



OPEN

Inhibition of the hexamerization of SARS-CoV-2 endoribonuclease and modeling of RNA structures bound to the hexamer

Duy Phuoc Tran¹, Yuta Taira^{1,2}, Takumi Ogawa^{1,2}, Ryoga Misu¹, Yoshiaki Miyazawa¹ & Akio Kitao¹

Non-structural protein 15 (Nsp15) of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) forms a homo hexamer and functions as an endoribonuclease. Here, we propose that Nsp15 activity may be inhibited by preventing its hexamerization through drug binding. We first explored the stable conformation of the Nsp15 monomer as the global free energy minimum conformation in the free energy landscape using a combination of parallel cascade selection molecular dynamics (PaCS-MD) and the Markov state model (MSM), and found that the Nsp15 monomer forms a more open conformation with larger druggable pockets on the surface. Targeting the pockets with high druggability scores, we conducted ligand docking and identified compounds that tightly bind to the Nsp15 monomer. The top poses with Nsp15 were subjected to binding free energy calculations by dissociation PaCS-MD and MSM (dPaCS-MD/MSM), indicating the stability of the complexes. One of the identified pockets, which is distinctively bound by inosine analogues, may be an alternative binding site to stabilize viral RNA binding and/or an alternative catalytic site. We constructed a stable RNA structure model bound to both UTP and alternative binding sites, providing a reasonable proposed model of the Nsp15/RNA complex.

Coronaviruses possess the largest genome of known RNA viruses¹ and have attracted significant attention since the severe acute respiratory syndrome coronavirus (SARS-CoV) outbreak in 2002, and the Middle East respiratory syndrome coronavirus (MERS-CoV) affected Arabic countries in 2012². Currently, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is spreading globally and has caused the largest pandemic of the twenty-first century. Understanding the behavior of the SARS-CoV-2 virus at the molecular level would thus help curb the pandemic and prevent coronavirus-related diseases.

The SARS-CoV-2 genome encodes spike, nucleocapsid, membrane, and envelope proteins, as well as 16 non-structural proteins (Nsps)³. The functions of the 16 Nsps were poorly understood until recently⁴. The highly conserved nidoviral RNA uridylylate-specific endoribonuclease (NendoU) activity of Nsp15 allows evasion of the immune response⁵. By using BLAST with the UniProtKB database⁶, we confirmed that SARS-CoV-2 Nsp15 shares significant amino acid sequence identity (more than 40%) with SARS, MERS, and bat, rat, bovine, shrew, porcine, canine and ferret coronaviruses Nsp15 and several other SARS-CoV-2 proteins can suppress primary interferon production and interferon signaling, thus possibly interfering with the body's defense against infections⁷. Nsp15 cleaves the polyuridine (polyU) of negative-sense RNAs of beta-CoV mouse hepatitis virus (MHV) strain A59 (MHV-A59) and alpha-CoV porcine epidemic diarrhea virus (PEDV), limits the abundance and length of polyU, and delays the type I interferon response in macrophages⁸. The structure of Nsp15, suggested to be a dimer of homo-trimers, is well-conserved between MERS-CoV, SARS-CoV, MHV, and MERS-CoV⁹. Also, hexamerization was shown to promote nidoviral uridylylate-specific endoribonuclease activity⁹. The crystal structures of apo and citrate-bound SARS-CoV-2 Nsp15 confirmed that both forms of the protein are hexamers¹⁰. Comparison of the apo and UTP-bound states of the Nsp15 hexamer obtained by cryo-EM reconstructions indicates conformational dynamics between these states¹¹. These results suggested that SARS-CoV-2 activity can be inhibited by preventing the hexamerization of Nsp15 and intervening NendoU activity through drug binding. Since interface residues in oligomeric proteins tend to be evolutionally conserved^{12,13}, targeting pockets around the Nsp15 oligomer

¹School of Life Sciences and Technology, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8550, Japan. ²These authors contributed equally: Yuta Taira and Takumi Ogawa. ✉email: tpduy@bio.titech.ac.jp; akitao@bio.titech.ac.jp

interface is a reasonable approach for reducing the possibility of drug resistance by mutation. In addition, a study of 27 SARS-CoV-2 proteins showed that mutation rate ranges are very high for spike, Nsp12 (around 1.0), NS9c, and nucleocapsid (>0.5), while those for the other proteins are very low, including for Nsp15 (≤ 0.03)¹⁴. Right after the translation from viral RNA to protein, Nsp15 should exist as a monomer. Then, each monomer should conduct conformational changes to adapt itself to the suitable oligomeric form. Therefore, if ligand binding to Nsp15 monomers occur before oligomerization, Nsp15 activities should be disturbed.

Inheriting the above idea of inhibition, we first explored stable conformations of the SARS-CoV-2 Nsp15 monomer based on the hexameric apo form crystal structure by using an enhanced molecular simulation method, the Parallel Cascade Selection Molecular Dynamics (PaCS-MD) simulation¹⁵. Analysis of the free energy landscape of the conformational space using the Markov state model (MSM)¹⁶ indicated that the conformations of the global free energy minimum in the monomeric state significantly differ from that in the hexameric state. We targeted these conformations and identified possible druggable pockets suitable for stabilizing the monomeric conformations through drug binding and inhibiting hexamerization, then virtually screened possible compounds confirmed to stably bind to the proposed pocket by binding free energy calculations. One of the identified pockets was distinctively bound by inosine analogues, suggesting an alternative UTP binding site. By constructing an RNA structure connecting the *prior*-knowledge and nearest alternative UTP binding sites on the Nsp15 hexamer surface, we propose a reasonable structure model of the Nsp15/RNA complex.

Results and discussion

Features of the hexameric structure and initial modeling of the monomeric conformation. The apo (PDB ID: 6VWW) and citrate-bound (6W01) states of the SARS-CoV-2 Nsp15 hexamer crystal structure form a ring-like complex as a dimer of trimers (Fig. 1a)¹⁰. The root-mean-square deviation (RMSD) between the apo and citrate-bound monomers is very small (0.026 nm) and thus we focused on the apo hexameric form (“apoH” hereafter). The UTP-bound form determined by cryo-EM (PDB ID: 7K0R¹¹) also shows the hexameric form (“UTPH” hereafter). The monomer RMSD value between UTPH and apoH is small (0.037 nm). Each monomer is L-shaped and consists of three domains: the N-terminal domain (residues 1–68 in blue, “N-term domain” hereafter); the middle domain (residues 69–202 in red, “Mid domain”); and the C-terminal domain, which contains a NendoU catalytic site (residues 203–347 in dark grey, “C-term domain”). N-term domain includes a linker to Mid domain in the C-terminal end, and Mid domain contains a linker to C-term domain in the C-terminal end. Nine pairs of residues that form inter-monomer ionic and hydrogen bonds (hereafter “electrostatic bonding”) in the hexamer (Fig. 1b–d) stabilize the hexamer. The monomer uses the large surface area of N-term domain as inter-subunit interfaces that mostly interact with the orange region to make a trimer. The dark grey region mainly interacts with another trimer.

In the initial step, we conducted 5 distinct trials of 1 μ s standard molecular dynamics (MD) simulations starting from the monomer structure taken from the apo Nsp15 hexamer crystal structure (PDB ID: 6VWW¹⁰). The backbone RMSD from the crystal structure plateaus at 0.18 ± 0.06 nm (hereafter, values after ‘ \pm ’ indicate standard deviation unless otherwise noted) after 200 ns relaxation (left panel of Fig. S1), showing that the monomer structure changed little within this time scale. However, this does not necessarily mean real convergence over longer time scales. To further examine possible larger conformational changes from the hexameric state, we conducted enhanced conformational sampling in the next step. As shown in the plot of the root-mean-square fluctuation of each residue (RMSF) using the last 500 ns of the 1 μ s MD trajectory (right panel of Fig. S1), there were large fluctuations up to 0.3 nm in the end of N-term domain that links to Mid domain.

Enhanced conformational sampling indicates large domain movement. Proteins often exhibit significant conformational change upon complex formation^{17,18}. We employed PaCS-MD to examine conformational differences between the monomeric and hexameric forms of Nsp15. This method is an enhanced conformational sampling simulation that generates conformational transition pathways using cycles of multiple independent MD simulations without applying bias to the system and can be used to observe events whose timescales are longer than that of standard MD^{15,19,20}. By integrating the trajectories obtained by PaCS-MD and analyzing them with the MSM¹⁶, we can obtain various quantities such as the free energy landscape of conformational change^{21,22}, binding free energy^{23,24}, and association/dissociation rate constants^{18,25}. To enhance conformational sampling, PaCS-MD requires a quantity for selecting the initial structures for the next cycle. Here, we employed $\text{RMSD}_{\text{init}}$, which is defined as the backbone RMSD from the initial structure of each PaCS-MD trial. The use of $\text{RMSD}_{\text{init}}$ allowed PaCS-MD to significantly enhance conformational changes further from the initial structure, and is here called *rmsdPaCS-MD*.

Thirty independent *rmsdPaCS-MD* trials were conducted starting from 30 distinct initial structures selected from the aforementioned five 1 μ s MDs. For each trial, we employed 30 replicas (the number of MD runs performed in parallel in each cycle), and performed 30 cycles of *rmsdPaCS-MD*, excluding the initial cycle (cycle 0) that generates the initial conformations of cycle 1. Figure 2a shows $\text{RMSD}_{\text{init}}$ as a function of the number of cycles. In cycle 1, $\text{RMSD}_{\text{init}}$ ranged from 0.09 to 0.19 nm except for two cases (green and red) and gradually increased as the cycle evolved, indicating significant enlargement of the sampled conformational space. After 30 cycles, $\text{RMSD}_{\text{init}}$ reached 0.2–0.78 nm, except for two cases with $\text{RMSD}_{\text{init}} > 1$ nm.

