

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

Blasco, R. and Coll, J.*

Julio Coll, Orcid: 0000-0001-8496-3493
Rafael Blasco, Orcid: 0000-0002-8819-5767

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 uniquely conserved small coronavirus proteins having a critical, yet intriguing participation on the replication of the long viral RNA genome after complexing with nsp8 and nsp12. Drugs with potential to interfere with nsp7s have not been described yet. Despite the difficulties of having no previously defined binding pocket, high-throughput blind screening of more than one hundred thousand natural compounds < 400 Dalton of molecular weight docked against the nsp7.1ysy conformer identified hundreds of leads displaying predicted high binding-affinities by AutoDockVina. The leads were then docked to 14 nsp7 available conformers by two different binding scoring algorithms (AutoDockVina-PyRx and HYDE-seeSAR), to identify consensus top-leads. Further predictive analysis of their physiological/toxicity ADMET criteria (chemical properties, adsorption, metabolism, toxicity) narrowed top-leads to a few drug-like ligands, most of them showing steroid-like structures closely related to some of those being actually used in clinical work. A final optimization by search for structural similarities to the drug-like top-lead, yielded a collection of novel steroid-like ligands with ~100-fold higher-affinity whose antiviral activity may be experimentally validated since they are available. Additionally, these 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 genomes¹.² 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, some of them form replication complexes with the RNA-dependent RNA polymerase (RdRp, nsp12). For successful viral RNA synthesis, the nsp12 polymerase requires nsp7 and nsp8 among other nsps²⁻⁵.

The recombinant nsp12 RdRp of coronaviruses renders 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 large ~30-kb positive stranded viral RNA genome. It was soon discovered that the addition of recombinant nsp7+nsp8 increased the polymerase activity of nsp12⁶. While pairwise combinations of nsp7, nsp8 and nsp12 displayed low RNA polymerase activity, that activity was higher only when, i) nsp7, nsp8 and nsp12 were present, ii) pre-incubated nsp7/nsp8 was added to nsp12, or iii) a nsp7-6polyH-nsp8 fusion protein was added to nsp12^{6,7}.

Mutations to Alanine in nsp7 residues K7, H36 and N37 reduced RNA binding by the nsp7+nsp8+nsp12 complexes, polymerase processivity and viral replication⁶. All these results suggested that nsp7 may play a crucial, yet intriguing role(s) in coronavirus replication⁸. On the other hand, for optimal RNA binding and processivity, the nsp12 requires not only the presence of nsp7+nsp8 but also other nsps. For instance, further association of the nsp7+nsp8+nsp12 complex with nsp14 provides proofreading of the new RNA sequence by correcting frequent mispairing assuring fidelity of RNA replication⁵.

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) it is unique and highly conserved within the *Coronaviridae* family⁹⁻¹¹, ii) it has no detectable orthologues outside coronaviruses, iii) it is a compact protein including four short α -helices, and iv) it may participate in lipid interactions during replication as suggested by its association with cytoplasmic membranes.

Models for 3D structures of recombinant nsp7 previously reported have been derived from either monomeric recombinants or from complexes 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: α 1-helix (residues 11-17), α 2-helix (residues 29-42), α 3-helix (residues 47-65), and α 4-helix (residues 71-81)⁹. The interdigitating α 2 / α 3 helices may help to stabilize the nsp7 structure as suggested by the strict conservation of most of their critical residues among different coronaviruses⁹. In contrast, small variations in the 3D conformations at 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^{1,10,12,13}. The enigmatic function of nsp7 and their coronavirus unique sequence/structure suggests a new role for RNA viral replication yet to be discovered⁹.

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 identified and mapped, respectively¹⁴. Work to find new anti-coronavirus drugs with more binding-affinity than those already known were mainly searched among approved compounds for other diseases (drug repurposing). These research have been targeted to the Angiotensin-converting enzyme II (ACE2) entry receptor, the RdRp polymerase, the S spike surface protein and/or the viral protease proteins nsp3 and/or nsp5¹⁴⁻¹⁶. The expectation being that it will be the faster way to detect more powerful drugs¹⁷. In contrast, because of the absence of known drugs and/or of identified binding pockets, targeting other coronavirus proteins like nsp7 have remained largely unexplored. Despite the difficulties to screen for any interacting compounds with these other viral protein surfaces, we focused on nsp7 conformers by blind searching for innovative ligands.

Rather than to the nsp7 surface, the closest study have been targeted to the nsp12 protein 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)¹⁴. However, to our knowledge, the nsp7 remains unexplored as a possible new target to interfere 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 for binding ligands, we have selected all their exposed surfaces (blind docking) and tested their 14 published conformations. The so called consensus screening to reduce false positives [14], has been employed with the expectation of improving experimental prediction success. For that, we have combined AutoDockVina-PyRx with HYDE-seeSAR software since these two predictive programs greatly differ in their ligand conformation (pose) optimization algorithms and scoring functions.

The workflow was organized as follows. First, a maximal number of ligands were screened using a high-throughput AutoDockVina of 158240, easier-to-screen <400 Daltons ligands against one unique 1ysy conformer. Second, 655 leads with mean binding energies < -8.1 kcal/mol equivalent to concentrations < 1100 nM, were selected for further work. Third, we used the leads as inputs to screen 14 nsp7 conformers by consensus docking with AutoDockVina 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 with nsp7s (drug-like top-lead). Fifth, a structural search for similar compounds yielded a commercially available steroid-like ligand collection with ~ 100 fold higher-binding affinities.

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, SNII.sdf) of natural ligands of the SuperNatural II database (http://bioinf-applied.charite.de/supernatural_new/index.php)¹⁸.

