In silico screening for natural ligands to non-structural nsp7 conformers of coronaviruses

Blasco, R. and Coll, J*.
Department of Biotechnology. Instituto Nacional Investigaciones y Tecnologías Agrarias y Alimentarias, INIA. Madrid, Spain.
short title: ligands to coronavirus nsp7
Keywords: nsp7; conformers; coronavirus; virtual screening; ligands; steroid-like scaffolds; SARS CoV-19
* Corresponding author Email: juliocollm@gmail.com (JC)

Abstract: The non-structural protein 7 (nsp7) of Severe Acute Respiratory Syndrome (SARS) coronaviruses was selected as a new target to potentially interfere with viral replication. The nsp7s are one of the most conserved, unique and small coronavirus proteins having a critical, yet intriguing participation on the replication of the long viral RNA genome after complexing with nsp8 and nsp12. Despite the difficulties of having no previous binding pocket, two high-throughput virtual blind screening of 158240 natural compounds > 400 Da by AutoDock Vina against nsp7.1ysy identified 655 leads displaying predicted binding affinities between 10 to 1100 nM. The leads were then screened against 14 available conformations of nsp7 by both AutoDock Vina and seeSAR programs employing different binding score algorithms, to identify 20 consensus top-leads. Further *in silico* predictive analysis of physiological and toxicity ADMET criteria (chemical properties, adsorption, metabolism, toxicity) narrowed top-leads to a few drug-like ligands many of them showing steroid-like structures. A final optimization by search for structural similarity to the top drug-like ligand that were also commercially available, yielded a collection of predicted novel ligands with ~100-fold higher-affinity whose antiviral activity may be experimentally validated. Additionally, these novel nsp7-interacting ligands and/or their further optimized derivatives, may offer new tools to investigate the intriguing role of nsp7 on replication of coronaviruses.

Introduction

Some of the non-structural proteins (nsps) of coronaviruses (CoV), including those causing Severe Acute Respiratory Syndrome (SARS) in 2019, are required for transcription/replication of their long positive-stranded RNA genomes [6, 25]. Coronavirus nsps are intracellularly synthesized during infection as cleavage products of larger viral polyproteins ORF1a and ORF1ab but they are absent in the virions released after infection. Once nsps are translated and processed into monomeric proteins by proteolytic cleavage, some of them form replication complexes with the RNA-dependent RNA polymerase (RdRp, nsp12). For successful viral RNA synthesis, the nsp12 polymerase requires co-factors nsp7 and nsp8 among other nsps [13, 19, 25, 26].

The recombinant nsp12 RdRp of coronaviruses shows an *in vitro* RNA synthesis with low processivity (number of nucleotides polymerized per RdRp/RNA template encounter), conflicting with the expected higher replication rate required for the largest known ~30-kb positive stranded viral RNA genome. It was soon discovered that the addition of recombinant nsp7+nsp8 increased the polymerase activity of nsp12 [27]. Both, nsp7 and nsp8 interact with each other forming hetero-oligomer complexes with restricted short RNA polymeric activity [10]. While pairwise combinations of recombinant nsp7, nsp8 and nsp12 displayed low processivity RNA polymerase activity, *in vitro* RNA synthesis was higher only when, **i)** nsp7, nsp8 and nsp12 were simultaneously present in the mixtures, **ii)** pre-incubated nsp7/nsp8 was added to nsp12, or **iii)** a nsp7-6polyH-nsp8 fusion protein was added to nsp12 [27].

Mutations to Alanine in nsp7 residues K7, H36 and N37 reduced both RNA binding by the nsp7+nsp8+nsp12 complexes and polymerase processivity as well as viral replication as detected by reduced viral plaque sizes [27]. All these results suggested that nsp7 may play a crucial yet intriguing role(s) in coronavirus replication. On the other hand, for optimal RNA binding and processivity, the nsp12 requires not only the presence of nsp7+nsp8 but also association with other nsps. For instance, further association of the nsp7+nsp8+nsp12 complex with nsp14 provides proofreading of the new RNA sequence correcting frequent mispairing to assur coronavirus replication fidelity [19].

Further to the characteristics mentioned above, the nsp7 is one of the smallest coronavirus–coded nsp sequences that translates into an active 83-amino acid protein. Previous studies on nsp7 have shown that, **i)** is highly conserved within the *Coronaviridae* family [8, 16], **ii)** has no detectable orthologues outside coronaviruses, **iii)** is a compact protein including four short α-helices, and **iv)** may participate in lipid interactions as suggested by its association with cytoplasmic membranes during replication.

Models for 3D structures of recombinant nsp7 reported earlier have been derived from either monomeric recombinants or from complexes of nsp7 with recombinant nsp8 or nsp8+nsp12. The isolated recombinant nsp7 3D model solved at pH 7.5 in reducing conditions contains four compact α-helices: α1-helix (residues 11-17), α2-helix (residues 29-42), α3-helix (residues 47-65), and α4-helix (residues 71-81) [16]. The interdigitating α2 and α3 helices may help to stabilize the nsp7 structure as suggested by the strict conservation of most of their critical residues among different coronaviruses [16]. In contrast, small variations in the 3D conformations in the amino and carboxy-terminal sequences occur at different pHs and/or when complexing with nsp8 or with nsp8-nsp12, as shown by structural data obtained by different authors [6, 8, 32, 33]. The enigmatic function of nsp7 and their coronavirus unique sequence/structure suggests a new role for RNA viral replication yet to be discovered [16].

Previous research to develop new anti-viral drugs against coronaviruses employed virtual screening by computational methods based on active drug-like compounds and binding pockets already identifed and mapped, respectively [20]. Work to find new anti-coronavirus drugs with more affinity than those already known were mainly searched among aproved compounds for other diseases (drug repurposing). These research have been targeted to the Angiotensin-converting enzyme II (ACE2) entry receptor, the RdRp, the S spike surface protein and/or the viral protease proteins nsp3 or nsp5 [9, 20, 29]. The expectation being that it will be the faster way to detect more powerful drugs [3]. In contrast, because of the absence of known drugs and/or of identified binding pockets, targeting other coronaviral proteins like nsp7 have remained largely unexplored. Despite the difficulties to screen for any interacting compounds with these viral protein surfaces, we focused on nsp7s by blind searching for innovative ligands.

Rather to the nsp7 surface, the closest study have been reported on the nsp12 surface interacting with nsp7. Thus, the nsp12-nsp7 interface was investigated by virtual screening of 7496 compounds from the ZINC data base (http://zinc15.docking.org/). Ten binding ligands to nsp12 were proposed for experimentation because of their lowest affinities ranging from ~ 350 to 2600 nM. The identified ligands of an average 536.9 Daltons were amongst those previously approved by the Food and Drug Administration (FDA) [20]. To our knowledge, the nsp7 remains unexplored as a new target for possible interferences with RNA replication. Whether nsp7 physiologically exist as different oligomeric structures, conformations and/or become specific for different ligand-binding or biological lipid-related functions remain also unknown.

To virtually explore the nsp7 surfaces for binding ligands, we have selected all their exposed surfaces (blind docking) and tested the 14 nsp7 published conformations. The so called consensus virtual screening strategy useful to reduce false positives [14], has been employed with the expectation of improving prediction success. For that, we have combined AutoDock Vina with seeSAR software since these two predictive programs greatly differ in their ligand conformation or pose optimization algorithms and score functions.

The workflow was organized as follows. First, the number of ligands to be screened were maximized using a high-troughput AutoDock Vina of 158240 non-minimized, easier-to-screen <400 Daltons ligands and one unique nsp7.1ysy conformer for target. Second, 655 leads with mean binding energies < -8.1 kcal/mol equivalent to concentrations < 1100 nM, were defined by using both non-minimized and energy minimized input ligands. Third, we used the output leads as inputs to screen 14 nsp7 conformers by consensus docking with AutoDock Vina and seeSAR to define 20 top-leads. Forth, we filtered the top-leads by its drug-like properties and commercial availability to define one pharmacophore-like molecule containing the minimal molecular skeleton required for specific interactions. Fifth, a structural search for similar compounds yielded an small ligand collection with ~ 100 fold higher-binding affinities among those commercially available ready for experimental validation.

Since we have found no previous reports on nsp7 interactions with small molecules in the literature, the functional significance of the above mentioned ligands requires further experimental validation studies. In addition to their possible antiviral activity, these new molecules may also contribute to clarify the functional role of these intriguing coronavirus-unique viral proteins.