To characterize the conformational space sampled by *rmsdPaCS-MD*, we first measured the inter- C_{α} distances between residues i and j of the last snapshot of each of the 30 trials and then calculated the distance changes ($|\Delta d_{ij}|$) compared to the distances in the *rmsdPaCS-MD* initial conformation (Fig. 2b). Three notable regions with low ($|\Delta d_{ij}|$) values (< 0.1 nm) agreed with the three domains already defined. The RMSD of each domain compared to the crystal structure ($\text{RMSD}_{\text{apoH}}$) versus the MD cycle index (Fig. 2c–e) remained around 0.2 nm or lower, except for the aforementioned two cases (cyan and green) in which Mid and C-term domains partially

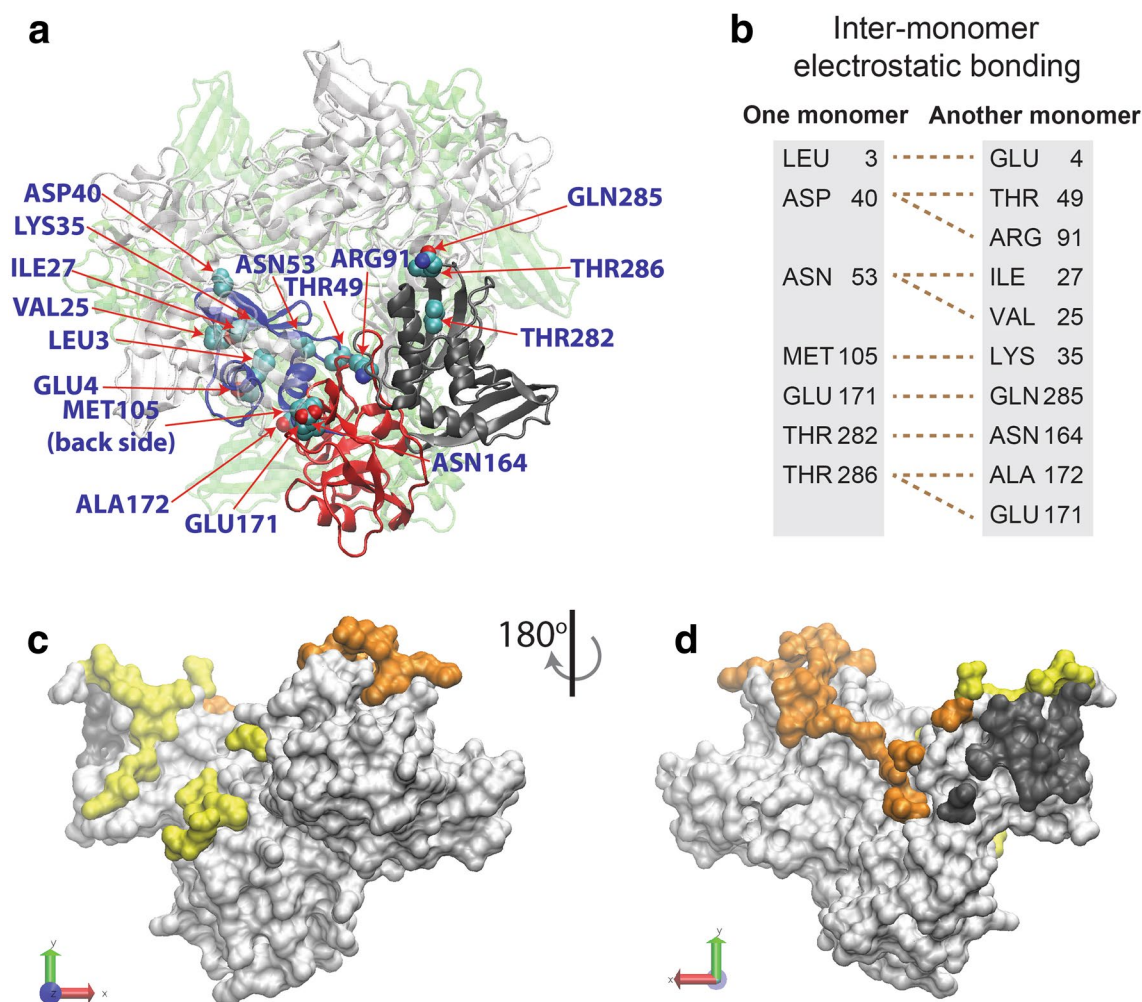


Figure 1. Structural features of the Nsp15 monomer in the apo hexamer. **(a)** A representative monomer structure (N-term domain: blue, Mid domain: red, and C-term domain: grey), the other dimers (transparent white) in the front trimer, and the backside trimers (pale green) are depicted. The residues shown by CPK models are involved in electrostatic bonding (salt bridges and hydrogen bonds) between the pairs of amino acids listed in **(b)**. In **(b)**, broken lines indicate the pattern of the electrostatic bondings. **(c, d)** The residues that form inter-monomer interfaces. The white area denotes non-interfacial residues, while other colors (yellow, orange, and dark grey) indicate the hexamer interface residues. In **(d)**, the protein is rotated 180° around the y axis. The structure images in this paper were created using VMD⁶⁰ unless otherwise specified.

unfolded. Consistent with the small-domain RMSDs shown in Fig. 2c–e, $\langle |\Delta d_{ij}| \rangle$ were mostly less than 0.4 nm in each domain, showing their rigidity. Even in the exceptional cases, N-term domain did not unfold in any of the rmsdPaCS-MD trials, indicating it has the greatest rigidity. Large distance changes between N-term and C-term domains (Fig. 2b) imply large motions between these domains. These motions can also be shown by best-fitting C-term domain and observing the position of N-term domain. These conformations include extremely large movements from the apoH structure (Fig. 2f), many of which may be unfeasible and should be excluded from the following analyses.

The open form is a stable conformation as a monomer. Although rmsdPaCS-MD implies possible motions of proteins, it does not directly indicate whether the observed motion is plausible. To identify stable conformations of the Nsp15 monomer, we calculated the free energy landscape in conformational space using rmsdPaCS-MD/MSM, in which merged rmsdPaCS-MD trajectories were analyzed by the MSM. rmsdPaCS-MD/MSM can obtain the protein free energy landscape spanned by representative coordinates of protein motion and find the lowest free energy conformation in the landscape^{18,22–25}. We calculated the free energy landscape spanned by two collective variables, namely, the two most important time-independent components (TIC 1 and TIC 2) obtained by time-independent component analysis (TICA)²⁶. Maximum-likelihood estimation was employed to construct the MSM by using the C_{α} coordinates of all residues projected onto the TICs after performing least-squares fitting of N-term domain. Since this domain exhibited the highest rigidity, this choice mainly focused the analysis on characterizing the possible motions of the other domains. After carefully check-

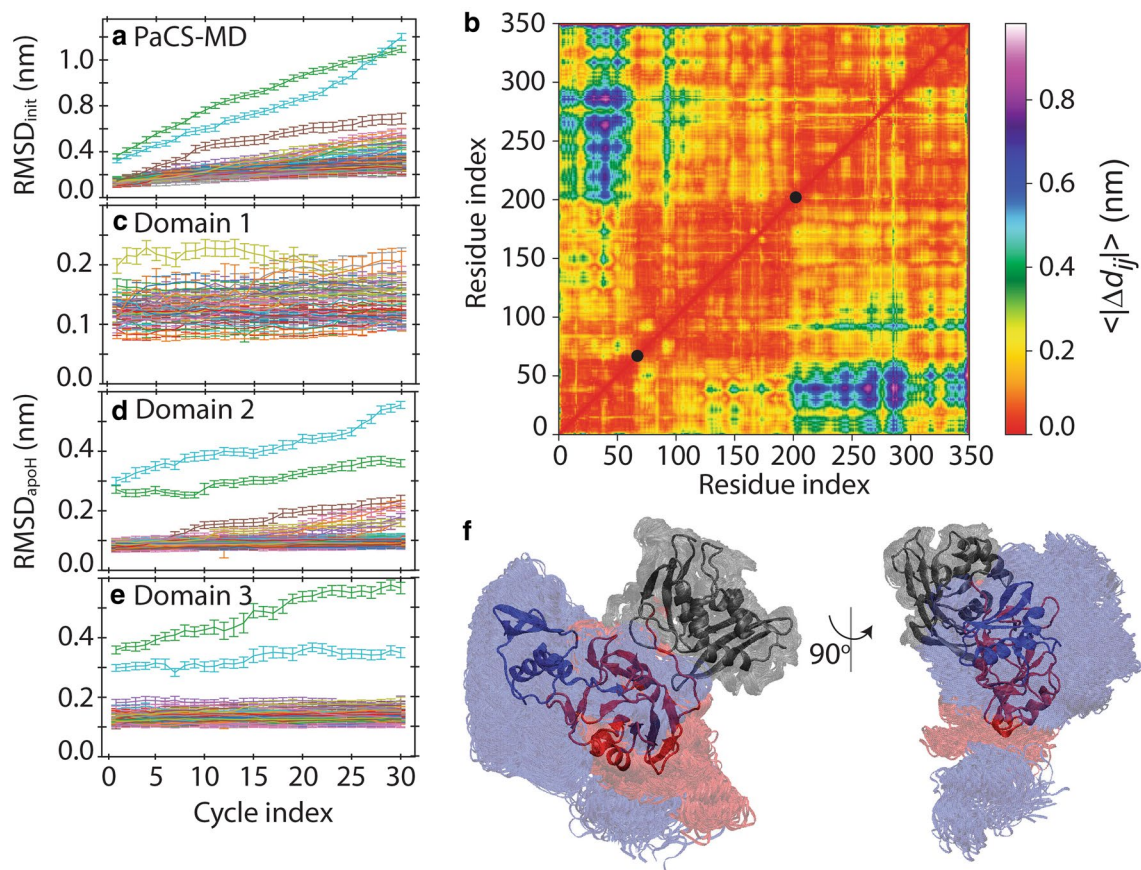


Figure 2. Conformational variation of the Nsp15 monomer. **(a)** Root-mean-square deviation ($\text{RMSD}_{\text{init}}$) and its standard deviations (error bars). **(b)** Change in the inter C_{α} distance between the initial and final snapshots of rmsdPaCS-MD averaged over trials, $\langle |\Delta d_{ij}| \rangle$. The color of a point specified by the residue indices i and j identifies the value of $\langle |\Delta d_{ij}| \rangle$, as shown in the color bar on the right. The two filled black circles indicate the borders of the three domains. **(c, d, e)** RMSD of each domain compared to that of the crystal structure ($\text{RMSD}_{\text{apoH}}$). **(f)** Visualization of the last frames at cycle 30 in all rmsdPaCS-MD trials after performing best-fitting of the backbone atoms of C-term domain. Each domain is specified by a distinct color, as in Fig. 1a.

ing the convergence of k -means clustering and the relation between lag time and the implied timescale via extensive MSM trials, we determined that the best lag time was 50 ps and used this value for further analyses. The evolution of the implied timescale as a function of lag time showed good quality of the MSM (Fig. S2), with 121 highly connected microstates identified. The disconnected states were related to unfolded structures.