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 1ysy conformer⁹ 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/>).

AutoDockVina virtual screening

The AutoDockVina program¹⁹ included in the PyRx 0.9.8. package²⁰ (<https://pyrx.sourceforge.io/>) was used in e7 desk computers to predict Gibbs free-energy (ΔG) as described before^{21,22} using $60 \times 40 \times 30$ 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 without and with energy minimized 158240 ligands to generate 655 consensus leads. In addition, predicted binding energies of the AutoDockVina leads were also investigated against 14 nsp7 conformers. When required, the output ΔG energies were converted to constant inhibition (K_i) values in molar concentrations (M), using the formula $K_i = \exp[(\Delta G \times 1000) / (R \times T)]$ ($R = 1.98$ 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 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 nsp7 conformers using the HYDE scoring function included in the BioSolveit seeSAR vs.10 package (<https://www.biosolveit.de/SeeSAR/>). The seeSAR package was chosen to complement the AutoDockVina because it employs the HYDE scoring function to evaluate HYdration and DESolvation (as calibrated with octanol/water partition data logP)^{24,25} while AutoDockVina relies on a Lamarckian genetic algorithm based on ligands previously protonated and charged¹⁹. Furthermore, to reduce false positives, the seeSAR FlexX function included into its calculations any unfavorable interactions²⁶. A 100-fold nM range of binding-affinity 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 among the lower boundary of estimated binding-affinities 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 $\Delta G = R \times T \times \ln(\text{lower boundary})$. 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 assessments²⁷ such as cellular, brain or intestinal absorption, tissue distribution, detoxifying 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 of nsp7 ligands within drug libraries

FDA-approved (<https://www.accessdata.fda.gov/scripts/cder/dafi/>) and NCGC Pharmaceutical Collection (<http://tripod.nih.gov/npc/>)²⁸ drugs were

retrieved from several different sources, and duplicates eliminated using OpenBabel (http://openbabel.org/wiki/Windows_GUI, vs 3.3.1).

Search for similar structures to SN00220679

To optimize the top-leads, a first search for similar chemical structures to the drug-like top-lead 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-lead drug-like ligand was made by the InterBioScreen Ltd company (<http://www.ibscreen.com>) with the help of Ms. E.Kabaeva.

Results

Overview

To target nsp7, 158240 ligands < 400 Dalton were selected from a wider library of random natural compounds (Figure 1). The resulting list was then high-throughput double screened against the 1ysy conformer to generate 655 leads by blind AutoDockVina. Consensus Dockings by AutoDockVina/SeeSAR were then performed to 14 nsp7 conformers extracted from the *.pdb files of their complexes with nsp8 and nsp12 (Table 1), in an attempt to mimic several possible nsp7 molecular dynamic situations. The consensus top-leads with putative drug-like characteristics were further studied to search for similar compounds with higher affinities to propose a final list of novel ligands with drug-like characteristics commercially available for validation assays.

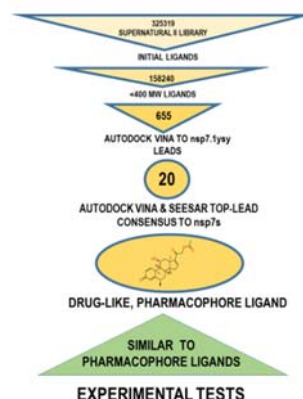


Figure 1. Scheme of the virtual screening search for drug-like ligands to nsp7. To target the nsp7 by virtual screening with high probability of detecting drug-like ligands, the SuperNatural II library of 325319 compounds was reduced to 158240 ligands of <400 Daltons of molecular weight. An AutoDockVina high-throughput molecular screening of non-minimized/energy minimized ligands against the 1ysy conformer further reduced the candidates to 655 leads. Consensus docking of the leads by both AutoDockVina / seeSAR was then performed against 14 nsp7 conformers to generate 20 top-leads. Drug-like characteristics and toxicity were then predicted by ADME/ADMET to select for the best drug-like top-lead. The search for similar compounds suggested a final list of steroid-like ligands with improved binding-scores for validation.

Characteristics of the nsp7 structural models selected for the study

The unique amino acid sequence of nsp7 is highly conserved among SARS coronaviruses [16]. The structure corresponding to the isolated recombinant 1ysy conformer contains 4 short α -helices and reveals that its 3 Cysteins are not involved in intermolecular nor in intramolecular disulphide bridges, at least under mild reducing conditions. Helices $\alpha 1$ and $\alpha 3$ are mostly hydrophobic, while $\alpha 2$ and $\alpha 4$ 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 these mutations were missense, most of them conservative (see Figure 2 for map details) [7].

Some characteristics of the nsp7 3D structure of the different conformers available were resumed in Table 1. The nsp7 conformers were solved by several authors for recombinant isolated nsp7 (1ysy, 2kys), and for nsp7 within *in vitro* complexes with nsp7+ nsp8 (6wiq, 6wqd, 6yhu, 2ahm) or nsp7+nsp8+nsp12 (in the absence: 7bw4, 7bv1, 7m71, 6nur or in the presence of RNA: 7bzf, 7bv2, 7c2k, without or with the remdisvir polymerase inhibitor, 7btf).

The Root Square Mean Differences (RMSD) between the carbon C α -atoms of nsp7 conformers estimated by superposing them by pairs to the 1ysy conformer, varied from 3.03 to 4.35 Å (not shown) or in other words from 56.4 to 75.9 % of their common amino acids (Table 1). The most similar to the 1ysy conformer⁹ were the isolated 2kys at pH6.5¹⁰ and the 6wiq extracted from the heterotetramer nsp7+ nsp8²⁹. The rest of the nsp7 conformers showed structural identities ranging from 56.4 to 68.7 % (Table 1).

