Identification of Potential Inhibitors of the SARS-CoV-2 NSP13 Helicase via Structure-Based Ligand Design, Molecular Docking and Nonequilibrium Alchemical Simulations

Giorgio Di Paco, Marina Macchiagodena , and Piero Procacci*

Dipartimento di Chimica "Ugo Schiff", Università degli Studi di Firenze, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy

E-mail: piero.procacci@unifi.it

Abstract

We have assembled a computational pipeline based on virtual screening, docking techniques, and nonequilibrium molecular dynamics simulations, with the goal of identifying possible inhibitors of the SARS-CoV-2 NSP13 helicase, catalyzing by ATP hydrolysis the unwinding of double or single-stranded RNA in the viral replication process inside the host cell. The druggable sites for broad-spectrum inhibitors are represented by the RNA binding sites at the 5' entrance and 3' exit of the central channel, a structural motif that is highly conserved across coronaviruses. Potential binders were first generated using structure-based ligand techniques. Their potency was estimated by using four popular docking scoring functions. Common docking hits for NSP13 were finally tested using advanced nonequilibrium alchemical techniques for binding

1

free energy calculations on a high-performing parallel cluster. Four potential NSP13 inhibitors with potency from submicrimolar to nanomolar were finally identified.

Introduction

Since their worldwide emergency approval between late 2020 and early 2021, RNA vaccines have been the major pharmaceutical weapon against the COVID-19 outbreak caused by the SARS-CoV-2 coronavirus. These vaccines are composed of the mRNA sequence encoding of the Spike structural protein of the virus, encapsulated in lipid nanoparticles for delivery. Once the RNA strand is released inside the human cells, the viral protein is expressed using the normal cell's manufacturing process and then expelled inducing the immune system to generate specific antibodies targeting the spike antigene. RNA vaccine efficacy was observed to rapidly wane after few months of the full administration cycle[1] as new variants[2] of the virus emerged. Impaired immunity was likely due to the mutational pressure induced by the vaccine itself on the Spike gene in the virus RNA genome.

Waning of the RNA vaccines based on Spike encoding calls for a renovated effort in identifying new antiviral agents to cure the SARS-CoV-2 infection. In the early stage of the pandemic, the research of antiviral agents for SARS-CoV-2 was mainly directed towards two non structural proteins as biological targets, namely the main protease $3CL^{pro}$,[3,4] cleaving the coronavirus polyprotein, and the RNA-dependent RNA polymerase (RdRp), catalyzing the replication of the RNA from the viral RNA template. The rationale for targeting these proteins lies in the fact that the mutation rate of the corresponding genes is believed to be lower than that of the gene involving the exposed structural spike protein.[5] These efforts led to the identification of two compounds, Remdesivir, a prodrug administered via vein injection inhibiting RdRp through its active metabolite GS-441524,[6] and Nirmatrelvir,[7] orally administered in combination with Ritonavir (Paxlovid) to block the $3CL^{pro}$ function by covalent binding. Both these drugs were approved by the FDA to treat patients with

mild-to-moderate COVID-19.

Concerning Paxlovid, recent studies[8–11] shows that SARS-CoV-2 can mutate in ways that make it less susceptible to the drug. In Ref.,[8] reverse genetics revealed that the combination of the E166V and L50F mutations on $3CL^{pro}$ conferred high resistance to the drug in infectious cell cultures. These findings were later confirmed in Ref.[9] where *in vitro* experiments showed that a triple mutation L50F, E166A, and L167F on $3CL^{pro}$, while still allowing binding of the substrate to the enzyme, was associated with a significant loss (6x to 72x) of the effectiveness of inhibition by Nirmatrelvir. Indeed, as reported in May in a biorxiv preprint,[10] some of these $3CL^{pro}$ mutations have been found already circulating in coronavirus-infected people. Moreover, in a recent randomized controlled trial on Paxlovid efficacy on a sample of vaccinated and non-vaccinated patients conducted by Pfizer,[11] "sustained alleviation of all symptoms for four consecutive days was not met, as previously reported," eventually leading to the discontinuation of the enrollment.

The clinical efficacy of Remdesivir remains contentious,[12–14] with World Health Organization (WHO) even recommending against using the drug for COVID-19 after results from a large randomized controlled trial.[15] Moreover, anectodical emergence of a Remdesivir resistance mutation during treatment of SARS-CoV-2 infection has been recently reported, identifying the E802D mutation of RdRp as the main cause for impairing drug efficacy.[16]

The capability of the SARS-CoV-2 virus to evade drug effectiveness by selecting mutations with good fitness on the two specific targets $3CL^{pro}$ and RdRp reduces the hope of using drugs targeting these non structural proteins as broad-spectrum effective antiviral agents in the treatment of COVID-19 as well as in other coronavirus infections that may arise in the future.

In this respect, the SARS-CoV-2 helicase NSP13, belonging to the helicase superfamily 1B, may represent a valid alternative anti-coronavirus biological target.[17] This non structural motor protein, as a component of the Replication Transcription Complex (RTC) of SARS-CoV-2 forming in the host cell upon infection, catalyzes the unwinding of the RNA activated

by the hydrolysis of a nucleotide triphosphate (ATP). NSP13 (see Figure 1) is a 67 kDa protein composed of five domains, a Zinc Binding Domain (ZBD), whose sole role is that of linking NSP13 to NSP8 in the RTC,[17,18] the 1A and 2A twin domains and the 1B domain, delimiting the central channel for RNA unwinding, and the so-called stalk domain. Domains 1A, 2A, 1B and stalk are essential for helicase activity.[19]
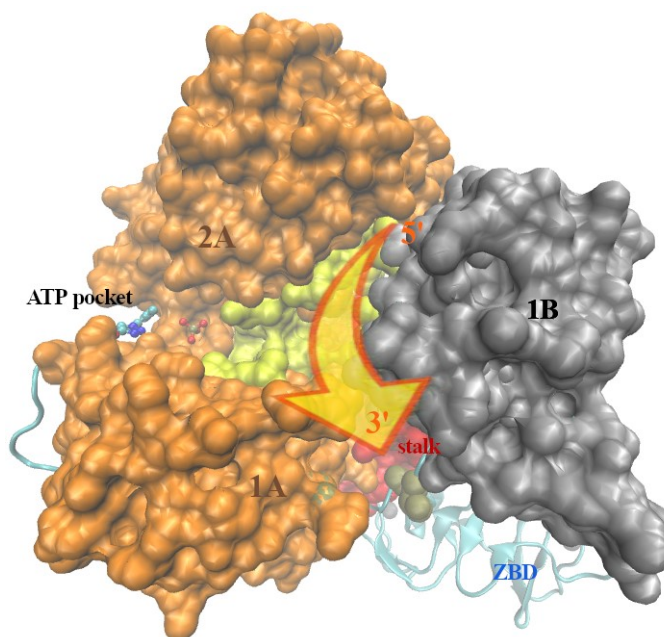

Figure 1: SARS-CoV-2 NSP13 helicase. 1A and 2A domains in orange, 1B domain in grey, stalk in red, ZBD in cyan and central channel residues in yellow. The arrow indicates the sense of sliding, from 5' to 3', of the RNA substrate.

The NSP13 helicase exhibits a remarkable sequence conservation across the coronaviridae family especially in the residues delimiting the central channel. Figure 2 shows[17] that 26 residues of the channel are conserved not only in SARS-CoV, SARS-CoV-2 and MERS human coronavirus, but also in many non-human beta as well as in alphacoronavirus with only two exceptions involving the point mutations R178K and A316S. Helicase inhibition is not a novel anti-viral strategy and potent inhibitors of herpes simplex virus, shingles, and hepatitis C viruses have been reported.[20] In contrast, coronavirus helicases inhibition has so far been underexplored.[21]

In this respect, in the summer of 2022, the Critical Assessment of Computational Hit-

finding Experiment[22] (CACHE), a joint private and governmental initiative for speeding up *in silico* hit-finding predictions have launched an international challenge to find ligands targeting the conserved RNA binding site of SARS-CoV-2 NSP13 using computational approaches. Each of the 23 computational groups joining the challenge was required to provide no more than 100 hit compounds. Nearly 2000 *in silico* hits compounds were experimentally assessed by CACHE and only preliminary results have been so far disclosed. The competition sorted out 48 compounds out of the 2000 hits with an inhibition power of no less than 50 μM against the SARS-CoV-2 helicase NSP13 in a Surface Plasmon Resonance (SPR) binding assay. These 48 compounds are advanced in the second round of the CACHE challenge, the so-called hit-expansion phase currently *in fieri*.

In this computational study, we have implemented a computational pipeline based on virtual screening, docking techniques, and nonequilibrium molecular dynamics simulations, with the aim of identifying possible NSP13 inhibitors blocking the entrance and/or exit of the RNA central channel. The approach started from the generation of 321 compounds using the LIGANN/PLAYMOLECULE interface[23] provided by the company Accelera, a tier Contract Research Organization (CRO) working and collaborating with more than 300 pharmaceutical and biotechnology companies around the world. Compounds were generated using a neural network algorithm based on the shape and chemical-physical matching of the 3' and 5' exit and entrance of NSP13 channel. Two alternative ligand positions, one internal and one external on each of the two RNA channel entry points were tested.
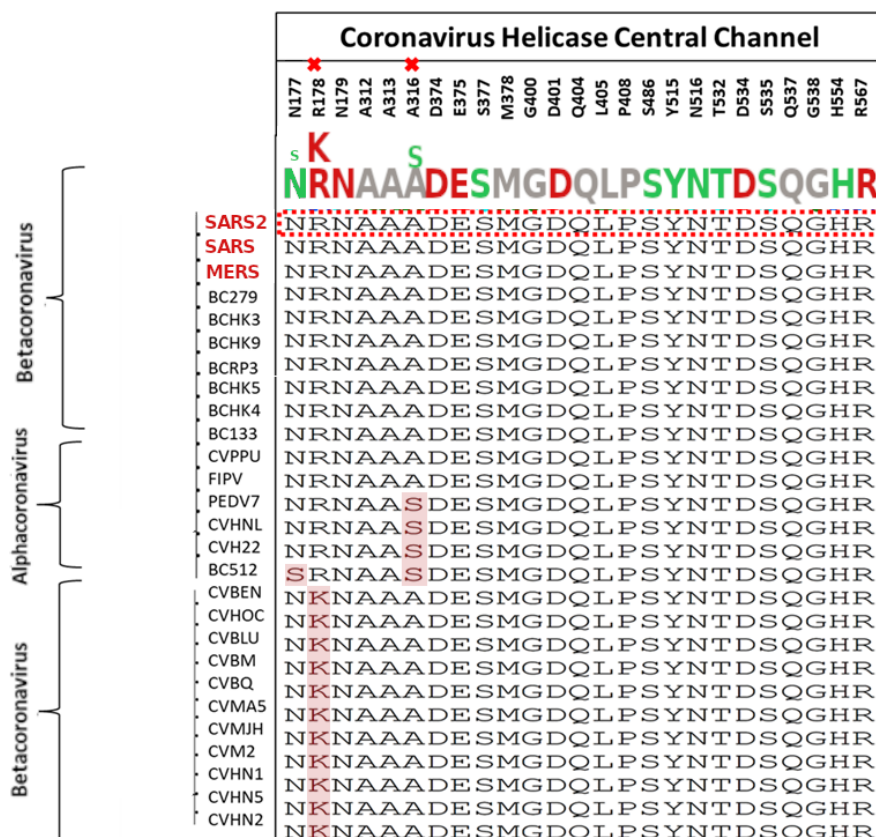
Figure 2: Conservation of NSP13 central channel across coronaviruses (in red BCXXX: bat coronavirus; CVHXX: human coronavirus; CVBXX: bovine coronavirus; CVMXXX: murine coronavirus; FIPV: feline coronavirus; for full abbreviation meaning see Ref.[24]).

