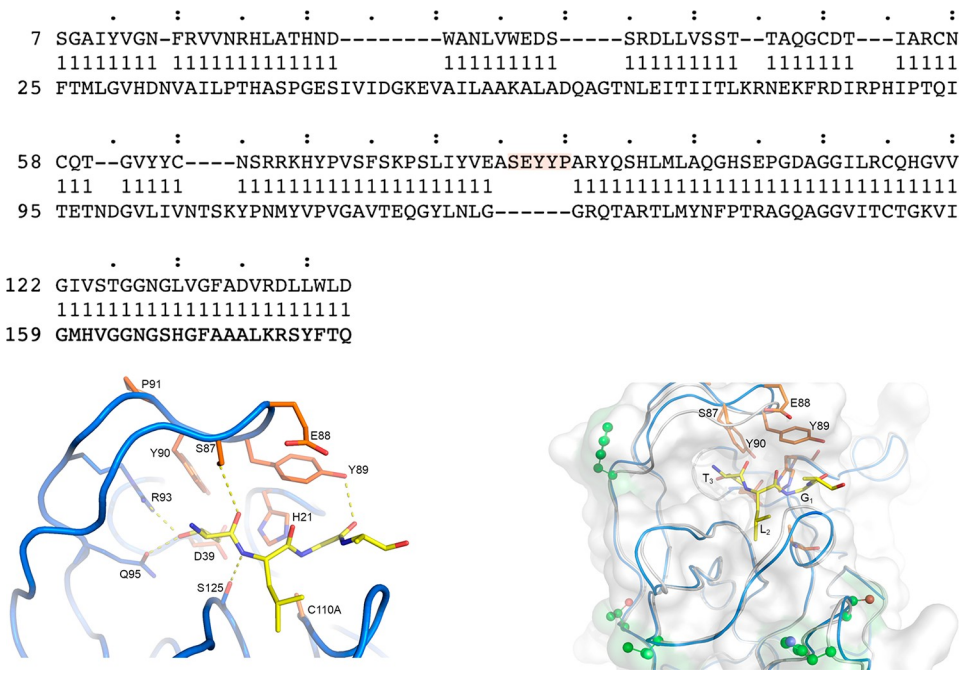
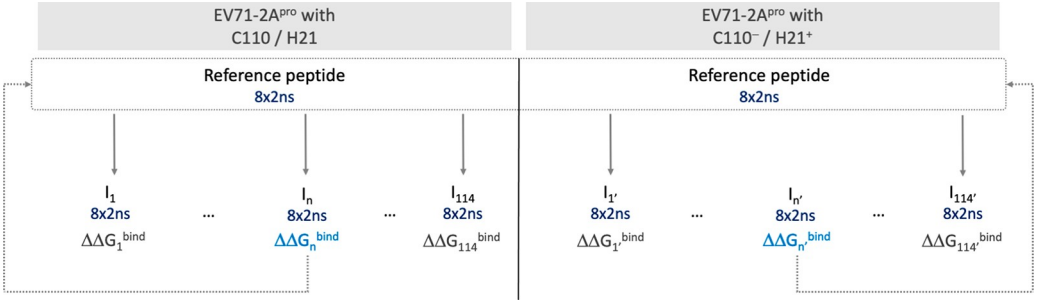


Figure 3. Viral drug target region. (a) Structural alignment of EV71-2A^{pro} (PDB 4fvb) and homologous human trypsin-like serine proteases reveal a unique region (^{87}SEYYP^{91}) in the bII2-cII loop near the substrate-binding site of the EV71-2A^{pro}. (b) EV71-2A^{pro} backbone C atoms are in blue, catalytic triad residues (H21, D39, and C110) and unique ^{87}SEYYP^{91} loop region are in orange, whereas the substrate is in yellow. Nitrogen atoms are in blue, and oxygen in red. Hydrogen-bonding interactions are indicated by dashed lines. (c) Superposition of the crystal structures of EV71-2A^{pro} (PDB 4fvd) and CVA16 (PDB 4mg3) whose backbone C atoms are in gray. The spheres highlight residues that differ between CVA16 and EV71-2A.

Scheme 1. Flowchart to Identify Peptide Candidates That Bind to EV71-2A^{pro} Assuming Neutral C110−H21 (Left) or C110^{-}/H21^{+} Ion Pair (Right)[a]



[a]The initial reference peptide was the substrate; single-point mutations of each aa generated peptide candidates, I_j, j = 1... 114. For a given charge state of the catalytic dyad, eight 2 ns simulations were performed for each EV71-2A^{pro}/peptide complex to generate a set of conformations for computing the binding free energy. A peptide that bound EV71-2A^{pro} more tightly than the substrate was used as a new reference for performing single-point mutations, followed by MD simulations and binding free energy calculations. This procedure was repeated until a peptide that could bind tighter to EV71-2A^{pro} than the current reference was not found.

Strategy to Identify Peptide Candidates

The starting point for designing peptides that can compete with the substrate in binding to EV71-2A^{pro} was the 1.66 Å crystal structure of EV71-2A^{pro} with the catalytic Cys110 replaced by Ala (C110A) in complex with a substrate, GSITTLGKFG (PDB 4fvd). In the 4fvd structure, only a partial substrate was visible, viz., TLG and the Lys backbone;[46] thus, we added a residue at either end to yield a substrate consisting of P4−P2′ residues (denoted by $T_4T_3L_2G_1K_{1'}F_{2'}$), which are located in subsites S4−S2′, respectively. We then mutated each peptide residue to the other 19 amino acid (aa) residues, yielding altogether 19 × 6 peptide candidates, I_j, j = 1... 114. Each EV71-2A^{pro}/peptide complex was subjected to eight 2 ns MD simulations to relax the structure and to obtain conformations for computing the binding free energy, ΔG^{bind}_{sln} according to eq 1 (see Scheme 1). Since experimental pK_a values for the catalytic C110−H21 dyad, which depend on the nature of the active-site peptide, are not available, we carried out two sets of eight 2 ns simulations for each EV71-2A^{pro}/peptide complex−one set with deprotonated C110^{-} and protonated H21^{+} and the other with both C110 and H21 neutral. For a given charge state of the catalytic dyad, the ΔG^{bind}_{sln} free energy of the peptide (including substrate) binding to EV71-2A^{pro} was computed.