Table 1. Characteristics of available nsp7 3D conformers

| nsp7 structure | Coronavirus | RBCS PDB | Method | Length solved | Identity to 1ysy, % | Ref |
|-------------------------|-------------|----------|--------|---------------|---------------------|-----|
| isolated, pH 7.5 | SARS-CoV-1 | 1ysy | NMR | 83 | 100.0 | 9 |
| isolated, pH 6.5 | SARS-CoV-1 | 2kys | NMR | 83 | 75.9 | 10 |
| +2 nsp8c heterotetramer | SARS-CoV-2 | 6wiq | X-ray | 71 | 71.8 | 29 |
| +2 nsp8c heterotetramer | SARS-CoV-2 | 6wqd | X-ray | 71 | 60.6 | 7 |
| +2 nsp8c heterotetramer | SARS-CoV-2 | 6yhu | X-ray | 71 | 60.6 | 30 |
| +8 nsp8 hexadecamer | SARS-CoV-1 | 2ahm | X-ray | 78 | 56.4 | 12 |
| +2 nsp8 +SNP12 | SARS-CoV-2 | 7bw4 | cryoEM | 64 | 68.7 | 31 |
| +2 nsp8 +SNP12 | SARS-CoV-2 | 7bv1 | cryoEM | 64 | 67.2 | 13 |
| +2 nsp8 +SNP12 -DTT | SARS-CoV-2 | 6m71 | cryoEM | 71 | 60.6 | 1 |
| +2 nsp8 +SNP12 | SARS-CoV-1 | 6nur | cryoEM | 71 | 60.6 | 32 |
| +2 nsp8 +SNP12 +RNA | SARS-CoV-2 | 7bzf | cryoEM | 68 | 67.6 | 33 |
| +2 nsp8 +SNP12 +RNA | SARS-CoV-2 | 7bv2 | cryoEM | 64 | 65.6 | 13 |
| +2 nsp8 +SNP12 +RNA | SARS-CoV-2 | 7c2k | cryoEM | 72 | 62.5 | 33 |
| +2 nsp8 +SNP12 +REM | SARS-CoV-2 | 7btf | cryoEM | 73 | 60.3 | 1 |

The nsp7 3D conformers were downloaded from the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB) as isolated or complexed with nsp8 or nsp8+nsp12 in *.pdb files. The 3D individual *.pdb files of nsp7 when in complexes were extracted using PyMOL. Structural similarity was estimated by pair superposition of the nsp7 conformers by the CCP4 Molecular Graphics program vs2.10.11 (<http://www.ccp4.ac.uk/MG>), with the 1ysy conformer⁹. **Length solved**, number of amino acids with reported 3D structure. **Identity to 1ysy**, calculated by the formula, $100 \times (\text{number of amino acids with similar location to the 1ysy conformer} / \text{number of total amino acids solved for nsp7 conformers})$. **-DTT**, absence of dithiothreitol. **nsp8c**, carboxy terminal segment of nsp8. **REM**, remdesivir RNA polymerase inhibitor. **White background**, nsp7 structure in isolated nsp7 conformers. **Gray background**, nsp7 structure in nsp7+nsp8 complexes. **Cyan background**, nsp7 structure in nsp7+nsp8+nsp12 complexes. **Orange background**, nsp7 structure in nsp7+nsp8+nsp12+RNA/REM complexes.

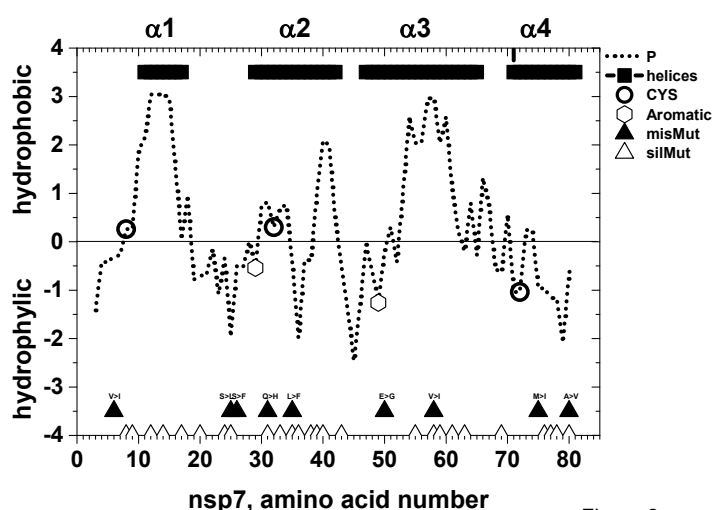


Figure 2

Figure 2. Properties and mutations of the amino acid sequence of the 1ysy conformer. The amino acid sequence of nsp7 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) [41][16]. Hydrophobic/hydrophilic plots were obtained by the corresponding Kyte & Doolittle values using Clone Manager vs9. The SARS-CoV2 amino acid mutations were as reported from ~4000 mutant genomic sequences³⁴. **P dashed line**, hydrophobic/hydrophilic profile. **Open circles**, cysteines. **Open Hexagons**, aromatic amino acids. **Filled triangles**, missense mutations in the single-letter amino acid code (misMUT). **Open triangles**, silent mutations (silMUT). **Horizontal black rectangles**, positions of the α -helices ($\alpha 1$ to $\alpha 4$) according to the 1ysy conformer solved structure⁹.