Materials and Methods

Ligands and tridimensional nsp7 models

A total of 158240 ligands <400 Daltons were extracted and splitted in sdf files of 10000 ligands each using a home-made Phyton script to the 325319 molecules random collection (~ 2G size in one spatial data file, sdf) of natural ligands of the SuperNatural II database (http://bioinf-applied.charite.de/supernatural_new/index.php).

To explore 14 available nsp7 conformers, those were downloaded from the RCSB PDB protein data bank (https://www.rcsb.org/) as individual or nsp8/nsp12 complexed *.pdb files before June-July of 2020 (Table 1). When appropriate, the nsp7 individual *.pdb files were extracted from the 3D complexes with either nsp8 or nsp8+nsp12 using PyMOL. Structural similarity to the nsp7.1ysy [16] was estimated by superposing the 3D models using the CCP4 Molecular Graphics program vs2.10.11 (http://www.ccp4.ac.uk/MG). Predicted binding pockets and α-helices were also investigated using the seeSAR vs.10 program (https://www.biosolveit.de/SeeSAR/).

AutoDock Vina virtual screening

The AutoDock Vina program [28] included in the PyRx 0.9.8. package [5] (https://pyrx.sourceforge.io/) was used in e7 desk computers to predict Gibbs free-energy (ΔG) using 60 × 40 × 30 Å grids including the whole nsp7 molecules. The *.sdf files were converted to *.pdbqt files by allowing ligand rotatable bonds, protonation and adding Gasteiger−Marsili partial atomic charges, without (non-minimized) and/or with a prior additional ffu energy minimization step (ffu minimized) using the Open Babel included into the PyRx package. Water molecules were not considered. Only the binding pose with the lowest binding energy of each *.out.pdbqt were retained for further analysis. Time to compute the binding scores of the *.sdf files of 10000 ligands each, took ~ 2 days. High-throughput docking was made both whitout and with energy minimized 158240 ligands to generate 655 consensus leads. In addition, predicted binding energies of the AutoDock Vina leads were also investigated against 14 nsp7 conformers. When required, the output ΔG energies were converted to constant inhibition (Ki) values in molar concentrations (M), using the formula Ki = exp([ΔG × 1000] / [R × T]) (R = 1.98 cal/mol, and T = 298 °C)[23]. Final values were converted to nM to compare them with seeSAR out puts. The predicted structures were visualized in PyRx and/or PyMOL (https://www.pymol.org/). Since the individual *.out.pdbqt files generated by AutoDock reduced the information content of the corresponding *.sdf files, for further analysis all the molecular characteristics of the leads were re-extracted from the initial SNII.sdf file using a home-made Phyton script.

SeeSAR virtual screening

Leads were also screened for binding to the 14 nsp7s using the BioSolveit seeSAR vs.10 package (https://www.biosolveit.de/SeeSAR/). The seeSAR package was chosen because it employs the HYDE scoring function to evaluate HYdration and DEsolvation (as calibrated with octanol/water partition data logP) [21, 22] while AutoDock Vina relies on protonation and charge distribution [28]. Furthermore, to reduce false positives, the seeSAR FlexX function included into its calculations unfavorable interactions [18]. A 100-fold nM range of concentrations (lower and higher boundaries) is estimated by seeSAR for each of 10 binding conformations (poses) per ligand. Only one pose per ligand corresponding to the smaller lower boundary of estimated affinity were retained for further analysis, using a home-made Excell Macro. When required, the output values in molar concentrations were converted to ΔG energies, using the formula ΔG = R × T × Ln(lower boundary). The predicted structures were visualized in seeSAR.

In silico analysis of pharmacokinetic parameters, and toxicity properties

To assess drug-like possibilities, the water solubility, partition coefficient between *n*-octanol and water (logP), hydrogen-bond acceptors, hydrogen-bond donors, violations of Lipinski's <5 rules, physiological absorption predictions, and some other ADME predictions were downloaded from the SwissADME web server for top-leads provided in their SMILES format (http://www.swissadme.ch/). ADMET risk parameters for toxicological assessments [1] such as cellular, brain or intestinal absorption, tissue distribution, detoxyfying metabolism, excretion, and toxicity to aquatic life, among others were also predicted by the admetSAR web server for top-leads provided in the SMILES format (http://lmmd.ecust.edu.cn/admetsar2/).

Search for similar structures to SN00220679

To optimize the top-leads, a first search for similar chemical structures to the top drug-like ligand (SN00220679) were made within the DrugBank chemical search tool to approved drugs of <450 Daltons with 0.7 similarity threshold (https://www.drugbank.ca/structures/search/small_molecule_drugs/structure#results). To increase the probabilities of finding commercially available compounds, a second search for similar chemical structures to the top drug-like ligand was made by the InterBioScreen Ltd company (http://www.ibscreen.com) with the help of E.Kabaeva.

Results

Overview

To target nsp7, 158240 ligands < 400 Dalton were selected from a library of natural compounds (Figure 1). The resulting list was then high-throughput double screened against nsp7.1ysy by AutoDock Vina to generate 655 consensus leads. Dockings by AutoDock Vina/SeeSAR were then performed to 14 nsp7 conformers extracted from the *.pdb files corresponding to their complexes with nsp8 and nsp12 (Table 1) in an attempt to mimic several possible nsp7 molecular dynamic situations. The 20 top-leads identified by their putative drug-like characteristics were further studied to define a common skeleton or pharmacophore to search for similar compounds to propose a final list of novel ligands for experimental studies.

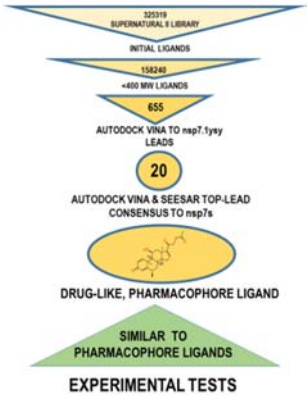


Figure 1. Scheme of the virtual screening for drug-like ligands to nsp7. To target the nsp7 by virtual screening with high probability of detecting drug-like ligands, the SuperNatural II library of 325319 compounds was reduced to 158240 possible ligands of <400 Daltons. An AutoDock Vina high-throughput molecular screening of non-minimized/energy minimized ligands against nsp7.1ysy further reduced the candidates to 655 leads. Consensus docking of the leads by both AutoDock Vina and seeSAR was then performed with 14 nsp7s to generate 20 top-leads. Drug-like characteristics and toxicity were then predicted by ADME/ADMET to define one pharmacophore. Similar compounds improved ~100-fold its binding affinity and suggested a final list of steroid-like ligands for experimental validation.

Table 1. Characteristics of published nsp7 3D conformers

nsp7 structure	Coronavirus	RBCS PDB	Method	Length solved	Identity to 1ysy, %	Reference
isolated, pH 7.5	SARS-CoV-1	1ysy	NMR	83	100.0	[16]
isolated, pH 6.5	SARS-CoV-1	2kys	NMR	83	75.9	[8]
+2 nsp8c heterotetra	SARS-CoV-2	6wiq	X-ray	71	71.8	[31]
+2 nsp8c heterotetra	SARS-CoV-2	6wqd	X-ray	71	60.6	[10]
+2 nsp8 heterotetra	SARS-CoV-2	6yhu	X-ray	71	60.6	[12]
+8 nsp8 hexadeca	SARS-CoV-1	2ahm	X-ray	78	56.4	[33]
+2 nsp8 +SNP12	SARS-CoV-2	7bw4	cryoEM	64	68.7	[15]
+2 nsp8 +SNP12	SARS-CoV-2	7bv1	cryoEM	64	67.2	[32]
+2 nsp8 +SNP12 -DTT	SARS-CoV-2	6m71	cryoEM	71	60.6	[6]
+2 nsp8 +SNP12	SARS-CoV-1	6nur	cryoEM	71	60.6	[11]
+2 nsp8 +SNP12 +RNA	SARS-CoV-2	7bzf	cryoEM	68	67.6	[30]
+2 nsp8 +SNP12 +RNA	SARS-CoV-2	7bv2	cryoEM	64	65.6	[32]
+2 nsp8 +SNP12 +RNA	SARS-CoV-2	7c2k	cryoEM	72	62.5	[30]
+2 nsp8 +SNP12 +REM	SARS-CoV-2	7btf	cryoEM	73	60.3	[6]