To determine if the peptide candidate bound EV71-2A^{pro} better than the reference substrate, we decomposed the binding free energy into the contributions from each residue (see Methods). This is illustrated in Figure 4, where we

https://doi.org/10.1021/acsbiomedchemau.2c00001
ACS Bio Med Chem Au 2022, 2, 437−449

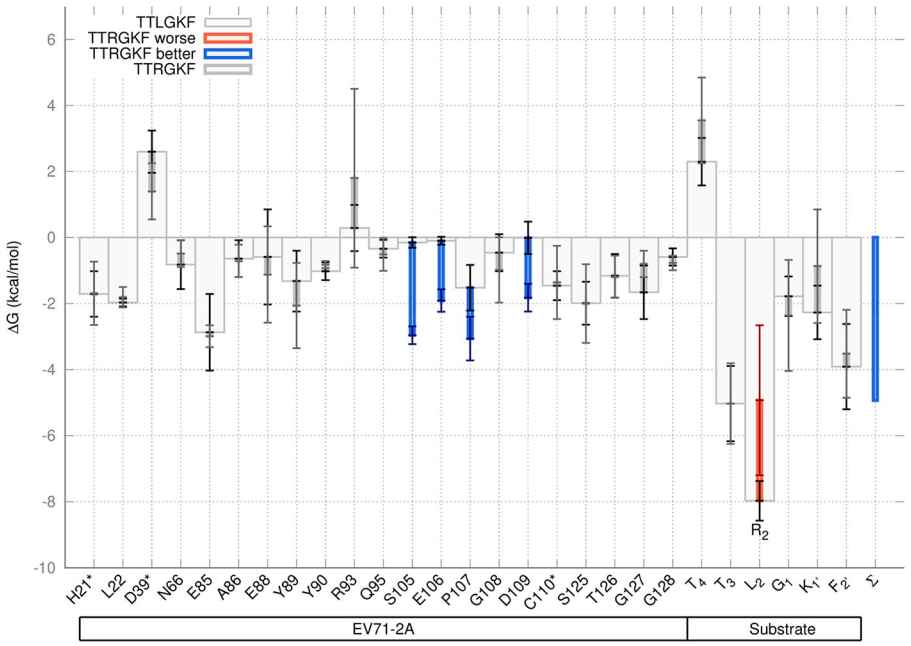
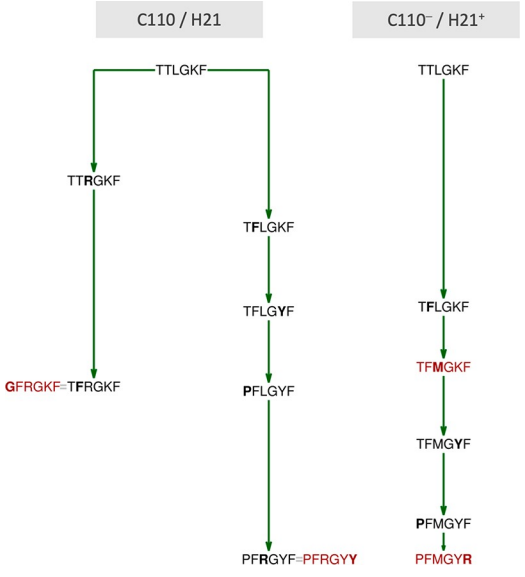


Figure 4. Free energy decomposition of mutant or reference peptide binding to EV71-2A^{pro} with neutral Cys110−H21. The per-residue free energy contributions of magnitude > 0.5 kcal/mol and corresponding standard deviations (shown by the error bars) of EV71-2A^{pro} binding to the reference $T_4T_3L_2G_1K_1{\cdot}F_2{\cdot}$ substrate (unfilled bars) and the peptide candidate, $T_4T_3\mathbf{R}_2G_1K_1{\cdot}F_2{\cdot}$ (filled narrow bars). Those that became more (or less) favorable by more than the sum of the standard deviations upon single-point mutation are shown by the blue (or red) bars. The rightmost bar is the sum of the blue and red bars; it is blue when there is an overall binding free energy gain upon single-point mutation.

mutated L_2 in the initial $T_4T_3\mathbf{L}_2G_1K_1{\cdot}F_2{\cdot}$ substrate to R_2 and decomposed the binding free energies for the $T_4T_3\mathbf{L}_2G_1K_1{\cdot}F_2{\cdot}$ substrate (unfilled bars) and the $T_4T_3\mathbf{R}_2G_1K_1{\cdot}F_2{\cdot}$ peptide candidate (filled bars) into contributions from each EV71-2A^{pro} or peptide residue. The per-residue binding free energy contributions that became more (or less) favorable by more than the sum of the standard deviations upon single-point mutation are shown by the blue (or red) bars in Figure 4, respectively. Compared to the substrate, the free energy contributions of the mutant R_2 as well as the surrounding peptide and virus residues have changed significantly. Although the R_2 mutant residue exhibited weaker binding than L_2 to the viral protease (red bar in Figure 4), it allowed enhanced interactions with S105, E106, P107, and D109 of EV71-2A^{pro} (blue bars), inducing an overall free energy gain, as indicated by the sum of the blue and red bars (rightmost bar in Figure 4). Hence, this mutant peptide was considered to be a potential high-affinity substrate.

One of the 114 peptides that resulted in tighter binding to EV71-2A^{pro} than the substrate was used as a new reference. For example, single-point mutations starting from the $T_4T_3L_2G_1K_1{\cdot}F_2{\cdot}$ substrate with neutral C110 and H21 yielded two peptides ($T_4T_3\mathbf{R}_2G_1K_1{\cdot}F_2{\cdot}$ and $T_4\mathbf{F}_3L_2G_1K_1{\cdot}F_2{\cdot}$) that bound EV71-2A^{pro} much tighter than the substrate (see Scheme 2 where the length of the arrow indicates the binding free energy gain). Using $T_4T_3\mathbf{R}_2G_1K_1{\cdot}F_2{\cdot}$ or $T_4\mathbf{F}_3L_2G_1K_1{\cdot}F_2{\cdot}$ as the new reference, we repeated the above procedure and *mutated each* residue in this peptide to the other 19 aa residues, followed by eight 2 ns MD simulations for each peptide complex and binding free energy decomposition analyses. The second round of single-point mutations resulted in a $T_4\mathbf{F}_3\mathbf{R}_2G_1K_1{\cdot}F_2{\cdot}$ or $T_4\mathbf{F}_3L_2G_1\mathbf{Y}_1{\cdot}F_2{\cdot}$ peptide that bound EV71-2A^{pro} tighter than the starting $T_4T_3\mathbf{R}_2G_1K_1{\cdot}F_2{\cdot}$ or $T_4\mathbf{F}_3L_2G_1K_1{\cdot}F_2{\cdot}$, respectively. This procedure of performing single-point mutations followed by MD simulations and binding free energy decomposition was repeated until we could not find a peptide that could bind tighter to EV71-2A^{pro} than the current reference.