High-throughput screening of the 1ysy conformer

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

The 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 high throughput full screening, the 548 ligands were energy minimized using all the possible energy minimization algorithms provided by the PyRx package (uff, gaff, ghemical, mmff94, mmff94s). Since similar results were obtained using any of those different algorithms (data not shown), the default uff energy minimization method was chosen to energy minimize all the input ligands for the rest of the analysis. Furthermore, a Pearson correlation coefficient of 0.74 between non-minimized and

uff-minimized binding scores (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), converted to *.pdbqt by Open Babel and docked to the 1ysy conformer. The ΔG distribution of relative frequencies from non-minimized and uff-minimized ligands showed a population of ~7 % of non-minimized ligands with > -2 Kcal/mol higher binding-scores (Figure S1,B). Study of some of the corresponding ligand molecular structures suggested that most discrepancies were due to ligands with a high number of rotatable bonds, but other characteristics such as carbon-ring rigidity may also be important. On the other hand, similar error estimations of ± 2.85 Kcal/mol have been previously reported for AutoDockVina predictions¹⁹. The uff-minimized high-throughput screening resulted in predictions of lowest ΔG from -11.2 to -8.1 Kcal/mol (~6 to 1100 nM). A conservative estimation of AutoDockVina ΔG binding-scores was chosen by averaging the data from non-minimized and uff-minimized ligands. Choosing a cut-off threshold at $\Delta G < 8.1$ Kcal/mol (1100 nM), 655 leads were selected for further analysis.

AutoDockVina docking of leads to nsp7 conformers

A preliminary comparison of the leads docked by AutoDockVina to 14 nsp7 conformers, showed ~1 Kcal/mol mean lower binding-scores (higher binding-affinities) of 1ysy compared to the rest of nsp7 conformers (Figure S2). These results prompt us to compare each of the nsp7 conformers binding-scores to the 1ysy conformer to best visualize the analysis. Figure 3 shows the top-leads predicted within the 1-1000 nM binding-score range. Such top-leads contained 3 ligands with steroid-like ring chemotypes, 2 with graphene-like rings, 3 with complex 4-5 ring structures and 2 with epoxide squalene-related linear molecules. The steroid-like NS00220679 showed the lowest binding-score and also appeared at this range (Figure 3 at X axis = 10 nM) in most other nsp7 conformers, 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 were predicted also in the binding-score low nM range.

SeeSAR docking of leads to nsp7 conformers

SeeSAR lead bindings to nsp7 conformers were compared to those of the 1ysy conformer, similarly to those mentioned above for AutoDockVina. Figure 4 shows the top-lead bindings predicted within the 1-1000 nM binding-scores from the seeSAR lower boundary estimations. These 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 were predicted in many other of the nsp7 conformers (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 AutoDockVina and seeSAR top-leads

Significant differences between binding-score ranking predictions were found between AutoDockVina 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 AutoDockVina binding-scores predicted also in some of the seeSAR lower binding-scores (consensus top-leads), corresponded to steroid-like SN00220679 which was present in many of the other nsp7 conformers (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 AutoDockVina (Figure 3) and seeSAR (Figure 4), respectively, as consensus top-leads for further analysis (see their corresponding 2D structures in Figure 6).

The consensus top-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.

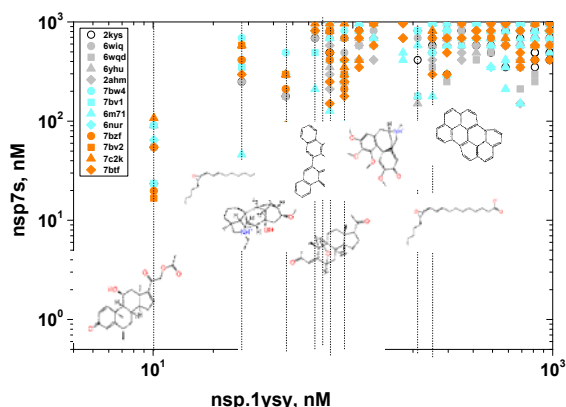


Figure 3

Figure 3. Comparison of AutoDockVina lead binding-scores of nsp7 conformers to 1ysy. Only the binding-scores between 1 to 10³ nM were represented. To facilitate comparison with seeSAR data (Figure 4,5), the AutoDockVina output binding-scores in Kcal/mol were converted to nM. The 2D chemical formula of some top-leads were included. The steroid-like compound rings were labeled with yellow circles. Open circles, 2kys. Gray circles, 6wiq. Gray squares, 6yhu. Gray triangles, 6m71. Gray diamonds, 2ahm. Cyan circles, 7bw4. Cyan squares, 7bv1. Cyan triangles, 6m71. Cyan diamonds, 6nur. Orange circles, 7bvf. Orange squares, 7bv2. Orange triangles, 7c2k. Orange diamonds, 7btf.

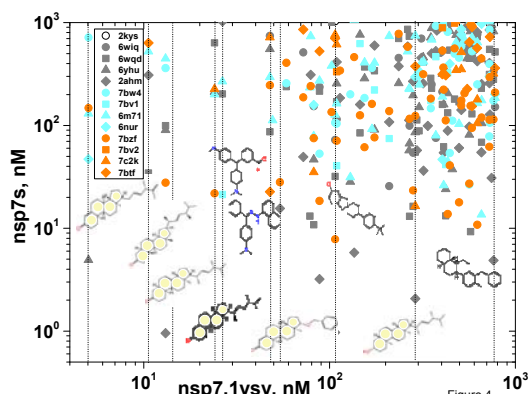


Figure 4

Figure 4. Comparison of seeSAR lead binding-scores of nsp7 conformers to 1ysy. Only the binding-scores between 1 to 10³ nM were represented. Other details and symbols as in Figure 3.