The nsp7 3D conformations were downloaded from the **R**esearch **C**ollaboratory for **S**tructural **B**ioinformatics (RCSB) **P**rotein **D**ata **B**ank (PDB) as isolated or complexed *.pdb files. 3D individual *.pdb files of nsp7 when in complexes were extracted using PyMOL. Structural similarity was estimated by superposing 3D nsp7s by the CCP4 Molecular Graphics program vs2.10.11 (http://www.ccp4.ac.uk/MG), to the 83 amino acid isolated recombinant nsp7.1ysy [16]. **Length solved**, number of amino acids with reported 3D structure. **Identity to 1ysy**, was calculated by the formula, 100 x (number of amino acids with similar location to nsp7.1ysy / number of total amino acids solved for nsp7s). **-DTT**, absence of dithiothreitol. **nsp8c,** carboxy terminal segment of nsp8. **REM**, remdisivir polymerase inhibitor. **White background**, isolated nsp7s. **Gray background**, nsp7+nsp8 complexes. **Cyan background,** nsp7+nsp8+nsp12 complexes**. Orange background,** nsp7+nsp8+nsp12+RNA/REM complexes

Characteristics of the nsp7 structural models selected for the study

The 83 amino acid sequence of nsp7s is highly conserved among SARS coronaviruses [16]. For instance, the reference amino acid nsp7 sequences of SARS- CoV2 [4] and -CoV1 [16] differ only in residue 70 (K and R, respectively). The structure corresponding to the first isolated recombinant nsp7.1ysy contains 4 characteristic short α-helices and reveals that the 3 Cysteins are not involved in intermolecular nor in intramolecular disulphide bridges. Helices α1 and α3 are mostly hydrophobic, while α2 and α4 contain both hydrophobic and hydrophilic segments [16] (Figure 2). An analysis of the genomic sequences from >4000 SARS-CoV2 isolates located 36 mutations at the nsp7 protein sequence. Only 27% of the mutations were missense mutations and most of those were conservative (Figure 2) [7].

Some characteristics of the nsp7 3D structures available at the RCSB protein data bank were resumed in Table 1. The nsp7 3D structures studied here were solved by other authors for recombinant isolated nsp7 (1ysy, 2kys), for complexes

obtained *in vitro* with recombinants nsp7+ nsp8 (6wiq, 6wqd, 6yhu, 2ahm) or nsp7+nsp8+nsp12 (in the absence: 7bw4, 7bv1, 7m71, 6nur or either in the presence of RNA: 7bzf, 7bv2, 7c2k or with the remdisivir polymerase inhibitor, 7btf).

The Root Square Mean Differences (RMSD) between the nsp7s Cα atoms estimated by superposing them to the original nsp7.1ysy, varied from 3.03 to 4.35 Å (not shown) or from 56.4 to 75.9% of the common amino acids as determined by paired 3D superposition (Table 1). The most similar to the isolated nsp7.1ysy at pH7.5 [16] were the isolated nsp7.2kys at pH6.5 [8] and the monomer nsp7.6wiq extracted from the heterotetramer nsp7+ nsp8 [31]. The rest of the nsp7s showed structural identities ranging from 68.7 to 56.4 % (Table 1).

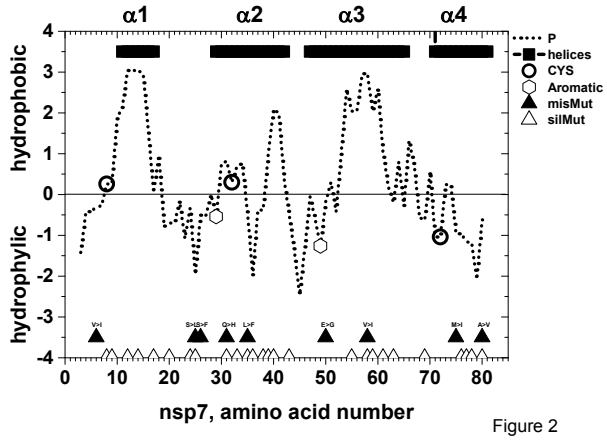


Figure 2

Figure 2. Properties and mutations in the amino acid sequence of nsp7.1ysy. The amino acid sequence of nsp7 analyzed in the figure was translated from the MN90897.3 reference (11843-12091 nucleotide numbers). The SARS-CoV2 nsp7 differs from the SARS-CoV1 by the 70th amino acid position (K in CoV2 and R in CoV1). Hydrophobic/hydrophilic plots were obtained by the Kyte & Doolittle values using Clone Manager vs9. The SARS-CoV2 amino acid mutations were reported from ~4000 mutant genomic sequences [7]. **P dashed line**, hydrophobic/hydrophilic profile. **Open circles**, mapped cysteins. **Open Hexagons**, mapped aromatic amino acids. **Filled triangles**, missense mutations in the single-letter amino acid code (misMUT). **Open triangles**, silent mutations (silMUT). **Horyzontal black rectangles**, positions of the α-helices (α1 to α4) according to the nsp7.1ysy solved structure [16].

High-throughput AutoDock Vina screening to nsp7.1ysy

Because there were no previous studies defining any drug binding pocket, to explore the nsp7 surface for binding we took advantage of its small size and compact conformation, to define an AutoDock Vina grid including the whole nsp7 molecule. As a first step, the initial SuperNatural II library was 48.6 % reduced by imposing a threshold of 400 Da. Further simplifications included using nsp7.1ysy as rigid (covalent lengths and angles constant) while the ligands were considered flexible according to their content of rotatable bonds. Furthermore, only one of the ligand conformations or poses, the one with the lowest Gibbs free-energy (ΔG) was retained for further analysis.

The potential ligands were first converted to individual *.pdbqt files without any energy minimization (non-minimized) by the Open Babel program within the PyRx 0.9.8 package. A preliminary analysis of the results obtained with the non-minimized ligands predicted the lowest ΔG from -13 to -10 Kcal/mol (~ 0.3 to 50 nM) for 548 ligands. To estimate the requirement for energy minimization before full screening, the 548 ligands were energy minimized using all the energy minimization algorithms provided by the PyRx package (uff, gaff, ghemical, mmff94, mmff94s). Since similar results were obtained using those different algorithms (data not shown), the default uff energy minimization method was chosen to minimize all the input ligands. Furthermore, a Pearson correlation coefficient of 0.74 between non-minimized and uff-minimized energy binding predictions (Figure S1 A), suggested that both could be averaged for a more conservative estimation.

Therefore, all the ligands were either non-minimized (first high-throughput) or uff minimized (second high-throughput) and converted to *.pdbqt by Open Bable and docked to nsp7.1ysy using AutoDock Vina. The second high-throughput screening resulted in predictions of lowest ΔG from -11.2 to -8.1 Kcal/mol (~ 6 to 1100 nM). The ΔG distribution of relative frequencies from non-minimized and minimized ligands showed a population of ~7% of ligands with > -2 Kcal/mol lower binding affinities when non-minimized (Figure S1 B). Study of some representants of the corresponding ligand molecular structures suggested that most discrepancies were due to ligands with a high number of rotatable bonds but other characteristics may also be important. On the other hand, similar error estimations of ± 2.85 Kcal/mol have been previously reported for AutoDock Vina [28]. Therefore, a conservative estimation of AutoDock Vina ΔG scores was chosen by averaging the data from non-minimized and uff-minimized ligands choosing a cut-off at ΔG < 8.1 Kcal/mol to yield a convenient size of 655 leads for further analysis.

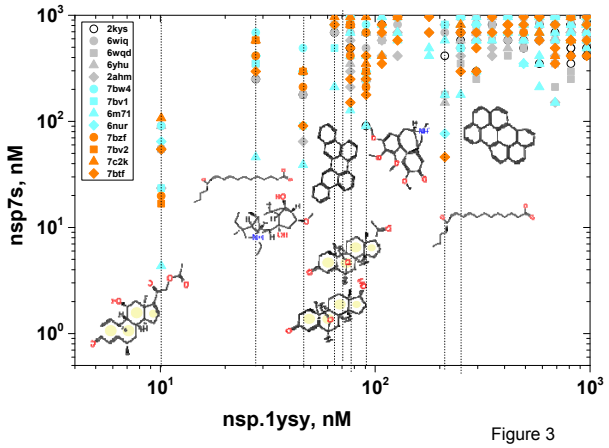


Figure 3

Figure 3. Comparison of the AutoDock Vina lead scores of nsp7s to nsp7.1ysy. Only the scores between ~1 to 1000 nM were compared in the figure. The AutoDock Vina output scores in Kcal/mol were converted to nM as indicated in methods, to facilitate comparison with seeSAR data (Figure 5). The nsp7 3D monomeric structures were extracted by PyMol from the *.pbd files of complexes.The 2D chemical formula of some top-leads were included. The steroid-like compound rings were labelled with yellow circles. **Open circles,** nsp7.2kys. **Gray circles**, nsp7.6wiq. **Gray squares**, nsp7.6yhu. **Gray triangles**, nsp7.6wqd. **Gray diamonds**, nsp7.2ahm. **Cyan circles**, nsp7.7bw4. **Cyan squares**, nsp7.7bv1. **Cyan triangles,** nsp7.6m71. **Cyan diamonds**, nsp7.6nur. **Orange circles**, nsp7.7bzf. **Orange squares**, nsp7.7bv2. **Orange triangles,** nsp7.7c2k. **Orange diamonds**, nsp7.7btf.

