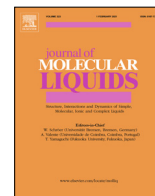




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



# Molecular basis of the new COVID-19 target neuropilin-1 in complex with SARS-CoV-2 S1 C-end rule peptide and small-molecule antagonists



Methus Klaewkla<sup>a,1</sup>, Thanapon Charoenwongpaiboon<sup>b,1</sup>, Panupong Mahalapbutr<sup>c,\*</sup>

<sup>a</sup>Structural and Computational Biology Research Unit, Department of Biochemistry, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand

<sup>b</sup>Department of Chemistry, Faculty of Science, Silpakorn University, Nakhon Pathom 73000, Thailand

<sup>c</sup>Department of Biochemistry, Faculty of Medicine, Khon Kaen University, Khon Kaen 40002, Thailand

## ARTICLE INFO

### Article history:

Received 29 January 2021

Revised 11 May 2021

Accepted 17 May 2021

Available online 20 May 2021

### Keywords:

COVID-19

Neuropilin-1

SARS-CoV-2 S1 CendR

EG00229

EG01377

## ABSTRACT

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), responsible for causing the current coronavirus 2019 (COVID-19) pandemic, uses its spike (S1) protein for host cell attachment and entry. Apart from angiotensin-converting enzyme 2, neuropilin-1 (NRP1) has been recently found to serve as another host factor for SARS-CoV-2 infection; thus, blocking S1–NRP1 interaction can be a potential treatment for COVID-19. Herein, molecular recognition between SARS-CoV-2 S1 C-end rule (CendR) heptapeptide including small-molecule antagonists (EG00229 and EG01377) and the NRP1 was investigated using molecular dynamics simulations and binding free energy calculations based on MM-PBSA method. The binding affinity and the number of hot-spot residues of EG01377/NRP1 complex were higher than those of CendR/NRP1 and EG00229/NRP1 systems, in line with the reported experimental data as well as with the lower water accessibility at the ligand-binding site. The (i) T316, P317, and D320 and (ii) S346, T349, and Y353 residues of NRP1 were confirmed to respectively form H-bonds with the positively charged guanidinium group and the negatively charged carboxyl moiety of all studied ligands. Moreover, Rosetta protein design was employed to improve the binding affinity between CendR peptide and NRP1. The newly designed peptides, especially R683G and A684M, exhibited higher binding efficiency than the native CendR heptapeptide as well as the small-molecule EG00229 by forming more H-bonds and hydrophobic interactions with NRP1, suggesting that these designed peptides could be promising NRP1 inhibitors to combat SARS-CoV-2 infection.

© 2021 Elsevier B.V. All rights reserved.

## 1. Introduction

The expanding of coronavirus disease 2019 (COVID-19) pandemic has had an unexpected impact on current society and humanity. In May 2021, the world health organization (WHO) declared over 167 million infections and 3.4 million deaths globally (<https://covid19.who.int>). Researchers have been attempted to develop an antiviral drug or a vaccine for controlling and reducing the COVID-19 cases, but now reliable pharmaceutical interventions against this disease have not been reported yet.

The COVID-19 is caused by a new coronavirus, namely severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [1–3]. Patients infected with SARS-CoV-2 may present mild to severe symptoms, including fever, cough, sore throat, tiredness, chest

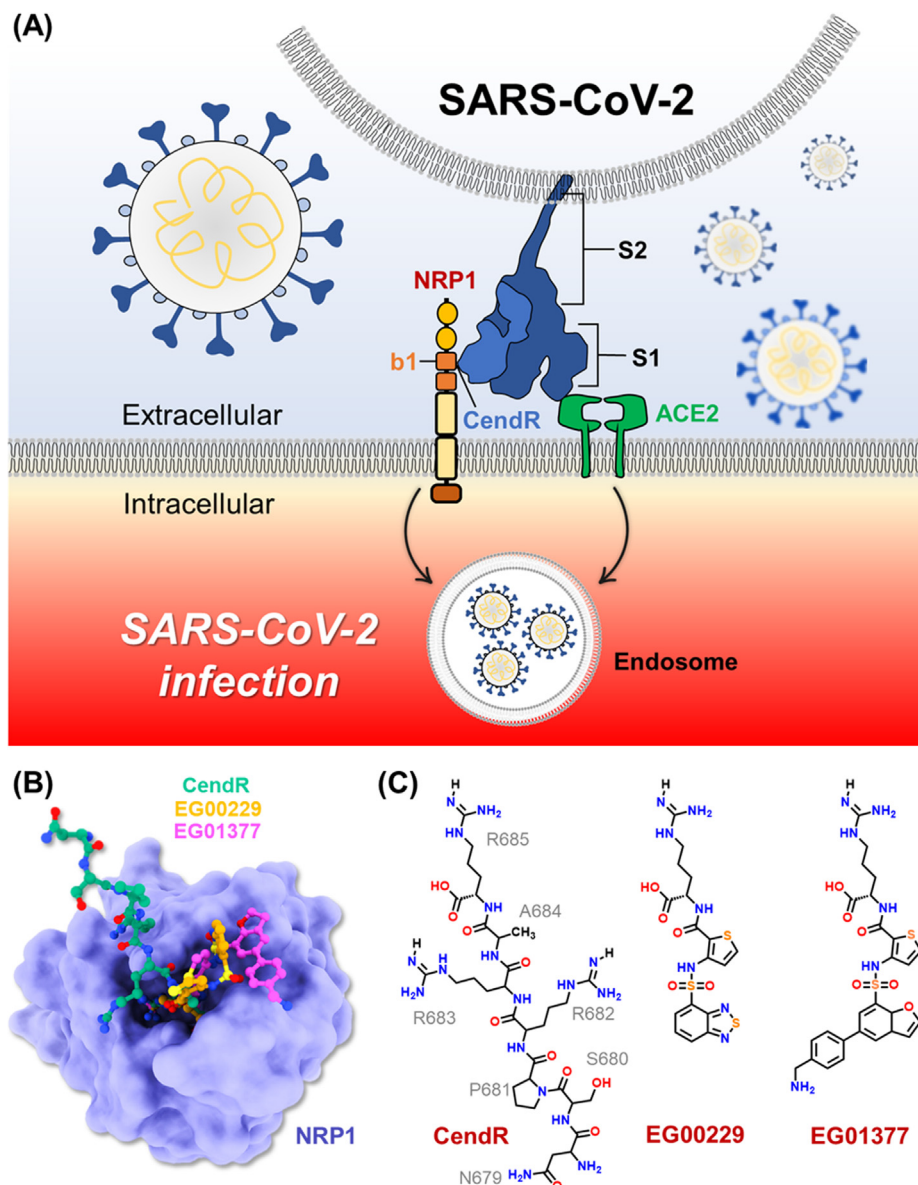
pain, and severe pneumonia [4,5]. Although both SARS-CoV-1 and SARS-CoV-2 are originated from bats, the human-to-human transmission of SARS-CoV-2 is easier than that of SARS-CoV-1 [1,6]. Structurally, SARS-CoV-2 consists of four major proteins, including the envelope (E), membrane (M), nucleocapsid (N), and spike (S) proteins [7,8]. The S protein of SARS-CoV-2 comprises S1 and S2 subunits that are responsible for the host cell attachment and entry. The S1 subunit of SARS-CoV-2 S protein directly binds to the angiotensin-converting enzyme 2 (ACE2) receptor of human cells, facilitating the viral infection [9–11]. Till now, many studies have been attempted to discover and develop the molecules that specifically disrupt the protein–protein interaction between S1 and ACE2 for preventing SARS-CoV-2 infection [12–14].