Scheme 2. Iterative Cycle of Peptide Design Using Single-Point Mutations Starting from the $T_4T_3L_2G_1K_1{\cdot}F_2{\cdot}$ Substrate with Neutral C110−H21 (left) and C110$^-$−H21$^+$ (right)



Thus, our strategy is based on maximizing the binding affinity of the peptide to EV71-2A^{pro} iteratively using single-point mutations. This led to the identification of four peptides with the most favorable binding free energies compared to the native substrate, but whose sequences were not too hydrophobic. They are $P_4F_3R_2G_1Y_1{\cdot}Y_2{\cdot}$ or $G_4F_3R_2G_1K_1{\cdot}F_2{\cdot}$ assuming neutral C110 and H21, and $P_4F_3M_2G_1Y_1{\cdot}R_2{\cdot}$ or $T_4F_3M_2G_1K_1{\cdot}F_2{\cdot}$ assuming C110$^-$/H21$^+$ ion pair. All 4 peptides form hydrogen-bonding or vdW interactions with the unique

https://doi.org/10.1021/acsbiomedchemau.2c00001
ACS Bio Med Chem Au 2022, 2, 437−449

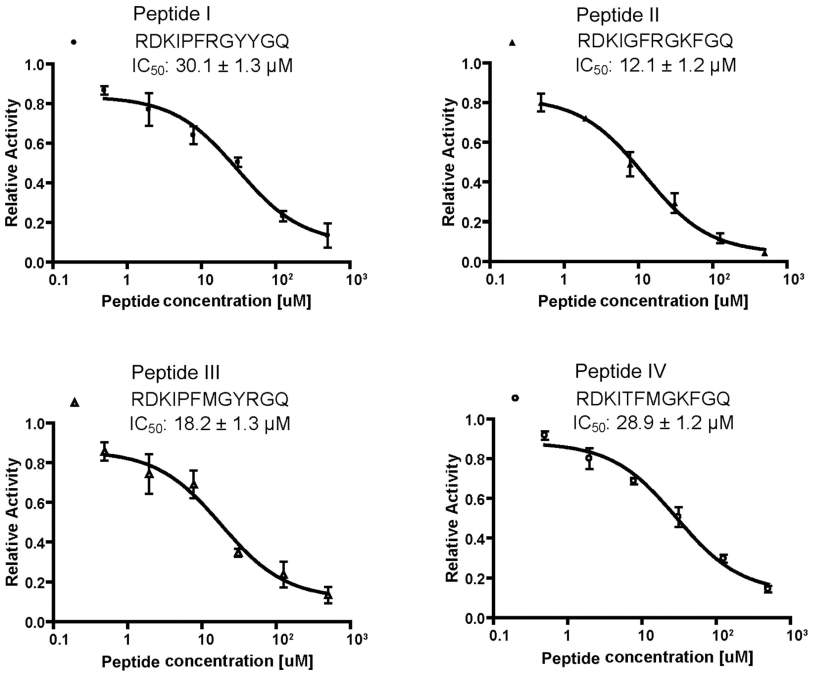


Figure 5. Inhibition of the protease activity of EV71-2A^{pro} by the four designed peptides. Four designed peptides (I) RDKIPFRGYYGQ, (II) RDKIGFRGKFGQ, (III) RDKIPFMGYRGQ, and (IV) RDKITFMGKFGQ significantly inhibited the protease activity of EV71-2A^{pro} in cleaving a fluorogenic peptide Dabcyl-RDKITTLGKFGQDE-Edans-NH2. Error bars represent the standard errors from at least three replicates of the experiment.

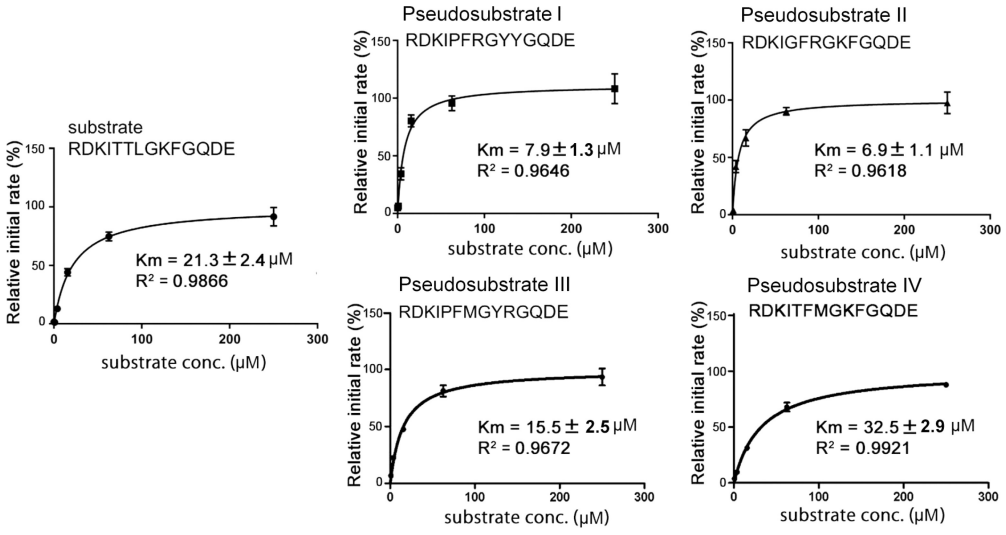


Figure 6. Four designed peptides function as pesudosubstrates to inhibit the protease activity of EV71-2A^{pro}. The four designed peptides with the same length (14 amino acids) and the same fluorophores labeled at the two ends as the native substrate Dabcyl-RDKI-TTLGKF-GQDE-Edans-NH_2 were cleaved by EV71-2A^{pro}. The Michaelis−Menten constant K_m values of the substrate and the four pseudosubstrate peptides were estimated as shown in the figure. The relative initial rate (%) of EV71-2A^{pro} in cleaving each peptide was estimated by the increase of relative fluorescent signal (RFU/s), which was normalized to the plateau of the initial rate (100%). Error bars represent the standard errors from at least three replicates of the experiment.

^{87}SEYYP^{91} motif of EV71-2A^{pro}: F_3 and R_2 pack against the Y89 and Y90 side chains, whereas the M_2 and P1′ backbones form hydrogen bonds with the S87 and Y90 side chains, respectively.