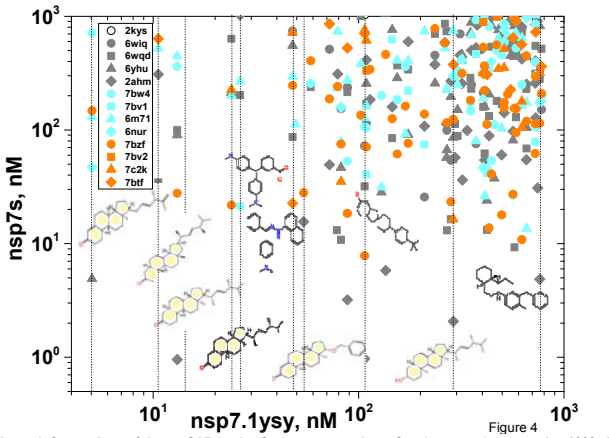


Figure 4

Figure 4. Comparison of the seeSAR leads of nsp7s to nsp7.1ysy. Ony the scores between ~1 to 1000 nM were compared in the figure. Other details and symbols as in Figure 3.

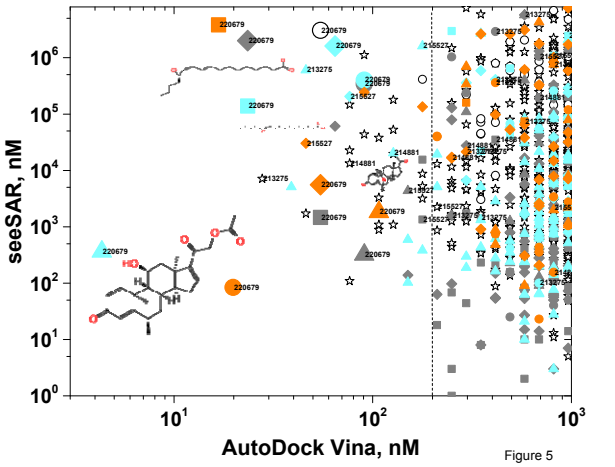


Figure 5

Figure 5. Comparison of AutoDock Vina and seeSAR binding affinities of leads to nsp7s. Only the scores between ~ 1 to 10^3 nM for AutoDock Vina and ~ 1 to 10^6 nM for seeSAR were compared in the figure (same data analyzed in Figures 3 and 4). The symbols corresponding to the most abundant SN00220679 ligand were represented with larger sizes to increase their visibility. The 2D chemical formulas of some of the top-lead ligands predicted with < 200 nM Vina scores were included in sizes proportional to their conformer abundances in the ranges of the figure (SN00220679, SN00213275, SN00215527 and SN00214881). Other details and symbols as in Figure 3.

The leads could be classified, at least, in 3 chemotypes or clusters, steroid-like (14.6%), small graphene-like (7.3%) and miscellaneous structures containing 4-8 rings as visualized by the PyRx, PyMol or ICM-MolSoft software. The most abundant steroid-like ligands contained typical ABCD-ring cores of 17C atoms in 4 fused rings: 3 cyclohexanes (A, B, C) and 1 cyclopentane (D). Many of those steroid-like ligands may be considered sterols since most of them contained one Oxygen in position 3.

AutoDock Vina docking of leads to nsp7s

A preliminary comparison of the 14 nsp7s AutoDock Vina scores of leads showed ~ 1 Kcal/mol mean lower energies in the nsp7.1ysy than in the rest of nsp7s (Figure S2), which prompt us to compare each of the nsp7s to the nsp7.1ysy scores to visualize the analysis. Figure 3 shows only the top-leads obtained within the 1-1000 nM range. The top-leads contained 3 ligands with steroid-like rings, 2 ligands with graphene-like rings, 3 complex 4-5 ring structures and 2 epoxide squalene-related linear ligands. The steroid-like NS00220679 showed the lowest binding concentrations and appeared in this range in most nsp7s (Figure 3 at X axis = 10 nM), followed by SN00213275 an epoxy squalene-related ligand (Figure 3 at X axis = 30 nM). Two other steroid- (SN00214881, SN00214827) and graphene-(SN00368224, SN00397406) like and another epoxide squalene-(SN00215527)- like and miscellaneous (SN00220673, SN00215633) ligands appeared also in such restricted range in the figure.

SeeSAR docking of leads to nsp7s

SeeSAR lead bindings to nsp7s were compared to those of nsp7.1ysy, similarly to those mentioned above for AutoDock Vina. Figure 4 shows only the top-lead bindings obtained within the 1-1000 nM lower boundary estimations. The top-leads contained 6 ligands with steroid-like rings, and 4 ligands with complex 3-5 ring structures. Among the steroid-like ligands, NS00404655, SN00284178, SN00266872, and SN00226453 appeared in many of the nsp7 conformations (Figure 4 at X axis = 0.5, 10, 12, 25 nM, respectively). The SN00130420, SN00137309 and SN00304362 (Figure 4 at X axis = 55 and 130 nM, respectively) belong also to the steroid-like chemotype. Other chemotypes with miscellaneous structures were SN00052896, SN00127794, SN00139654 and SN00142212.

Consensus of AutoDock Vina and seeSAR top-leads

Significant differences between score ranking predictions were found between AutoDock Vina and seeSAR (Figure 5). However, some of the top-leads were predicted in both programs despite the fundamental differences between their pose functions and scoring algorithms. Those were the cases of the steroid-like NS00220679 and to a lower extent of SN00214881, SN00213275 and SN00215527 (Figure 5). Expectations are that despite their differences, combination of the results obtained by the two programs may predict experimental binding capacities better than by each prediction alone. The top-leads predicted within the < 200 nM range of Vina scores predicted also in some of the seeSAR lower scores, corresponded to steroid-like SN00220679 which was present in many of the nsp7s (Figure 5, larger symbols) and less significantly to epoxy squalene-related SN00213275. To expand the possibilities of finding drug-like compounds, we chose 9 and 11 of the leads from both AutoDock Vina (Figure 3) and seeSAR (Figure 4), respectively as consensus 20 top-leads for further analysis (see their chemical 2D structures in Figure 6).

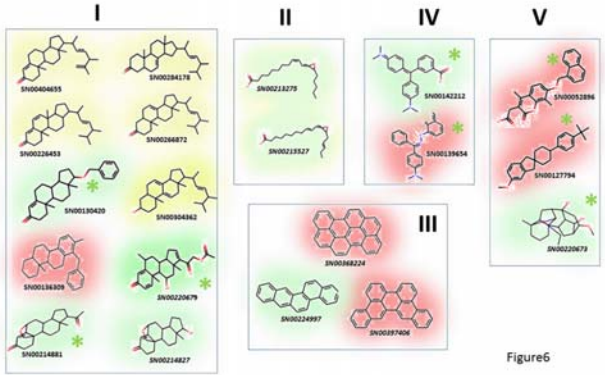


Figure 6. 2D representation of top-leads. The 20 top-leads defined according to the consensus results represented in Figures 3, 4 and 5, have been clustered in those having steroid-like rings (**I**), peroxide squalene-like (**II**), small graphene-like (**III**), N-containing rings (**IV**) and miscellaneous structures containing 4-8 rings (**V**). Each of the molecules have been background-colored according to their ADME/ADMET drug-like property predictions (Tables S1 and S2). **Green**, highly favored by predictions. **Yellow**, moderately favored by predictions. **Reddish**, not favored by predictions**. Green** *, SN00130420, SN00214881, SN00220679, SN00142212, SN00139654, SN0052896, SN00127794 and SN00220673 were commercially available according to the SNII data base.