Recent studies have shown that SARS-CoV-2 S1 protein not only binds to the ACE2 but also the neuropilin-1 (NRP1) receptor (Fig. 1A) [8,15]. Interestingly, suppression of NRP1 expression in ACE2-expressing cells significantly reduced SARS-CoV-2 infection [16]. Based on X-ray crystal structure, the SARS-CoV-2 S1 protein

\* Corresponding author.

E-mail address: [panupma@kku.ac.th](mailto:panupma@kku.ac.th) (P. Mahalapbutr).

<sup>1</sup> These authors contributed equally to this work.



**Fig. 1.** (A) Cartoon representation of SARS-CoV-2 infection. (B) Superimposed structures of NRP1 b1 domain in complex with CendR heptapeptide (PDB entry 7JJC) [19], EG00229 (PDB entry 3I97) [19], and EG01377 (PDB entry 6FMF) [20]. (C) Chemical structure of all studied ligands.

binds to NRP1 receptor via its Arg-Arg-Ala-Arg C-terminal sequence ( $^{682}\text{RRAR}^{685}$ ), called “C-end rule (CendR)” motif [16]. Site-directed mutagenesis indicated that C-terminal R685 of S1 protein is crucial for the binding of NRP1 [16]. Moreover, blocking this interaction using selective inhibitor, EG00229, can inhibit the viral infection, suggesting that S1–NRP1 interaction serves as a potential anti-SARS-CoV-2 target. Therefore, in this study, we aimed to investigate the binding pattern and susceptibility of SARS-CoV-2 S1 CendR heptapeptide as well as small-molecule inhibitors (EG00229 and EG01377; Fig. 1C) [17,18] against NRP1 receptor using all-atom molecular dynamics (MD) simulations and free energy calculations. Moreover, computational protein design with Rosetta was employed to improve the binding affinity between the SARS-CoV-2 S1 CendR and the NRP1 b1 domain. We hope that the information obtained from this study could be useful for the development of more effective drugs targeting NRP1 in the fight against COVID-19.

## 2. Computational methods

### 2.1. System preparation and molecular dynamics simulations

The crystal structures of human NRP1 b1 domain in complex with SARS-CoV-2 S1 CendR heptapeptide (PDB ID: 7JJC, chain C) [16], EG00229 (PDB ID: 3I97, chain A) [19], and EG01377 (PDB ID: 6FMF, chain A) [21] were obtained from the RSCB Protein Data Bank. The H++ server [22] was used to determine the protonation state of all amino acid residues at pH 7.4. The electrostatic potential (ESP) charges of EG00229 and EG01377 were calculated with the HF/6-31G\* level of theory using Gaussian09 program [23]. The antechamber module of AMBER20 was employed to generate the restrained ESP (RESP) charges of the two small-molecule inhibitors. The AMBER ff14SB force field was applied for the protein, while other parameters were obtained from the generalized AMBER force field version 2 (GAFF2) [24]. Remaining missing

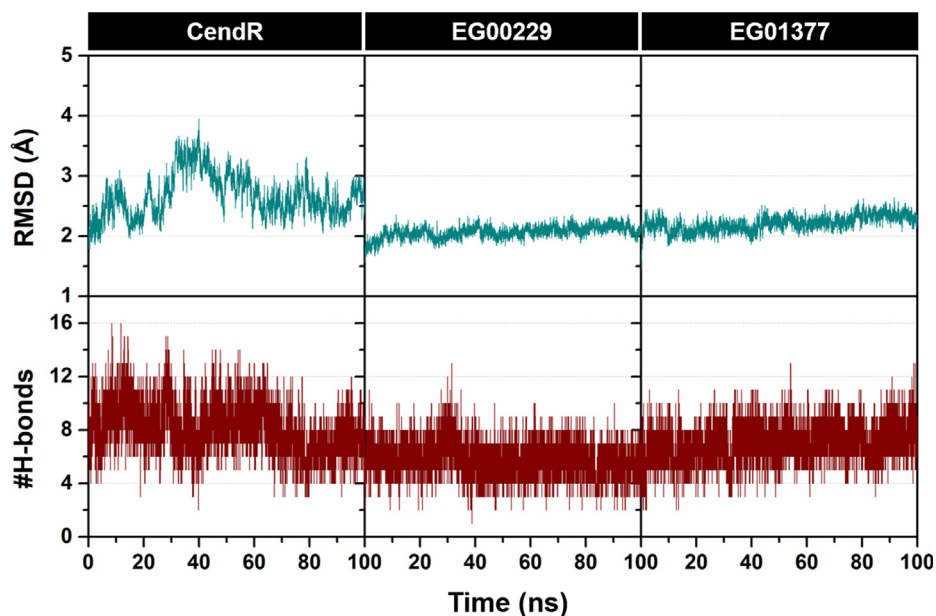


Fig. 2. Time evolution of (top) all-atom RMSD and (bottom) #H-bonds of CendR heptapeptide (left), EG00229 (middle), and EG01377 (right) in complex with NRP1.

Table 1

Average  $\Delta G_{\text{bind}}$  and its energy components (kcal/mol) of CendR, EG00229, and EG01377 in complex with NRP1 calculated with the MM-PBSA method. Data are shown as mean  $\pm$  standard error of mean.

Energy component (kcal/mol)	CendR	EG00229	EG01377
<b>Gas Term</b>			
$\Delta E_{\text{vdW}}$	$-23.81 \pm 0.18$	$-25.60 \pm 0.14$	$-31.29 \pm 0.15$
$\Delta E_{\text{ele}}$	$-328.35 \pm 1.97$	$-203.22 \pm 0.67$	$-223.28 \pm 0.70$
$\Delta E_{\text{MM}}$	$-352.16 \pm 2.00$	$-228.82 \pm 0.65$	$-254.58 \pm 0.70$
$-T\Delta S$	$35.91 \pm 1.07$	$27.71 \pm 0.82$	$33.51 \pm 0.57$
<b>Solvation Term</b>			
$\Delta G_{\text{solv}}^{\text{polar}}$	$314.75 \pm 1.83$	$197.10 \pm 0.55$	$213.56 \pm 0.56$
$\Delta G_{\text{solv}}^{\text{nonpolar}}$	$-5.91 \pm 0.02$	$-4.76 \pm 0.01$	$-5.72 \pm 0.01$
$\Delta G_{\text{solv}}$	$308.84 \pm 1.82$	$192.34 \pm 0.54$	$207.84 \pm 0.56$
<b>Binding Free Energy</b>			
$\Delta G_{\text{bind}}$	$-7.41 \pm 0.34$	$-8.76 \pm 0.24$	$-13.23 \pm 0.28$
$K_{\text{d}}$ ( $\mu\text{M}$ )	20.3 [41]	5.1 [41]	1.3 [20]

parameters were generated by using parmchk2 modules. LEaP module in AMBER20 was used to add missing hydrogen atoms and solvate each complex in an isomeric truncated octahedral box of TIP3P water model [25] with a buffer distance of 13 Å. Chloride ions were added to neutralize the systems. The systems were then minimized to remove unfavorable interactions. Each procedure composed of steepest descent (SD) 2500 steps followed by conjugated gradient (CG) 2500 steps, in which the heavy atoms of proteins were first restrained with the force constant of  $5.0 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ . Then, the protein backbones were restrained with the force constants of 10, 5, and  $1 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ , respectively. At last, the systems were fully minimized without any restraint. After that, the systems were simulated under the periodic boundary condition [26–28] with a simulation time step of 0.002 ps using PMEMD module. The SHAKE algorithm [29] was used to constrain all bonds involving hydrogen atoms. The temperatures were controlled using Langevin dynamic technique [30] with the collision frequency of  $1 \text{ ps}^{-1}$ . For the heating step, each system was gradually heated from 0 to 310 K for 200 ps, while the protein backbones were restrained with the force constant of  $10 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$  in the NVT ensemble. In the equilibrating step, the systems were simulated at 310 K for 300 ps in the NVT ensemble. After that, the systems were simulated at 310 K and 1 atm in the NPT ensemble until reaching 100 ns.