The LIGANN-matched compounds for the four possible binding sites were then fed to four popular docking scoring functions based on the Autodock code.[25] Common docking hits were further validated by assessing the stability of the ligand poses using standard molecular dynamics simulations extending for tens of nanoseconds. The selected ligand-receptor pairs were finally advanced to nonequilibrium alchemical simulations on the Marconi/CINECA HPC cluster using the recently proposed virtual double system single box (vDSSB) methodology for absolute binding free energy (BFE) estimation.[26] The computational pipeline eventually delivered four potential inhibitors of the NSP13 helicase with predicted activity from low micromolar to nanomolar.

Results and Discussion

Ligand design by adversarial generative neural networks

The starting featurization of compounds for NSP13 inhibition was performed using the software LIGANN,[23] a structure-based *de novo* drug design tool based on generative neural-networks. Given a binding pocket shape, sensed around the user-provided box center of an input receptor PDB file, LIGANN implements a generative adversarial network[27] producing in a multimodal fashion complementary ligand shapes that are eventually decoded into SMILES strings.[28] Starting from the protein chain B of the 7NNG X-ray PDB structure of NSP13,[17] we hence selected four possible locations for the ligand pose. To this end, through the use of the molecular graphics toolkit VMD,[29] we identified the 5' 3' endpoints of the NSP13 central channel of interest (see Figure 1), and obtained the coordinates of four points along the axis of the channel using pairs of residues as a reference (see Section Experimental) to be fed as box centers for four independent LIGANN runs. These four ligand locations along the channel are labeled as 3'(est), 3'(int) 5'(int), and 5'(est), as they approximately mark the external and internal position of the the 3' and 5' entrances of the NSP13 central channel for RNA substrate.

In Table 1 we report the main chemical features of the LIGANN-generated compounds in the four channel locations. In general, we note that LIGANN-generated compounds are bulkier when posed on the external ends of the channel. At the 5'(est) endpoint, for example, ligands are found with a mean MW of about 300 Da, bearing at least two rings and exhibiting significant flexibility, as measured by the number of rotatable bonds. The internal poses of the RNA channel, both on the 3' and 5' sides, are occupied in general by lighter ligands with reduced flexibility. The water/octanol partition coefficients are found in an optimal range for absorption (1.5-3) for all four locations, reflecting the amphiphilic character of the central channel (see Figure 2).

Table 1: Ligand main features obtained with LIGANN. LogP: o/w partition coefficient; $n_{\mathrm{Ha/d}}$: number of Hbond acceptor or donors; MW: molecular weight (in kDa); n_{ring}: number of rings; n_{rot}: number of rotatable bonds; n: total number of compounds generated by LIGANN.

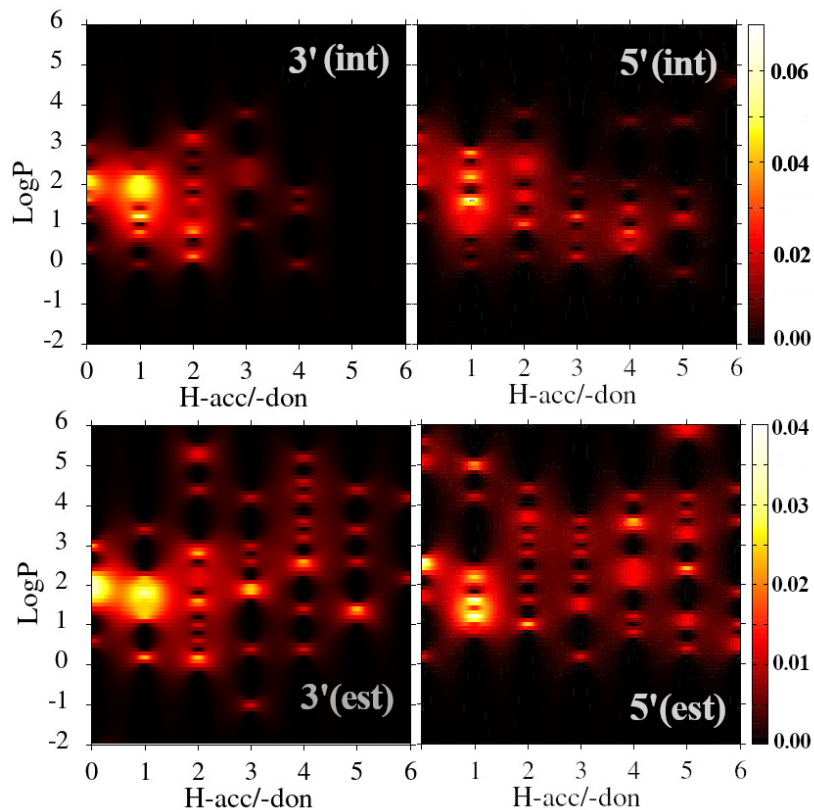|	3'(est)	3'(int)
$\langle \mathrm{Log}P \rangle$	2.2 ± 1.4	1.8 ± 1.1
$\langle n_{\mathrm{Ha/d}} \rangle$	2.3 ± 2.1	$1.4 \pm 1.$
$\langle \mathrm{MW} \rangle$	0.25 ± 0.11	0.18 ± 0.62
$\langle n_{\mathrm{ring}} \rangle$	2.3 ± 1.2	1.4 ± 0.7
$\langle n_{\mathrm{rot}} \rangle$	4.1 ± 2.0	1.9 ± 2.0
n	81	78
	5'(est)	5'(int)
$\langle \mathrm{Log}P \rangle$	2.8 ± 1.7	1.9 ± 1.0
$\langle n_{\mathrm{Ha/d}} \rangle$	2.8 ± 2.1	2.2 ± 1.8
$\langle \mathrm{MW} \rangle$	0.30 ± 0.121	0.23 ± 0.96
$\langle n_{\mathrm{ring}} \rangle$	2.5 ± 1.1	1.7 ± 1.1
$\langle n_{\mathrm{rot}} \rangle$	6.2 ± 4.9	3.8 ± 3.6
n	75	87

Figure 3: 2D maps of LIGANN-generated compounds on the four binding site of the RNA channel as a function of number of H-bond acceptor/donors and XLOGP3.

In the 2D color maps, we report the incidence of LIGANN-generated compounds for each of the four poses as a function of the water-octanol partition coefficient LogP, calculated using the XLOGP3 software,[30] and of the number of H-bond acceptors or donors, $n_{\mathrm{Ha/d}}$. On the external sites, we notice a much larger variability of the LogP and $n_{\mathrm{Ha/d}}$ of the LIGANN-generated compounds. Major peaks are found at LogP=1-2 and $n_{\mathrm{Ha/d}} \simeq 1$ while many secondary peaks are scattered in a wide region of the LogP/$n_{\mathrm{Ha/d}}$ semi-plane with LogP > 0. When compounds are generated at the internal sites of the 5' 3' endpoints of the channel, we observe a much more peaked distribution with maximum probability at LogP $\simeq 1.5 - 2$ and $n_{\mathrm{Ha/d}} \simeq 1$. This lack of variability for the two essential chemical-physical features describing the capability of H-bond formation and the solubility in aqueous and non-aqueous environment together with the reduced dimension (see Table 1) seems to pinpoint a quite specific pharmacophore for the internal binding sites 5'(int) and 3'(int) of the NSP13 channel.

Docking of the LIGANN-generated compounds

Docking was performed for each LIGANN-generated compound on all four binding sites of the RNA central channel using four different Autodock-derived scoring functions, namely Autodock4, Smina, Vina and Vinardo (see Experimental section for details), hence producing a total of 5136 docking calculations. Results are shown in Figure 4. The shaded areas below the curves indicate the compounds that are found with $\Delta G < -8.2$ kcal/mol ($K_d < 1$ μM). Autodock4 yields 17% of LIGANN compounds with micromolar dissociation constant or less. Vina and Smina produce similar results with on average weaker activity, with only 8% of the compounds with K_d less than 1 μM. The Vinardo scoring function gives most of the 1284 binding free energies between -9 and -4 kcal/mol with only 4% of compounds of micromolar activity or less.
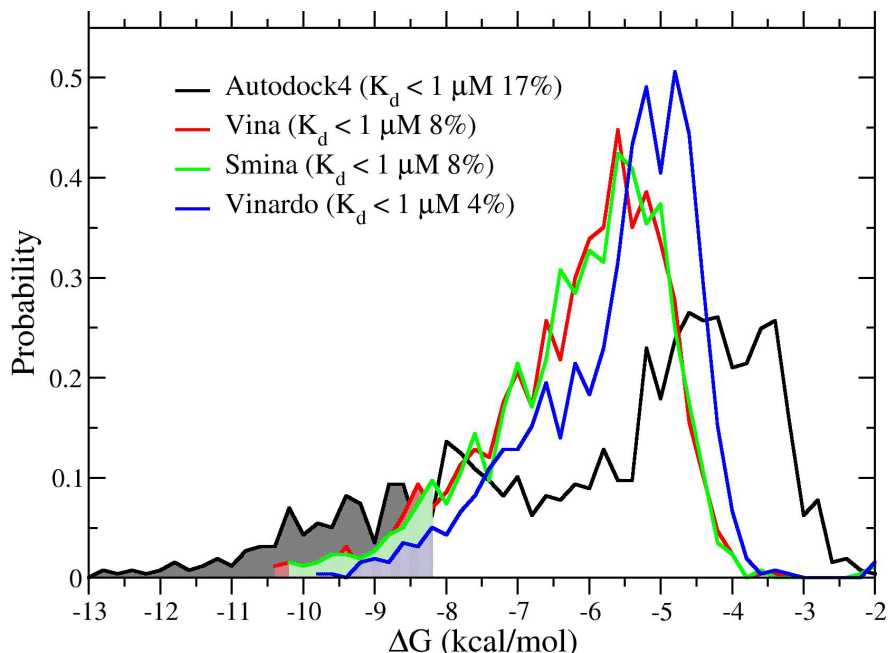
Figure 4: Free energy distribution for various docking scoring functions on all 4 sites of the RNA channel.