In silico **analysis of pharmacokinetic parameters and toxicity properties of top-leads**

The corresponding *in silico* physicochemical and drug-like characteristics of the 20 top-leads, were examined by ADME computer predictions (Table S1) to define possible drug-like pharmacophores to search for similar compounds with improved affinities. Among the steroid-like compounds, the SN00220679, SN00214881 and SN00214827 were soluble, complied with Lipinski rules and have enough permeability predictions. Although less soluble, the SN00130420 could be also incorporated to the drug-like list. Nevertheless, because of their higher permeability predicted to the brain barrier, the SN00214881 and SN00214827 may only be physiologically recommended for particular cases. The SN00136309 may be discarded because of a high logP value. The rest of highly similar steroid-like ligands differing only in their double bounds (SN00404655, SN00284178, SN00226453, SN00266872 and SN00304362), although classified as of poor solubility/low permeability, may be also included for further drug optimization. The epoxy-squalene-like SN00213275 and SN00215527 ligands differing only in one double bond, were less soluble but complied with drug-like characteristics. However, the SN00215527 appeared as inhibitor of detoxifying cytochromes in some of the predictions (data not shown) and may not be preferred for physiological studies. The small graphene-like group of ligands should be discarded as drug-like because they all are of very poor solubility, high lipophilicity, low gastrointestinal permeability and potentially toxic ring alerts. Nevertheless, these ligands may be important as *in vitro* tools to study *in vitro* interactions with nsp7 because of their predicted high binding scores. With respect to the miscellaneous group of ligands, only the SN00052896 may be included in a final drug-like list since SN00127794 was poorly soluble and permeable and NS00220673 penetrates the brain barrier.

The steroid-like and epoxy-squalene-like ligands showed absence of ADMET toxic alerts such as mutagenesis, carcinogenicity, hapatotoxicity, eye irritation, high acute oral toxicity and avian toxicity, in contrast to the graphene-like, N-rings group and miscellaneous ligands (Table S2). However, many of the top-leads predicted also high aquatic toxicity by multiple tests and high binding to physiological receptors to molecules containing steroid-rings (estrogen, glucocorticoids, thyroid, aromatase). Although toxic for aquatic environments, the epoxy-squalene-like SN00213275 and SN00215527 ligands were not recognized by steroid-ring receptors.

According to the ADME/ADMET criteria, the top-leads which may be proposed as drug-like were narrowed to SN00220679 and SN00213275. As mentioned above, these two ligands were also among those with the lowest scores by both AutoDock Vina and seeSAR. Therefore, to better understand their possibilities for drug development, we explored in more detail their molecular interactions with nsp7s.

Predicted interactions between SN00220679 / SN00213275 with nsp7s

To map their predicted nsp7s amino acid neighbors, the interactions of the top poses of the 2 drug-like candidates were studied with AutoDock Vina. While α2/α3 helixes were present with small variations in all nsp7s (Figure S3), their amino-terminal parts (~ residues 2-22) shows no α1-helix structures in isolated nsp7 (1ysy, 2kys) in contrast to their presence in the rest of nsp7 when complexed with nsp8 or nsp12. In contrast, their carboxy-terminal parts (~ residues 68-80) only become α4-helix at pH 6.5, while being partially coiled at pH 7.5 or remaining unsolved in complexed nsp7. A systematic mapping study of the top poses of the 2 drug-like+nsp7s showed a high variation among the amino acid neighbors for each ligand+nsp7s (Figure S3, yellow circles). Neighbor amino acids belonging to or outside of α-helixes were interacting with any of the drug-like ligands**.** The α-helices of nsp7s with > 1 amino acid been neighbor were at α1 (50.0%), α2 (28.5%), α3 (78.5%), and α4 (21.4%) for SN00220679 or at α1 (64.2%), α2 (21.4%), α3 (85.7%), or α4 (35.7%) for SN00213275. No rules for amino acid neighbor mapping could be drawn from the above mentioned analysis.

Mapping AutoDock Vina interactions of the 2 drug-like ligands with nsp7 amino acids when within the nsp8 and the nsp8+nsp12 complexes were also tested by using grids surrounding the whole complexes. Only in the case of complexes with nsp8, the bound SN00213275 and SN00220679 identified nsp7 amino acid neighbors (yellow circles in the figure). In all these complexes, all the nsp7 neighbor amino acids mapped were found at their carboxy-terminal part (Figure S4). Additional amino acids mapped also to nsp8s (data not shown). Similar nsp7 amino acid neighbors could not be detected in any of the nsp7+nsp8+nsp12 complexes analysed. In these complexes, the bound drug-like ligands mapped mostly to nsp12, even when reducing grid sizes (data not shown).

Optimization of SN00220679 by similar search

Because this work identified the steroid-like SN00220679 as the preferred nsp7s ligand and is commercially available (Figure 6), its structure was used as SMILE skeleton to search for similar chemical structures among aproved drugs or any other commercially available compounds to explore other possible alternatives with higher binding affinities.

A first search among approved drugs with similar structures made within the DrugBank chemical search tool, yielded 18 commertially available compounds. The drugs that were most similar to SN00220679 were prednisolone/prednisone acetates, presently available as anti-inflammatory corticosteroids. Other less similar compounds such as cortisone/hydrocortisone, desoxycorticosterone, desametaxone, or hydroxyprogesterone were also found in this search. Most of these approved drugs differed from SN00220679 in additional OHs at positions 17/18 and/or different linked hydrocarbon tail(s) at position 17. Most of these approved drugs activate glucocorticoid receptor-mediated gene expression, some of them reducing inflammation by downregulating IL-6 (C-reactive protein, CRP), among a plethora of other gene expression effects. This may suggest that similar mechanisms of action could be also possible for SN00220679. However, despite their similar structures and mapping to similar amino acid neighbors in nsp7, none of the 18 drugs bound to nsp7s with lower affinities than SN00220679 (drug binding affinities > 100 to 1000 nM), as estimated by studying their AutoDock Vina poses and scores, respectively (data not shown).

In contrast, a second search based on similar chemical structures, yielded 74 commercially available compounds, most of them complying with ADME properties (data not shown). Most important, 25.6 % of these compounds showed 10-1000 fold lower or similar binding affinities than SN00220679 (Figure 7A) and mapped to similar amino acid neighbors in nsp7s (insert in Figure 7A). These compounds differed from SN00220679 in the presence of additional OHs, double bonds and/or differently linked hydrocarbon tail(s) to the common steroid rings. A 44.4% of the new top-scoring compounds were capable of low bindings to all the 14 nsp7 conformers studied (Figure 7B) increasing their possibilities to be active against nsp7.

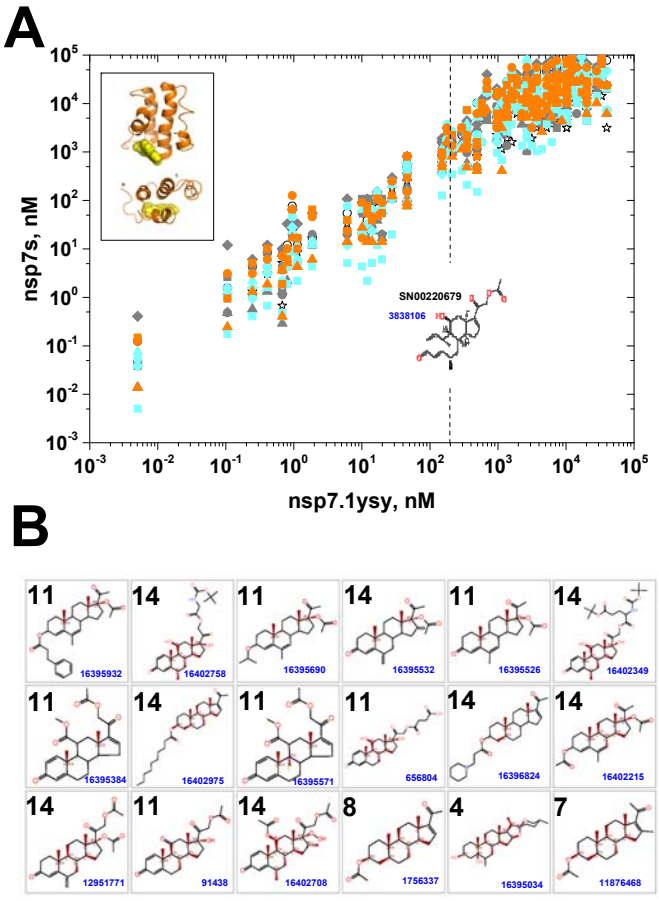


Figure 7. Comparison of AutoDock Vina scores of nsp7s and nsp7.1ysy for ligands with similar structures to SN00220679 (A) and top-score drug-like structures proposed for experimental tests (B). **A)** The top- scores of 74 structurally similar available compounds were compared. The AutoDock Vina outputs in Kcal/mol were converted to nM as indicated in methods. The vertical dashed line marks the cut-off value of 200 nM chosen to select for the top-score similar compounds. Other details were as in Figure 3. The insert shows the predicted mapping of the ligands (yellow) against the nsp7.1ysy model (side and above views). The N and C labels, identify the amine and carboxil ends of nsp7. **B)** The black larger numbers indicate the different nsp7 conformers recognized with binding affinities < 200nM by the 2D chemical formula of the similar top-score structures scored in A. The 2D structures were ordered by their lowest binding affinities according to the data shown in A from left to right and up to low. The blue smaller numbers in each down-right corner indicate the corresponding PubChem ID. Other details and symbols as in Figure 3. Complete information could be supplied upon request.