## 2.2. Analysis

The root-mean-square displacement (RMSD) of complex and the number of hydrogen bonds (#H-bonds) between protein and ligand were calculated using cpptraj module [31] of AMBER20. The H-bond interactions were calculated using the following criteria: (i) distance between the H-bond donor (HD) and H-bond acceptor (HA)  $\leq 3.0 \text{ \AA}$  and (ii) the angle of HD–H...HA  $\geq 135^\circ$ . The solvent-accessible surface area (SASA) was calculated using the NRP1 residues within 5 Å of small-molecule antagonists as solvent-exposed area. All energy calculations were performed using MMPBSA.py module [32] of AMBER20. The total binding free energy ( $\Delta G_{\text{bind}}$ ) and per-residue decomposition free energy ( $\Delta G_{\text{bind}}^{\text{residue}}$ ) were calculated using molecular mechanics-Poisson-Boltzmann surface

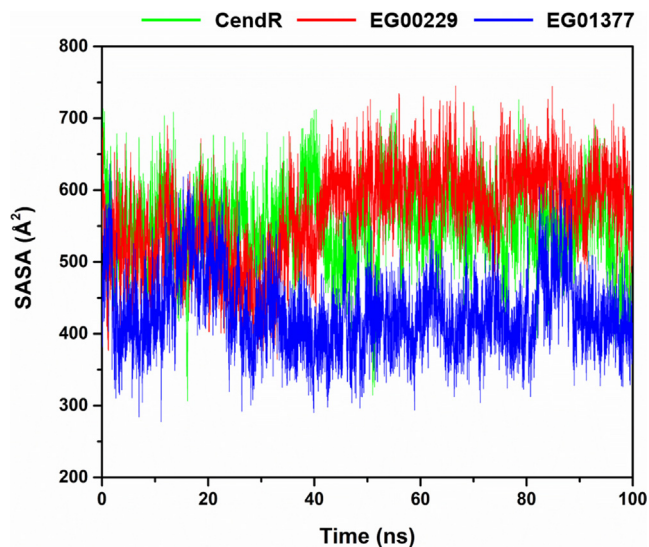
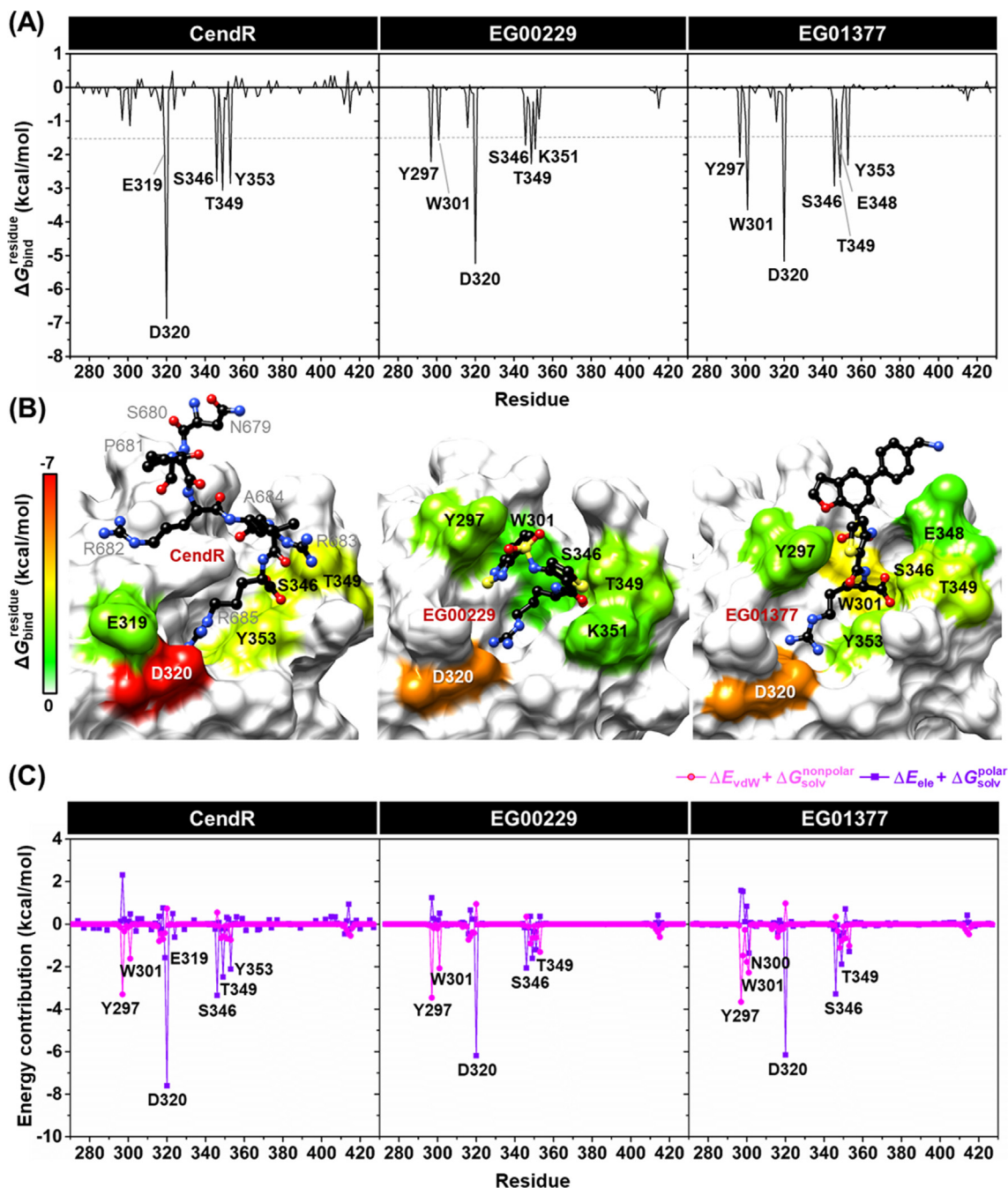


Fig. 3. SASA plots along the simulation time of the three studied systems.