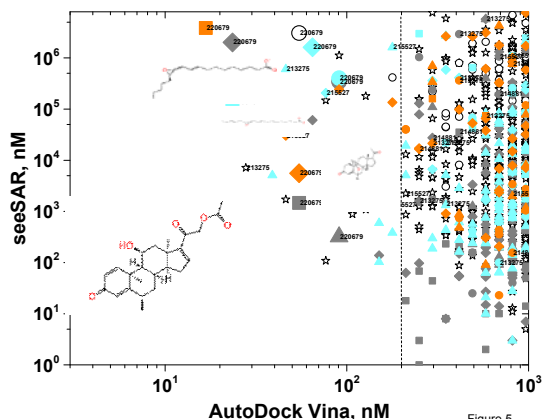


Figure 5

Figure 5. Comparison of AutoDockVina and seeSAR binding-scores of leads to nsp7 conformers. Only the binding-scores between 1 to 10³ nM for AutoDockVina and 1 to 10⁶ nM for seeSAR were shown in the figure (same data analyzed in Figures 3 and 4). Note that the seeSAR values represented were their lower boundary binding scores of a 100-fold range estimation. 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 AutoDockVina binding-scores were included in sizes proportional to their conformer abundances in the ranges selected in the figure (SN00220679, SN00213275, SN00215527 and SN00214881). Other details and symbols as in Figure 3.

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

The corresponding physicochemical and drug-like characteristics of the top-leads, were examined by ADME computer predictions (Table S1) to define possible drug-like common skeletons or pharmacophores to search for similar compounds with improved binding-scores. Among the steroid-like ligands of group I (Figure 6, I), the SN00130420, SN00220679, SN00214881 and SN00214827 were soluble, complied with Lipinski rules and have enough membrane permeability predictions (greenish backgrounds in the figure). Furthermore,

SN00130420 (less soluble), SN00220679 and SN00214881 were commercially available (labeled with green stars*) which may facilitate *in vitro* experimentation. Because of their higher permeability predicted to the brain barrier, the SN00214881 and SN00214827 may be physiologically recommended for particular cases. The SN00136309 may be discarded (reddish background) because of a high logP value. The epoxy-squalene-like SN00213275 and SN00215527 ligand of group II (Figure 6, II) differing only in one double bond, were less soluble but complied with other 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 III (Figure 6, III) of ligands should be discarded as drug-like because they have very poor solubility, high lipophilicity, low gastrointestinal permeability and carbon rings with potentially toxic alerts. Nevertheless, these ligands, specially SN00224997, may be important as *in vitro* tools to study *in vitro* interactions with nsp7 because of their predicted low binding-scores. With respect to the group having N rings (Figure 6, IV) and the miscellaneous group V (Figure 6, V) of ligands, only the SN00142212 or the NS00220673 may be labeled as drug-like.

The steroid-like (group I) and epoxy-squalene-like (group II) ligands showed absence of ADME toxic alerts such as mutagenesis, carcinogenicity, hepatotoxicity, 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-lead structures also predicted high aquatic toxicity by multiple tests and high binding to physiological receptors of molecules containing steroid-rings (estrogen, glucocorticoids, thyroid, aromatase). In contrast, the epoxy-squalene-like ligands of group II were not recognized by steroid-ring receptors although they were also toxic for aquatic environments.

According to the ADME/ADMET criteria, the top-leads which may be proposed as drug-like ligands were narrowed to SN00220679 (group I) and SN00213275 (group II). As mentioned above, these two drug-like top-leads were also among those with the lowest binding-scores by both AutoDockVina and seeSAR. Therefore, to better understand their possibilities for drug development, we explored in more detail their molecular interactions with nsp7 conformers.

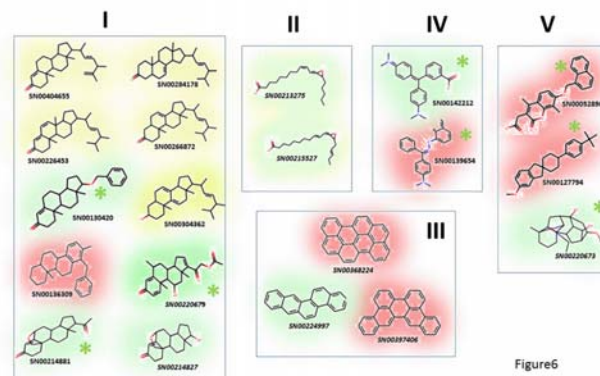


Figure 6

Figure 6. Two dimensional 2D representation of top-leads. The 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 stars*, commercially available according to the SNI data base.

Predicted interactions between SN00220679 / SN00213275 and nsp7 conformers (amino acid neighbors)

To map nsp7 conformers amino acid neighbors, differences in 3D structures among nsp7 conformers were studied first (Figure S3), followed by interactions of the two drug-like top-leads selected (SN00220679, SN00213275) with all nsp7 conformers (Figure S3), and with nsp7 conformers within replication complexes such as nsp7+nsp8 (Figure S4) or nsp7+nsp8+nsp12 (not shown).

Structural comparison of amino acid sequences and their predicted 3D structures among nsp7 conformers, showed that α2/α3 helices were present in all of them with small variations (Figure S3). However, their amino-terminal segments (~ residues 2-22) showed no α1-helix structures in isolated nsp7 (1ysy, 2kys), while being present in the rest of nsp7 conformers extracted from replication complexes. The nsp7 carboxy-terminal segments (~ residues 68-80) only become part of the α4-helix at pH 6.5, while being partially coiled at pH 7.5 or remaining unsolved when coming from the replication complexes.