Correlation plots between the four docking scoring functions are shown in Figure 5. Correlation, as measured by the Pearson correlation coefficient is good going from a minimum of 0.86 (Autodock4-Vinardo) to a maximum value of 0.99 (Vina-Smina). Ranking, as expressed by the Kendall ranking coefficient τ, is also consistent in all docking scoring functions, ranging from a minimum value of 0.64 (Autodock4-Vinardo) to a maximum of 0.83 (Vina-Smina). BFE ranking, correlation and mean unsigned errors (MUE) are excellent among the Vina, Smina and Vinardo triplet, and nearly ideal for the Smina-Vina pair, showing that the corresponding scoring functions produce in general very similar BFE's. On the other hand, the correlation diagrams with Autodock on the abscissa are all characterized by a slope a that is significantly less than 1, crossing the perfect correlation line at around -6 kcal/mol, and with a strongly left-shifted intercept. These common features indicate that Vina, Smina and Vinardo tend to systematically underestimate the BFE of strong binders and overestimate that of weak binders with respect to Autodock4, hence clustering the BFE's in a narrow range of [-9,-4] kcal/mol. In Ref.,[31] clustering of Vina-calculated binding free energies in

a much narrower range with respect to that obtained using Autodock4 was also observed. Besides, recent benchmarks on extensive protein-ligand database[32] or host-guest systems involving macrocyclic receptor,[31] showed that autodock-derived scoring functions such as those used in Vina, Smina, or Vinardo, while often producing binding poses closer to the experimental counterpart were outperformed by Autodock4 in the binding affinity. Consistently, MUE exceeding 1 kcal/mol are observed between the Autodock4 BFE and the other scoring functions.
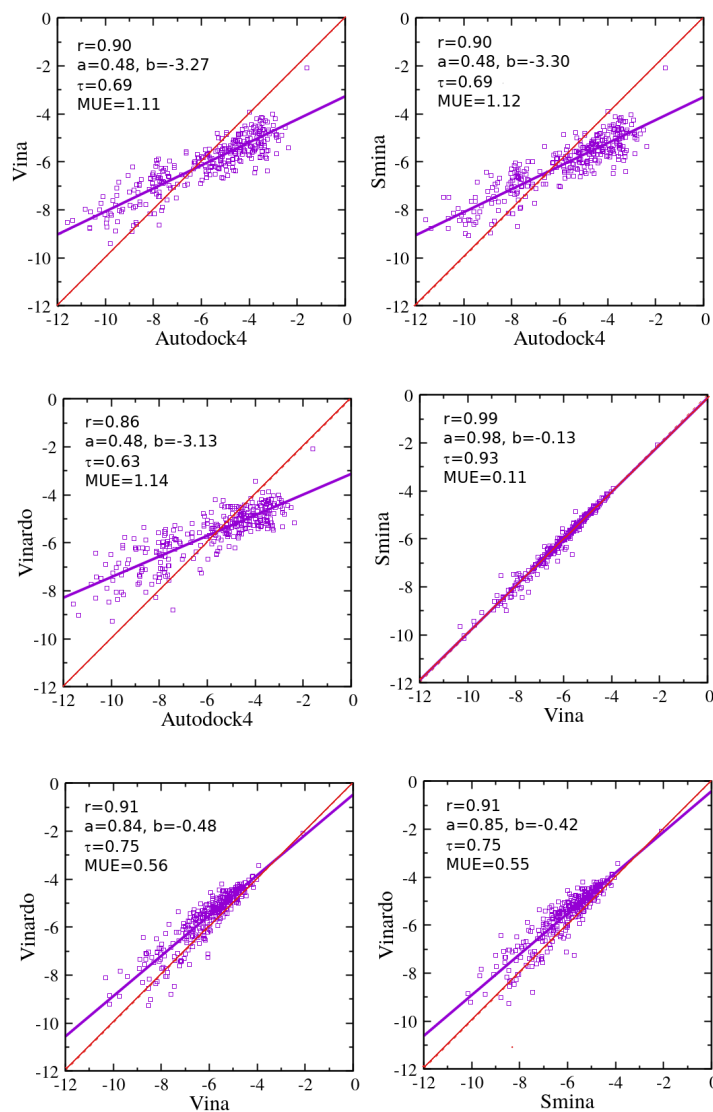

Figure 5: Correlation plot for free energies (in kcal/mol) computed with Autodock4, Vina, Smina and Vinardo scoring functions.

In Figure 6, we show the binding free energy distribution obtained on the four selected binding sites of the central RNA channel using the Autodock4 program. Each of the solid black line distributions is obtained on the subset of LIGANN ligands by centering the docking search box in *the same point* used in the LIGANN search box. Comparison of the distributions in Figure 6 with the distributions of the LIGANN-generated compounds on *all* sites (dashed black line), i.e. irrespectively of the original location selected in the LIGANN calculation, should give a measure of the capability of LIGANN to generate ligands with shape and chemical features matching that of the selected binding site and at the same time, should provide indications of the best binding pose as predicted by the LIGANN-docking pipeline. Inspection of Figure 6 reveals that the external sites 3' and 5' are those where the LIGANN-generated compounds have the best docking performance compared to the same compounds tested on all four sites. Remarkably, more than 60% of the 75 compounds generated by LIGANN on the 5'(est) binding site (see Table 1) have micromolar binding affinity or less according to Autodock4, compared to a value of $\simeq 30\%$ when these compounds are tested with the same docking program on all four binding sites.
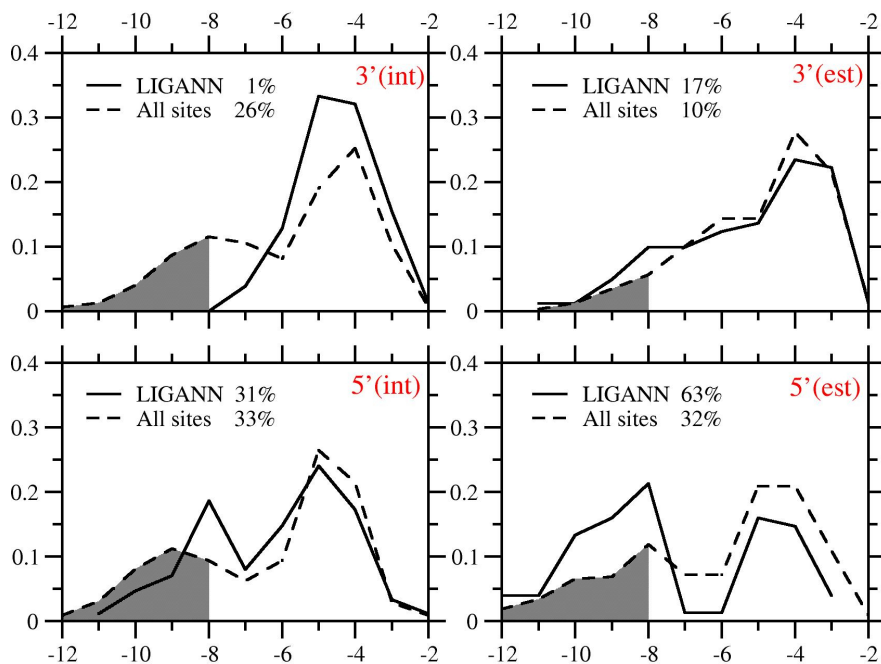
Figure 6: Normalized free energy distribution for the Autodock4 docking scoring function on the 4 binding poses labeled 3'(int), 3'(est), 5'(int), and 5'(est) for the LIGANN-generated compounds. In the plots, the percentage of ligands with $K_d < 1\mu M$ are reported for the LIGANN-generated compounds docked in the LIGANN pose (solid line) and for the same compounds on all four sites (dashed line). The shaded area yields the fraction of compounds with $K_d \leq 1 \ \mu M$.

On the site 3'(int), we found only 1% of the 78 LIGANN-generated compound with Autodock4-predicted micromolar activity or less against a 26% value when we dock the same 78 compounds on all four binding sites. The latter value seems to indicate that, rather than a deficiency of the LIGANN-docking pipeline, compounds that are moderately fit for the internal site 3'(int) according to LIGANN (i.e. those clustered around the point [$n_{Ha/d}$=1, LogP=2] in the 2D reported in Figure 3) are likely good binders for the external sites also.

In Table 2, we finally report the ten compounds that are found with $K_d < 1\mu M$ ($\Delta G_{bind} < -8.2$ kcal/mol) when averaged over all four docking scoring functions. None of these ten compounds are commercially available according to the PubChem database.[33] The BFEs reported in bold were obtained by docking the corresponding compound on the same box center used in the LIGANN calculation. As inferred by Figure 6, and given the high correla-

13

tion between the four scoring functions (see Figure 5), we note that 5'(est) appears to be the best druggable site for all docking scoring functions irrespective of the LIGANN generation site. So, for example, the compound 79 generated by LIGANN when centering the box on the internal location at the RNA entry site 5'(int) has a BFE of only -6.8 kcal/mol according to Autodock4 when posed in the same LIGANN site, while its BFE rises to -11.4 kcal/mol when docking is performed by centering the box at the 5'(est) location.