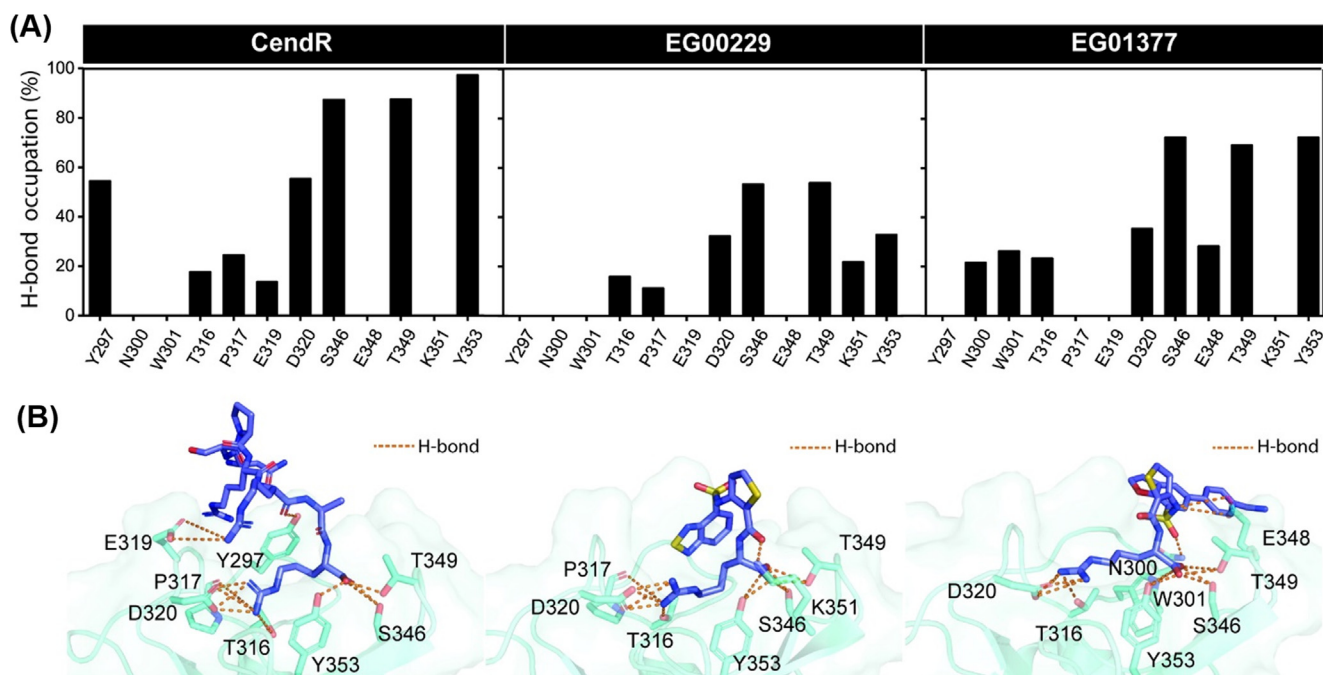
**Fig. 4.** (A)  $\Delta G_{\text{bind}}^{\text{residue}}$  of CendR heptapeptide (left), EG00229 (middle), and EG01377 (right) in complex with NRP1 b1 coagulation factor domain. (B) Representative structures showing the ligand orientation in NRP1 b1 domain drawn from the last MD snapshot. The contributing residues of NRP1 involved in the binding of all studied ligands are colored according to their  $\Delta G_{\text{bind}}^{\text{residue}}$  values, where the highest to lowest free energies are shaded from white to red, respectively. (C) Electrostatic ( $\Delta E_{\text{ele}} + \Delta G_{\text{solV}}^{\text{polar}}$ , purple) and vdW ( $\Delta E_{\text{vdW}} + \Delta G_{\text{solV}}^{\text{nonpolar}}$ , magenta) energy contributions from each residue of NRP1 b1 domain to the binding of CendR (left), EG00229 (middle), and EG01377 (right).

area (MM-PBSA) method [33] on 100 frames taken from the production phase.

### 2.3. Peptide design

Rosetta 3.12 [34] was employed with the beta\_nov16 energy function to design new mutants of SARS-CoV-2 S1 CendR peptide. The FastDesign protocol [35,36] based on the FastRelax was used in this study. This protocol not only designed target residues but also attempted to optimize structure by allowing the movement of sidechains and backbones. The crystal structure of NRP1/SARS-

CoV-2 S1 CendR complex was used as a template. To design peptides with improved binding affinity, positions 683 and 684 of CendR were selected as target residues due to their locations that were close to the arginine-binding cleft and could form potential interactions with the NRP1 protein. In each run, the residues within 5 Å of the target residue were allowed to move. The 50 independent runs [37] of design were performed on each position. The 50 independent runs without design were also performed and used as a reference for further analysis. The free energy ( $\Delta G$ ) in Rosetta was calculated in Rosetta Energy Units (REU). Each of the designed peptides with the lowest free energy was selected. The binding free



**Fig. 5.** (A) Percentage of H-bond occupation of NRP1 contributing to the binding of (left) CendR heptapeptide, (middle) EG00229, and (right) EG01377. (B) Ligand orientation in the NRP1 b1 domain showing H-bond formations between protein and ligand(s). Orange dashed line indicates H-bond formation.

energy ( $\Delta G_{\text{bind(Rosetta)}}$ ) of each complex was then evaluated using InterfaceAnalyzer module [38] of Rosetta. Finally, these complexes with designed peptides of CendR were subjected to MD simulation.

### 3. Results and discussion

#### 3.1. System stability

The stability of all simulated systems was determined using the calculations of root-mean-square displacement (RMSD) and the number of intermolecular H-bonds (#H-bonds) along the simulation time. The obtained results are plotted and summarized in Fig. 2. We found that the structural flexibility of EG00229/NRP1 and EG01377/NRP1 complexes (RMSD of  $\sim 2.0$  Å) was lower than that of CendR/NRP1 model (RMSD of  $\sim 2.0$ – $3.0$  Å), since the number of degrees of freedom of CendR heptapeptide is higher than that of EG00229 and EG01377. In the case of time evolution of #H-bonds, the CendR system exhibited higher fluctuation than the other two models, especially during the first 80 ns. Notably, #H-bonds of EG01377 system ( $7.32 \pm 1.35$  over the last 20 ns) was higher than that of the CendR ( $6.99 \pm 1.31$ ) and EG00229 ( $5.40 \pm 1.14$ ) models, suggesting that EG01377 exhibited the highest binding affinity with NRP1 (discussed in more detail later).

#### 3.2. Inhibitory efficiency

The binding free energy ( $\Delta G_{\text{bind}}$ ) of all studied ligands toward NRP1 was estimated using MM-PBSA method. As shown in Table 1, the molecular mechanics energy ( $\Delta E_{\text{MM}}$ ) in gas phase revealed that electrostatic interaction ( $\Delta E_{\text{ele}}$ ) is the main force inducing protein–ligand complexation ( $\Delta E_{\text{ele}}$  of  $-328.35 \pm 1.97$ ,  $-203.22 \pm 0.67$ , and  $-223.28 \pm 0.70$  kcal/mol for CendR, EG00229, and EG01377, respectively), which is  $\sim 7$ – $14$ -fold stronger than the van der Waals (vdW) interaction ( $\Delta E_{\text{vdW}}$  of  $-23.81 \pm 0.18$ ,  $-25.60 \pm 0.14$ , and  $31.29 \pm 0.15$  kcal/mol for CendR, EG00229, and EG01377, respectively). This electrostatic-driven complexation is in good agreement with the reported binding of a prototypic CendR peptide,