A systematic mapping of the two drug-like top-leads bound to nsp7 conformers showed a high variation among the amino acid neighbors for each

ligand (Figure S3, yellow circles). Neighbor amino acids belonging to or outside of α -helices were interacting with any of the two drug-like top-leads. The α -helices of nsp7 conformers with > 1 neighbor amino acid for SN00220679 were at $\alpha 1$ (50.0%), $\alpha 2$ (28.5%), $\alpha 3$ (78.5%), and $\alpha 4$ (21.4%) or for SN00213275 at $\alpha 1$ (64.2%), $\alpha 2$ (21.4%), $\alpha 3$ (85.7%), or $\alpha 4$ (35.7%). No rules for amino acid neighbor mapping could be drawn from the above mentioned analysis.

Mapping AutoDockVina interactions of the two drug-like top-leads with nsp7 amino acids when located within the nsp8 and/or the nsp8+nsp12 replication complexes were also tested by using grids surrounding the whole replication complex macromolecule. Only in the case of replication complexes with nsp8, the binding of the two drug-like top-leads identified nsp7 amino acid neighbors (Figure S4, yellow circles). Interestingly, in all these replication complexes, neighbor amino acids mapped at their nsp7 carboxy-terminal segments (Figure S4). Additional amino acid neighbors mapped also to the nsp8 molecules (data not shown). Similar nsp7-neighbor amino acids could not be detected in any of the nsp7+nsp8+nsp12 replication complexes analyzed. In these replication complexes, the bound drug-like top-leads mapped mostly to nsp12, even when restricting grids to the nsp7 (data not shown).

Optimization of SN00220679-like structures by similar chemical search

Because this work identified the steroid-like SN00220679 as the best drug-like top-lead for nsp7 as well as some other steroid-like top-leads (Figure 6), search for similar chemical structures among approved drugs or any other commercially available compounds was performed to explore other possible alternatives with lower binding-scores.

A first search among approved drugs with similar structures made within the DrugBank chemical search tool, yielded 18 steroid-like 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, desametasone, or hydroxyprogesterone were also found. Most of these were FDA-approved drugs that differed from SN00220679 in additional OHs at positions 17/18 and/or different linked hydrocarbon tail(s) at position 17. Additionally most of these approved drugs activate glucocorticoid receptor-mediated gene expression, some of them reducing inflammation by downregulating IL-6 (a C-reactive protein inducer), 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 nsp7 conformers with lower binding-scores than SN00220679 (drug binding-scores > 100-1000 nM), as estimated by their AutoDockVina poses and binding-scores, respectively (data not shown). Similar results were obtained by screening 1700 FDA-approved and 7879 NCGC drugs using the 1ysy conformer (data not shown).

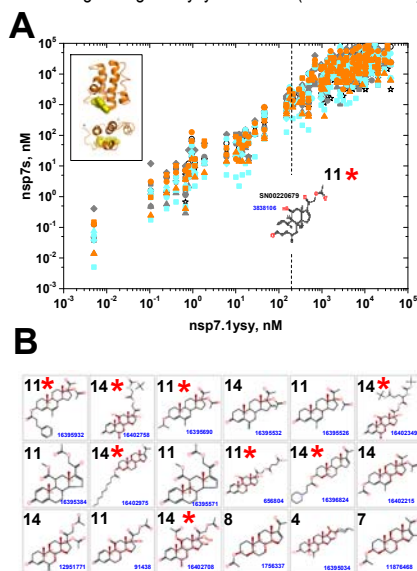


Figure 7. Comparison of AutoDockVina binding-scores of nsp7 conformers and 1ysy for SN00220679-like structures (A) and resulting top-binding-score drug-like structures proposed for experimental validation (B).

A) The binding-scores of 74 structurally similar and available compounds were compared for nsp7 conformers versus the 1ysy conformer. The vertical dashed line marks the cut-off value of 200 nM chosen to select for the lowest binding-scores. Other details were as in Figure 3. The insert shows the predicted mapping of the bound ligands (yellow) against the 1ysy conformer structure (orange in side and above views). The N and C labels identify the amine and carboxyl ends of nsp7.

B) Black numbers, number of nsp7 conformers with binding-scores < 200 nM by the lowest binding-score molecules of A. The 2D structures were ordered by their lowest binding-scores (higher binding-affinities) according to the data shown in A from left to right and up to down. Red stars*, similar compounds showing both AutoDockVina lowest binding-scores and seeSAR <20000 nM binding-scores to > 2 conformers. Blue numbers, PubChem ID. Other details and symbols as in Figure 3.

In contrast, a company search based on similar chemical structures, yielded 74 available compounds, most of them complying with ADME properties (data not shown). Most important, 25.6 % of these compounds showed 10-1000 fold lower binding affinities than SN00220679 (Figure 7A) and mapped to similar amino acid neighbors in nsp7 conformers (insert in Figure 7A). Additionally, 50% were amongst those predicted not only by AutoDockVina but also by seeSAR (Figure 7, red stars). A 44.4% of the new top-scoring compounds were capable of being bound by all the 14 nsp7 conformers (Figure 7B, black numbers), increasing their possibilities to be active against nsp7. These compounds differed from SN00220679 in the presence of additional OHs, double bonds and/or differently linked hydrocarbon tail(s) (see their 2D chemical formula in Figure 7B).

Discussion

To target nsp7 conformers, hundred thousands of ligands were first screened against a selected nsp7 conformer to generate hundreds of leads. AutoDockVina and SeeSAR were then used against 14 nsp7 conformers to select top-leads with the best consensus binding-scores and drug-like possibilities to finally search for similar steroid-like compounds to one of them.