Mutation of P3, P2, and P1′ Residues Yields the Most Binding Free Energy Gain

The binding free energies of the peptides to EV71-2A^{pro} show that the protonation state of the catalytic dyad affects the preference for an aa residue at a given position: This is especially the case for the P2 position that consistently yields an Arg in the optimized peptide when C110 is neutral, but a Met when C110 is deprotonated (see Scheme 2). The same residue was preferred at the other subsites regardless of the catalytic dyad charge state. At the P3 position, Phe is favored because its aromatic side chain can interact favorably with the two Tyr side chains of the ^{87}SEYYP^{91} motif. At the P1′ position, Lys/Tyr is preferred, as the NH_3^+/OH side chain can form hydrogen bonds with the E88 carboxylate group. Since

https://doi.org/10.1021/acsbiomedchemau.2c00001
ACS Bio Med Chem Au 2022, 2, 437−449

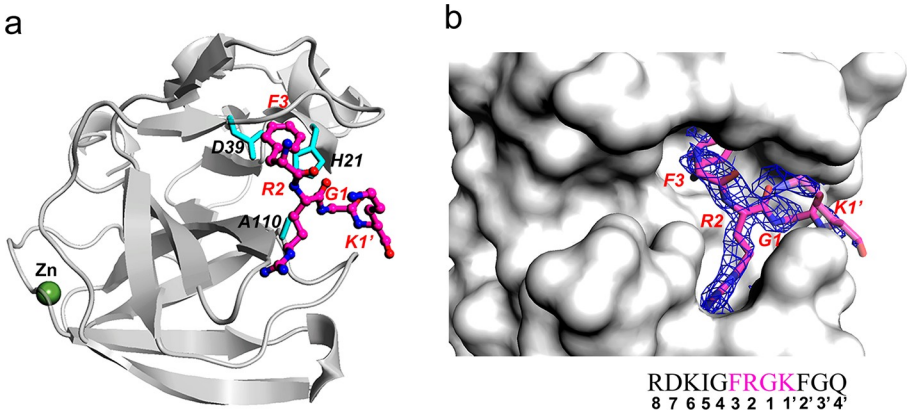


Figure 7. Crystal structure of EV71 2A^{pro}-C110A in complex with the pseudosubstrate II. (a) Overall crystal structure of the EV71 2A^{pro}-C110A/pseudosubstrate II complex. The peptide (magenta ball-and-stick format) is bound in the EV71 2A^{pro} active site lined by the catalytic residues, H21, D39, and C110A (in cyan stick format). (b) Designed peptide (RDKIG-**FRGK**-FGQ) bound in the substrate-binding groove of EV71 2-A^{pro}. The 2mFo-DFc maps contoured at 0.8σ reveal the location of four amino acids $F_3R_2G_1K_{1'}$ bound in the active site. Electron density is well-defined for the R_2 side chain but is ill-defined for the F_3 and $K_{1'}$ side chains.

residues at the terminal P4 and P2′ positions were absent in the starting 4fvd crystal structure, predictions for the preferred residues at these positions are less reliable than those at the nonterminal positions. Nevertheless, the per-residue free energy contributions suggest that the substrate Phe at the P2′ position was already optimal, although Tyr or Arg may serve as alternatives. At the P4 position, only mutations to Gly or Pro led to a gain in the binding free energy.

Designed Peptides Inhibit EV71-2A^{pro} Protease Activity

In initial protease activity tests, the designed 6-mer peptide sequences could not compete with the longer 14-mer native substrate (RDKI-TTLGKF-GQDE); hence, we extended them by adding the native substrate residues, RDKI and GQ, at the N- and C terminus, respectively. This yielded the following four 12-mer peptides: (I) RDKI**PFRGYY**GQ, (II) RDKI-**GFRGKF**GQ, (III) RDKI**PFMGYR**GQ, and (IV) RDKI-**TFMGKF**GQ. To see if these four candidates can inhibit the EV71-2A^{pro} activity, we measured the activity of EV71-2A^{pro} to cleave a fluorogenic peptide substrate (Dabcyl-RDKI-TTLGKF-GQDE-Edans-NH_2) in the presence or absence of the designed peptide. Increasing of the relative fluorescent signal (RFU) during the time interval, 100−600 s, was calculated as the initial velocity (RFU/s). Upon adding a designed peptide of a given concentration, the fluorescent signals produced from substrate cleavage decreased, resulting in reduced initial velocity. Increasing the concentration of each designed peptide decreased the activity of EV71-2A^{pro} to cleave the fluorogenic peptide substrate, indicating that all four designed peptides can indeed inhibit the EV71-2A^{pro} activity: In particular, peptide II displayed the most efficient inhibition among the four candidates with a half maximal inhibitory concentration (IC_{50}) of 12.1 ± 1.2 μM (Figure 5). The other three peptides also had low IC_{50} values ranging from 18.2 to 30.1 μM.

Designed Peptides Act as Higher-Affinity Substrates

To determine if these peptides inhibited EV71-2A^{pro} activity by outcompeting the native substrate, we synthesized four 14-mer designed peptides with the same fluorophores labeled at the two ends as compared to the native substrate Dabcyl-RDKI-TTLGKF-GQDE-Edans-NH_2. In analogy to the substrate, the four designed peptides can be cleaved by the EV71-2A^{pro} enzyme, as manifested by the fluorescent signals, indicating that they are effective EV71-2A^{pro} substrates or pseudosubstrates (Figure 6). The Michaelis−Menten constant K_m of each peptide was then derived from the relative initial rate of EV71-2A^{pro} in cleaving the pseudosubstrate peptide as a function of the peptide concentration. Compared to the native substrate peptide (K_m = 21.3 ± 2.4 μM), pseudosubstrate peptides *I* (K_m = 7.91 ± 1.26 μM) and II (K_m = 6.92 ± 1.11 μM) bound EV71-2A^{pro} with higher affinity, whereas pseudosubstrate peptides III (K_m = 15.51 ± 2.48 μM) and IV (K_m = 32.53 ± 2.87 μM) bound EV71-2A^{pro} with comparable affinity (Figure 6). These results confirm that the designed peptides I and II indeed bind EV71-2A^{pro} with higher affinity than the native substrate (by 3-fold). They suggest that the designed peptides can compete with the substrate for the same binding site, resulting in the inhibition of protease activity.