The obtained free energy landscape of the Nsp15 monomer is shown in Fig. 3a. We found multiple free energy minima, global free energy minima (denoted as GM), and three intermediates (I1, I2, and I3), all of which are significantly different from apoH. The C_{α} RMSD between GM and apoH is 0.53 nm if the whole monomer is superimposed. The free energy of the microstate closest to apoH is higher than that of GM by 7.4 kcal/mol, which indicates that the GM structure is significantly more stable than the apoH structure as a monomer. GM, I1 (+0.9 kcal/mol compared to GM), and I2 (+1.3 kcal/mol) can be considered to belong to the same free energy basin (see the close-up view of the landscape, shown as an inset in Fig. 3a). Of these microstates, the I3 conformation is closest to that of apoH. Figure 3b shows the GM conformation (red) superimposed onto one of the Nsp15 hexamer subunits (blue) by best-fitting Mid and C-term domains and indicates the large relative movement of N-term domain outward of the hexamer ring. Since the GM conformation tends to open the trimer ring, we regard GM as a more “open” conformation compared to apoH. We obtained deeper insights to the structural differences between GM and I1–I3 by performing least-squares fitting of Mid and C-term domains (Fig. 3c). While N-term domain of GM (red), I1 (green), and M2 (orange) from the same basin highly overlapped, M3 (blue), which is closest to apoH, took a more compact form compared to the other conformations. The GM conformation is more exposed to solvent as a monomer, showing an increase in solvent accessible surface area of $12.0 \pm 2.1 \text{ nm}^2$ compared to apoH.

The conformational difference between the conformations of GM and apoH was further analyzed by the dynamic domain analysis method DynDom3D²⁷, which identifies “dynamic domains” from two structures and characterizes the rotation of a dynamic domain relative to others around an axis. First, we applied DynDom3D to the GM and apoH conformations (Fig. 3d). DynDom3D identified two domains: large (yellow) and small (cyan). The large domain consists of the entire C-term domain and most of Mid domain, while the small domain comprises N-term domain (except for the side chain of ASN63 and PRO66–PRO68) and small fragments of

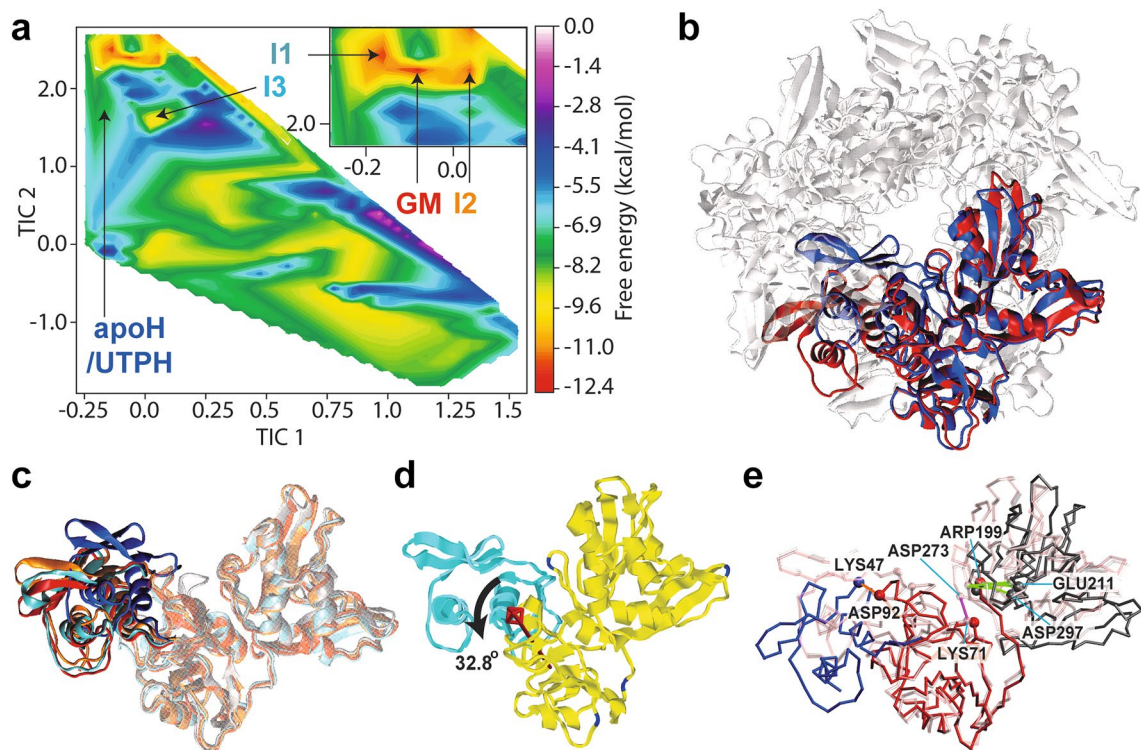


Figure 3. Stable conformation of the Nsp15 monomer. **(a)** Free energy landscape obtained by MSM. The inset shows a close-up view around the global free energy minimum (GM). I1, I2, and I3 represent three intermediates, and apoH indicates the crystal conformation. **(b)** Visualization of GM (red) superimposed onto one monomer of apoH (blue) in the hexameric state by best-fitting Mid and C-term domains. **(c)** Comparison of the GM (red), I1 (cyan), I2 (orange), and I3 (blue) conformations superimposed onto the Mid and C-term domains. **(d)** Result of DynDom3D analysis between GM and apoH, shown by the apoH conformation. The red arrow indicates the rotational axis of the dynamic domain (cyan) rotation, which corresponds to N-term domain. **(e)** Inter-domain salt bridges in the GM, apoH, and UTPH conformations. The apoH conformation is shown in white. N-term (blue), Mid (red), and C-term (dark grey) domains of the GM conformation and UTPH conformation (pink) are superimposed on the apoH conformation (white) by best-fitting Mid domain. The green lines connecting ARG199 and LYS205/ASP297 indicate the salt bridges maintained in all the conformations. The magenta lines between LYS71 and ASP273 represent the salt bridge formed in apoH and UTPH, and the orange line between LYS71 and ASP92 shows the salt bridge established solely in GM. **(d)** was created using Rasmol⁶¹.

Mid domain (parts of VAL85–THR99 and GLY101–THR106). The small domain rotates around the axis shown in Fig. 3d by 32.8° with a negligibly small translation (0.004 nm), yielding a 75.0% closure motion. This result means that conformational change upon hexamer formation can be considered as a rigid-body rotation of one domain relative to the other.

Electrostatic interactions to stabilize the monomeric conformation. We investigated the conformational difference between the hexameric and monomeric forms by analyzing the change in electrostatic bonding (hydrogen bonds and salt bridges) between apoH and GM within a monomer. The salt bridges are shown in Fig. 3e. Hydrogen bonds formed by ASN63 and PRO66 of N-term domain with their counterparts in Mid domain maintain electrostatic bonding in both apoH and GM. Although ASN63 and PRO66 are situated in N-term domain, they are assigned as part of the large dynamic domain by DynDom3D, indicating that the cyan dynamic domain in Fig. 3d (mostly N-term domain) can rotate while maintaining these interactions. ASN63 exchanged a hydrogen bonding partner with Mid domain from THR84 (apoH) to TYR89 in GM. An additional salt bridge between LYS47 (N-term domain) and ASP92 (Mid domain) is formed in GM, helping stabilize the open N-term domain arrangement relative to Mid domain.

Between Mid and C-term domains, 12 residue pairs maintain electrostatic bonding, including salt bridges between ARG199 (Mid domain) and GLU211/ASP297 (C-term domain) in both apoH and GM. These many interactions are considered to stabilize the arrangement between Mid and C-term domains, consistent with the result that these two domains are considered as one dynamic domain. However, the contribution of the ARG199–GLU211/ARG199–ASP297 interactions may be limited, as ARG199 is situated in the end linker between Mid and C-term domains. Three interactions found in apoH are not formed in GM but two new interactions are created. This result implies that the interactions between Mid and C-term domains are slightly weakened in the monomer. In addition, the salt bridge between LYS71 (Mid domain) and ASP273 (C-term domain) maintained in apoH is

lost in GM. This loss contributes to loosening the packing between the two domains around LYS71–ASP273, consistent with the presence of a larger space in GM, as shown in the next section: Pocket B located between Mid and C-term domains in Fig. 4b in the monomer is significantly larger than the corresponding pocket in apoH in Fig. 4e. As mentioned earlier, the monomer conformation of UTPH is very similar to that of apoH, and the salt bridges in apoH are also maintained in UTPH.

Druggable pockets around the hexamer interface. Significant conformational difference between GM and apoH/UTPH indicates that large conformational change is essential for oligomerization. If the GM conformation is stabilized by ligand binding, ligand binding should prevent hexamerization. For this purpose, we predicted druggable pockets in the GM conformation using PockDrug. In this method, each pocket is defined by the atoms that form the pocket. We applied this method to 100 structures selected from the GM microstates and identified 150 druggable pockets. The residues that form the pockets are highly overlapped in all 150 cases, and interestingly, TRP89 in Mid domain was without exception part of the pocket. The identified pockets were grouped into 25 clusters using the overlap ratio of the pocket-forming residues. Of the clusters situated around the hexamer interfaces, we selected the three most-populated clusters of pockets (hereafter Pockets A, B, and C), whose populations were 19/150 (12.7%), 16/150 (10.7%), and 13/150 (8.7%), respectively. The table in Fig. 4 shows the residues common in each of Pockets A–C, and Table S1 provides the PockDrug descriptors. Pocket A (druggable score: 0.73) is mostly formed by Mid domain, while Pocket B (0.65) spreads across all the domains. All the pocket-A-forming residues are included in the pocket-B-forming residues because Pocket B is assigned slightly more open structures. Pocket C (0.76) is mainly formed by the interface between N-term and Mid domains, implying that the binding of a compound to this pocket can prevent conformational change of N-term domain relative to the other domains. Of these pockets, Pocket A is the smallest in volume and Pockets B and C are 3- and 1.5-fold larger compared to Pocket A, respectively (Fig. 4a–c). The ratio of hydrophobic residues is similar between the pockets, but according to “hydrophobic kyte” obtained by PockDrug, Pockets A (-0.13 ± 0.21) and C (-0.11 ± 0.32) are less hydrophilic than Pocket B (-0.43 ± 0.18). Considering the ratio of polar residues, Pocket B showed slightly higher polarity (0.61 ± 0.03) compared to Pockets A and C, whose polarities coincidentally have the same value (0.52 ± 0.05). Furthermore, Pocket C is formed by the three positively charged residues HIS15, ARG62, and LYS65. Overall, we judged that Pocket C, whose druggable score is the highest, was a suitable target for subsequent virtual screening and thus we examined it by two screening methods. Furthermore, we conducted virtual screening of Pockets A and B by one screening method. Pockets A, B, and C are distant from the UTP-binding pocket (Fig. 4d). The main mutated residues of Nsp15 are THR34 (residue mutation rate = 0.03), LYS13 (0.01), ARG207 (0.01), and THR115 (0.01)¹⁴, which are not included in these pockets (Fig. 4).