Table 2: BFE of the best 10 hits (in kcal/mol) obtained from the four docking programs on each of the four sites. The first column refers to the LIGANN index. BFE referring to the original LIGANN sites are marked in bold.

n	ATD4	VI	SM	VNR	mean
	3'(est)				
52	**-10.8**	**-8.0**	**-7.5**	**-7.7**	-8.5
23	-9.7	-8.2	-8.1	-8.3	-8.6
43	-9.3	-8.2	-8.2	-7.2	-8.2
79	**-8.7**	**-8.6**	**-7.5**	**-7.1**	-8.0
62	-8.4	-7.9	-7.7	-7.3	-7.8
44	**-8.1**	**-8.2**	**-8.2**	**-8.3**	-8.2
57	**-8.6**	**-7.8**	**-8.0**	**-8.1**	-8.1
12	**-9.5**	**-7.1**	**-7.5**	**-6.1**	-7.6
7	**-7.4**	**-7.6**	**-7.5**	**-8.8**	-7.8
	3'(int)				
52	-11.7	-8.3	-9.3	-8.3	-9.4
23	**-10.4**	**-9.4**	**-9.4**	**-8.1**	-9.3
43	-10.9	-10.1	-10.1	-8.8	-10.0
79	-10.0	-9.3	-9.7	-8.6	-9.4
62	-9.7	-9.1	-9.1	-7.5	-8.8
44	-8.9	-9.0	-8.9	-8.1	-8.7
57	-9.8	-8.3	-8.7	-7.2	-8.5
12	-10.7	-8.8	-8.8	-7.2	-8.9
7	-8.8	-8.7	-8.3	-8.6	-8.6
	5'(est)				
79	-11.5	-10.5	-10.2	-9.8	-10.5
52	-13.8	-10.3	-10.1	-7.8	-10.5
3	-12.6	-9.7	-9.6	-7.8	-9.9
43	-12.7	-10.0	-10.0	-8.5	-10.3
62	-11.2	-10.1	-10.1	-8.5	-10.0
12	-12.0	-9.6	-9.7	-8.0	-9.8
44	-11.1	-9.8	-9.8	-8.6	-9.8
57	-10.7	-9.2	-9.4	-9.0	-9.6
23	-9.0	-7.2	-7.2	-6.2	-7.4
	5'(int)				
79	-10.4	-9.3	-9.1	-7.3	-9.0
52	-10.9	-8.6	-8.9	-8.3	-9.2
3	-11.4	-8.7	-8.3	-9.6	-9.5
43	**-9.9**	**-8.4**	**-9.0**	**-7.3**	-8.7
62	**-10.6**	**-8.7**	**-8.8**	**-6.4**	-8.6
12	-11.9	-7.8	-7.9	-6.7	-8.6
44	-8.6	-8.7	-8.7	-6.8	-8.2
57	-8.9	-8.3	-8.5	-8.2	-8.5
23	-12.4	-8.3	-8.3	-6.6	-8.9

MD preliminary assessment

To proceed further in our computational pipeline, we selected from the docking stage the best four binders by sorting all the 321 LIGANN-generated compounds according to their average BFE score taken over the combination of the four docking functions and the four binding sites. These compounds, shown in Figure 7 have labels 43 (generated with LIGANN on site 5'(int)), 52 (generated with LIGANN on site 3'(est)), 62 (generated with LIGANN on site 5'(int)), and 79 (generated with LIGANN on site 3'(est)). At pH=7, 43, 53, and 79 have one net positive charge, while 62 has two. All compounds are characterized by an extensive linear network of connected aromatic or non-aromatic rings with poor ramification (e.g. 43 and 62), connected by rotatable bonds. Chiral centers are present only on compound 43. Reported pharmacokinetic parameters such ESOL (estimated water solubility), XLOGP3, gastrointestinal (GI) absorption, Hbond donors or acceptors, blood-brain barrier permeability (BBB), and synthetic accessibility, were estimated using the SWISSADME online tool.[34]
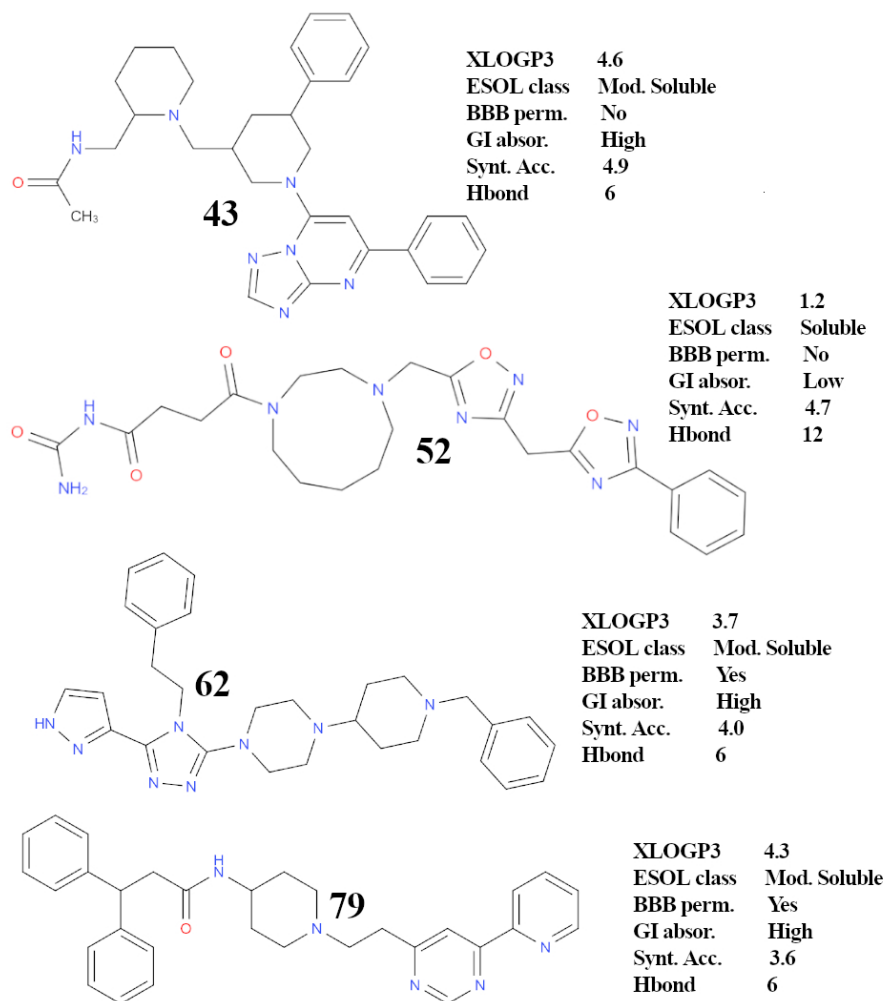
Figure 7: Docking best hit compounds.

As previously discussed, docking calculations in several instances have shown that rather strong BFEs can be obtained when the compound is docked on a different site than that used in the LIGANN calculation. As an example, in Figure 8, we show the best scoring Vinardo poses of the 79 compound when docked using as box center all the four external and internal binding sites along the central channel. Inspection to Table 2 shows that the best Vinardo score for compound 79 (-9.8 kcal/mol) is obtained when centering on 5'(est) rather than on the site 3'(est) (-7.1 kcal/mol) that was originally used for generating the compound with LIGANN. From Figure 8, it should be further noted that the center of mass

(COM) of the best docking pose of 79, while close to the box center 3'(est) used in LIGANN for its generation can be quite far from the alternate box centers as in 3'(int) and 5'(int).
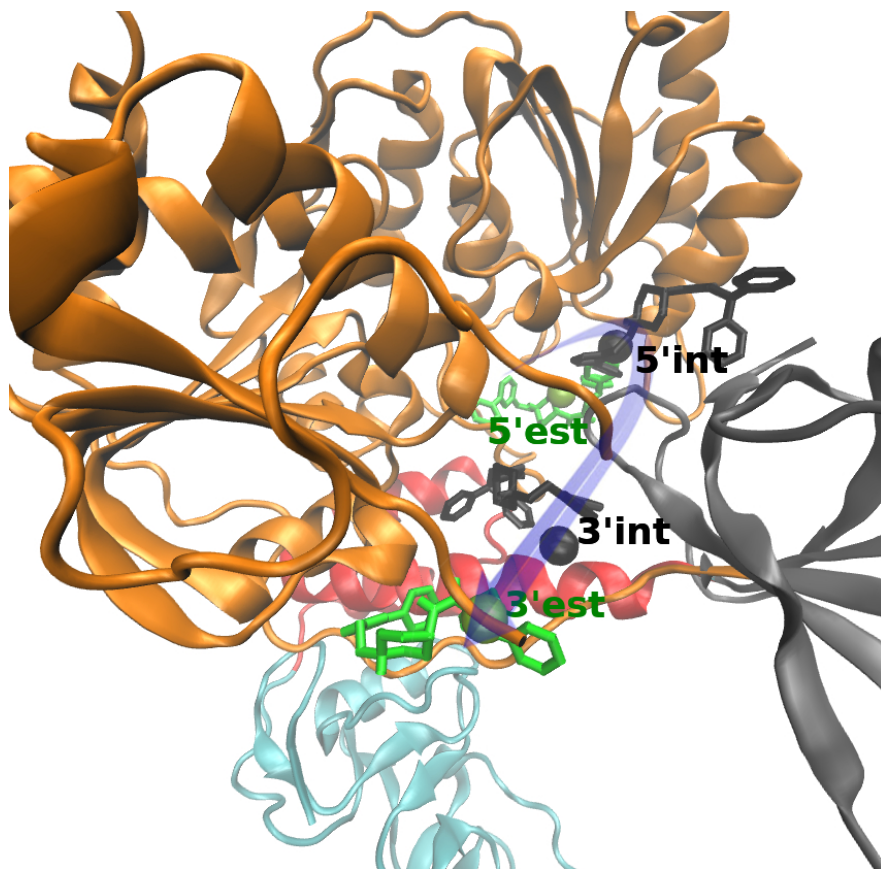

Figure 8: Best Vinardo pose of ligand 79 obtained with the docking search box centered at the 5'(int), 3'(int) (black) 3'(est), and 5'(est) (green). The violet arrow (sense from 5' to 3') marks the central channel of NSP13. 1A and 2A domains in orange, 1B domain in grey, stalk in red, ZBD in cyan, and central channel residues in yellow.