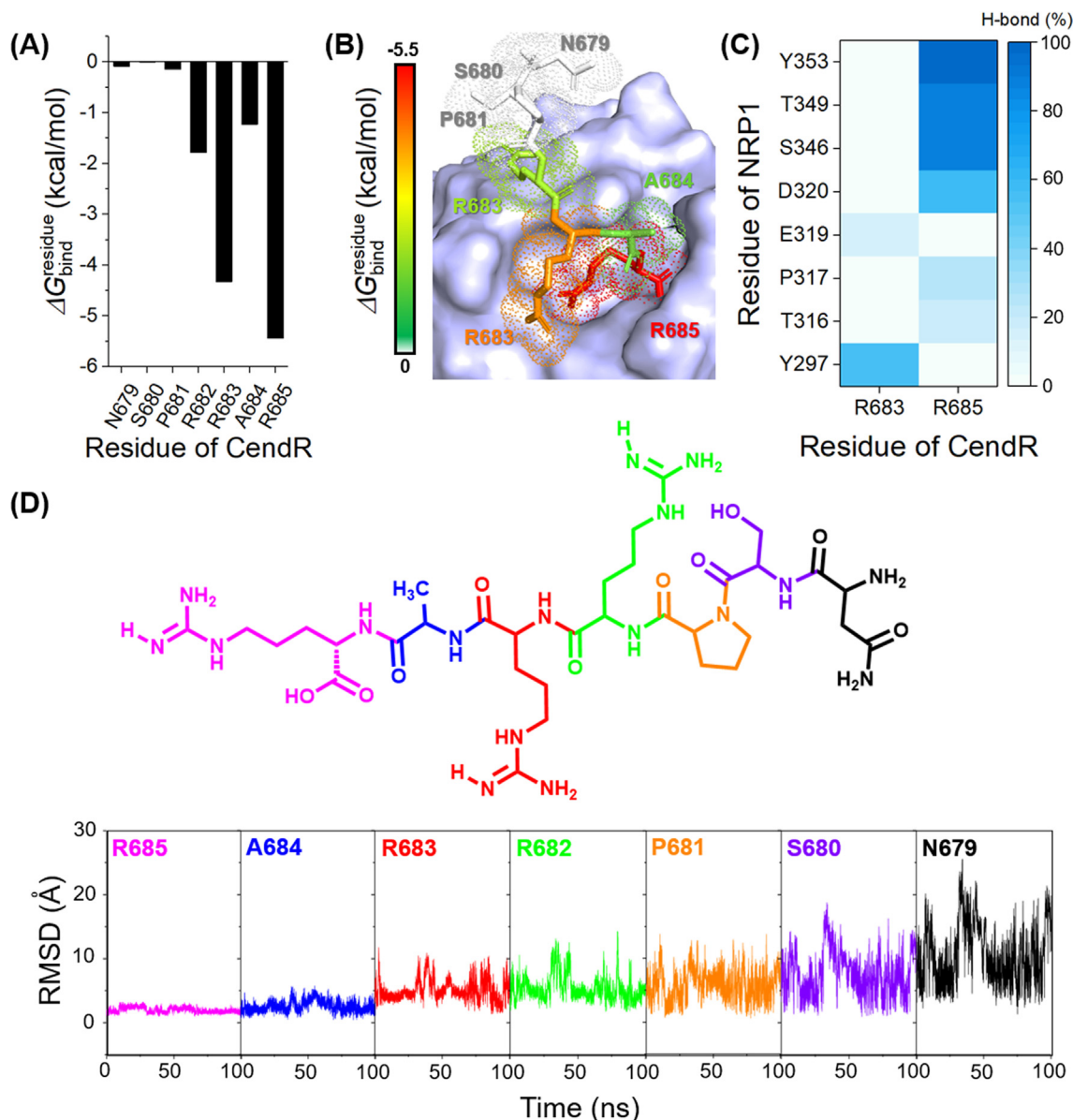
RPAR to NRP1 b1 domain [39] as well as the electrostatic surface of compound 3 (Lys(Har)-GlyΨ[Trl]GlyΨ[Trl]Arg) at the NRP1 b1 domain [40]. The electrostatic contribution of CendR system was higher than that of the two focused small-molecules antagonists, since the CendR has several polar moieties that can form H-bonds with NRP1 (discussed in more details later). Even though the CendR system exhibited the lowest  $\Delta E_{\text{MM}}$  value among the three systems, the high values of entropic term ( $T\Delta S$ ) and solvation free energy ( $\Delta G_{\text{solv}}$ ) increased the total  $\Delta G_{\text{bind}}$ , leading to the low binding affinity with NRP1 b1 domain compared to the other two small-molecule inhibitors.

In the case of  $\Delta G_{\text{bind}}$  results, we found that EG01377 displayed the highest binding affinity with NRP1 ( $\Delta G_{\text{bind}}$  of  $-13.23 \pm 0.28$  kcal/mol) followed by EG00229 ( $-8.76 \pm 0.24$  kcal/mol) and CendR ( $-7.41 \pm 0.34$  kcal/mol), respectively. These results are in line with (i) the reported experimental data showing that the inhibitory activity against NRP1 of EG01377 ( $K_d$  of  $1.3$   $\mu\text{M}$  [20]) is higher than that of EG00229 ( $K_d$  of  $5.1$   $\mu\text{M}$  [41]) and CendR ( $K_d$  of  $20.3$   $\mu\text{M}$  [41]) and (ii) the solvation effect around the ligand-binding site (Fig. 3) demonstrating that the average SASA values of EG01377/NRP1 complex ( $430.62 \pm 51.98$  Å<sup>2</sup>) were lower than that of the other two systems (CendR/NRP1 complex ( $550.76 \pm 0.53$  Å<sup>2</sup>) and EG00229/NRP1 complex ( $566.28 \pm 0.63$  Å<sup>2</sup>)).

#### 3.3. Key binding residues

To investigate hot-spot residues involved in the binding of all studied ligands to NRP1, the  $\Delta G_{\text{bind}}^{\text{residue}}$  calculation based on MM-PBSA method was conducted. The total energy contribution from each residue associated with ligand binding of all studied systems is shown in Fig. 4, where the negative and positive  $\Delta G_{\text{bind}}^{\text{residue}}$  values are energy stabilization and destabilization, respectively. It should be noted that only residues exhibiting the energy stabilization of  $\leq -1.5$  kcal/mol were taken into account.

The obtained results demonstrated that there were five (E319, D320, S346, T349, and Y353), six (Y297, W301, D320, S346, T349, and K351), and seven (Y297, W301, D320, S346, T349, E348, and