Discussion

To blind target nsp7s we sequentially combined a first high-throughput virtual screening of a large collection of ligands by AutoDock Vina to one nsp7 conformer to generate hundreds of leads, with a second consensus docking score with AutoDock Vina and SeeSAR of 14 nsp7s conformers to select 20 top-leads.

By starting with a large collection of small ligands, combining 2 docking programs, and targeting 14 nsp7s, we expected to have improved the success of leads for subsequent experimental predictions. Also, a final optimization step by searching for similar structures to top-leads revealed a collection of alternative molecules with higher binding affinities. Further optimization may even be obtained by redesigning those molecules, but that was not attempted in this work. It should be noted that despite these efforts, all virtual predictions revealed here remain hypothetical, in particular the correspondance between the higher virtual scores and experimental bindings. Therefore, only experimental assays may or may not validate the activity of the proposed compounds.

Despite the many success of virtual screening, conclusions about the possible experimental success of any predictions are not firm due to reports on poor correlations between both [14]. For instance, correlations between virtual predictions and experimental data ranged from 0.1-0.4 or 0.3-0.7 when comparative evaluating seven [17] or nineteen [24] different scoring algorithms, respectively. In other words, a unique input of ligands, may output different binding poses and scores depending on the algorithm used. Those examples highlight the limitations of actual molecular scoring programs. One of the reasons for some weak performances could be due to generating predictions based on a single pose per ligand, disregarding a more complete pose analysis, such as it has been suggested by multipose docking to yield moderate improvements in experimental predictions [2]. However, when faced to high-throughoutput screening, it was difficult to handle multipose results avoiding prohibitive time penalties. There are also a reduced amount of available programs to analyze, compare and/or interpret multipose docking in a quantitative manner [2]. Still other kind of uncertainties appeared in the present studies because of differences between mapped interacting positions and their docking scores throughout many of the “dinamic” 14 nsp7s conformers studied. Thus, contrary to other cases were the binding site appears defined, apparently there was no preference for any nsp7 unique binding pocket. Nevertheless, in isolated nsp7.1ysy all the top-score drug-like ligands amino acid neighbors mapped to the amino terminal part of the protein molecule between α1, α2 and α3 helices (Figure 7A insert and not shown data), while in nsp7+nsp8s complexes, amino acid neighbors to drug-like ligands could be mapped to the carboxy-terminal nsp7 segment, including amino acids in its α4-helix (Figure S4) as well as in nsp8 (not shown). In contrast, similar attempts with grids containing the whole nsp7+nsp8+nsp12 complexes did not show any binding pose of significant binding to the complexed nsp7. In this case, most of the poses were distributed through nsp12 (not shown).

For many ligands, the highest binding affinities were found for isolated recombinant nsp7.1ysy (therefore in the absence of nsp8 or nsp12), including the top-lead candidate “winner” NS00220679. Although other possible poses may remain undetectable when using only one pose for prediction, these results may suggest that the nsp7 experimental target could be more accessible right after translation or shortly before its complexing with nsp8+nsp12.

In silico screening and ADME/ADMET analysis are only the first steps to predict drug-like candidates. Despite the above mentioned characteristics may have predicted successfully some drug-like properties in the top-leads, they only reflect probabilities obtained by computer programs in isolated molecular situations and therefore they may be completely wrong in the much more complicated *in vitro* or *in vivo* molecular mixtures. Therefore, these hypothetical drug-like ligands and the optimized list of similar compounds derived from them must be validated for biological activity to be of some utility.

In vitro solid-phase assays for testing the binding to isolated or complexed recombinant nsp7s and possible blocking of *in vitro* coronavirus cell infection, may indicate whether any of these newly described molecules have possibilities to be relevant for coronavirus-causing human diseases. In addition to their possible antiviral activities, any of these new molecules may also be used as tools to continue the study of the role(s) these nsp7 coronavirus-unique viral proteins may have in RNA replication.

SUPPORTING INFORMATION

Table S1. Physicochemical characteristics of top-leads predicted by the SwissADME web server

	Molecule	Solubility Class	mg/mL	mol/L	LIPK	#H1	#H2	LogP	#At	TPSA, Å²	GIA	BBB	A	B
1	SN00404655	Poor	3.0E-05	7.7E-08	1	1	0	6.7	29	17.1	Low	-	0	1
2	SN00284178	Poor	8.7E-05	2.2E-07	1	1	0	6.6	29	17.1	Low	-	0	1
3	SN00226453	Poor	3.9E-05	9.9E-08	1	1	0	6.6	29	17.1	Low	-	0	1
4	SN00266872	Poor	8.7E-05	2.2E-07	1	1	0	6.6	29	17.1	Low	-	0	1
5	SN00130420	Medium	1.2E-03	3.0E-06	1	2	0	5.1	28	26.3	High	-	0	0
6	SN00304362	Poor	1.1E-04	2.8E-07	1	1	1	6.4	29	20.2	Low	-	0	1
7	SN00136309	Poor	5.6E-06	1.6E-08	1	0	0	7.1	27	0.0	Low	-	0	0
8	SN00220679	Soluble	4.7E-02	1.2E-04	0	5	1	2.9	29	80.7	High	-	0	0
9	SN00214881	Soluble	1.4E-01	4.2E-04	0	3	0	3.2	24	43.4	High	+	0	0
10	SN00214827	Soluble	4.3E-01	1.4E-03	0	3	1	2.6	22	46.5	High	+	0	0
11	SN00213275	Medium	1.4E-02	4.8E-05	0	3	1	4.7	21	49.8	High	+	0	1
12	SN00215527	Medium	2.9E-02	9.8E-05	0	3	1	4.4	21	49.8	High	+	0	2
13	SN00368224	Poor	1.4E-06	4.1E-09	1	0	0	7.2	28	0.0	Low	-	0	2
14	SN00397406	Poor	3.7E-06	1.0E-08	1	0	0	7.1	28	0.0	Low	-	0	2
15	SN00224997	Poor	4.6E-05	1.6E-07	1	0	0	5.9	22	0.0	Low	-	0	2
16	SN00142212	Soluble	6.0E+00	1.6E-02	0	2	0	1.0	28	46.4	High	+	1	2
17	SN00139654	Poor	8.5E-05	2.3E-07	1	1	1	5.2	28	27.6	High	+	0	1
18	SN00052896	Medium	2.6E-03	6.6E-06	0	5	0	4.0	29	79.6	High	-	0	1
19	SN00127794	Poor	3.7E-05	1.1E-07	1	1	0	6.2	26	9.2	Low	-	0	0
20	SN00220673	Soluble	1.6E-01	4.5E-04	0	4	2	2.4	26	52.9	High	+	0	0

The ligand molecular structures were grouped as steroid-like (numbers 1-10), epoxy squalene-like (11-12), small graphene-like (13-15), N-rings (16-17) and miscellaneous (18-20). The corresponding 2D structures to the SuperNatural II SN numbers can be consulted at Figure 6. **Solubility**, solubilities in water classified in general classes, mg/mliter, and mol/liter. **LIPK**, number of violations of Lipinski rules that would make the ligand less likely to be an orally administrable drug if >5. LIPK counts the number of Nitrogen (N) and oxygen (O) Hydrogen (H)-bond acceptors (best to have <10) and H-bond donors (best to have <5), the molecular weight (best if < 500) and the logP (best to be <5). **#H1**, number of H-bond acceptors. **#H2**, number of H-bond donors. **LogP,** consensus value of multiple predictions of lipophilicity. **#At**, number of heavy atoms per ligand. **TPSA,** estimates of the amount of topological polar molecular surface area, lowest values improve permeation of cell membranes (best to be <90 Å²). **GIA**, prediction of gastro-intestinal adsorption. **BBB**, prediction of brain barrier permeation. In the present context, a high brain permeation will be considered as a disadvantage. **A**, Pan Assay Interference Structures (PAINS) number, alerting of the number of chemical fragments that return false positive signals in virtual binding. **B**, Brenk number, alerting of the number of chemical moieties that are toxic and/or unstable. **Green**, favorable. **Yellow**, moderate. **Reddish**, unfavorable.