By starting with a large collection of small ligands, combining 2 docking programs, and targeting several nsp7 conformers, improvement of successful leads for subsequent experimental predictions was expected. A final optimization step by searching for similar structures to the best drug-like top-lead revealed the existence of alternative molecules with higher binding-affinities and with steroid-like novel structures closely related to some of those being actually used in clinical work. 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 remain hypothetical, in particular the correspondence between the lower binding-scores and successful experimental bindings. It should also be mentioned that *in silico* screening and ADME/ADMET analysis are only the first steps to predict drug-like candidates. The drug-like properties predicted for the top-leads, only reflect probabilities in isolated molecular situations which may be erroneous in the *in vitro* or *in vivo* molecular mixtures. Therefore, hypothetical drug-like top-leads must be validated for biological activity to be of some utility.

Despite the many success of virtual screening predictions, conclusions about their possible experimental success based solely in computational results are not firm due to reports on poor correlations between both ³⁵. For instance, such correlations ranged from 10-40 % or 30-70% when evaluating seven ³⁶ or nineteen ³⁷ different scoring algorithms, respectively. Those examples highlight the limitations of actual molecular docking scoring programs. One of the reasons for some weak performances could be due to the way most predictions are generated based on a single pose per ligand. This way a more complete multipose analysis is disregarded, such as those that obtained moderate improvements in experimental predictions ³⁸. However, when faced to high-throughput 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 ³⁸. Other kind of uncertainties appeared in the present studies because of differences between blind mapped interacting positions and their docking binding-scores throughout many of the 14 nsp7 conformers studied. Thus, contrary to other cases where the binding site was well defined and confirmed by crystallographic data, there was no apparent preference for any nsp7 unique binding-pocket in our results. Nevertheless, in the 1ysy conformer all amino acid neighbors to two drug-like top-leads mapped to an unique amino terminal region in between $\alpha 1$, $\alpha 2$ and $\alpha 3$ helices (Figure 7A insert and data not shown), while in all the nsp7 conformers extracted from nsp7+nsp8s complexes, amino acid neighbors could be mapped to the carboxy-terminal nsp7 segment, including amino acids in its $\alpha 4$ -helix (Figure S4). In contrast, similar attempts with grids containing the whole nsp7+nsp8+nsp12 complexes did not show any binding pose of significant binding-score to their complexed nsp7.

For many ligands, the highest binding-affinities (lower binding-scores) were found for the 1ysy conformer modeled in the absence of nsp8 and nsp12, including the top-lead drug-like candidate SN00220679. Although other possible poses may remain undetectable when using only one pose per ligand for prediction, these results may suggest that the nsp7 physiological target could be more accessible right after translation or shortly before its complexing with nsp8+nsp12.

In vitro solid-phase assays for testing of experimental binding to isolated or complexed recombinant nsp7 conformers and possible blocking *in vitro* assays for 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

| | | Solubility | | LIPK | LogP | #At | TPSA, Å ² | | GIA | BBB | PAINS | B |
|----------|------------|------------|---------|------|------|-----|----------------------|------|-----|-----|-------|---|
| Molecule | Class | mg/mL | | | | | | | | | | |
| 1 | SN00404655 | Poor | 3.0E-05 | 1 | 6.7 | 29 | 17.1 | Low | - | - | 0 | 1 |
| 2 | SN00284178 | Poor | 8.7E-05 | 1 | 6.6 | 29 | 17.1 | Low | - | - | 0 | 1 |
| 3 | SN00226453 | Poor | 3.9E-05 | 1 | 6.6 | 29 | 17.1 | Low | - | - | 0 | 1 |
| 4 | SN00266872 | Poor | 8.7E-05 | 1 | 6.6 | 29 | 17.1 | Low | - | - | 0 | 1 |
| 5 | SN00130420 | Medium | 1.2E-03 | 1 | 5.1 | 28 | 26.3 | High | - | - | 0 | 0 |
| 6 | SN00304362 | Poor | 1.1E-04 | 1 | 6.4 | 29 | 20.2 | Low | - | - | 0 | 1 |
| 7 | SN00136309 | Poor | 5.6E-06 | 1 | 7.1 | 27 | 0.0 | Low | - | - | 0 | 0 |
| 8 | SN00220679 | Soluble | 4.7E-02 | 0 | 2.9 | 29 | 80.7 | High | - | - | 0 | 0 |
| 9 | SN00214881 | Soluble | 1.4E-01 | 0 | 3.2 | 24 | 43.4 | High | + | - | 0 | 0 |
| 10 | SN00214827 | Soluble | 4.3E-01 | 0 | 2.6 | 22 | 46.5 | High | + | - | 0 | 0 |
| 11 | SN00213275 | Medium | 1.4E-02 | 0 | 4.7 | 21 | 49.8 | High | + | - | 0 | 1 |
| 12 | SN00215527 | Medium | 2.9E-02 | 0 | 4.4 | 21 | 49.8 | High | + | - | 0 | 2 |
| 13 | SN00368224 | Poor | 1.4E-06 | 1 | 7.2 | 28 | 0.0 | Low | - | - | 0 | 2 |
| 14 | SN00397406 | Poor | 3.7E-06 | 1 | 7.1 | 28 | 0.0 | Low | - | - | 0 | 2 |
| 15 | SN00224997 | Poor | 4.6E-05 | 1 | 5.9 | 22 | 0.0 | Low | - | - | 0 | 2 |
| 16 | SN00142212 | Soluble | 6.0E+00 | 0 | 1.0 | 28 | 46.4 | High | + | - | 1 | 2 |
| 17 | SN00139654 | Poor | 8.5E-05 | 1 | 5.2 | 28 | 27.6 | High | + | - | 0 | 1 |
| 18 | SN00052896 | Medium | 2.6E-03 | 0 | 4.0 | 29 | 79.6 | High | - | - | 0 | 1 |
| 19 | SN00127794 | Poor | 3.7E-05 | 1 | 6.2 | 26 | 9.2 | Low | - | - | 0 | 0 |
| 20 | SN00220673 | Soluble | 1.6E-01 | 0 | 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, and mg/mliter. **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). **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. **PAINS**, Pan Assay Interference Structures 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 fragments 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. **Total favorable percentages**, as calculated by the formula 100 x number of "greens" / total number of parameter estimations: 14. **Green**, favorable. **Yellow**, moderate. **Reddish**, unfavorable.