Crystal Structure of EV71-2A^{pro} in Complex with Pseudosubstrate Peptide

To reveal the molecular basis for the observed inhibition EV71-2A^{pro} by the tested peptides, we cocrystallized EV71-2A^{pro}-C110A with the pseudosubstrate II (RDKI-GFRGKF-GQ) and determined the crystal structure of the EV71-2A^{pro}-C110A/pseudosubstrate II complex at a resolution of 2.2 Å (see Table 1 for structure refinement statistics). Pseudosubstrate II was bound in the substrate-binding groove of EV71-2A^{pro}-C110A. However, the electron density was well-defined only for $F_3R_2G_1K_{1'}$ (Figure 7a,b), in particular the long R_2 side chain, but not for the rest of the peptide. Although we could model only 4 of the 12-mer peptide into the density map, it is evident that these four residues are bound in the enzyme active site. Interestingly, the *4fvd* crystal structure of EV71-2A^{pro}-C110A bound to a 10-mer substrate (GSITT LGKGF) also showed electron density for four residues in the same subsites. This suggests that the P3, P2, P1, P1′ residues account for most of the peptide binding free energy, consistent with free energy decomposition analyses showing that mutation at these residue positions resulted in the most binding free gain. Hence, mutational screening (including simultaneous multiple mutations) of these four residues may further improve the affinity of pseudosubstrate II for EV71-2A^{pro}. Although the structure of the EV71-2A^{pro}-C110A/pseudosubstrate II complex shows only a partial peptide, it, nevertheless, clearly indicates that

https://doi.org/10.1021/acsbiomedchemau.2c00001
ACS Bio Med Chem Au 2022, 2, 437−449

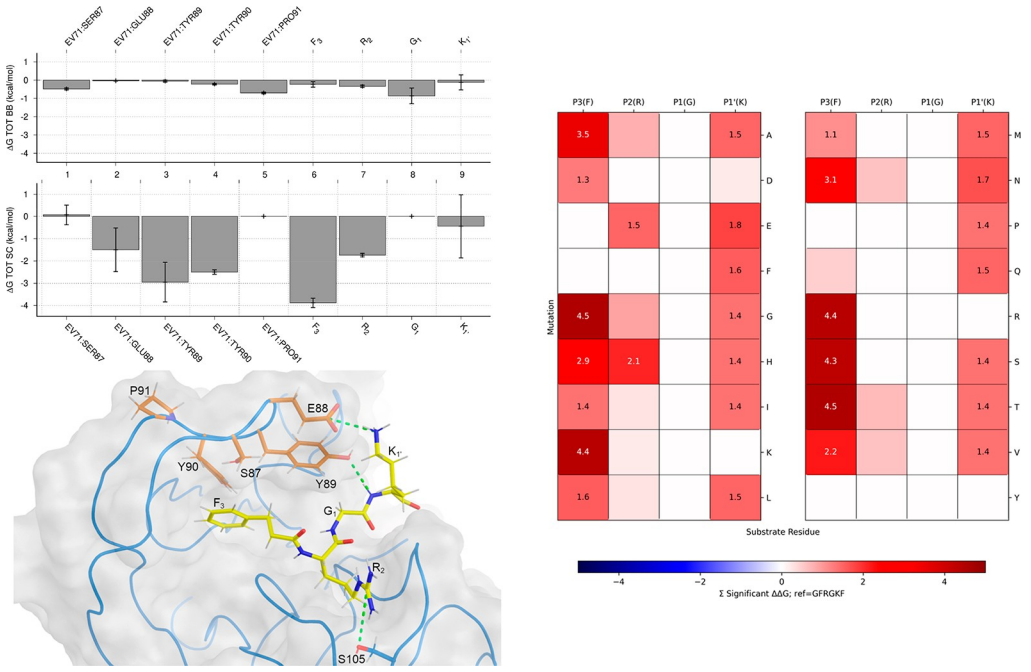


Figure 8. Interactions of the ^{87}SEYYP^{91} loop of EV71 2A^{pro} and $F_3R_2G_1K_{1'}$ of pseudosubstrate II. Free energy decomposition between the EV71 2A^{pro} ^{87}SEYYP^{91} loop and pseudosubstrate II residues, $F_3R_2G_1K_{1'}$ (top left). Interactions between the EV71 2A^{pro} ^{87}SEYYP^{91} loop and the pseudosubstrate II residues seen in the complex structure (bottom left). Effect of single point mutations on the ^{87}SEYYP^{91}−$F_3R_2G_1K_{1'}$ interactions, depicted with a color gradient where red indicates a loss of binding (right).

pseudosubstrate II inhibits the protease activity of EV71-2A^{pro} by blocking the active site.

Pseudosubstrate II Interacts with the ^{87}SEYYP^{91} Loop of EV71-2A^{pro}

Although the crystal structure of the EV71-2A^{pro}-C110A/pseudosubstrate II complex shows pseudosubstrate II in the substrate-binding groove near the unique ^{87}SEYYP^{91} loop region, it does not reveal (i) the strength of the pseudosubstrate II−^{87}SEYYP^{91} loop interactions, (ii) which peptide/loop residue makes the most favorable contributions to the binding free energy, and (iii) whether pseudosubstrate II can be further optimized to interact with the ^{87}SEYYP^{91} loop. Thus, to verify that the designed peptides not only would inhibit protease activity but may also exhibit specificity for viral 2A proteases, we carried out eight 2 ns simulations for the EV71-2A^{pro}/pseudosubstrate II complex with neutral C110 and H21 starting from the corresponding EV71-2A^{pro}/pseudosubstrate complex structure. The resulting simulation structures were stable, as evidenced by backbone RMSDs of 0.8−1.0 Å compared to the starting crystal structure. Using the simulation structures, we computed the binding free energy and decomposed it into contributions from the backbone and side chains of the peptide ($F_3R_2G_1K_{1'}$) and the ^{87}SEYYP^{91} loop. The per-residue free energy contributions in Figure 8 (left panel) show that the side chains of the two Tyr in the ^{87}SEYYP^{91} loop and the peptide F_3 make the most favorable contributions, as they form favorable vdW interactions. S87 and P91 do not make significant free energy contributions to binding the peptide, but they may help to orient the ^{87}SEYYP^{91} loop to interact with peptide.

To determine how efficient pseudosubstrate II is at targeting the ^{87}SEYYP^{91} motif, we evaluated the effect of single-point mutations of peptide residues F_3, R_2, and $K_{1'}$ on the binding free energy with the ^{87}SEYYP^{91} residues. The results in Figure 8 (right panel) show that none of the mutations tested could further strengthen the interactions with ^{87}SEYYP^{91}, as they resulted in no gain (white squares) or even a loss (red squares) of the binding free energy, especially at position P3 and P1′. These results indicate that the designed peptide is well optimized to interact favorably with the ^{87}SEYYP^{91} loop.