We also investigated druggable pockets for the monomer conformation of apoH and identified pockets corresponding to Pockets A–C (Fig. 4e–g). The pockets in apoH were considerably smaller than Pockets A–C, and Pocket B was divided into two pockets in the crystal form of apoH, indicating that the GM conformation is more suitable as a drug target. PockDrug did not find any pocket around the UTP pocket in the UTPH conformation (Fig. 4h), likely due to the induced fit of the residues surrounding 5'-UMP eliminating any available space.

To understand conservation of the pocket-A- and -C-forming residues across corona viruses, we performed BLAST on Uniprot database, searching for sequences similar to the SARS-CoV-2 NSP15 sequence. Then, the selected sequences whose pairwise sequence identities with the SARS-CoV-2 NSP15 > 40% were subjected to multiple sequence alignment (Fig. S3). We found that TRP110 and TYR112 of pocket A and ASP148 and Gly188 of pocket C are highly conserved across the coronavirus species. Those residues may be considered as key residues for the future drug screening.

Virtual screening of antiviral compounds to find hexamerization inhibitors. We performed ligand docking with 49,430 antiviral compounds provided by The American Chemical Society (CAS COVID-19 Antiviral Candidate Compounds Dataset) by using the docking tool AutoDock Vina²⁸, targeting the centers of mass of Pockets A and C as the centers of the docking box. Since the docking box of Pocket A also completely contains Pocket B, docking to Pocket A simultaneously considers docking to Pocket B. Of 49,430 compounds, the screening yielded 22,945 poses per Pocket, except for 26,485 compounds that contain atoms not supported by AutoDock Vina or were not converted into the AutoDock PDBQT format (Table S2). Of these, 7 poses of 6 compounds showed an AutoDock Vina score lower than -12.0 kcal/mol (Fig. 5a). The compound with the best Vina score (CAS Registry Number[®], CAS RN[®]: 156210-14-9), which was originally designed as an ion-selective electrode²⁹, binds to Pockets A/B (pose 1) and C (pose 3). This compound is very flexible and relatively large (805.01 Da) compared to Lipinski's rule of five (500 Da)^{30,31}. The second compound (CAS RN[®]: 108037-59-8) that tightly binds to Pocket A/B (pose 2) also binds to Pocket C as pose 9 (Table S2). The compound of pose 4 (1883795-10-5), which is much smaller (591.54 Da) compared to the previous two compounds, was reported to have antimicrobial activity³². Poses 5–7 have equivalent Vina scores. The compound of pose 7 (1710363-55-5) is the smallest of the compounds shown in Fig. 5a (556.58 Da). Considering Lipinski's rule, the compounds of poses 4 and 6 are close to 500 Da and the others are much larger. Next, we investigated the stability of poses 1, 3 and 4 with Nsp15 using five 100 ns classical MD simulations (Fig. 6a, c, d). Pose 2 (Fig. 6b) containing three zinc ions was not optimized because classical MD simulation of zinc-containing compounds is not straightforward. Poses 1, 3 and 4 maintained stable binding with Nsp15, and Nsp15 maintained the open conformation that prevents hexamerization. We further examined the standard binding free energy ΔG° of poses 1, 3, and 4 by dissociation PaCS-MD and MSM (dPaCS-MD/MSM) (Fig. S4). Of these, pose 4 (1883795-10-5) showed the lowest ΔG° of -15.4 ± 0.2 kcal/mol ($\Delta W = 17.3 \pm 0.3$ and $\Delta G_V = 1.9 \pm 0.2$ kcal/mol (the values after ‘ \pm ’ show the standard errors), which is equivalent to a dissociation constant K_d of 6.0 pM at 300 K. Although AutoDock Vina

Pocket	Common residues
A	VAL67 VAL70 ILE86 TRP87 TYR89 SER162 GLY165
B	VAL67 PRO68 GLU69 VAL70 LYS71 ILE86 TRP87 TYR89 SER162 GLY165 SER198 ARG199 LEU252 ASP273 SER274 LYS277 ASP297
C	HIS15 ARG62 ASN63 ILE64 LYS65 ASN83 THR84 ILE86 PHE123 ASP125 VAL128

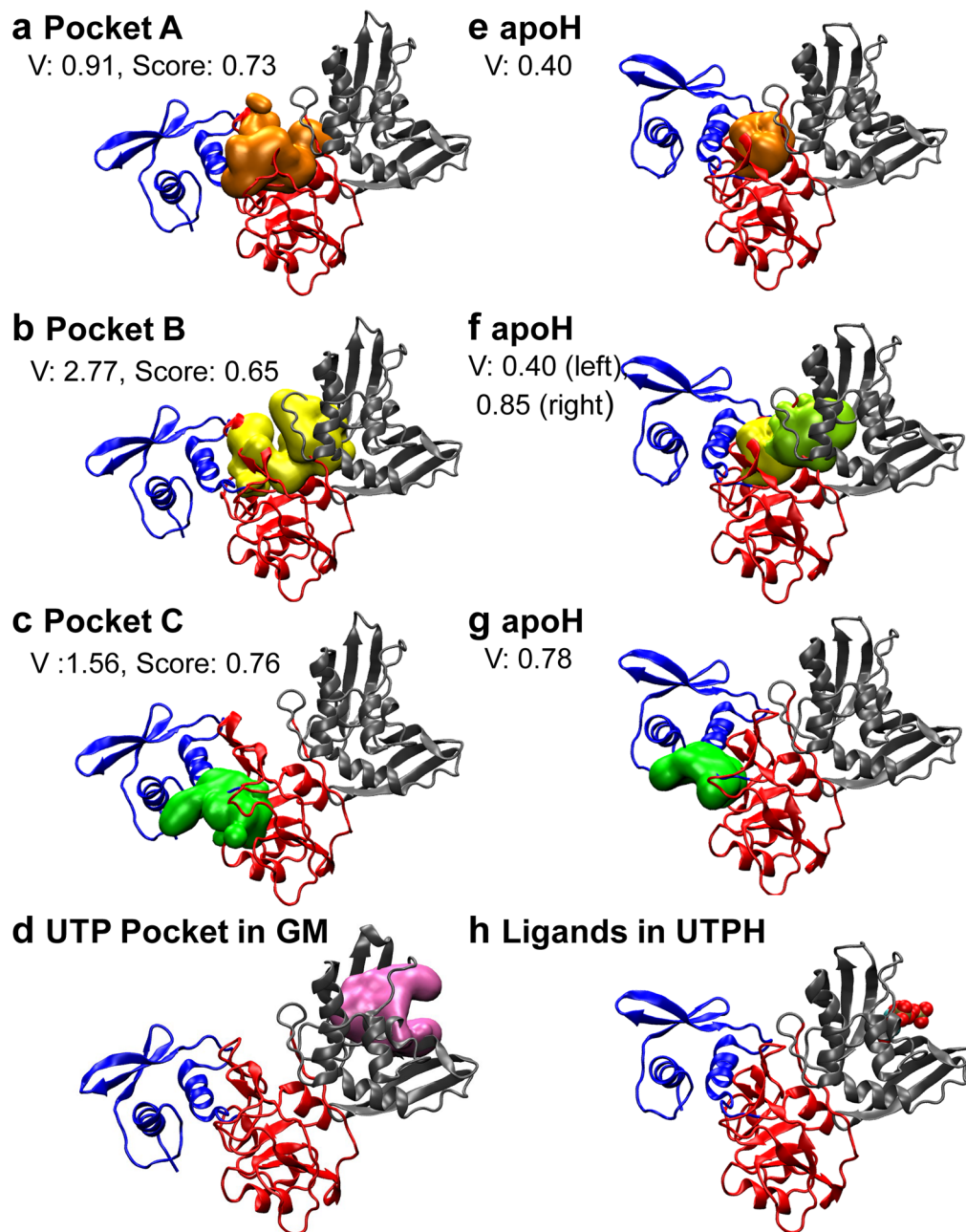
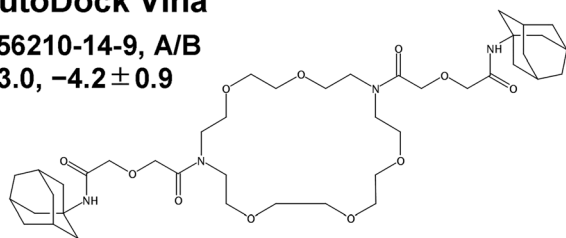


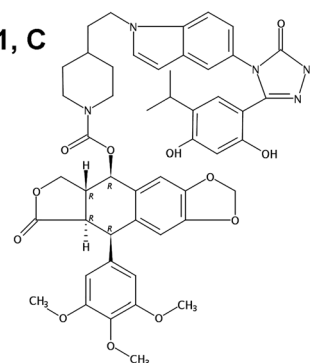
Figure 4. Major druggable pockets around the hexamer interfaces. The table at the top shows the residues common to the pockets and are assigned as Pockets A–C. The residue names shown in blue, red, and black indicate that they belong to N-term, Mid, or C-term domain, respectively. (a) Pockets A (green), (b) B (yellow), and (c) C (orange) identified for the monomer GM conformation. Druggable score of the pocket (Score) and pocket volume (V. unit: nm³) are shown. (d) UTP-binding pocket found in GM. (e–g) The pockets most overlapped with pockets A–C in apoH are depicted. (h) UTPH structure and 5'-UMP shown as a CPK model. Each domain is specified by a distinct color, as in Fig. 1a.