For each of the four compounds of Figure 7, we hence performed four preliminary MD simulations at constant pressure and constant temperature starting from the best Vinardo pose on the sites 3'(int), 3'(est), 5'(int), and 5'(est). Full simulation details can be found in the Methods section. We use the Vinardo pose as the starting structure for the ligand-NSP13 complex since this scoring function is known to provide docked structures that are closer to their experimental counterpart when compared to other Autodock-derived scoring functions[35] such as Smina, Vina or Autodock4. These simulations were aimed at assessing the

stability of the binding states of the ligand-receptor complexes in standard thermodynamic conditions.

In Figure 9, we show the probability distributions of the ligand-receptor COM-COM distance in the eight MD simulations. The significant shift of the COM of the ligand from the original LIGANN box centers, indicated by straight lines in the Figure is a combined effect of the Vinardo and MD calculations. Such shifts are more pronounced on the 5' end, where we even observe an inversion of the distributions of the internal and external sites for compound 52. Several complexes are characterized by distributions with double or multiple peaks, indicating either that the initial Vinardo location of the ligand COM was not optimal or the ligand is swapping between alternate poses of comparable strength in the channel. On the 5' end, the internal and external COM-COM distributions of compound 52 cover a range of more than 10 Å with significant overlap, indicating that the Vinardo poses for this compound at the 5' end are not stable.

Remarkably, compound 79 appears to be stable in all 4 poses with no appreciable merging of the narrow COM-COM distributions. Compound 43 exhibits two stable single poses internal and external at the 3' end, while the mobility of the ligand is more pronounced at the 5' end. Compound 52 exhibits a stable pose at the 5' end inverted internal position (black trait), while the 62-receptor exhibits two stable conformations at the 3' internal location.
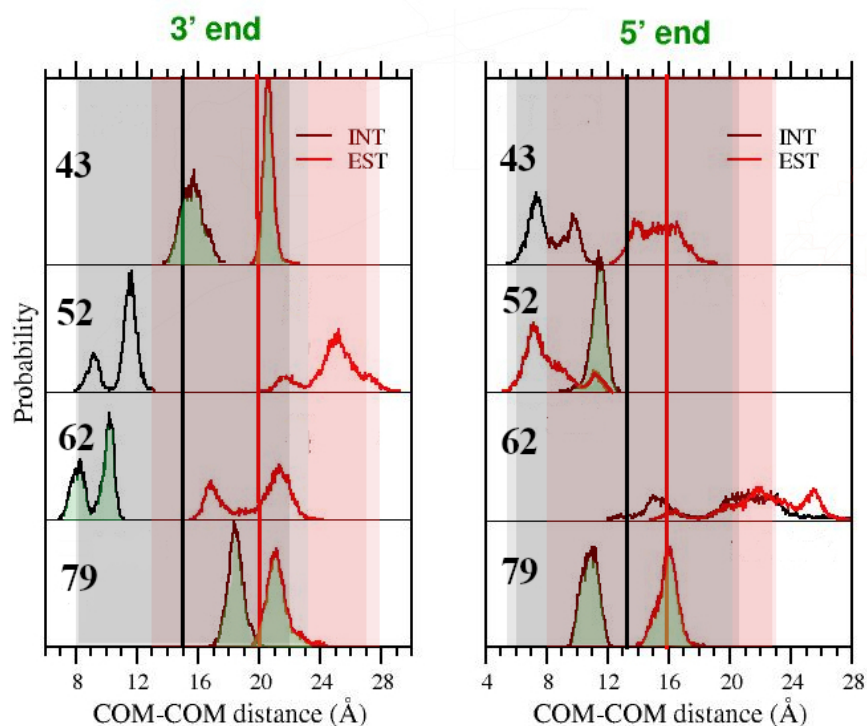
Figure 9: Ligand-receptor COM-COM distribution of the four selected ligands computed in 20 ns of MD simulation. The plots on the left and the right refer to ligands placed in the 3' end and 5' end, respectively. The original LIGANN sites where the compounds were initially generated can be inferred from Table 2. The straight lines mark the distance between the LIGANN box centers and the COM of the receptor. The width of the shaded (red and black) areas corresponds to the LIGANN box size. The green shaded areas mark the distributions of the selected ligand-receptor complexes for nonequilibrium alchemical simulation.

We hence retained eight ligand-receptor complexes to proceed further in our computational pipeline, namely 79 in all four poses, the 62 at 3'(int), 52 at 5'(int), and 43 at 3'(int) and 3'(est). Note that only two (43 at 3'(int) and 79 at 3'(est)) of the 321 original LIGANN hits survived in the docking-MD computational pipeline.

vDSSB alchemical calculations of the BFE

The eight selected complexes based on the MD preliminary assessment (see Figure 9) were further advanced to the last step of the virtual computational pipeline, consisting of a combi-

nation of enhanced sampling MD simulations for generating a canonical sample of the initial equilibrium states,[36] followed by swarms of hundreds of nonequilibrium (NE) alchemically-driven MD trajectories connecting the end-states to finally recover the absolute BFE as a functional of the work distribution using the Jarzynski theorem.[37,38] Both these sequential computational steps, which are based on the generation of concurrent and *independent* MD runs, can be effectively executed on massively parallel facilities with basically no communication overhead.

More specifically, in the first step we collected hundreds of uncorrelated equilibrium configurations of the solvated complex by running replicates of MD simulations for a total time of $\simeq 0.2$ microseconds for each of the eight selected ligand-receptor systems. Independent MD replicates have been shown in several instances[36,39] to provide a sampling as accurate as that of the expensive Replica Exchange Method (HREM).[39,40] For the decoupled ligand, canonical sampling is efficiently obtained performing an HREM simulation of the gas-phase solute molecule alone, combined with equilibrium configurations of the solvent.

In the second alchemical step, two independent parallel jobs, corresponding to NE fast decoupling trajectories of the initially equilibrated ligand in the bound state, and to the NE re-coupling trajectories of the initially equilibrated ligand, concur to the determination of the standard dissociation free energy as a functional, $\mathcal{E}$, of the convolution of the bound-state and unbound state alchemical work distributions, i.e.:

$$
\begin{aligned}
\Delta G^0_{\text{vDSSB}} &= -RT \ln \left[\frac{1}{n_b n_u} \sum_b \sum_u e^{-\beta(W_b + W_u)} \right] \\
&+ RT \ln(V_{\text{site}}/V_0)
\end{aligned}
\tag{1}
$$

where the indices b, u run on the n_b bound state and n_u unbound state NE alchemical work values, and $\beta = 1/k_B T$. The last term on the RHS is a standard state correction term depending on the ratio between the translational volume V_{site} of the ligand in the bound state and of the standard state volume V_0 as thoroughly discussed in Refs.[41–43]

The full process, termed virtual double system single box (vDSSB),[26] corresponds to a non physical alchemical path whereby the ligand is transferred from the bound state to the bulk solvent as if the two independent processes occurred in the same simulation box, with one ligand disappearing on the host and the other materializing in the far distant bulk. NE alchemical simulations and vDSSB have been successfully tested against experimental BFEs in several SAMPL blind challenges[43–47] and in the absolute binding free energy estimation of non-covalent ligands of the SARS-CoV-2 main protease.[26,48] Technical details on the vDSSB implementation are provided in Section Experimental.
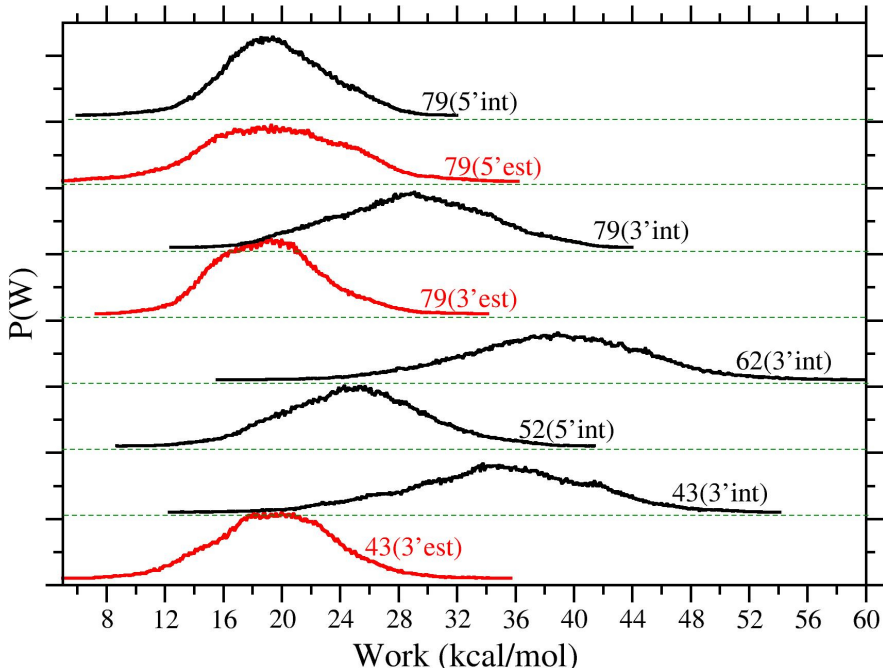

Figure 10: Convolution of the bound state and unbound state work distributions for selected ligands.

In Figure 10, we show the convolution of the unbound (growth) and bound (annihilation) work distributions, $(P_b * P_u)(W) = \int dw P_b(W) P_u(W - w)$ obtained using vDSSB for the eight ligand-receptor poses. The dissociation free energy, ΔG^0, is a function of the mean work value μ and of the variance σ^2 of the convolution. More specifically, for Gaussian or quasi-Gaussian distribution, $\Delta G^0 \simeq \mu - \frac{1}{2}\beta\sigma^2$ increases linearly with the mean work μ and decreases linearly with the variance σ. For non-Gaussian distribution, we use the Jarzynski

22

exponential average corrected for a positive systematic bias proportional to the variance and inversely proportional to the number sampled points $n_u \times n_b$.[44,49,50]

Table 3: vDSSB estimate of the dissociation free energy (in kcal/mol) for the eight NSP13 complexes. V_{corr} is the standard state correction computed from the equilibrium MD replicates of the bound state (see Eq. 2). K_d (M) is the dissociation constant.