Table S2. Toxicity estimations of top-leads predicted by the ADMETSAR web server

		Ames mutagenesis test	Carcinogenicity	Hepatotoxicity	Eye irritation	Acute Oral Toxicity: I high, IV none	Avian toxicity	crustacea aquatic toxicity	Fish aquatic toxicity	Honey bee toxicity	Estrogen receptor binding	Glucocorticoid receptor binding	Thyroid receptor binding	Androgen receptor binding	Aromatase binding	Total favorable percentage %
1	SN00404655	-	-	-	-	III	-	+	+	+	+	+	+	+	+	42.8
2	SN00284178	-	-	-	-	III	-	+	+	+	+	+	+	-	-	57.1
3	SN00226453	-	-	-	-	III	-	+	+	+	+	+	+	+	+	42.8
4	SN00266872	-	-	-	-	III	-	+	+	+	+	+	+	-	-	57.1
5	SN00130420	-	-	-	-	III	-	+	+	+	+	+	+	+	+	42.8
6	SN00304362	-	-	-	-	III	-	+	+	+	+	+	+	+	-	50.0
7	SN00136309	-	-	-	-	IV	-	+	+	+	+	+	+	+	+	42.8
8	SN00220679	-	-	-	-	III	-	-	+	+	+	+	-	+	+	57.1
9	SN00214881	-	-	-	-	III	-	+	+	+	+	+	+	+	+	42.8
10	SN00214827	-	-	-	-	IV	-	+	+	+	+	+	+	+	+	42.8
11	SN00213275	-	-	-	-	III	-	+	+	+	-	-	+	-	-	71.4
12	SN00215527	-	-	-	-	III	-	+	+	+	-	-	+	-	-	71.4
13	SN00368224	+	±	+	+	III	-	+	+	+	+	+	+	+	+	14.2
14	SN00397406	+	±	+	+	III	-	+	+	+	+	+	+	+	+	14.2
15	SN00224997	+	±	+	+	III	-	+	+	+	+	+	+	+	+	14.2
16	SN00142212	-	-	+	+	III	-	-	+	-	+	+	+	+	+	42.8
17	SN00139654	+	+	+	+	II	-	+	+	-	+	+	+	+	+	14.2
18	SN00052896	+	-	+	-	III	-	-	+	+	+	+	+	+	-	42.8
19	SN00127794	-	-	-	-	III	-	+	+	+	+	+	+	+	+	42.8
20	SN00220673	-	-	-	-	III	-	+	-	+	+	+	+	+	+	50.0

Data for inhibitor/substrate of detoxifying cytochromes were omitted because of high divergences between the predictions obtained from the ADME and ADMETSAR web servers. The total favorable percentages were calculated by the formula 100 x number of "greens" / total number of parameter estimations (14). **Green**, favorable. **Yellow**, moderate. **Reddish**, unfavorable.

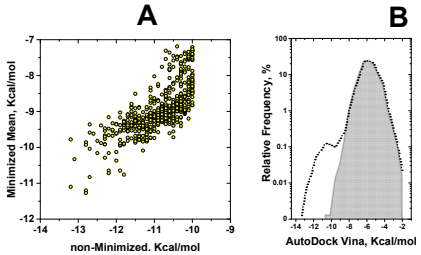


Figure S1. Comparison of AutoDock Vina binding scores of non-minimized and minimized ligands (A) and relative frequencies of ΔG from high-throughput screening (B). A) Scores of 548 non-minimized and uff energy minimized ligands were compared, showing a correlation Pearson coefficient of 0.74. To note the different scales of the scores in the X and Y axes. **B)** The relative frequencies of the high-troughput scoring in Kcal/mol by AutoDock Vina to nsp7.1ysy were independently obtained and compared in the figure from the non-minimized and uff energy minimized <400 Dalton, 158240 ligands.

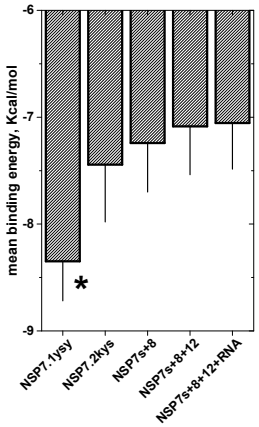
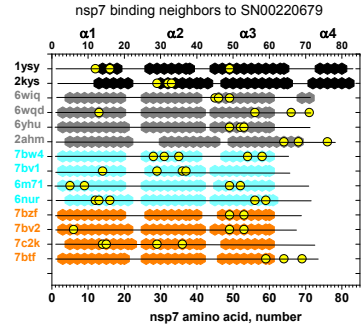


Figure S2. Comparison of the AutoDock Vina scores of 655 leads with nsp7s groups. The nsp7s were grouped into isolated 1ysy and 2kys, and hetero-oligomers nsp7 8 (6wiq, 6yhu, 6wqd, 2ahm), and nsp7+8+12 in the absence (7bw4, 7bv1, 6m71, 6nur) or in the presence of RNA or remdisivir (7bzf, 7bv2, 7c2k, 7btf). The score values of 655 energy uff energy minimized leads were averaged for each group and means ± standard deviations represented. *, significant by Student's T. According to this preliminary analysis, the nsp7.1ysy was selected for comparison of scores to the rest of the nsp7s.



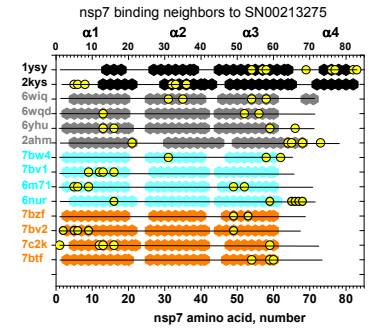


Figure S3. nsp7 amino acid neighbors to virtually bound drug-like ligands. The monomeric nsp7s were extracted in *. pdb format from the corresponding nsp7+nsp8 or nsp7+nsp8+nsp12 complexes from the RCSB data bank (Table 1). Location of their α-helixes were identified from the 3D visual tool provided at RCSB. The RCSB code of each of the nsp7 models was represented in the Y-axis. Only nsp7.1ysy and 2kys were solved in solution as isolated molecules. The rest of the nsp7 structures were solved in hetero-oligomers SNP7+SNP8 (α-helixes represented as gray hexagons) or nsp7+ nsp8+ nsp12 whether in the absence (Cyan hexagons) or in the presence of RNA or remdisivir (Orange hexagons). Amino acid neighbors to the virtually bound SN00213275/SN00220679 to each nsp7s were identified by analyzing their corresponding *.out.pdqt files by the "show interactions" option included in the PyRx package. **Yellow circles**, neighbor nsp7 amino acids predicted by AutoDock Vina. **Hexagons**, α-helixes. **Black hexagons**, isolated nsp7s. **Gray hexagons**, monomeric nsp7s extracted from nsp7+nsp8 complexes. **Cyan hexagons**, monomeric nsp7s extracted from nsp7+nsp8+nsp12 complexes. **Orange hexagons**, monomeric nsp7s extracted from nsp8+nsp12+RNA/remdisivir complexes. **--------**, not solved structure.

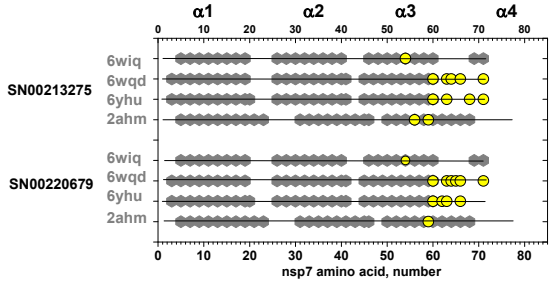


Figure S4. nsp7 amino acid neighbors in nsp7+nsp8 complexes to virtually bound drug-like ligands. The α-helixes of nsp7 were drawn as described in Figure S3. In the SN00213275 and SN00220679 bound to nsp7 within nsp7+ nsp8 hetero-oligomers, amino acid neighbors were identified by analyzing their corresponding *.out.pdqt files by the "show interactions" option included in the PyRx package. Amino acid neighbors could not be detected in nsp7 when within nsp7+nsp8+nsp12 complexes. **Yellow circles**, neighbor nsp7 amino acids predicted by AutoDock Vina. **Gray hexagons**, α-helices. **--------**, not solved structure.

List of abbreviations
SARS, Severe Acute Respiratory Syndrome. **SN**, SuperNatural library of natural compounds. **RCSB**, Research Collaboratory for Structural Bioinformatics. **PDB**, Protein Data Bank. nsp, non-structural proteins. **ADME**, absorption, distribution, metabolism and excretion. **ADMET**, absorption, distribution, metabolism, excretion and toxicity. **TPSA**, topological polar molecular surface area. **GIA**, gastro-intestinal adsorption. **BBB**, brain barrier permeation. **PAINS**, Pan Assay Interference Structures. **REM**, remdisivir polymerase inhibitor. **ORF**, Open reading frame. **RdRp**, RNA-dependent RNA polymerase. **RMSD**, Root Square Mean Differences.