a AutoDock Vina

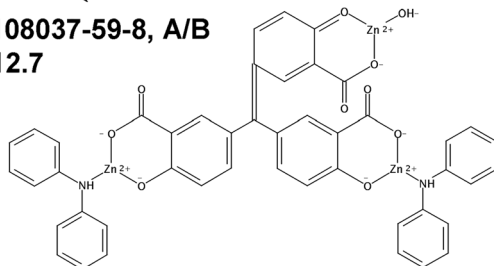
1. 156210-14-9, A/B
-13.0, -4.2 ± 0.9



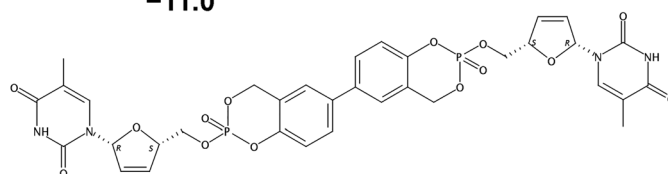
5. 1680212-58-1, C
-11.0



2. 108037-59-8, A/B
-12.7

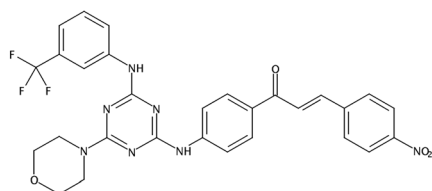


6. 1160003-72-4, A/B
-11.0

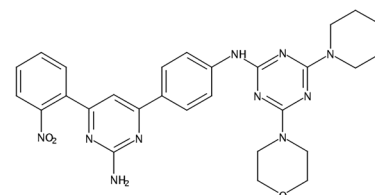


3. 156210-14-9, C
-12.2, -5.9 ± 0.5

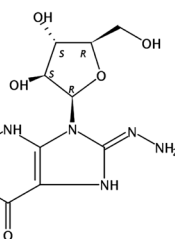
4. 1883795-10-5, A/B
-11.2, -15.4 ± 0.5



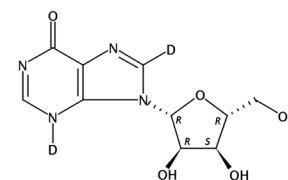
7. 1710363-55-5, A/B
-11.0

**b Glide**

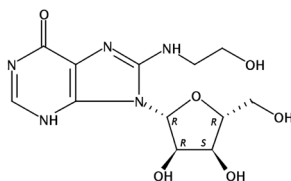
1. 69301-99-1, C
-6.81, -5.2 ± 0.4



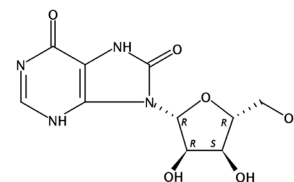
5. 160059-42-7, C
-6.03 (-6.2)



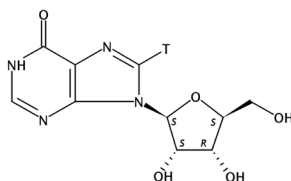
2. 1358577-79-3, C
-6.20 (-6.7)



6. 63699-77-4, C
-6.02



3. 2086327-90-2, C
-6.11



4. 2086327-90-2, C
-6.04

Figure 5. Top compounds identified by virtual screening. These compounds were screened from the antiviral compound library provided by The American Chemical Society (Mar. 2020)⁶² using (a) AutoDock Vina²⁸ and (b) Schrödinger Glide³³. Their ranking numbers, CAS Registry Number®, pockets (A/B or C), Vina or Glide scores, and standard binding free energies (ΔG°) obtained by dPaCS-MD/MDM and 2D structures are shown. In (b), Vina scores are also shown in parentheses if AutoDock Vina scores were calculated. The Vina and Glide scores and ΔG° are in kcal/mol. 2D structures were created by SciFinder.

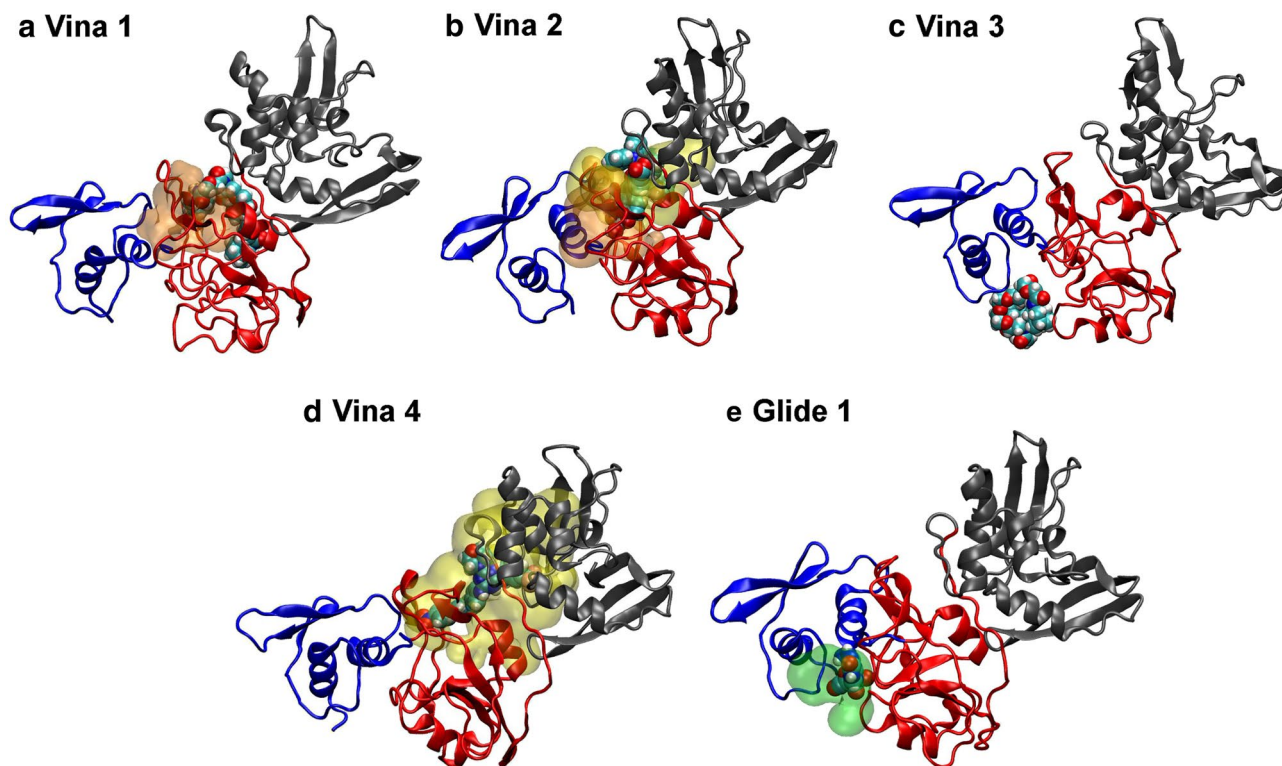


Figure 6. Complex structures of the top poses. (a–d) The top 4 poses identified by AutoDock Vina²⁸ and (e) the top pose identified by Schrödinger Glide³³. Representative complex structures after free 100 ns MD are shown except for (b), in which the output of AutoDock Vina is shown.

scores were very low for poses 1 and 3, the values of ΔG° were significantly higher (pose 1: $\Delta G^\circ = -4.2 \pm 0.4$, $\Delta W = 6.0 \pm 0.5$, and $\Delta G_V = 1.8 \pm 0.3$ kcal/mol; pose 3: $\Delta G^\circ = -5.9 \pm 0.2$, $\Delta W = 7.6 \pm 0.2$, and $\Delta G_V = 1.7 \pm 0.2$ kcal/mol), suggesting that the binding of this compound (156210-14-9) is not as strong as expected from the AutoDock Vina scores.

We performed ligand docking using the docking tool Glide^{33–35} in the Schrödinger® package, targeting Pocket C that was judged as the best target in the previous section. All the compounds were considered. The screening yielded 225 poses of 24 compounds with Glide scores below -2.8 kcal/mol (Table S3), which is the default cutoff value of Glide. Of these, 6 poses of 5 compounds showed a Glide score lower than -6.0 kcal/mol, implying relatively strong binding to Pocket C (Fig. 5b). Interestingly, all 5 compounds are inosine-related compounds and interact with HIS15, ILE64, and LYS65 in N-term domain (Fig. S5). Most of the other compounds are nucleoside analogues (Table S3). The 6 binding poses with Nsp15 maintained stable binding during 100 ns MD simulations (Fig. 6e, and S6), and Nsp15 maintained the open conformation. Vina scores were obtained for the compounds of poses 2 and 5 (values in parentheses in Fig. 5b) which agreed with the corresponding Glide scores, indicating that the Vina and Glide scores are comparable. We further examined the standard binding free energy ΔG° of the top pose with the compound (69301-99-1) by dPaCS-MD/MSM (Fig. 3d), and obtained a standard binding free energy ΔG° of -5.2 ± 0.4 kcal/mol ($\Delta W = 6.5 \pm 0.3$ and $\Delta G_V = 1.3 \pm 0.2$ kcal/mol) that is similar to the Glide score. These results indicate that the compounds shown in Fig. 5b are significantly weaker binders than the others.

During the revising process of this paper, Schultz et al. validated antiviral activity and selectivity of 122 drugs against SARS-CoV-2 and found that 16 of these are nucleoside analogues, including antivirals remdesivir and molnupiravir approved for COVID-19³⁶. As shown above, all five compounds that we found by GLIDE are nucleotide analogues as well. Very recently, Kumar et al. carried out in silico and in vitro screening for active compounds from their in-house libraries targeting to Nsp15 and identified ‘Compound IV’ ((2S,3S)-3-amino-1-(4-(4-(tert-butyl)benzyl)piperazin-1-yl)-4-phenylbutan-2-ol) as the best one in in vitro assays³⁷. Of these, functional groups of Compound IV share analogous interactions with Nsp15 in Pocket A/B with those of the compound of pose 4 that shows the lowest ΔG° . Choi et al. used high-throughput assay to screen compounds against Nsp15, and inhibition was confirmed for three hits in vitro and showed that Exebryl-1 (β -amyloid anti-aggregation molecule for Alzheimer’s therapy) has antiviral activity³⁸. They also showed that the most plausible binding residues with Exebryl-1 significantly overlap with those of Pocket B. Interestingly, the compounds that we found by AutoDock Vina shares some similarity in functional groups with Exebryl-1. This suggests the potentials of those predicted compounds as a starting point for further drug developments.