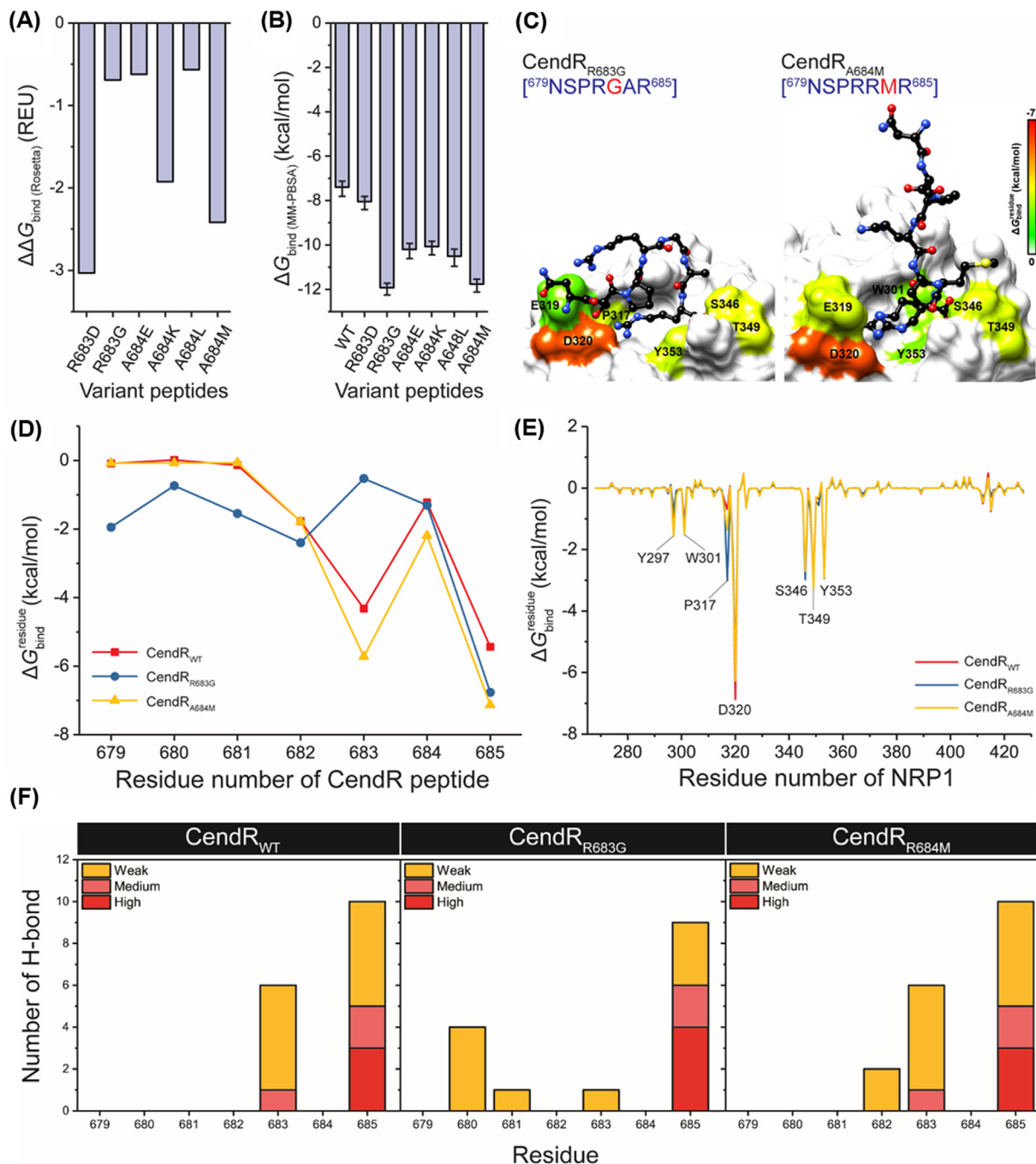
**Fig. 6.** (A)  $\Delta G_{\text{bind}}^{\text{residue}}$  of SARS-CoV-2 S1 CendR toward the binding of NRP1 b1 domain. (B) Representative structures showing the ligand orientation in NRP1 b1 domain drawn from the last MD snapshot. The contributing residues involved in the binding of CendR are colored according to its  $\Delta G_{\text{bind}}^{\text{residue}}$  values, where the highest to lowest free energies are shaded from white to red, respectively. (C) Percentage of H-bond occupation of NRP1 contributing to the binding of CendR peptide. (D) RMSD of each residue of CendR peptide along the simulation time.

Y353) NRP1 residues associated with the binding of CendR, EG00229, and EG01377, respectively. These key binding residues are strongly supported by the alanine mutagenesis showing that Y297A, W301A, D320A, S346A, T349A, K351A, and Y353A result in a complete loss of biotinylated VEGF-A binding to NRP1 [19]. The highest contributing residues found in the EG01377/NRP1 complex was in good agreement with the  $\Delta G_{\text{bind}}$  results (Table 1) as mentioned earlier. Among nine hot-spot residues, D320 showed the lowest  $\Delta G_{\text{bind}}^{\text{residue}}$  for all systems ( $-6.87$ ,  $-5.24$ , and  $-5.17$  kcal/mol for CendR, EG00229, and EG01377, respectively) via electrostatic attraction and H-bond formation with a positively charged guanidinium group of all studied ligands (Fig. 5), suggesting D320 as the most important key binding residue. This result is consistent with the mutational analysis mentioned above as well as the work of J. Daly *et al.* reported that the binding between the NRP1 and the alanine-substituted C-terminal R685 of CendR (R685A), which is the position that interacts with D320 residue of NRP1, was not observed [41].

In terms of the energy contribution from electrostatic ( $\Delta E_{\text{ele}} + \Delta C_{\text{sol}}^{\text{polar}}$ , purple line) and the vdW ( $\Delta E_{\text{vdW}} + \Delta C_{\text{sol}}^{\text{nonpolar}}$ , magenta line) energies from each NRP1 residue (Fig. 4C), we found that the main energy contribution for stabilizing all studied ligands was the electrostatic energy (up to  $\sim -8.0$  kcal/mol), especially for the D320 (electrostatically interacted with the positively charged guanidinium group of ligands) as well as the S346 and T349 (electrostatically interacted with the negatively charged carboxyl group of ligands), whereas the vdW contribution was observed in the range of  $\sim 1.0$  to  $-3.5$  kcal/mol as related to the  $\Delta E_{\text{MM}}$  results (Table 1).

### 3.4. Protein-ligand hydrogen bonding

Since electrostatic attraction was the main force inducing protein-ligand complexations (Table 1), we further investigated the structural insights into the intermolecular H-bond formation between ligand(s) and NRP1 during the last 20 ns of simulation using the defining criteria described in material and method



**Fig. 7.** (A)  $\Delta\Delta G_{\text{bind}}^{\text{(Rosetta)}} \text{ (REU)}$  of all designed peptides obtained from Rosetta. Note that the  $\Delta\Delta G_{\text{bind}}$  was calculated using the following equation:  $\Delta\Delta G_{\text{bind}} = \Delta G_{\text{bind}}^{\text{(mutant)}} - \Delta G_{\text{bind}}^{\text{(wild-type)}}$ . (B)  $\Delta G_{\text{bind}} \text{ (kcal/mol)}$  of all designed peptides obtained from MM-PBSA method. (C) Representative structures of the two most promising peptides CendR<sub>R683G</sub> and CendR<sub>A684M</sub> showing the ligand orientation in NRP1 b1 domain drawn from the last MD snapshot. (D)  $\Delta G_{\text{bind}}^{\text{residue}}$  of CendR<sub>WT</sub>, CendR<sub>R683G</sub>, and CendR<sub>A684M</sub> contributing to the binding of NRP1. (E)  $\Delta G_{\text{bind}}^{\text{residue}}$  of NRP1 contributing to the binding of (left) CendR<sub>WT</sub>, (middle) CendR<sub>R683G</sub>, and (right) CendR<sub>A684M</sub>. (F) Percentage of H-bond occupation of NRP1 contributing to the binding of (left) CendR<sub>WT</sub>, (middle) CendR<sub>R683G</sub>, and (right) CendR<sub>A684M</sub>. %H-bond occupation (%H-bond<sub>oc</sub>) was classified into four levels: (i) strong H-bond (%H-bond<sub>oc</sub> of > 75%), (ii) medium H-bond (75% ≥ %H-bond<sub>oc</sub> > 50%), and (iii) weak H-bond (50% ≥ %H-bond<sub>oc</sub> > 10%).

section. The percentage of H-bond occupation (%H-bond) is plotted in Fig. 5A, whereas the ligand orientation showing H-bond formations is illustrated in Fig. 5B. Note that only NRP1 residues that exhibited %H-bond of > 10% were taken into account.

As expected, all studied ligands formed H-bonds with several polar and charged residues in the NRP1 b1 coagulation factor domain. There were eight (Y297, T316, P317, E319, D320, S346, T349, and Y353), seven (T316, P317, D320, S346, T349, K351, and

Y353), and eight (N300, W301, T316, D320, S346, E348, T349, and Y353) residues responsible for H-bond formations in CendR, EG00229, and EG01377 systems, respectively. The (i) T316, P317, and D320 and (ii) S346, T349, and Y353 residues were found to respectively form H-bonds with the positively charged guanidinium group and the negatively charged carboxyl moiety of all studied ligands, consistent with the crystalized form of NRP1 in complex with CendR heptapeptide [16], EG00229 [19], EG01377



[20], and tuftsin [42]. It should be noted that the number of strong H-bonds (>70% occupancy) of CendR (3 residues: S346, T349, and Y353) was higher than that of EG01377 (2 residues: S346 and Y353) and EG00229 (0 residue) models, respectively, in line with the  $\Delta E_{\text{ele}}$  results (Table 1).