To explain why pseudosubstrate II can bind the EV71-2A^{pro} better than the native substrate, we superimposed the structures of the EV71-2A^{pro} complexed with the substrate, $T_3L_2G_1K_{1'}$ (PDB 4fvd), and pseudosubstrate II, $F_3R_2G_1K_{1'}$ (PDB 7da6). At the P3 position, F_3 is preferred over the native T_3 because it can stabilize a hydrophobic pocket formed by the EV71-2A^{pro} Y89 and Y90 residues as well as form stacking interactions with the catalytic H21 side chain. At the P2 position, the L_2 side chain of the substrate is sandwiched by T126/G127 on one side and P107/C110 on the other, whereas the long R_2 side chain of pseudosubstrate II can additionally form a stable hydrogen bond with the S105 hydroxyl group of the virus. Thus, the favorable packing and electrostatic interactions formed by pseudosubstrate II with EV71-2A^{pro} compared to the native substrate enabled it to acquire higher affinity, leading to inhibition of protease activity.

DISCUSSION

Prior to this work, 2A protease inhibitors have been discovered by using high-throughput yeast two-hybrid screening (e.g., the human rhinovirus-2A^{pro} LVLQTM peptide inhibitor)[47] or by mimicking the substrate peptide sequences (e.g., the CVB3-2A^{pro} 16-mer peptide inhibitor mimics the human eIF4G cleavage site).[48] Here, we have presented an efficient in silico strategy using well-established methods to identify 2A protease pseudosubstrate inhibitors. The first step of our strategy is to identify an appropriate viral region to target by comparative sequence and structure analyses. Such a region should be highly conserved among viruses but should ideally be absent

https://doi.org/10.1021/acsbiomedchemau.2c00001
ACS Bio Med Chem Au 2022, 2, 437−449

from human proteins. Having identified a "unique" viral region to target, the second step of our strategy is to identify peptides that can bind the target region more tightly than the substrate peptide by performing single-point mutations followed by MD simulations and free energy calculations iteratively.

Our peptide design strategy, summarized in Scheme 1, offers a reasonable trade-off between accuracy and cost to allow an assessment of a large number of peptide candidates. It provides a relatively quick way of choosing candidates that would likely bind more tightly than the substrate to EV71-2A^{pro} (rather than an exhaustive list of candidates), while allowing specific tuning of interactions with a selected region of the drug target protein. It can be used as a filter to select a smaller number of peptide candidates for more accurate, but compute-intensive alchemical simulations[42] to evaluate the relative binding free energy prior to experimental verification or to make a small library of synthetic peptides for screening. Previous studies generally assessed two putative inhibitors using the total binding free energy change, which is likely accurate to within 2 kcal/mol.[38,49] In contrast, we take into account the inherent errors of the relative binding free energy calculations by neglecting changes of individual binding free energies that are smaller than the sum of standard deviations; i.e., we focus on mutations that alter the per-residue free energy contributions beyond the standard deviations (Figure 4). Furthermore, the individual free energy contributions indicate which and why a peptide substituent could enhance binding over the substrate to the viral protease; e.g., Phe is favored in the hydrophobic S3 subsite in EV71-2A^{pro}, as its side chain can pack against the aromatic Tyr side chains of the ^{87}SEYYP^{91} motif as well as the catalytic H21.

We have validated our strategy on the multifunctional EV71-2A^{pro}. The first step of our strategy led to the discovery of a unique loop region (^{87}SEYYP^{91}) near the substrate-binding site that is conserved in enterovirus 2A proteases but is not found with the same secondary structure in human protease sequences to-date. By interacting with the unique viral ^{87}SEYYP^{91} structural motif, a peptide inhibitor would be expected to minimize toxicity in human. On the other hand, it may also target other viral proteases that share nearly identical structure with EV71-2A^{pro}; e.g., the CVA16-2A^{pro} and EV71-2A^{pro} structures differ by a C$^{\alpha}$ RMSD of only 0.6 Å (see Figure 3c). The second step of our strategy led to the discovery of four peptides that were predicted and subsequently experimentally confirmed to compete with the substrate in binding to the target region. One of the four peptide candidates (GFRGKF) inhibited EV71-2A^{pro} protease activity (IC_{50} = 12.1 μM) better than the furoquinoline alkaloid derivative, CW-33 (IC_{50} ~ 53 μM, see Introduction).[18] Note that the human rhinovirus-2A^{pro} LVLQTM cannot be compared with our pseudosubstrate peptide as its structure in complex with EV71-2A^{pro} has not been solved so its binding mode is unclear. Our designed EV71-2A^{pro} peptide, by binding to the EV71-2A^{pro} active site and inhibiting protease activity, may not only reduce viral load but also unleash antiviral cellular responses by blocking EV71-2A^{pro}-mediated cleavage of antiviral signaling proteins and interferon receptor 1 (see Introduction).[7,50] This would allow our pseudosubstrate peptide to be combined with type 1 interferon or other broad-spectrum antivirals to achieve higher antiviral potency and barrier to drug resistance.[18] Furthermore, it could be converted into an inhibitor that will not be hydrolyzed by EV71-2A^{pro} by modifying the scissile bond into an uncleavable one using a keto-methlyene (−CH_2−C═O−)[51] or an alkyne (−CH_2−NH−) group.[52]

A limitation of our drug design strategy is the assumption that the single-point mutation of the native substrate significantly alters neither the configurational entropy nor the binding mode seen in the protein−peptide crystal structure; i.e., the binding mode of the designed peptide is similar to that of the native substrate. Furthermore, reliable predictions are restricted to peptide residues whose backbone atoms have been well-resolved in the starting protein−peptide complex structure. On the other hand, the computational protocol presented herein can be improved to include double or triple mutations depending on the system and computing resources. Although our peptide design strategy has been applied herein to determine mutants that can bind more tightly than the substrate to an enzyme, it may also be applied to determine mutants that bind more weakly than the native peptide to a target protein to avoid digestion and prolong the peptide's lifetime.[20]

In conclusion, we have described an efficient strategy to design peptides that can bind specifically to a "unique" region of a viral drug target. We have shown how our strategy can be applied to EV71-2A^{pro} to identify pseudosubstrate peptides, which would exhibit minimal off-target effects against host proteases by interacting with an "unique" viral region but may also target the other key HFMD pathogen, CVA16-2A^{pro}, whose active-site structure is virtually identical to the EV71-2A^{pro} one.