CMP	ΔG^0	V_{corr}	K_d
43(3'est)	7.4 ± 1.6	-1.6	4×10^{-6}
43(3'int)	5.6 ± 3.1	-1.9	8×10^{-5}
52(5'int)	8.0 ± 0.8	-2.4	1×10^{-6}
62(3'int)	9.4 ± 1.3	-1.9	1×10^{-7}
79(3'est)	9.7 ± 0.7	-1.4	8×10^{-8}
79(3'int)	10.9 ± 1.5	-1.7	1×10^{-8}
79(5'est)	2.6 ± 1.0	-1.9	1×10^{-2}
79(5'int)	7.9 ± 2.0	-2.5	2×10^{-6}

In Table 3 we finally report the dissociation free energy computed using Eq. 2 for the eight ligand-receptor systems. Six out of the eight combinations compound/location are predicted of micromolar or submicromolar activity, in basic agreement with the docking calculations (see Table 2). Two of the combinations, 79 (5'(est)) and 43 (3'(int)), appear to be false positive, identified in the last stage, vDSSB, of our computational pipeline.

The best binder according to vDSSB is compound 79 at the 3' end of the channel with the external and internal position yielding predicted dissociation constant in the nanomolar range. Not surprisingly, compound 62, with similar chemical-physics features (see Figure 7) is also predicted to be a strong binder with submicromolar activity at the 3'(int) end. Compounds 79 and 62, according to the SWISSADME online platform (see Figure 7), are among the four scrutinized ligands, those with the lowest (most facile) synthetic accessibility.

Binding strength at the 5' end is also notable with 79 and 52 resulting in the low micromolar range. Compound 43 at the 3' end is a much weaker binder although in this case, the spread of the corresponding work distribution functions (see Figure 10) do not allow a sufficiently precise measurement of the vDSSB free energy (first term on the RHS of Eq. 2). At the upper end of the 95% confidence interval for the vDSSB dissociation free energy,

43 is nonetheless predicted of sub-micromolar activity at both the 3'(int) and the 3'(est) locations.

vDSSB, pinpointing the 3' end as the best druggable site, appears to be at variance with earlier predictions based on the ICM software by MOLSOFT L.L.C. (molsoft.com), yielding the 5' end of the central channel as the site with the highest druggable score in the SARS-CoV-2 NSP13 helicase.[17] The ICM method (Pocketfinder)[51] relies on a purely geometric approach on static structures, scanning putative ligand envelopes based on a grid potential constructed using a Lennard-Jones atomic probe. This methodology has the same limitation of the common high-throughput docking approaches with flexible ligand and *rigid* receptor whereby the dynamical adjustments (induced fit) of the protein pocket upon binding is disregarded altogether. Not surprisingly, our docking calculations also seem to indicate the 5' end at the best druggable site (see Table 2). We stress that in the MD-based vDSSB method, induced-fit as well as microsolvation phenomena are explicitly accounted for in standard thermodynamic conditions using accurate atomistic models, hence providing a reliable tool of identifying false positives obtained from pure geometric approaches such as structure-based ligand or pocket probing (LIGANN or ICM), or docking technologies. In this regard, in the Supporting Information (SI), we provide a detailed analysis for the eight selected compounds/locations by comparing the binding pattern of the best docking poses (according to Vinardo) to those obtained from the equilibrium MD simulations of the bound state. While the static docking binding pattern is illustrated by LigPlot diagrams relying on a purely geometric criterion, the strength of specific residue-ligand interactions in the MD equilibrium sampling is represented by heat-maps. Significant difference of the binding modality between the docking approach (lock and key model) and the MD approaches (induced fit) are apparent in most of the instances. Full details can be found in the SI.

Conclusion

We set up a funnel-like computational pipeline aimed at identifying potential inhibitors of the SARS-CoV-2 helicase NSP13 binding to a highly conserved central channel. We initially generated hundreds of fit ligands relying on a structure-based *de novo* drug design based on generative neural networks. We then proceeded with extensive docking screening using various scoring functions. We finally assessed the resulting docking hits using advanced MD-based nonequilibrium alchemical simulations delivering accurate absolute dissociation free energies with credible confidence intervals. We ended up with four compounds (see Figure 7) with predicted micromolar or sub-micromolar activity on at least one of the four identified binding sites along the RNA channel. Although none of the predicted active compounds is commercially available, two of them are in principle easily synthetically accessible. Compound 79, binding at internal 3'(int) location of the NSP13 channel with nanomolar activity, has remarkable structure similiarities with one of the best binders according to preliminary data recently disclosed in round #2 of the ongoing CACHE initiative.[52]

Indeed, the availability of a molecule capable of inhibiting NSP13 by blocking the highly conserved RNA channel could be a good candidate of a broad-spectrum drug, hence potentially providing a powerful asset for the treatment of SARS-CoV-2 infections as well as of future infections mediated by the Coronavirus family. To this end, in the context of the EU-funded National Recovery and Resilience Plan, Tuscany Health Ecosystem (THE), Nanotechnologies for diagnosis and therapy, we are planning to synthesize and test, in wet-lab experiments, the vDSSB hits identified in the present study with the final aim of developing nanosystems for the delivery of a broad-spectrum antiviral agent against pulmonary infections with facile administration.

Methods

LIGANN calculations: The 3D structure of NSP13 in its open form[17] (after ATP hydrolysis into ADP) was downloaded from the protein data bank (PDB code 7NNG). The 7NNG X-ray structure contains two symmetry-related molecules. We have chosen the second molecule (chain B) and fixed the gaps using the Chimera software[53] followed by a minimization of the inserted residues using the VMD toolkit. The resulting PDB file was fed to the LIGANN program. Box centers for the LIGANN calculations are specified in Table 4.

Table 4: Coordintes of the box centers (chain B of 7NNB) for the LIGGAN calculations are defined as the midpoint vectors connecting the C_α of residue 1 to the C_α of residue 2.

	$C_\alpha(1)$	$C_\alpha(2)$	x	y	z
3'(int)	K139	E197	-22.0	16.5	-38.5
3'(est)	T231	P335	-16.5	21.5	-44.5
5'(int)	T413	Y515	-32.0	19.7	-24.8
5'(est)	R178	Y515	-27.4	13.2	-21.0

For the generation of the ligands, we used 10 shapes and 10 decoding smiles for a maximum of 100 generated ligands in each of the four LIGANN runs. The effective number of ligands after removal of duplicate or invalid SMILES code is reported in Table 1 for the four locations. Results of LIGANN calculations, including panels, shape shape-matching outcube files (that can visualized by loading the 7NNG-B coordinates into VMD), are available on the Zenodo open repository.[54]

Docking Calculations: SMILES codes from the LIGANN output were converted to pdb and pdbqt files using OpenBabel toolkit.[55] The chain B of the 7NNG file was used to produce the pdbqt file of the receptor. Docking was performed using the programs Autodock4, Vina, Smina, and Vinardo. Autodock4[25] is based on an empirical scoring function explicitly including atom-atom Lennard-Jones potentials and atomic charges. We used the default atomic charges (Gesteiger) generated using the MGLtools suite.[56] Vina[57,58] uses a simplified approach for electrostatic interactions replaced by knowledge-based hydrophobic and hydrogen bonding terms. Smina,[59] a fork of Vina, supports the customization of the scor-

ing function improving the performance of scoring and minimization workflows. We used the default scoring function in Smina, which is equivalent to Vina with improved empirical parameters.[59] Vinardo,[60] is available as an option in the Smina-fork. The Vinardo scoring function is in essence a modulation of the Vina simplified score, trained on updated state-of-the-art datasets such as PDBBind.[61]

In all docking calculations, the receptor was kept rigid. Docking was performed placing the 321 LIGANN compounds in the four channel locations at 3'(int), 3'(est), 5'(int), 5'(est), and using the four scoring functions hence producing a total of 5136 docking calculations. In all calculations, the search box was cubic with a sidelength of 18 Å and the box centers are those given in Table 4. Complete docking results are available on the Zenodo archive.[54]

MD calculations Preliminary MD calculations of the NSP13 complex with the four ligands, shown in Figure 7, are done using the GROMACS code (2022 release).[62] In all 16 MD simulations, we started from the Vinardo best docking pose of the complex, generating a cubic MD box of $\simeq 96$ Å sidelength filled with $\simeq 27600$ TIP3P[63] water molecules using the standard GROMACS ancillary scripts, eventually resulting in a system containing $\simeq 100000$ atoms. We adopted the AMBER99SB-ILDN force field[64] for the protein and the GAFF2 general force field for the ligand with parameters generated using the PrimaDORAC web interface.[65] We introduced a weak harmonic restraint potential between the COM of the ligand (Vinardo pose) and the COM of the receptor, with the equilibrium distance computed on the initial structure of the complex, and force constant of 120 kJ/nm^2. The systems were initially minimized at 0 K with the steepest descent procedure and subsequently equilibrated at constant volume and at constant temperature (NVT) T=298.15 K, enforced by way of a Bussi thermostat[66] with an integration time step of 1 fs and a thermal friction coupling constant of 4.18 ps for 250 ps. The production stage of 20 ns (preliminary MD assessment) was performed switching to the constant pressure constant temperature ensemble (NPT), using the Parrinello-Rahman barostat,[67] rigid constraints only on the X-H bonds (with X being any heavy atom) and setting the timestep to 2.0 fs. Electrostatic interactions were

treated using particle-mesh Ewald (PME) method[68] with a grid spacing of 1.2 Å and a spline interpolation of order 4. MD assessment data are available in the Zenodo archive.[54]

vDSSB: canonical sampling The NPT canonical (equilibrium) sampling of the bound state for the eight survivor complexes after stability assessment of the docking pose (79 in all four channel locations, 62 at 3'(int), 52 at 5'(int), and 43 at 3'(int) and 3'(est)) was carried out on the Marconi heterogeneous cluster provided by CINECA,[69] by running in parallel, using the `multidir` GROMACS option, 12 replicates of independent NPT simulations lasting 15 ns each and with initial velocities generated by a random seed, hence totaling 180 ns. Due to the chaotic nature of the system, the rate of separation of trajectories starting from the same set of initial coordinates but with different initial velocities is extremely fast with Lyapunov exponents of the order of the ps^{-1}. In each of the replicates, we collected 75 system configurations taken at regular intervals hence harvesting a canonical (equilibrium) sample of 900 space phase points of the bound state whence we randomly extracted 400 initial configurations for the subsequent ligand annihilation in the bound state.