### 3.5. Key binding residues in SARS-CoV-2 S1 CendR peptide

To evaluate the hot-spot residues of CendR peptide upon the binding of NRP1, the  $\Delta G_{\text{bind}}^{\text{residue}}$  and RMSD calculations were conducted. As shown in Fig. 6, the positively charged R685 residue of CendR showed the lowest  $\Delta G_{\text{bind}}^{\text{residue}}$  ( $-5.44$  kcal/mol) and the highest stability (RMSD of  $1.98 \pm 0.41$  Å averaged over 100 ns). The guanidinium group of R685 electrostatically interacted with the negatively charged D320 (Fig. 6B) and formed H-bonds with T316, P317, and D320 residues, while its C-terminal carboxyl group formed H-bonds with S346, T349, and Y353 of NRP1 (Fig. 6C). The R682, R683, and A684 residues of CendR exhibited the  $\Delta G_{\text{bind}}^{\text{residue}}$  of  $\sim -1.0$  to  $-4.0$  kcal/mol and the moderate stability (RMSD of  $5.38 \pm 1.97$  Å,  $5.05 \pm 1.55$  Å, and  $2.57 \pm 0.85$  Å for R682, R683, and A684, respectively), whereas the other residues, including N679, S680, and P681 (located far away from the arginine-binding cleft of NRP1) showed the highest  $\Delta G_{\text{bind}}^{\text{residue}}$  ( $\sim 0$  kcal/mol) and the lowest stability (RMSD of  $10.52 \pm 4.54$ ,  $7.65 \pm 3.32$ , and  $5.91 \pm 2.3$  Å for N679, S680, and P681, respectively). Altogether, the R685 residue of SARS-CoV-2 S1 CendR peptide is suggested as the most important hot-spot residue against NRP1.

### 3.6. Computational design of NRP1 peptide binders

To enhance the binding affinity between the SARS-CoV-2 S1 CendR and the NRP1 b1 domain, computational protein design with Rosetta was employed. The R683 and A684 residues of CendR were selected as target positions, since they located near the arginine-binding pocket of NRP1 and exhibited the moderate stability and binding efficiency against NRP1. The R685 was not selected for peptide design because (i) it showed strong binding affinity with NRP1 and (ii) its guanidinium group was a key functional group for the binding of NRP1 via electrostatic attraction (Figs. 4-6). Rosetta provided six designed peptides, including R683D, R683G, A684E, A684K, A684L, and A684M. Interestingly, the  $\Delta G_{\text{bind}}$  (Rosetta) of all designed peptides was lower than that of the wild-type CendR ( $\Delta \Delta G_{\text{bind}}$  (Rosetta) of  $< 0$ ) (Fig. 7A). To validate this, MD simulations and free energy calculations based on MM-PBSA method were then performed. The obtained results (Fig. 7B-C) revealed that all six designed peptides showed higher binding efficiency than the wild-type CendR. Among them, R683G ( $\Delta G_{\text{bind}}$  of  $-11.94 \pm 0.26$  kcal/mol) and A684M ( $\Delta G_{\text{bind}}$  of  $-11.78 \pm 0.29$  kcal/mol) were the two most promising NRP1 inhibitors, since their predicted binding efficiency is better than the other designed peptides as well as the small-molecule antagonist EG00229. Replacement of arginine by glycine (R683G) may reduce the steric effect on the peptide structure, allowing its N-terminal residues ( $^{679}\text{NSPR}^{682}$ ) to interact with NRP1 (Fig. 7D-E), while replacement of alanine by methionine (A684M) increased the hydrophobic interaction between peptide and NRP1. Notably, both designed peptides also formed more H-bonds with NRP1 residues 680-682 compared to the parent CendR heptapeptide (Fig. 7F). Altogether, these designed peptides are suggested as promising inhibitors of SARS-CoV-2 S1 CendR motif interaction with NRP1

## 4. Conclusion

In this work, the binding mechanism and susceptibility of the SARS-CoV-2 S1 CendR peptide as well as the small-molecule antag-

onists EG00229 and EG01377 in complex with the NRP1 b1 domain were fully revealed by all-atom MD simulations and free energy calculations based on MM-PBSA method. From  $\Delta G_{\text{bind}}$  calculation, the susceptibility against the NRP1 of all studied ligands was ranked in the order of EG01377 > EG00229 > CendR, in line with the reported experimental data. The molecular complexation of all studied systems was driven mainly by electrostatic attraction rather than vdW interaction. Among nine hot-spot residues, D320 showed the lowest  $\Delta G_{\text{bind}}^{\text{residue}}$  via electrostatic attraction and H-bond formation with the positively charged guanidinium group of all studied ligands. According to computational protein design with Rosetta, all six designed CendR peptides (R683D, R683G, A684E, A684K, A684L, and A684M) exhibited higher binding efficiency with NRP1 than the wild-type CendR heptapeptide, and, among them, R683G and A684M are suggested as promising peptide inhibitors of SARS-CoV-2 S1 CendR motif interaction with NRP1. Altogether, the obtained structural and energetic information at the atomic level from this work could be helpful for further design and development of more specific NRP1 inhibitors in the fight against COVID-19.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

We would like to thank the Department of Chemistry, Faculty of Science, Silpakorn University for computing resources.

## Funding

This research was supported by the Invitation Research Fund from the Faculty of Medicine, Khon Kaen University (IN64227) to P.M.

## References

- [1] J.F.-W. Chan, S. Yuan, K.-H. Kok, K.K.-W. To, H. Chu, J. Yang, F. Xing, J. Liu, C.C.-Y. Yip, R.W.-S. Poon, *The Lancet* 395 (2020) 514.
- [2] R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, W. Wang, H. Song, B. Huang, N. Zhu, *The Lancet* 395 (2020) 565.
- [3] B. Nutho, P. Mahalapbutr, K. Hengphasatporn, N.C. Pattarangoon, N. Simanon, Y. Shigeta, S. Hannongbua, T. Rungrotmongkol, *Biochemistry* 59 (2020) 1769.
- [4] N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han, Y. Qiu, J. Wang, Y. Liu, Y. Wei, *The Lancet* 395 (2020) 507.
- [5] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, *The Lancet* 395 (2020) 497.
- [6] M. Hoffmann, H. Kleine-Weber, S. Schroeder, N. Krüger, T. Herrler, S. Erichsen, T.S. Schiergens, G. Herrler, N.-H. Wu, A. Nitsche, *Cell* (2020).
- [7] G. Li, Y. Fan, Y. Lai, T. Han, Z. Li, P. Zhou, P. Pan, W. Wang, D. Hu, X. Liu, *J. Med. Virol.* 92 (2020) 424.
- [8] M. Kielian, *Science* 370 (2020) 765.
- [9] J. Lan, J. Ge, J. Yu, S. Shan, H. Zhou, S. Fan, Q. Zhang, X. Shi, Q. Wang, L. Zhang, *Nature* 581 (2020) 215.
- [10] D.J. Benton, A.G. Wrobel, P. Xu, C. Rouston, S.R. Martin, P.B. Rosenthal, J.J. Skehel, S.J. Gamblin, *Nature* (2020) 1.
- [11] H. Hadi-Alijanvand, M. Rouhani, *J. Proteome Res.* (2020).
- [12] M. Prajapat, N. Shekhar, P. Sarma, P. Avti, S. Singh, H. Kaur, A. Bhattacharyya, S. Kumar, S. Sharma, A. Prakash, *J. Mol. Graph. Model.* 101 (2020) 107716.
- [13] C.G. Benítez-Cardoza, J.L. Vique-Sánchez, *Life Sci.* 256 (2020) 117970.
- [14] Y. Han, P. Král, *ACS nano* 14 (2020) 5143.
- [15] L. Cantuti-Castelvetri, R. Ojha, L.D. Pedro, M. Djannatian, J. Franz, S. Kuivanen, F. van der Meer, K. Kallio, T. Kaya, M. Anastasina, *Science* 370 (2020) 856.
- [16] J.L. Daly, B. Simonetti, K. Klein, K.-E. Chen, M.K. Williamson, C. Antón-Plágaro, D.K. Shoemark, L. Simón-Gracia, M. Bauer, R. Hollandi, U.F. Greber, P. Horvath, R.B. Sessions, A. Helenius, J.A. Hiscox, T. Teesalu, D.A. Matthews, A.D. Davidson, B.M. Collins, P.J. Cullen, Y. Yamauchi, *Science* 370 (2020) 861.
- [17] A. Jarvis, C.K. Allerston, H. Jia, B. Herzog, A. Garza-García, N. Winfield, K. Ellard, R. Aqil, R. Lynch, C. Chapman, *J. Med. Chem.* 53 (2010) 2215.