AUTHOR INFORMATION

Corresponding Authors

Carmay Lim − *Institute of Biomedical Sciences, Academia Sinica, Taipei 115, Taiwan; Department of Chemistry, National Tsing Hua University, Hsinchu 300, Taiwan;* orcid.org/0000-0001-9077-7769; Email: carmay@gate.sinica.edu.tw

Hanna S. Yuan − *Institute of Molecular Biology, Academia Sinica, Taipei 115, Taiwan*; Email: hanna@sinica.edu.tw

Authors

Ting Chen − *Institute of Biomedical Sciences, Academia Sinica, Taipei 115, Taiwan;* orcid.org/0000-0001-8055-3104

Cédric Grauffel − *Institute of Biomedical Sciences, Academia Sinica, Taipei 115, Taiwan;* orcid.org/0000-0002-9478-4699

Wei-Zen Yang − *Institute of Molecular Biology, Academia Sinica, Taipei 115, Taiwan*

Yi-Ping Chen − *Institute of Molecular Biology, Academia Sinica, Taipei 115, Taiwan*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsbiomedchemau.2c00001

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank the staff members of TPS beamline BL05A in the National Synchrotron Radiation Research Center, Hsin-Chu, Taiwan, a national user facility supported by the Ministry of Science and Technology of Taiwan. The Synchrotron Radiation Protein Crystallography Facility is supported by

the National Core Facility Program for Biotechnology. We also acknowledge the Biophysics Core of the Institute of Molecular Biology for the work on fluorescence-based assays. This work was supported by the Ministry of Science & Technology (MOST-107-2113-M-001-018 to C.L.) and Academia Sinica (AS-IA-107-L03 to C. L. and AS-IA-110-L02 to H.S.Y.) Taiwan.

REFERENCES

(1) Nikonov, O. S.; Chernykh, E. S.; Garber, M. B.; Nikonova, E. Y. Enteroviruses: Classification, diseases they cause, and approaches to development of antiviral drugs. *Biochemistry (Moscow)* **2017**, *82* (13), 1615−1631.
(2) Lin, J. Y.; Kung, Y. A.; Shih, S. R. Antivirals and vaccines for Enterovirus A71. *J. Biomed. Sci.* **2019**, *26*, 65.
(3) Wang, S. M.; Liu, C. C.; Tseng, H. W.; Wang, J. R.; Huang, C. C.; Chen, Y. J.; Yang, Y. J.; Lin, S. J.; Yeh, T. F. Clinical Spectrum of Enterovirus 71 Infection in Children in Southern Taiwan, With an Emphasis on Neurological Complications. *Clin. Infect. Dis.* **1999**, *29* (1), 184−190.
(4) Ryu, W. S.; Kang, B.; Hong, J.; Hwang, S.; Kim, A.; Kim, J.; Cheon, D. S. Enterovirus 71 Infection With Central Nervous System Involvement, South Korea. *Emerg. Infect. Dis.* **2010**, *16* (11), 1764−1766.
(5) Li, R.; Liu, L.; Mo, Z.; Wang, X.; Xia, J.; Liang, Z.; Zhang, Y.; Li, Y.; Mao, Q.; Wang, J.; et al. An inactivated enterovirus 71 vaccine in healthy children. *N. Eng. J. Med.* **2014**, *370* (9), 829−837.
(6) Mao, Q.; Wang, Y.; Bian, L.; Xu, M.; Liang, Z. EV-A71 vaccine licensure: a first step for multivalent enterovirus vaccine to control HFMD and other severe diseases. *Emerg. Microbes Infect.* **2016**, *5* (1), 1−7.
(7) Yi, E.-J.; Shin, Y.-J.; Kim, J.-H.; Kim, T.-G.; Chang, S.-Y. Enterovirus 71 infection and vaccines. *Clin. Exp. Vaccine Res.* **2017**, *6*, 4−14.
(8) Baggen, J.; Thibaut, H. J.; Strating, J. R.; van Kuppeveld, F. J. The life cycle of non-polio enteroviruses and how to target it. *Nat. Rev. Microbiol.* **2018**, *16* (6), 368−381.
(9) Martinez-Gualda, B.; Sun, L.; Martí-Marí, O.; Noppen, S.; Abdelnabi, R.; Bator, C. M.; Quesada, E.; Delang, L.; Mirabelli, C.; Lee, H.; et al. Scaffold simplification strategy leads to a novel generation of dual human immunodeficiency virus and enterovirus-A71 entry inhibitors. *J. Med. Chem.* **2020**, *63* (1), 349−368.
(10) Bedard, K. M.; Semler, B. L. Regulation of picornavirus gene expression. *Microbes Infect.* **2004**, *6* (7), 702−713.
(11) Anasir, M. I.; Zarif, F.; Poh, C. L. Antivirals blocking entry of enteroviruses and therapeutic potential. *J. Biomed. Sci.* **2021**, *28*, 10.
(12) Gradi, A.; Svitkin, Y. V.; Imataka, H.; Sonenberg, N. Proteolysis of human eukaryotic translation initiation factor eIF4GII, but not eIF4GI, coincides with the shutoff of host protein synthesis after poliovirus infection. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 11089−11094.
(13) Lin, J.-Y.; Chen, T.-C.; Weng, K.-F.; Chang, S.-C.; Chen, L.-L.; Shih, S.-R. Viral and host proteins involved in picornavirus life cycle. *J. Biomed. Sci.* **2009**, *16*, 103.
(14) Lu, J.; Yi, L.; Zhao, J.; Yu, J.; Chen, Y.; Lin, M. C.; Kung, H. F.; He, M. L. Enterovirus 71 Disrupts Interferon Signaling by Reducing the Level of Interferon Receptor 1. *J. Virol.* **2012**, *86* (7), 3767−3776.
(15) Pathinayake, P.; Hsu, A.; Wark, P. Innate immunity and immune evasion by enterovirus 71. *Viruses* **2015**, *7*, 6613−6630.
(16) Wang, J.; Hu, Y.; Zheng, M. Enterovirus A71 antivirals: Past, present, and future. *Acta Pharm. Sin. B* **2021**, DOI: 10.1016/j.apsb.2021.08.017.
(17) Falah, N.; Montserret, R.; Lelogeais, V.; Schuffenecker, I.; Lina, B.; Cortay, J.-C.; Violot, S. b. Blocking human enterovirus 71 replication by targeting viral 2A protease. *J. Antimicrob. Chemother.* **2012**, *67*, 2865−2869.
(18) Wang, C.-Y.; Huang, A.-C.; Hour, M.-J.; Huang, S.-H.; Kung, S.-H.; Chen, C.-H.; Chen, I.; Chang, Y.-S.; Lien, J.-C.; Lin, C.-W. Antiviral potential of a novel compound CW-33 against enterovirus A71 via inhibition of viral 2A protease. *Viruses* **2015**, *7* (6), 3155−3171.
(19) Guidotti, G.; Brambilla, L.; Rossi, D. Cell-penetrating peptides: from basic research to clinics. *Trends Pharmacol. Sci.* **2017**, *38*, 406−424.
(20) Drucker, D. J. Advances in oral peptide therapeutics. *Nat. Rev. Drug Discovery* **2020**, *19*, 277.
(21) Marchler-Bauer, A.; Derbyshire, M. K.; Gonzales, N. R.; Lu, S.; Chitsaz, F.; Geer, L. Y.; Geer, R. C.; He, J.; Gwadz, M.; Hurwitz, D. I.; et al. CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **2015**, *43* (D1), D222−D226.
(22) Thompson, J. D.; Higgins, D. G.; Gibson, T. J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignments through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **1994**, *22*, 4673−4680.
(23) Berman, H.; Henrick, K.; Nakamura, H. Announcing the worldwide Protein Data Bank. *Nat. Struct. Mol. Biol.* **2003**, *10* (12), 980.
(24) Roy, A.; Yang, J.; Zhang, Y. COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. *Nucleic Acids Res.* **2012**, *40* (W1), W471−W477.
(25) Zhang, C.; Freddolino, P. L.; Zhang, Y. COFACTOR: improved protein function prediction by combining structure, sequence and protein-protein interaction information. *Nucleic Acids Res.* **2017**, *45* (W1), W291−W299.
(26) Ye, Y.; Godzik, A. FATCAT: a web server for flexible structure comparison and structure similarity searching. *Nucleic Acids Res.* **2004**, *32*, W582−W585.
(27) Webb, B.; Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinf.* **2016**, *54* (1), 5.6.1−5.6.37.
(28) Krivov, G. G.; Shapovalov, M. V.; Dunbrack, R. L., Jr Improved prediction of protein side-chain conformations with SCWRL4. *Proteins* **2009**, *77* (4), 778−95.
(29) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M. R.; Jensen, J. H. PROPKA3: Consistent treatment of internal and surface residues in empirical pKa predictions. *J. Chem. Theory & Comput.* **2011**, *7* (2), 525−537.
(30) Brooks, B. R.; Brooks III, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545−1615.
(31) Huang, J.; MacKerell, A. D., Jr CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J. Comput. Chem.* **2013**, *34* (25), 2135−2145.
(32) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(33) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26* (16), 1781−1802.
(34) Andersen, H. C. RATTLE: a "velocity" version of the SHAKE algorithm for molecular dynamics calculations. *J. Comput. Phys.* **1983**, *52*, 24−34.
(35) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577−8592.
(36) Homeyer, N.; Gohlke, H. Free energy calculations by the molecular mechanics Poisson- Boltzmann surface area method. *Mol. Inform.* **2012**, *31*, 114−122.
(37) Wright, J. D.; Chu, H.-M.; Huang, C.-H.; Ma, C.; Chang, T. W.; Lim, C. Structural and physical basis for anti-IgE therapy. *Sci. Rep.* **2015**, *5*, 11581.