To obtain the initial unbound state configurations with the fully decoupled ligand in bulk water, 400 configurations of the gas-phase isolated ligand were generated at 300 K *via* an HREM simulation lasting 8 ns with a maximum scaling factor of 0.1 for the intramolecular potential energy (corresponding to a "potential temperature" of 3000 K) and 12 replicas with standard scaling protocol based on the geometric progression. These gas-phase HREM simulations on the four compounds in Figure 7 were done using the ORAC software.[70] The HREM-generated 400 gas-phase configurations were inserted in a random position within an equilibrated water box composed of $\simeq 2500 - 3000$ (depending on the ligand size) TIP3P molecules, hence generating 400 initial configurations for the subsequent ligand growth in bulk solvent.

vDSSB NE alchemical calculations. Annihilation of the ligand in the bound state was performed in a single parallel run on the Marconi HPC platform, using again the `multidir` option starting from the bound state canonical sample of 400 configurations. The ligand

was therefore annihilated in 400 NE independent trajectories in the NPT ensemble (each lasting 1.125 ns) linearly switching off sequentially the electrostatic and the Lennard-Jones (LJ) interactions using the GROMACS free energy module. Similarly, the growth process of the ligand in bulk solvent consisted in a parallel job running 400 NE recoupling trajectories with an inverted protocol (switching on LJ first followed by electrostatic interactions), each lasting 1.05 ns. For both annihilation and growth processes we used a soft-core regularization[71] in the LJ stage to avoid catastrophic instabilities during the final (annihilation) and initial (growth) stages of the NE MD trajectories. No *a posteriori* charge correction due to charge annhilation or generation is needed when using GROMACS, as the direct-lattice contribution due to the neutralizing background[72] (the reciprocal lattice contribution is automatically included when using PME in GROMACS) is evaluated on-the-fly during the transition in the `ewald.cpp` routine by a call to the function `ewald_charge_correction`. Input `mdp` files as well as examples of batch submission scripts for the bound and unbound jobs are available in the Zenodo archive.[54]

The bound state and unbound state work values were computed by integrating the corresponding GROMACS-generated `dhdl` files (printed every 15 fs) using the trapezoidal rule. The dissociation free energy was computed from the convolution of the two work samples using Eq. 2, correcting the exponential average for the bias as reported in Refs.[44,49,50] V_{site} in Eq. 2 was computed from the variance of the COM-COM distribution during the equilibrium stage as prescribed in Refs.[41–43]

In the Zenodo archive,[54] work values are provided along with a `bash` script to compute the standard dissociation free energies.

Supporting Information

Analysis of the binding pattern of the NSP13 for the four compounds shown in Figure 7 as obtained in docking and in MD simulations. Additional references cited within the

Supporting Information.[73]

Acknowledgements

Conflict of Interest

The authors declare no conflict of interest.

References

(1) Menegale, F.; Manica, M.; Zardini, A.; Guzzetta, G.; Marziano, V.; d'Andrea, V.; Trentini, F.; Ajelli, M.; Poletti, P.; Merler, S. Evaluation of Waning of SARS-CoV-2 VaccineInduced Immunity: A Systematic Review and Meta-analysis. *JAMA Network Open* **2023**, *6*, e2310650–e2310650.

(2) Variants of the Virus. 2020; https://www.cdc.gov/coronavirus/2019-ncov/variants/, accessed November 13 2023.

(3) Chodera, J.; Lee, A. A.; London, N.; von Delft, F. Crowdsourcing drug discovery for pandemics. *Nat. Chem.* **2020**,

(4) COVID Moonshot. 2020; https://postera.ai/covid, accessed November 13 2023.

(5) Lu, H.; Li, J.; Yang, P.; Jiang, F.; Liu, H.; Cui, F. Mutation in the RNA-Dependent

RNA Polymerase of a Symbiotic Virus Is Associated With the Adaptability of the Viral Host. *Front. Microbiol.* **2022**, *13*.

(6) Yan, V. C.; Muller, F. L. Advantages of the Parent Nucleoside GS-441524 over Remdesivir for Covid-19 Treatment. *ACS Med. Chem. Lett.* **2020**, *11*, 1361–1366.

(7) Halford, B. Pfizer unveils its oral SARS-CoV-2 inhibitor. *Chem. Eng. news* **2021**, *99*.

(8) Zhou, Y.; Gammeltoft, K. A.; Ryberg, L. A.; Pham, L. V.; Tjrnelund, H. D.; Binderup, A.; Hernandez, C. R. D.; Fernandez-Antunez, C.; Offersgaard, A.; Fahne, U.; Peters, G. H. J.; Ramirez, S.; Bukh, J.; Gottwein, J. M. Nirmatrelvir-resistant SARS-CoV-2 variants with high fitness in an infectious cell culture system. *Sci. Adv.* **2022**, *8*, eadd7197.

(9) Jochmans, D.; Liu, C.; Donckers, K.; Stoycheva, A.; Boland, S.; Stevens, S. K.; Vita, C. D.; Vanmechelen, B.; Maes, P.; Treb, B.; Ebert, N.; Thiel, V.; Jonghe, S. D.; Vangeel, L.; Bardiot, D.; Jekle, A.; Blatt, L. M.; Beigelman, L.; Symons, J. A.; Raboisson, P.; Chaltin, P.; Marchand, A.; Neyts, J.; Deval, J.; Vandyck, K. The Substitutions L50F, E166A, and L167F in SARS-CoV-2 3CLpro Are Selected by a Protease Inhibitor In Vitro and Confer Resistance To Nirmatrelvir. *mBio* **2023**, *14*, e02815–22.

(10) Sedova, M.; Jaroszewski, L.; Iyer, M.; Godzik, A. Monitoring for SARS-CoV-2 drug resistance mutations in broad viral populations. *bioRxiv* **2022**,

(11) Inc., P. Pfizer Reports Additional Data on PAXLOVID Supporting Upcoming New Drug Application Submission to U.S. FDA. 2023; sec.gov website, tiny URL https://tinyurl.com/yr9zrkn9, acceded 20 November 2023.

(12) Yan, V. C.; Muller, F. L. Why Remdesivir Failed: Preclinical Assumptions Overestimate the Clinical Efficacy of Remdesivir for COVID-19 and Ebola. *Antimicrob. Agents Chemother.* **2021**, *65*, 10.1128/aac.01117–21.

(13) Amirizadeh, M.; Kharazmkia, A.; Sharifi abdoli, K.; Hayati abbarik, H.; Azimi, G. The effect of remdesivir on mortality and the outcome of patients with COVID-19 in intensive care unit: A casecontrol study. *Health Sci. Rep.* **2023**, *6*, e1676.

(14) Grundeis, F.; Ansems, K.; Dahms, K.; Thieme, V.; Metzendorf, M.-I.; Skoetz, N.; Benstoem, C.; Mikolajewska, A.; Griesel, M.; Fichtner, F.; Stegemann, M. Remdesivir for the treatment of COVID19. *Cochrane Database Syst. Rev.* **2023**,

(15) WHO Guideline Development Group advises against use of remdesivir for covid-19. 2021; https://www.bmj.com/company/newsroom/who-guideline-development-group-advises-against-use-of-remdesivir-for-covid-19/, accessed November 13 2023.

(16) Gandhi, S.; Klein, J.; Robertson, A. J.; Peña-Hernández, M. A.; Lin, M. J.; Roychoudhury, P.; Lu, P.; Fournier, J.; Ferguson, D.; Mohamed Bakhash, S. A. K.; Catherine Muenker, M.; Srivathsan, A.; Wunder, E. A.; Kerantzas, N.; Wang, W.; Lindenbach, B.; Pyle, A.; Wilen, C. B.; Ogbuagu, O.; Greninger, A. L.; Iwasaki, A.; Schulz, W. L.; Ko, A. I. De novo emergence of a remdesivir resistance mutation during treatment of persistent SARS-CoV-2 infection in an immunocompromised patient: a case report. *Nat. Commun.* **2022**, *13*, 1547.

(17) Newman, J. A.; Douangamath, A.; Yadzani, S.; Yosaatmadja, Y.; Aimon, A.; Brandão-Neto, J.; Dunnett, L.; Gorrie-stone, T.; Skyner, R.; Fearon, D.; Schapira, M.; von Delft, F.; Gileadi, O. Structure, mechanism and crystallographic fragment screening of the SARS-CoV-2 NSP13 helicase. *Nature Communications* **2021**, *12*, 4848.

(18) Yan, L.; Zhang, Y.; Ge, J.; Zheng, L.; Gao, Y.; Wang, T.; Jia, Z.; Wang, H.; Huang, Y.; Li, M.; Wang, Q.; Rao, Z.; Lou, Z. Architecture of a SARS-CoV-2 mini replication and transcription complex. *Nature Communications* **2020**, *11*, 5874.

(19) Yue, K.; Yao, B.; Shi, Y.; Yang, Y.; Qian, Z.; Ci, Y.; Shi, L. The stalk domain of SARS-CoV-2 NSP13 is essential for its helicase activity. *Biochem. Biophys. Res. Commun.* **2022**, *601*, 129–136.

(20) Shadrick, W. R.; Ndjomou, J.; Kolli, R.; Mukherjee, S.; Hanson, A. M.; Frick, D. N. Discovering New Medicines Targeting Helicases: Challenges and Recent Progress. *J. Biomol. Screen.* **2013**, *18*, 761–781.

(21) Yazdi, A. K.; Pakarian, P.; Perveen, S.; Hajian, T.; Santhakumar, V.; Bolotokova, A.; Li, F.; Vedadi, M. Kinetic Characterization of SARS-CoV-2 nsp13 ATPase Activity and Discovery of Small-Molecule Inhibitors. *ACS Infect. Dis.* **2022**, *8*, 1533–1542.

(22) FINDING LIGANDS TARGETING THE CONSERVED RNA BINDING SITE OF SARS-CoV-2 NSP13 https://cache-challenge.org/competitions/competition-2. 2022; `https://cache-challenge.org/competitions/competition-2`, Accessed 20 February 2024.

(23) LiGANN software by Accelera. https://www.playmolecule.com/LiGANN/ (accessed 19 October 2023).

(24) Yazdani, S.; De Maio, N.; Ding, Y.; Shahani, V.; Goldman, N.; Schapira, M. Genetic Variability of the SARS-CoV-2 Pocketome. *J. Proteome Res.* **2021**, *20*, 4212–4215.

(25) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **2009**, *30*, 2785–2791.