Possible alternative RNA binding site and RNA structure bound to the Nsp15 hexamer. As described above, the top 6 poses in Pocket C identified by Glide are inosine analogs, which suggests possible

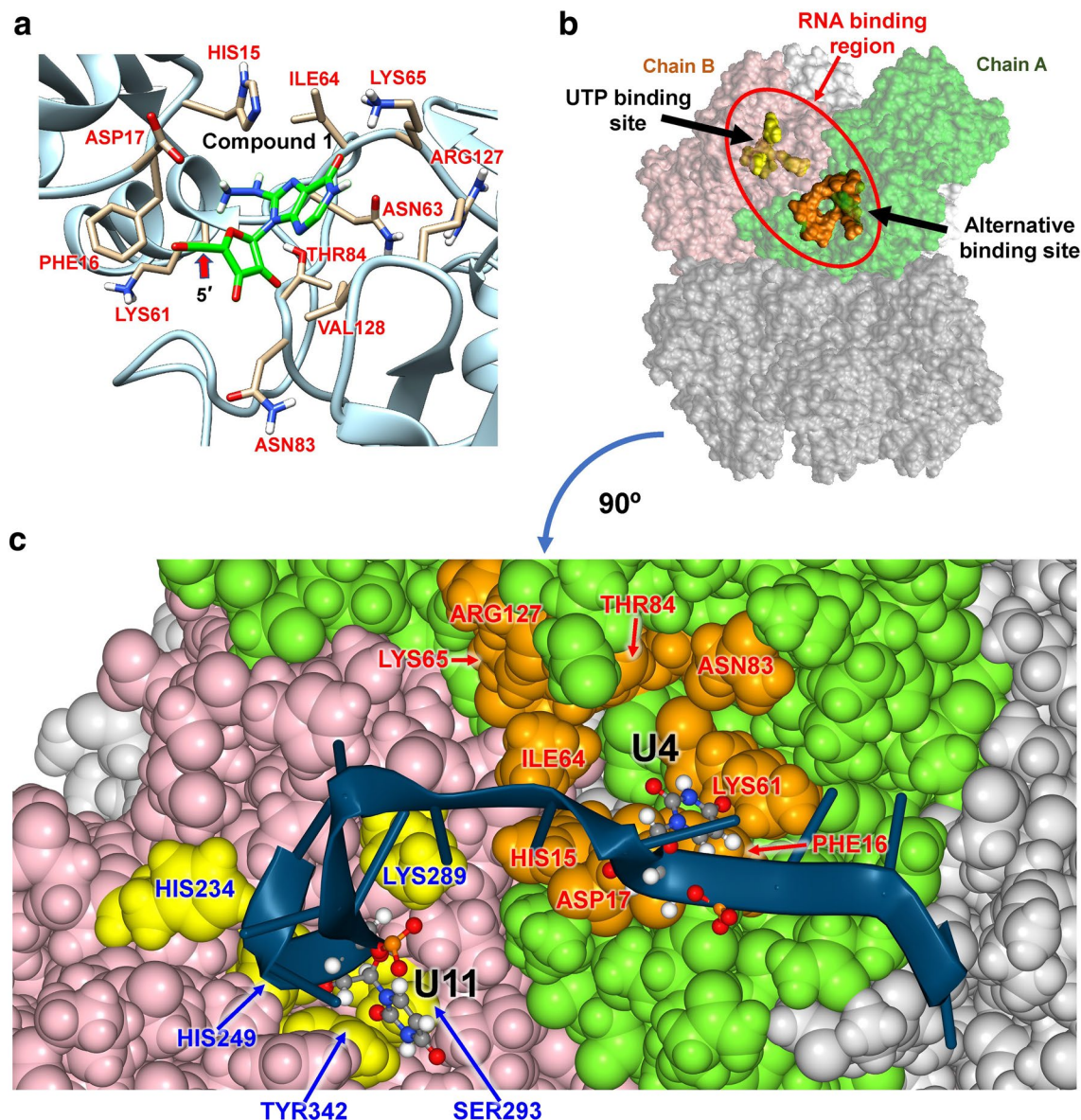


Figure 7. Nucleoside binding in Pocket C and possible RNA structure bound to the Nsp15 hexamer. **(a)** The best compound identified by Glide (CAS RN^o: 69301-99-1. 1H-Purine-6,8-dione, 9- β -D-arabinofuranosyl-7,9-dihydro-, 8-hydrazone) and surrounding residues. The orientation is slightly changed from Fig. 6e to better show coordination of the surrounding residues. The 5'-carbon of the ribose ring is indicated by a red arrow. **(b)** Spatial arrangement of the UTP binding site and alternative nucleotide binding site (Pocket C) mapped onto UTPH. The residues shown in yellow are HIS235, HIS250, LYS290, SER294, and TYR343, and those shown in orange are those presented in (a). **(c)** A representative tridecaU RNA structure bound to the Nsp15 hexamer after 100 ns equilibration by MD. The coloring is the same as in (b). (a and c) were created using Chimera⁶³ and NGLview⁶⁴, respectively.

binding of other nucleosides and nucleotides. We used PockDrug to examine if space remained around the compounds in the complex structures and found that it did, except for Glide poses 4 and 6 (Fig. 6 and S6), indicating the possibility of improving binding by adding chemical groups to the compounds. Figure 7a shows the coordinated residues around the compound of pose 1. The 5'-end of the ribose ring is exposed to solvent, reserving space for an additional group (e.g., a phosphate group), suggesting that Pocket C may be an alternative nucleotide binding site.

Interestingly, the pocket around this region is partly exposed to solvent even in the hexamer (Fig. 7b), although the pocket has half the volume of the pocket in the monomer (compare Fig. 4c, g). Also, Pocket C of one monomer is situated near the UTP binding site of the nearest monomer in the hexamer (Fig. 7b). If this pocket is an alternative RNA binding site, it can stabilize the viral RNA bound to Nsp15 during NendoU catalysis. Another possibility is that Pocket C has a catalytic site because the residues of Pocket C resemble the catalytic residues of

ribonucleases. Histidine, aspartic acid and lysine are typical amino acid residues located around ribonuclease catalytic sites¹¹ and HIS15, ASP17, LYS61, and LYS65 in Pocket C are situated near the ribose ring (Fig. 7a).

To further examine this idea, we constructed single-stranded RNA models bound to the Nsp15 hexamer based on UTPH (the UTP-bound hexameric form) and the arrangement of the two binding sites on the Nsp15 hexamer surface (Fig. 7b). First, three RNA chains consisting of tridecaU (5'-(U)₁₃-3') were modeled, so that each 11th U (U11) binds to one of the UTP binding sites. We prepared relatively long RNA chains so that one of the RNA residues can fit into the alternative binding site. After simulated annealing, the RNA chain closest to the alternative binding site (the 3rd RNA chain) was selected, and the closest U residue (initially U3, and U4 in the final step) was pulled toward the alternative binding site by Steered MD (SMD). After modeling and equilibration, five independent 100 ns MD simulations of the RNA-bound Nsp15 hexamer were conducted. We confirmed that the RNAs stably bound to Nsp15 during the MDs, and U4 and U11 maintained binding to the alternative and UTP binding sites, respectively (Fig. 7c, and S7). Therefore, single stranded RNA chains consisting of 8 nucleic acids or longer can bind to both pockets. RNA binding is stabilized not only by the interactions of U4 and U11 with the pocket residues but also by the following Nsp15/RNA interactions shown as (Nsp15 residues)/(RNA residue): SER2, ASN5, GLN19/U1; GLU22/U2; GLN245, LYS290/U10; TRP333/U12. These amino acid residues have no overlap with the aforementioned mutated residues, suggesting that the mutations have no direct effect on RNA binding. Interestingly, U4 of the 1st RNA chain, which was not subjected to SMD, also reached near the alternative binding site with a flipped-out uridine group, although the uridine group was not perfectly situated in the pocket. Since RNA binding should be stabilized by binding to the alternative binding site, Nsp15 hexamerization is a key for NendoU catalysis, suggesting that preventing hexamerization may efficiently inhibit the catalytic activation of Nsp15. After initial submission of this paper, Frazier et al. determined the post-cleavage cryo-EM structure of Nsp15, indicating that AUA trinucleotide actually binds to Pocket C. They also showed that mutating HIS15 disrupts RNA cleavage and Nsp15 oligomerization³⁹. These findings show that our calculation results are consistent with these experiments.

Conclusion

We proposed that SARS-CoV-2 activity can be prevented by preventing the hexamerization of endoribonuclease Nsp15 with drug binding. We first explored the stable conformation of the Nsp15 monomer as the global free energy minimum conformation by using rmsdPaCS-MD/MSM. Compared to the hexamer form, N-term domain rotates by 32.8°, creating larger druggable pockets on the surface of Nsp15. Targeting the pockets with high drug-gability scores, we conducted ligand docking and identified compounds that tightly bind to the Nsp15 monomer. Nsp15s complexed with the top compounds were subjected to binding free energy calculations by dPaCS-MD/MSM, indicating the stability of the complexes. The binding of the compounds maintained the open conformation of Nsp15, which can prevent hexamerization. These compounds may provide leads for drug development against COVID-19. Of these, the compound of AutoDock Vina pose 4, whose binding free energy is the lowest and whose molecular weight is relatively low, is the best candidate. Pocket C is suggested to be an alternative binding site to stabilize viral RNA binding and/or an alternative catalytic site. Further, we constructed a structure model of RNA-bound Nsp15 and demonstrated the stability of the complex by MD simulation, thereby proposing a reasonable model of the Nsp15/RNA complex during NendoU activity.

Methods

Model preparation, equilibration, and rmsdPaCS-MD. The AMBER ff19SB force field⁴⁰ was used for the protein. Nsp15 monomer was solvated in a 15.9 × 14.7 × 14.7 nm³ box with OPC water molecules⁴¹. Potassium and chloride ions were added to mimic a 0.15 M ion concentration and charge neutrality. The relaxation simulations were performed using AMBER18⁴² and PaCS-MD simulations were performed using GROMACS 2019.4⁴³.