- [18] J. Powell, F. Mota, D. Steadman, C. Soudy, J.T. Miyauchi, S. Crosby, A. Jarvis, T. Reisinger, N. Winfield, G. Evans, *J. Med. Chem.* 61 (2018) 4135.
- [19] A. Jarvis, C.K. Allerston, H. Jia, B. Herzog, A. Garza-Garcia, N. Winfield, K. Ellard, R. Aqil, R. Lynch, C. Chapman, B. Hartzoulakis, J. Nally, M. Stewart, L. Cheng, M. Menon, M. Tickner, S. Djordjevic, P.C. Driscoll, I. Zachary, D.L. Selwood, *J. Med. Chem.* 53 (2010) 2215.
- [20] J. Powell, F. Mota, D. Steadman, C. Soudy, J.T. Miyauchi, S. Crosby, A. Jarvis, T. Reisinger, N. Winfield, G. Evans, A. Finniear, T. Yelland, Y.-T. Chou, A.W.E. Chan, A. O'Leary, L. Cheng, D. Liu, C. Fotinou, C. Milagre, J.F. Martin, H. Jia, P. Frankel, S. Djordjevic, S.E. Tsirka, I.C. Zachary, D.L. Selwood, *J. Med. Chem.* 61 (2018) 4135.
- [21] J. Powell, F. Mota, D. Steadman, C. Soudy, J.T. Miyauchi, S. Crosby, A. Jarvis, T. Reisinger, N. Winfield, G. Evans, A. Finniear, T. Yelland, Y.-T. Chou, A.W.E. Chan, A. O'Leary, L. Cheng, D. Liu, C. Fotinou, C. Milagre, J.F. Martin, H. Jia, P. Frankel, S. Djordjevic, S.E. Tsirka, I.C. Zachary, D.L. Selwood, *J. Med. Chem.* 61 (2018) 4135.
- [22] R. Anandakrishnan, B. Aguilar, A.V. Onufriev, *Nucleic Acids Res.* 40 (2012) W537.
- [23] M. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. Petersson, Gaussian, Inc., Wallingford CT, 2009.
- [24] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, *J. Comput. Chem.* 25 (2004) 1157.
- [25] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, *The Journal of Chemical Physics* 79 (1983) 926.
- [26] P. Mahalapbutr, M. Sangkhawasi, J. Kammarabutr, S. Chamni, T. Rungrotmongkol, *Curr Top Med Chem* 20 (2020) 2046.
- [27] P. Mahalapbutr, V.S. Lee, T. Rungrotmongkol, *J. Agric. Food Chem.* 68 (2020) 7974.
- [28] P. Mahalapbutr, N. Kongtaworn, T. Rungrotmongkol, *Comput. Struct. Biotechnol. J.* 18 (2020) 2757.
- [29] D.M. York, T.A. Darden, L.G. Pedersen, *The Journal of Chemical Physics* 99 (1993) 8345.
- [30] X. Wu, B.R. Brooks, *Chem. Phys.Lett.* 381 (2003) 512.
- [31] D.R. Roe, T.E. Cheatham, *J. Chem. Theory Comput.* 9 (2013) 3084.
- [32] B.R. Miller, T.D. McGee, J.M. Swails, N. Homeyer, H. Gohlke, A.E. Roitberg, *J. Chem. Theory Comput.* 8 (2012) 3314.
- [33] S. Genheden, U. Ryde, *Expert Opin Drug Discov* 10 (2015) 449.
- [34] A. Leaver-Fay, M. Tyka, S.M. Lewis, O.F. Lange, J. Thompson, R. Jacak, K. Kaufman, P.D. Renfrew, C.A. Smith, W. Sheffler, I.W. Davis, S. Cooper, A. Treuille, D.J. Mandell, F. Richter, Y.-E.A. Ban, S.J. Fleishman, J.E. Corn, D.E. Kim, S. Lyskov, M. Berrondo, S. Mentzer, Z. Popović, J.J. Havranek, J. Karanicolas, R. Das, J. Meiler, T. Kortemme, J.J. Gray, B. Kuhlman, D. Baker, P. Bradley, *MethodsEnzymol* 487 (2011) 545.
- [35] G. Bhardwaj, V.K. Mulligan, C.D. Bahl, J.M. Gilmore, P.J. Harvey, O. Cheneval, G. W. Buchko, S.V.S.R.K. Pulavarti, Q. Kaas, A. Eletsky, P.-S. Huang, W.A. Johnsen, P. Greisen, Jr., G.J. Rocklin, Y. Song, T.W. Linsky, A. Watkins, S.A. Rettie, X. Xu, L. P. Carter, R. Bonneau, J.M. Olson, E. Coutsiyas, C.E. Correnti, T. Szyperski, D.J. Craik, D. Baker, *Nature* 538 (2016) 329.
- [36] A.L. Loshbaugh, T. Kortemme, *Proteins* 88 (2020) 206.
- [37] C.S. Poultney, G.L. Butterfoss, M.R. Gutwein, K. Drew, D. Gresham, K.C. Gunsalus, D.E. Shasha, R. Bonneau, *PLoS ONE* 6 (2011) e23947.
- [38] P.B. Stranges, B. Kuhlman, *Protein Sci.* 22 (2013) 74.
- [39] N. Haspel, D. Zanuy, R. Nussinov, T. Teesalu, E. Ruoslahti, C. Aleman, *Biochemistry* 50 (2011) 1755.
- [40] B. Fedorczyk, P.F.J. Lipiński, A.K. Puszek, D. Tymecka, B. Wilenska, W. Dudka, G. Y. Perret, R. Wieczorek, A. Misicka 24 (2019) 1756.
- [41] J.L. Daly, B. Simonetti, K. Klein, K.-E. Chen, M.K. Williamson, C. Antón-Plágaro, D.K. Shoemark, L. Simón-Gracia, M. Bauer, R. Hollandi, U.F. Greber, P. Horvath, R.B. Sessions, A. Helenius, J.A. Hiscox, T. Teesalu, D.A. Matthews, A.D. Davidson, B.M. Collins, P.J. Cullen, Y. Yamauchi 370 (2020) 861.
- [42] A. Starzec, M.A. Miteva, P. Ladam, B.O. Villoutreix, G.Y. Perret, *Bioorg. Med. Chem.* 22 (2014) 4042.