(26) Macchiagodena, M.; Pagliai, M.; Karrenbrock, M.; Guarnieri, G.; Iannone, F.; Procacci, P. Virtual double-system single-box: A nonequilibrium alchemical technique for absolute binding free energy calculations: Application to ligands of the SARS-CoV-2 main protease. *J. Chem. Theory Comput.* **2020**, *16*, 7160–7172.

(27) Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Commun. ACM* **2020**, *63*, 139–144.

(28) Skalic, M.; Jiménez, J.; Sabbadin, D.; De Fabritiis, G. Shape-based generative modeling for de novo drug design. *J. Chem. Inf. Model.* **2019**, *59*, 1205–1214.

(29) Humphrey, W.; Dalke, A.; Schulten, K. VMD – Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38.

(30) Cheng, T.; Zhao, Y.; Li, X.; Lin, F.; Xu, Y.; Zhang, X.; Li, Y.; Wang, R.; Lai, L. Computation of Octanol−Water Partition Coefficients by Guiding an Additive Model

with Knowledge. *J. Chem. Inf. Model.* **2007**, *47*, 2140–2148.

(31) Casbarra, L.; Procacci, P. Binding free energy predictions in host-guest systems using Autodock4. A retrospective analysis on SAMPL6, SAMPL7 and SAMPL8 challenges. *J. Comput.-Aided Mol. Des.* **2021**,

(32) Nguyen, N. T.; Nguyen, T. H.; Pham, T. N. H.; Huy, N. T.; Bay, M. V.; Pham, M. Q.; Nam, P. C.; Vu, V. V.; Ngo, S. T. Autodock Vina Adopts More Accurate Binding Poses but Autodock4 Forms Better Binding Affinity. *J. Chem. Inf. Model.* **2020**, *60*, 204–211.

(33) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; Zaslavsky, L.; Zhang, J.; Bolton, E. E. PubChem 2023 update. *Nucl. Acids Res.* **2022**, *51*, D1373–D1380.

(34) Daina, A.; Michielin, O.; Zoete, V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* **2017**, *7*, 1–13.

(35) Quiroga, R.; Villarreal, M. A. Vinardo: A Scoring Function Based on Autodock Vina Improves Scoring, Docking, and Virtual Screening. *PLOS ONE* **2016**, *11*, 1–18.

(36) Bieniek, M. K.; Bhati, A. P.; Wan, S.; Coveney, P. V. TIES 20: Relative Binding Free Energy with a Flexible Superimposition Algorithm and Partial Ring Morphing. *J. Chem. Theory Comput.* **2021**, *17*, 1250–1265.

(37) Jarzynski, C. Nonequilibrium equality for Free energy differences. *Phys. Rev. Lett.* **1997**, *78*, 2690–2693.

(38) Crooks, G. E. Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems. *J. Stat. Phys.* **1998**, *90*, 1481–1487.

(39) Procacci, P. Dealing with Induced Fit, Conformational Selection, and Secondary Poses in Molecular Dynamics Simulations for Reliable Free Energy Predictions. *J. Chem. Theory Comput.* **2023**, *19*, 8942–8954.

(40) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141–151.

(41) Procacci, P. I. Dissociation free energies of drug-receptor systems via non-equilibrium alchemical simulations: a theoretical framework. *Phys. Chem. Chem. Phys.* **2016**, *18*, 14991–15004.

(42) Nerattini, F.; Chelli, R.; Procacci, P. II. Dissociation free energies in drug-receptor systems via nonequilibrium alchemical simulations: application to the FK506-related immunophilin ligands. *Phys. Chem. Chem. Phys.* **2016**, *18*, 15005–15018.

(43) Procacci, P.; Guarnieri, G. SAMPL7 blind predictions using nonequilibrium alchemical approaches. *J. Comput.-Aided Mol. Des.* **2021**, *35*, 37–47.

(44) Procacci, P.; Guarnieri, G. SAMPL9 blind predictions using nonequilibrium alchemical approaches. *J. Chem. Phys.* **2022**, *156*, 164104.

(45) Amezcua, M.; El Khoury, L.; Mobley, D. L. SAMPL7 Host-Guest Challenge Overview: assessing the reliability of polarizable and non-polarizable methods for binding free energy calculations. *J. Comput.-Aided Mol. Des.* **2021**, *35*, 1–35.

(46) Amezcua, M.; Setiadi, J.; D.L., M. The SAMPL9 Host-Guest Blind Challenge: An Overview of Binding Free Energy Predictive Accuracy. 2023; Cambridge: Cambridge Open Engage; This content is a preprint and has not been peer-reviewed.

(47) Procacci, P.; Guarnieri, G. SAMPL9 blind predictions for toluene/water partition coefficients using nonequilibrium alchemical approaches. *J. Chem. Phys.* **2023**, *158*, 124117.

(48) Macchiagodena, M.; Karrenbrock, M.; Pagliai, M.; Procacci, P. Virtual Double-System Single-Box for Absolute Dissociation Free Energy Calculations in GROMACS. *J. Chem. Inf. Model.* **2021**, *61*, 5320–5326.

(49) Gore, J.; Ritort, F.; Bustamante, C. Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements. *Proc. Natl. Acad. Sci.* **2003**, *100*, 12564–12569.

(50) Procacci, P.; Macchiagodena, M. On the NS-DSSB unidirectional estimates in the SAMPL6 SAMPLing challenge. *J. Comput.-Aided Mol. Des.* **2021**, *35*, 1055–1065.

(51) An, J.; Totrov, M.; Abagyan, R. Pocketome via Comprehensive Identification and Clas-

sification of Ligand Binding Envelopes*. *Mol. Cell. Proteom.* **2005**, *4*, 752–761.

(52) CACHE#3 Compounds Advancing to Round #2 (hit expansion). 2022; `https://cache-challenge.org/sites/default/files/downloadable/forms/Passed_to_CACHE3_Round2.pdf`, Accessed 20 February 2024.

(53) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF ChimeraA visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.

(54) Di Paco, G.; Macchiagodena, M.; Procacci, P. Identification of Potential inhibitors of the SARS-CoV2 NSP13 Helicase via Structure-Based Ligand Design, Molecular Docking and Nonequilibrium Alchemical Simulations. https://zenodo.org/records/10716268, accessed on 27 February 2024.

(55) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *J. Cheminform.* **2011**, *3*, 1–14.

(56) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comp. Chem.* **2009**, *30*, 2785–2791.

(57) 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.* **2010**, *31*, 455–461.

(58) `https://vina.scripps.edu/manual/`, AutoDock Vina manual https://vina.scripps.edu/manual.

(59) Koes, D. R.; Baumgartner, M. P.; Camacho, C. J. Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. *J. Chem. Inf. Model.* **2013**, *53*, 1893–1904.

(60) Quiroga, R.; Villarreal, M. A. Vinardo: A scoring function based on autodock vina improves scoring, docking, and virtual screening. *PloS one* **2016**, *11*, e0155183.

(61) Li, Y.; Han, L.; Liu, Z.; Wang, R. Comparative Assessment of Scoring Functions on

an Updated Benchmark: 2. Evaluation Methods and General Results. *J. Chem. Inf. Model.* **2014**, *54*, 1717–1736.

(62) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19–25.

(63) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(64) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78*, 1950–1958.

(65) Procacci, P. PrimaDORAC: A Free Web Interface for the Assignment of Partial Charges, Chemical Topology, and Bonded Parameters in Organic or Drug Molecules. *J. Chem. Inf. Model.* **2017**, *57*, 1240–1245.

(66) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling Through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.

(67) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.

(68) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(69) Consorzio Interuniversitario del Nord Est per il Calcolo Automatico (Interuniversity Consortium High Performance Systems) http://www.cineca.it (accessed 07 September 2022).

(70) Procacci, P. Hybrid MPI/OpenMP Implementation of the ORAC Molecular Dynamics Program for Generalized Ensemble and Fast Switching Alchemical Simulations. *J. Chem. Inf. Model.* **2016**, *56*, 1117–1121.

(71) Beutler, T.; Mark, A.; van Schaik, R.; Gerber, P.; van Gunsteren, W. Avoiding sin-

gularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem. Phys. Lett.* **1994**, *222*, 5229–539.

(72) Darden, T.; Pearlman, D.; Pedersen, L. G. Ionic charging free energies: Spherical versus periodic boundary conditions. *J. Chem. Phys.* **1998**, *109*, 10921–10935.

(73) Wallace, A. C.; Laskowski, R. A.; Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng. Des. Sel.* **1995**, *8*, 127–134.

Entry for the Table of Contents

SARS-CoV-2 NSP13 HELICASE
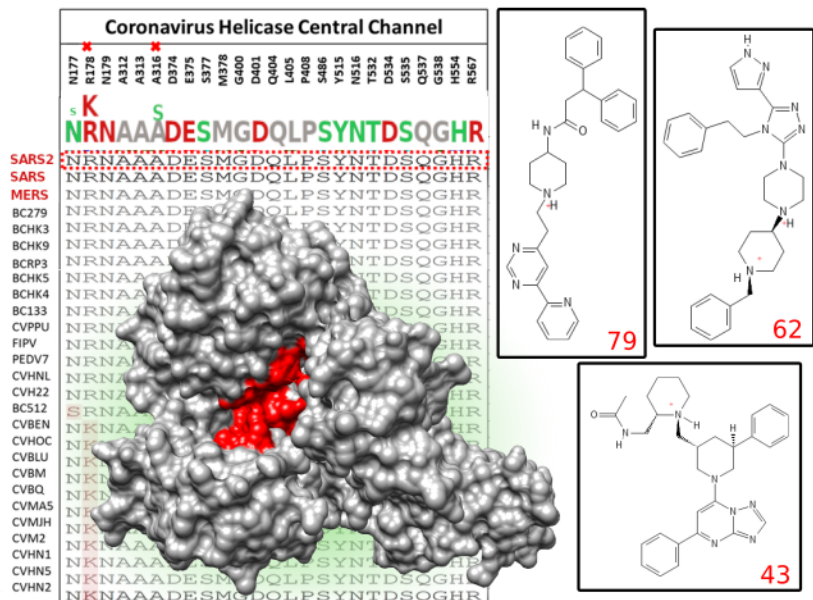
The channel for RNA unwinding of NSP13 SARS-CoV-2 helicase is highly conserved in alpha and beta mammals coronavirus. Exploiting a funnel-like computational pipeline, we identified several micromolar or submicromolar NSP13 ligands blocking the central channel of NSP13. These compounds could potentially be good candidates for the development of a broad-spectrum drug for coronavirus infections.