We carried out equilibration as follows. (1) The solvated models were energy-minimized by the steepest descent method followed by the conjugate gradient method with positional restraints applied on the heavy atoms of Nsp15 (force constant: 1000 kJ/mol nm²). (2) The system with the same restraints was heated from 0 to 300 K within 1 ns and thermalized at 300 K for another 1 ns using an NVT ensemble simulation. Five different trials were carried out with different randomly-generated initial velocities to provide statistics for the simulations. (3) MD simulation with the NPT ensemble was conducted for the next 100 ps at 300 K with a relaxation time of 0.1 ps for heat bath coupling, and at 1.0 atm with a relaxation time of 2.0 ps for isotropic pressure coupling. (4) The force constant of the positional restraints was reduced by 100 kJ/mol nm² every 100 ps until it vanished (total 0.9 ns). (5) The system was equilibrated for 1 μs using the NPT ensemble. (6) The selected conformations were converted for GROMACS to conduct 30 cycles of PaCS-MD. The equation of motion was integrated using the velocity Verlet method⁴⁴ with bond constraints by the SHAKE method⁴⁵ (steps 2–5) and without bond constraints (PaCS-MD step). The isothermal condition was established by Langevin dynamics^{46,47} in steps 3, 4, and 5, and by the Nosé-Hoover thermostat in the PaCS-MD step. The isobaric condition was achieved using the Berendsen (steps 3, 4)⁴⁸, Monte-Carlo (step 5)⁴⁹, and Parrinello-Rahman barostats (step 6)⁵⁰. The snapshots sampled in the second half (500 ns each) of the 5 runs of step 5 were grouped into 50 clusters. We selected the 30 highest populated structures from step 5 and used them in step 6. Upon conversion from AMBER to GROMACS, we equilibrated the system for 10 ns with 1000 kJ/mol nm² positional restraints of the protein backbone. For each of these structures, 1 ns MD was conducted (hereafter cycle 0), and 30 snapshots of the first rmsdPaCS-MD cycle were selected. We used 30 replicas and 0.1 ns MD simulations for each cycle and recorded the snapshots every 0.5 ps for analysis. All the MD trajectories generated by PaCS-MD, including cycle 0, were merged and used in MSM. The total simulation cost of rmsdPaCS-MD was 2.73 μs MD: [(1 ns MD for cycle 0 + 0.1 ns MD × 30 replicas × 30 cycles) × 30 trials]. Together with the 1.8 μs relaxation MDs, the total computational cost of this step

was 4.53 μ s. The same timestep, barostat, and thermostat settings were applied to the following MD simulations of the complexes and dPaCS-MD.

Markov state model. The initial dataset for the MSM was constructed as all C_{α} coordinates of Nsp15 after performing least-squares fitting of N-term domain with the crystal conformation. The dataset was then discretized into 1000 microstates by k-means clustering⁵¹ with the k-means++ algorithm⁵². After carefully checking the convergence of the cluster centers over multiple trials with different lag times, the dataset was projected onto the time-lagged independent components space (TIC space)⁵³. The original dimension of 1044 was reduced to 562, keeping 95% of the fluctuations in the reduced space. We built the MSM by using maximum-likelihood estimators with a sliding count algorithm to fulfill the detailed balance condition. PyEMMA package⁵⁴ was used to construct the MSM.

Domain motion analysis. We used DynDom3D for domain motion analysis²⁷. Default DynDom3D parameter values were used: grid size 0.4 nm; block factor 2; occupancy 0.6; and minimum domain size 200.

Prediction of druggable pockets. We used PockDrug-Server⁵⁵ to predict druggable pockets. The pocket estimation method fpocket⁵⁶ was applied to 100 structures randomly selected from the GM microstate with a ligand proximity threshold of 5.5 Å. Pockets with a druggability score greater than 0.5 were regarded as druggable pockets. PockDrug can identify small pockets but we only considered pockets formed by more than 13 residues because smaller pockets tend to be less druggable. Of the identified druggable pockets, those frequently found for different conformers were selected by clustering using pocket similarity measured by the overlap ratio, which was defined as the number of common pocket-forming residues included in both pockets divided by the number of pocket-forming residues included in one pocket or the other. The clustering method UPGMA (unweighted pair group method with arithmetic mean) was employed.

Binding free energy calculation by dPaCS-MD/MSM. First, we conducted 5 trials of the 100 ns relaxation MD simulations for the aforementioned 4 complexes formed between the selected compounds and Nsp15 protein. Then, we carried out 5 trials of dPaCS-MD for each complex. The initial structure of each trial was taken from the last frame of the 5 MDs. We used 30 replicas in dPaCS-MD, with each replica 100 ps long. We used the AMBER ff19SB force field⁴⁰ for Nsp15 and GAFF2⁵⁷ to determine the ligand force field. The partial charges of the ligand were parameterized using the Gaussian package⁵⁸. The OPC water model was applied in these simulations. The dPaCS-MD trials were carried out until the inter-COM distance was over 7 nm, then we constructed the MSM using the inter-Center of Mass (COM) distances obtained by dPaCS-MD and calculated the volume correction as described in the literature^{23,25}. We calculated the standard binding free energy of the best compound as follows:

$$\Delta G^{\circ} = -\Delta W + \Delta G_V \quad (1)$$

where ΔW is the free energy difference (potential of mean force: PMF) from the bound state to the unbound state (Fig. S4) and ΔG_V is the volume correction of the free energy difference.

Modeling of RNA structure bound to Nsp15 hexamer. The settings for modeling were the same as those in the aforementioned simulations, unless otherwise specified. For RNA, the DESRES potential⁵⁹ was used. In the following MD simulations, the isothermal condition was established by Langevin dynamics^{46,47} and the force constant for the positional restraints was 100 kcal/molÅ² for steps 2–10, and 13. Modeling was conducted using the following procedure. (1) An RNA model consisting of tridecaU (5'-(U)₁₃-3') was constructed by using the 5'-UMP structure of the cryo-EM as U11 and by adding 10 U in the 5'-end and two U in the 3'-end. One of the Nsp15 trimers from the UTP-bound hexamer was employed, and U11 of tridecaU was placed in each UTP-binding pocket. (2) The system was solvated into a 15.8 × 15.2 × 12.9 nm³ box and energy minimized. Then, 200 ps MD simulation was conducted using an NVT ensemble at 300 K with positional restraints imposed on the C_{α} atoms of Nsp15 and the heavy atoms of RNA, followed by 10 ns MD using an NPT ensemble with the Berendsen barostat and 40 ns MD with the Monte Carlo barostat. (3) The system was heated to 400 K during 100 ps with positional restraints on the C_{α} atoms of Nsp15 and U11, and (4) 100 ps NVT MD was performed at 400 K. (5) The system was heated to 500 K during 100 ps with the same positional restraints, (6) 100 ps restrained NVT MD was conducted at 500 K, and (7) 2.0 ns MD at 500 K was continued. (8) 100 ps NPT MD at 500 K and 1.0 atm was performed with the Berendsen barostat. (9) 20 ns NPT MD was conducted with the Monte Carlo barostat. (10) Simulated annealing down to 300 K was performed starting from a selected structure whose RNAs were relatively close to the alternative binding site. (11) A structure whose RNA was the closest to the alternative binding site (the 3rd RNA chain) was selected and 100 ns MD was conducted with reduced positional restraints (force constant: 10 kcal/molÅ²). (12) Another trimer was added to the system to reconstruct the Nsp15 hexamer with 3 RNA chains, and the system was solvated into a 14.6 × 15.3 × 15.9 nm³ box. A relaxation procedure similar to (2) was conducted and then 200 ns NPT MD was performed at 300 K and 1 atm without positional restraints. (13) 10 ns steered MD (SMD) was performed by pulling the O4 atom of U3 (the residue closest to the binding site at this point) toward the NE2 atom of HIS17 of the alternative binding site with positional restraints on Nsp15 and U11. (14) 10 ns MD was conducted by decreasing the force constant by 10 kcal/molÅ² per 1 ns until it vanished. At this stage, positional restraints for the O4 and NE2 were also applied with a force constant of 100 kcal/molÅ². (15) 10 ns MD was performed by reducing the positional restraint for the O4 and NE2, and (16) free MD was conducted for 50 ns. This resulted in U4 of the 3rd RNA chain being the closest to the target

pocket. (17). Similar to (13), 10 ns SMD was performed by pulling the O4 of U4 toward the C_α atom of LYS67 with the same setting as in (15), so that the RNA tightly bound deeper into the alternative binding site. (18) 10 ns MD was performed by reducing the positional restraints for O4 and NE2, and 19) five independent free MD simulations were conducted for 100 ns.

Data availability

The data supporting this study are included in this published article and its Supplementary Information.