Competing interests
The authors declare that they have no competing interests

Authors' contributions
RB, collaborated in home-made Phyton scripts and general discussion. JC performed and analyzed the dockings, coordinated the work and drafted the manuscript. Both authors read and approved the manuscript.

Acknowledgements
Dr.Jose Antonio Encinar from the IBMC-UMH, Elche (Spain) provided the initial SuperNatural II sdf file of natural compounds.

References

1.Alam S, Khan F (2018) Virtual screening, Docking, ADMET and System Pharmacology studies on Garcinia caged Xanthone derivatives for Anticancer activity. Sci Rep 8: 5524
2.Atkovska K, Samsonov SA, Paszkowski-Rogacz M, Pisabarro MT (2014) Multipose binding in molecular docking. Int J Mol Sci 15: 2622-45
3.Bergman SJ (2020) treatment of coronavirus disease 2019 (COVID-19): investigational drugs and other therapies. https://emedicine.medscape.com/article/2500116-overview
4.Chan JF, Kok KH, Zhu Z, Chu H, To KK, Yuan S, Yuen KY (2020) Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. Emerg Microbes Infect 9: 221-236
5.Dallakyan S, Olson AJ (2015) Small-molecule library screening by docking with PyRx. Methods Mol Biol 1263: 243-50
6.Gao Y, Yan L, Huang Y, Liu F, Zhao Y, Cao L, Wang T, Sun Q, Ming Z, Zhang L, Ge J, Zheng L, Zhang Y, Wang H, Zhu Y, Zhu C, Hu T, Hua T, Zhang B, Yang X, Li J, Yang H, Liu Z, Xu W, Guddat LW, Wang Q, Lou Z, Rao Z (2020) Structure of the RNA-dependent RNA polymerase from COVID-19 virus. Science 368: 779-782
7.Gomez-Carballa A, Bello X, Pardo-Seco J, Martinón-Torres F, Salas A (2020) The impact of super-spreaders in COVID-19: mapping genome variation worldwide. bioRxiv preprint doi: https://doi.org/10.1101/2020.05.19.097410
8.Johnson MA, Chatterjee A, Neuman BW, Wuthrich K (2010) SARS coronavirus unique domain: three-domain molecular architecture in solution and RNA binding. J Mol Biol 400: 724-42
9.Kandeel M, Al-Nazawi M (2020) Virtual screening and repurposing of FDA approved drugs against COVID-19 main protease. Life Sci 251: 117627
10.Kim Y, Wilamowski M, Jedrzejczak R, Maltseva N, Endres M, Godzik A, Michalska K, Joachimiak A (2020) The 1.95 A Crystal Structure of the Co-factor Complex of NSP7 and the C-terminal Domain of NSP8 from SARS CoV-2. RCSB protein Data Bank (to be published) DOI: 10.2210/pdb6WQD/pdb
11.Kirchdoerfer RN, Ward AB (2019) Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. Nat Commun 10: 2342
12.Konkolova E, Klima M, Boura E (2020) Crystal structure of the nsp7-nsp8 complex of SARS-CoV-2. RCSB protein Data Bank (to be published) DOI: 10.2210/pdb6YHU/pdb
13.Konkolova E, Klima M, Nencka R, Boura E (2020) Structural analysis of the putative SARS-CoV-2 primase complex. J Struct Biol 211: 107548
14.Maia EHB, Medaglia LR, da Silva AM, Taranto AG (2020) Molecular Architect: A User-Friendly Workflow for Virtual Screening. ACS Omega 5: 6628-6640
15.Peng Q, Peng R, Yuan B, Zhao J, Wang M, Wang X, Wang Q, Sun Y, Fan Z, Qi J, Gao GF, Shi Y (2020) Structural and Biochemical Characterization of the nsp12-nsp7-nsp8 Core Polymerase Complex from SARS-CoV-2. Cell Rep: 107774
16.Peti W, Johnson MA, Herrmann T, Neuman BW, Buchmeier MJ, Nelson M, Joseph J, Page R, Stevens RC, Kuhn P, Wuthrich K (2005) Structural genomics of the severe acute respiratory syndrome coronavirus: nuclear magnetic resonance structure of the protein nsP7. J Virol 79: 12905-13
17.Plewczynski D, Lazniewski M, Augustyniak R, Ginalski K (2011) Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database. J Comput Chem 32: 742-55
18.Reau M, Langenfeld F, Zagury JF, Montes M (2018) Predicting the affinity of Farnesoid X Receptor ligands through a hierarchical ranking protocol: a D3R Grand Challenge 2 case study. J Comput Aided Mol Des 32: 231-238
19.Romano M, Ruggiero A, Squeglia F, Maga G, Berisio R (2020) A Structural View of SARS-CoV-2 RNA Replication Machinery: RNA Synthesis, Proofreading and Final Capping. Cells 9
20.Ruan Z, Liu C, Guo Y, He Z, Huang X, Jia X, Yang T (2020) SARS-CoV-2 and SARS-CoV: Virtual Screening of Potential inhibitors targeting RNA-dependent RNA polymerase activity (NSP12). J Med Virol
21.Schneider N, Hindle S, Lange G, Klein R, Albrecht J, Briem H, Beyer K, Claussen H, Gastreich M, Lemmen C, Rarey M (2012) Substantial improvements in large-scale redocking and screening using the novel HYDE scoring function. J Comput Aided Mol Des 26: 701-23
22.Schneider N, Lange G, Hindle S, Klein R, Rarey M (2013) A consistent description of HYdrogen bond and DEhydration energies in protein-ligand complexes: methods behind the HYDE scoring function. J Comput Aided Mol Des 27: 15-29
23.Shityakov S, Forster C (2014) In silico predictive model to determine vector-mediated transport properties for the blood-brain barrier choline transporter. Adv Appl Bioinform Chem 7: 23-36
24.Smith RD, Dunbar JB, Jr., Ung PM, Esposito EX, Yang CY, Wang S, Carlson HA (2011) CSAR benchmark exercise of 2010: combined evaluation across all submitted scoring functions. J Chem Inf Model 51: 2115-31
25.Snijder EJ, Decroly E, Ziebuhr J (2016) The Nonstructural Proteins Directing Coronavirus RNA Synthesis and Processing. Adv Virus Res 96: 59-126
26.Srinivasan S, Cui H, Gao Z, Liu M, Lu S, Mkandawire W, Narykov O, Sun M, Korkin D (2020) Structural Genomics of SARS-CoV-2 Indicates Evolutionary Conserved Functional Regions of Viral Proteins. Viruses 12
27.Subissi L, Posthuma CC, Collet A, Zevenhoven-Dobbe JC, Gorbalenya AE, Decroly E, Snijder EJ, Canard B, Imbert I (2014) One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. Proc Natl Acad Sci U S A 111: E3900-9
28.Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31: 455-61
29.Tsuji M (2020) Potential anti-SARS-CoV-2 drug candidates identified through virtual screening of the ChEMBL database for compounds that target the main coronavirus protease. FEBS Open Bio 10: 995-1004
30.Wang Q, Wu J, Wang H, Gao Y, Liu Q, Mu A, Ji W, Yan L, Zhu Y, Zhu C, Fang X, Yang X, Huang Y, Gao H, Liu F, Ge J, Sun Q, Yang X, Xu W, Liu Z, Yang H, Lou Z, Jiang B, Guddat LW, Gong P, Rao Z (2020) Structural Basis for RNA Replication by the SARS-CoV-2 Polymerase. Cell
31.Wilamowski M, Kim Y, Jedrzejczak R, Maltseva N, Endres M, Godzik A, Michalska K, Joachimiak A (2020) Crystal structure of the co-factor complex of NSP7 and the C-terminal domain of NSP8 from SARS CoV-2. RCSB protein Data Bank (to be published) DOI: 10.2210/pdb6WIQ/pdb
32.Yin W, Mao C, Luan X, Shen DD, Shen Q, Su H, Wang X, Zhou F, Zhao W, Gao M, Chang S, Xie YC, Tian G, Jiang HW, Tao SC, Shen J, Jiang Y, Jiang H, Xu Y, Zhang S, Zhang Y, Xu HE (2020) Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. Science
33.Zhai Y, Sun F, Li X, Pang H, Xu X, Bartlam M, Rao Z (2005) Insights into SARS-CoV transcription and replication from the structure of the nsp7-nsp8 hexadecamer. Nat Struct Mol Biol 12: 980-6