(38) Wang, E.; Sun, H.; Wang, J.; Wang, Z.; Liu, H.; Zhang, J. Z.; Hou, T. End-point binding free energy calculation with MM/PBSA and MM/GBSA: Strategies and applications in drug design. *Chem. Rev.* **2019**, *119*, 9478−9508.
(39) Forouzesh, N.; Mishra, N. An Effective MM/GBSA Protocol for Absolute Binding Free Energy Calculations: A Case Study on SARS-CoV-2 Spike Protein and the Human ACE2 Receptor. *Molecules* **2021**, *26*, 2383.
(40) King, E.; Aitchison, E.; Li, H.; Luo, R. Recent Developments in Free Energy Calculations for Drug Discovery. *Front. Mol. Biosci.* **2021**, *8*, 712085.
(41) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham III, T. E. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33*, 889−897.
(42) Cournia, Z.; Allen, B.; Sherman, W. Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. *J. Chem. Inf. Model.* **2017**, *57*, 2911−2937.
(43) Baker, D.; Sali, A. Protein structure prediction and structural genomics. *Science* **2001**, *294* (5540), 93−96.
(44) Sun, Y.; Wang, X.; Yuan, S.; Dang, M.; Li, X.; Zhang, X. C.; Rao, Z. An open conformation determined by a structural switch for 2A protease from coxsackievirus A16. *Protein Cell* **2013**, *4*, 782−792.
(45) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Zidek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583−589.
(46) Cai, Q.; Yameen, M.; Liu, W.; Gao, Z.; Li, Y.; Peng, X.; Cai, Y.; Wu, C.; Zheng, Q.; Li, J.; Lin, T. Conformational Plasticity of the 2A Proteinase from Enterovirus 71. *J. Virol.* **2013**, *87* (13), 7348−7356.
(47) Falah, N.; Violot, S.; Décimo, D.; Berri, F.; Foucault-Grunenwald, M.-L.; Ohlmann, T.; Schuffenecker, I.; Morfin, F.; Lina, B.; Riteau, B.; et al. Ex vivo and in vivo inhibition of human rhinovirus replication by a new pseudosubstrate of viral 2A protease. *J. Virol.* **2012**, *86* (2), 691−704.
(48) Maghsoudi, N.; Tafreshi, N. K.; Khodagholi, F.; Zakeri, Z.; Esfandiarei, M.; Hadi-Alijanvand, H.; Sabbaghian, M.; Maghsoudi, A. H.; Sajadi, M.; Zohri, M.; et al. Targeting enteroviral 2A protease by a 16-mer synthetic peptide: inhibition of 2Apro-induced apoptosis in a stable Tet-on HeLa cell line. *Virology* **2010**, *399* (1), 39−45.
(49) Yang, X.; Wang, Z.; Xiang, Z.; Li, D.; Hu, Z.; Cui, W.; Geng, L.; Fang, Q. Peptide probes derived from pertuzumab by molecular dynamics modeling for HER2 positive tumor imaging. *PLoS Comput. Biol.* **2017**, *13* (4), No. e1005441.
(50) Kuo, R.-L.; Shih, S.-R. Strategies to develop antivirals against enterovirus 71. *Virol. J.* **2013**, *10*, 28.
(51) Budnjo, A.; Narawane, S.; Grauffel, C.; Schillinger, A. S.; Fossen, T.; Reuter, N.; Haug, B. E. Reversible ketomethylene-based inhibitors of human neutrophil proteinase 3. *J. Med. Chem.* **2014**, *57* (22), 9396−408.
(52) Van Kersavond, T.; Konopatzki, R.; Chakrabarty, S.; Blank-Landeshammer, B.; Sickmann, A.; Verhelst, S. H. L. Short Peptides with Uncleavable Peptide Bond Mimetics as Photoactivatable Caspase-3 Inhibitors. *Molecules* **2019**, *24*, 206.

https://doi.org/10.1021/acsbiomedchemau.2c00001
ACS Bio Med Chem Au 2022, 2, 437−449