Received: 20 July 2021; Accepted: 24 February 2022

Published online: 09 March 2022

References

1. Woo, P. C. Y., Huang, Y., Lau, S. K. P. & Yuen, K.-Y. Coronavirus genomics and bioinformatics analysis. *Viruses* **2**, 1804–1820 (2010).
2. Cui, J., Li, F. & Shi, Z.-L. Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* **17**, 181–192 (2019).
3. Fung, T. S. & Liu, D. X. Human coronavirus: Host–pathogen interaction. *Annu. Rev. Microbiol.* **73**, 529–557 (2019).
4. V'kovski, P. *et al.* Determination of host proteins composing the microenvironment of coronavirus replicase complexes by proximity-labeling. *Elife* **8**, e42037. <https://doi.org/10.7554/eLife.42037> (2019).
5. Kindler, E. *et al.* Early endonuclease-mediated evasion of RNA sensing ensures efficient coronavirus replication. *PLOS Pathog.* **13**, e1006195. <https://doi.org/10.1371/journal.ppat.1006195> (2017).
6. UniProt Consortium. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515. <https://doi.org/10.1093/nar/gky1049> (2019).
7. Yuen, C.-K. *et al.* SARS-CoV-2 nsp13, nsp14, nsp15 and orf6 function as potent interferon antagonists. *Emerg. Microbes Infect.* **9**, 1418–1428 (2020).
8. Hackbart, M., Deng, X. & Baker, S. C. Coronavirus endoribonuclease targets viral polyuridine sequences to evade activating host sensors. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 8094–8103 (2020).
9. Zhang, L. *et al.* Structural and biochemical characterization of endoribonuclease Nsp15 encoded by middle east respiratory syndrome coronavirus. *J. Virol.* **92**, e00893 (2018).
10. Kim, Y. *et al.* Crystal structure of Nsp15 endoribonuclease NendoU from SARS-CoV-2. *Protein Sci.* **29**, 1596–1605 (2020).
11. Pillon, M. C. *et al.* Cryo-EM structures of the SARS-CoV-2 endoribonuclease Nsp15 reveal insight into nuclease specificity and dynamics. *Nat. Commun.* **12**, 636. <https://doi.org/10.1038/s41467-020-20608-z> (2021).
12. Janin, J., Bahadur, R. P. & Chakrabarti, P. Protein–protein interaction and quaternary structure. *Q. Rev. Biophys.* **41**, 133–180 (2008).
13. Aoto, S. & Yura, K. Case study on the evolution of hetero-oligomer interfaces based on the differences in paralogous proteins. *Biophys. Physicobiol.* **12**, 103–116 (2015).
14. Vilar, S. & Isom, D. G. One year of SARS-CoV-2: How much has the virus changed?. *Biology (Basel)*. **10**, 91 (2021).
15. Harada, R. & Kitao, A. Parallel cascade selection molecular dynamics (PaCS-MD) to generate conformational transition pathway. *J. Chem. Phys.* **139**, 035103 (2013).
16. Bowman, G. R., Pande, V. S. & Noé, F. An introduction to Markov state models and their application to long timescale molecular simulation. In Bowman, G. R., Pande, V. S. & Noé, F. (eds.) *Springer* **797**, 148 (Springer Netherlands, 2014).
17. Kitao, A. & Takemura, K. High anisotropy and frustration: The keys to regulating protein function efficiently in crowded environments. *Curr. Opin. Struct. Biol.* **42**, 50–58 (2017).
18. Tran, D. P. & Kitao, A. Kinetic selection and relaxation of the intrinsically disordered region of a protein upon binding. *J. Chem. Theory Comput.* **16**, 2835–2845 (2020).
19. Harada, R. & Kitao, A. Nontargeted parallel cascade selection molecular dynamics for enhancing the conformational sampling of proteins. *J. Chem. Theory Comput.* **11**, 5493–5502 (2015).
20. Takaba, K., Tran, D. P. D. P. & Kitao, A. Edge expansion parallel cascade selection molecular dynamics simulation for investigating large-amplitude collective motions of proteins. *J. Chem. Phys.* **152**, 225101 (2020).
21. Kitao, A., Harada, R., Nishihara, Y. & Tran, D. P. Parallel cascade selection molecular dynamics for efficient conformational sampling and free energy calculation of proteins. *AIP Conf. Proc.* **1790**, 020013 (2016).
22. Inoue, Y. *et al.* Structural insights into the substrate specificity switch mechanism of the type III protein export apparatus. *Structure* **27**, 965–976 (2019).
23. Tran, D. P., Takemura, K., Kuwata, K. & Kitao, A. Protein–ligand dissociation simulated by parallel cascade selection molecular dynamics. *J. Chem. Theory Comput.* **14**, 404–417 (2018).
24. Hata, H. *et al.* High pressure inhibits signaling protein binding to the flagellar motor and bacterial chemotaxis through enhanced hydration. *Sci. Rep.* **10**, 2351. <https://doi.org/10.1038/s41598-020-59172-3> (2020).
25. Tran, D. P. & Kitao, A. Dissociation process of a MDM2/p53 complex investigated by parallel cascade selection molecular dynamics and the Markov state model. *J. Phys. Chem. B* **123**, 2469–2478 (2019).
26. Naritomi, Y. & Fuchigami, S. Slow dynamics in protein fluctuations revealed by time-structure based independent component analysis: The case of domain motions. *J. Chem. Phys.* **134**, 065101 (2011).
27. Girdlestone, C. & Hayward, S. The DynDom3D webserver for the analysis of domain movements in multimeric proteins. *J. Comput. Biol.* **23**, 21–26 (2016).
28. Trott, O. & Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461 (2010).
29. Suzuki, K. *et al.* Design and synthesis of calcium and magnesium ionophores based on double-armed diazacrown ether compounds and their application to an ion sensing component for an ion-selective electrode. *Anal. Chem.* **67**, 324–334 (1995).
30. Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **46**, 3–26 (2001).
31. Lipinski, C. A. Lead- and drug-like compounds: The rule-of-five revolution. *Drug Discov. Today Technol.* **1**, 337–341 (2004).
32. Solankee, A. & Tailor, R. Synthesis, characterisation, antimicrobial evaluation of chalcones and its cyclised product: Phenyl pyrazolines and benzodiazepines. *Chem. Sci. Trans.* **4**, 1057–1065 (2015).
33. Friesner, R. A. *et al.* Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein–ligand complexes. *J. Med. Chem.* **49**, 6177–6196 (2006).
34. Halgren, T. A. *et al.* Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **47**, 1750–1759 (2004).
35. Friesner, R. A. *et al.* Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **47**, 1739–1749 (2004).
36. Schultz, D. C. *et al.* Pyrimidine inhibitors synergize with nucleoside analogues to block SARS-CoV-2. *Nature* <https://doi.org/10.1038/s41586-022-04482-x> (2022).

37. Kumar, S. *et al.* A novel compound active against SARS-CoV-2 targeting uridylyate-specific endoribonuclease (NendoU/NSP15): *In silico* and *in vitro* investigations. *RSC Med. Chem.* **12**, 1757–1764 (2021).
38. Choi, R. *et al.* High-throughput screening of the ReFRAME, pandemic box, and COVID Box drug repurposing libraries against SARS-CoV-2 nsp15 endoribonuclease to identify small-molecule inhibitors of viral activity. *PLoS ONE* **16**, e0250019. <https://doi.org/10.1371/journal.pone.0250019> (2021).
39. Frazier, M. N. *et al.* Characterization of SARS2 Nsp15 nuclease activity reveals it's mad about U. *Nucleic Acids Res.* **49**, 10136–10149 (2021).
40. Tian, C. *et al.* Ff19SB: Amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *J. Chem. Theory Comput.* **16**, 528–552 (2020).
41. Izadi, S., Anandakrishnan, R. & Onufriev, A. V. Building water models: A different approach. *J. Phys. Chem. Lett.* **5**, 3863–3871 (2014).
42. Case, D. A., Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., Cheatham, III, T. E., Cruzeiro, V. W. D., Darden, T. A., Duke, R. E., Ghoreishi, D., Giambasu, G., Giese, T., Gilson, M. K., Gohlke, H., Goetz, A. W., Greene, D., Harris, R., Homeyer, N., Huang, Y., Izadi, S., Kovalenko, A., D. M. Y., P. A. K. AMBER 2019. (2019).
43. Abraham, M. J., van der Spoel, D., Lindahl, E., Hess, B., & the G. development team. GROMACS User Manual version 2019 (2019).
44. Swope, W. C., Andersen, H. C., Berens, P. H. & Wilson, K. R. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J. Chem. Phys.* **76**, 637–649 (1982).
45. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341 (1977).
46. Uberuaga, B. P., Anghel, M. & Voter, A. F. Synchronization of trajectories in canonical molecular-dynamics simulations: Observation, explanation, and exploitation. *J. Chem. Phys.* **120**, 6363–6374 (2004).
47. Sindhikara, D. J., Kim, S., Voter, A. F. & Roitberg, A. E. Bad seeds sprout perilous dynamics: Stochastic thermostat induced trajectory synchronization in biomolecules. *J. Chem. Theory Comput.* **5**, 1624–1631 (2009).
48. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).
49. Åqvist, J., Wennerström, P., Nervall, M., Bjelic, S. & Brandsdal, B. O. Molecular dynamics simulations of water and biomolecules with a Monte Carlo constant pressure algorithm. *Chem. Phys. Lett.* **384**, 288–294 (2004).
50. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
51. Lloyd, S. Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **28**, 129 (1982).
52. Arthur, D. & Vassilvitskii, S. k-means++: The advantages of careful seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms* 1027–1035 (2007).
53. Pérez-Hernández, G., Paul, F., Giorgino, T., De Fabritiis, G. & Noé, F. Identification of slow molecular order parameters for Markov model construction. *J. Chem. Phys.* **139**, 015102 (2013).
54. Scherer, M. K. *et al.* PyEMMA 2: A software package for estimation, validation, and analysis of Markov models. *J. Chem. Theory Comput.* **11**, 5525–5542 (2015).
55. Borrel, A., Regad, L., Xhaard, H., Petitjean, M. & Camproux, A.-C. PockDrug: A model for predicting pocket druggability that overcomes pocket estimation uncertainties. *J. Chem. Inf. Model.* **55**, 882–895 (2015).
56. Le Guilloux, V., Schmidtke, P. & Tuffery, P. Pocket: An open source platform for ligand pocket detection. *BMC Bioinform.* **10**, 168. <https://doi.org/10.1186/1471-2105-10-168> (2009).
57. Vassetz, D., Pagliai, M. & Procacci, P. Assessment of GAFF2 and OPLS-AA General Force Fields in Combination with the Water Models TIP3P, SPCE, and OPC3 for the Solvation Free Energy of Druglike Organic Molecules. *J. Chem. Theory Comput.* **15**, 1983–1995 (2019).
58. Frisch, M. *et al.* Gaussian 09, revision D. 01. (2009).
59. Tan, D., Piana, S., Dirks, R. M. & Shaw, D. E. RNA force field with accuracy comparable to state-of-the-art protein force fields. *Proc. Natl. Acad. Sci.* **115**, E1346–E1355 (2018).
60. Humphrey, W., Dalke, A. & Schulten, K. V. M. D. Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
61. Sayle, R. A. & Milner-White, E. J. RASMOL: Biomolecular graphics for all. *Trends Biochem. Sci.* **20**, 374–376 (1995).
62. American Chemical Society. CAS COVID-19 antiviral candidate compounds dataset. <https://www.cas.org/covid-19-sar-dataset>.
63. Pettersen, E. F. *et al.* UCSF Chimera?A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
64. Nguyen, H., Case, D. A. & Rose, A. S. NGLview-interactive molecular graphics for Jupyter notebooks. *Bioinformatics* **34**, 1241–1242 (2018).

Acknowledgements

This research was supported by MEXT/JSPS KAKENHI Nos. JP19H03191, JP20H05439 and JP21H05510 to A.K. and JP19K23721 to D.P.T., and by MEXT as a “Program for Promoting Researches on the Supercomputer Fugaku” (Application of Molecular Dynamics Simulation to Precision Medicine Using Big Data Integration System for Drug Discovery, JPMXP1020200201 and Biomolecular Dynamics in a Living Cell, JPMXP1020200101) to A.K. This work mainly used computational resources of the supercomputer TSUBAME provided by Tokyo Institute of Technology through the HPCI System Research Project (Project ID: hp200152) and TSUBAME Young and Female Users Support Program (Project ID: tge-20IJ0046). This work partly used computational resources of the supercomputer TSUBAME provided by Tokyo Institute of Technology, FUGAKU through the HPCI System Research Project (Project IDs: hp210029, hp210172, and hp210177), Research Center for Computational Science, The National Institute of Natural Science, and The Institute for Solid State Physics, The University of Tokyo.

Author contributions

D.P.T., Y.T., T.O., R.M., Y.M., and A.K. contributed to design of the work and wrote the manuscript; D.P.T. performed MD simulations, molecular docking, free energy analysis, and RNA modeling; R.M. conducted dynamic domain analysis; T.O. analyzed hydrogen bonds and salt bridges; Y.T. analyzed druggable pockets; D.P.T. and A.K. conducted overall analysis.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-07792-2>.

Correspondence and requests for materials should be addressed to D.P.T. or A.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022