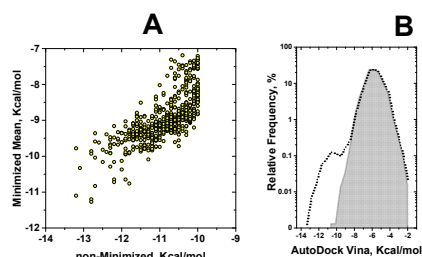


Figure S1. Comparison of AutoDockVina binding-scores of non-minimized and minimized ligands (A) and relative frequencies of ΔG from high-throughput screening (B).

A) Binding-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 binding-scores in the X and Y axes. B) The relative frequencies of the high-throughput scoring in Kcal/mol by AutoDockVina to the 1ysy conformer were independently obtained and compared in the figure from the non-minimized and uff energy minimized (gray background) <400 Dalton, 158240 ligands.

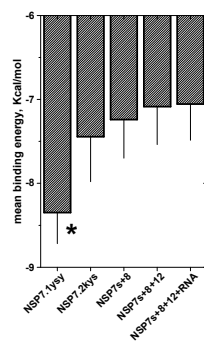


Figure S2. Comparison of the AutoDockVina binding-scores of 655 leads with nsp7 conformers groups. The nsp7 conformers were grouped into isolated 1ysy and 2kys, and hetero-oligomers nsp7 8 (6wiq, 6yhu, 6wqd, 2ahm), and nsp7+8+12 in the absence (7bw4, 7bv1, 6m71, 6nur) or in the presence of RNA or remdisivir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

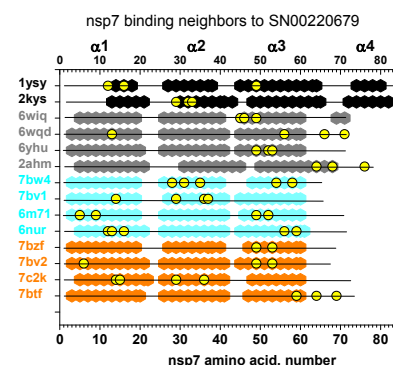


Figure S3. Amino acid neighbors of nsp7 conformers bound to 2 drug-like top-lead ligands. The nsp7 conformers were those from Table 1 (Y-axis). Location of their α -helices were identified from the 3D visual tool provided at RCSB. Only the 1ysy conformer and 2kys were solved in solution as isolated molecules. The rest of the nsp7 structures were solved in hetero-oligomers nsp7+nsp8 (α -helices 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 bound SN00220679 / SN00213275 were identified from their *.out.pdb files by the "show interactions" option included in the PyRx package. **Yellow circles**, neighbor nsp7 amino acids predicted by AutoDockVina. **Hexagons**, α -helices. **Black hexagons**, isolated nsp7 conformers. **Gray hexagons**, monomeric nsp7 conformers extracted from nsp7+nsp8 replication complexes. **Cyan hexagons**, monomeric nsp7 conformers extracted from nsp7+nsp8+12 replication complexes. **Orange hexagons**, monomeric nsp7 conformers extracted from nsp8+12+RNA/remdisivir replication complexes. ---, not solved structures.

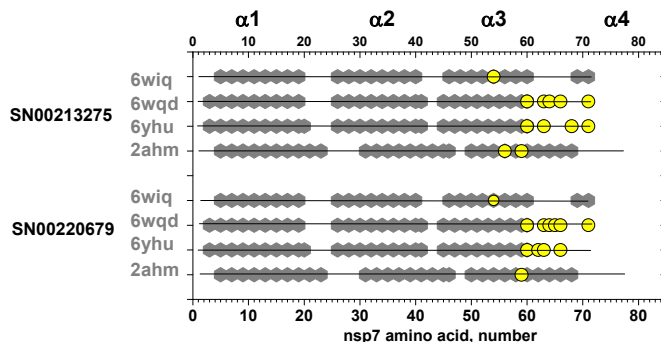


Figure S4. Amino acid neighbors of nsp7 conformers in nsp7+nsp8 replication complexes bound to 2 drug-like top-lead ligands. The α -helices of nsp7 were drawn as described in Figure S3. Amino acid neighbors were identified as in Figure S3. Amino acid neighbors could not be detected in nsp7 when within nsp7+nsp8+12 replication complexes. **Yellow circles**, neighbor nsp7 amino acids predicted by AutoDockVina. **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 computational designs and general discussion. JC performed and analyzed the dockings, coordinated the work and drafted the manuscript. Both authors read and approved the manuscript.

Funding

This work was supported in part by grant RTA2017-0066 from Ministerio de Ciencia e Innovación and grant COV20-0901 from Instituto de Salud Carlos III.

Acknowledgements

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

References

- Gao, Y., L. Yan, Y. Huang, F. Liu, Y. Zhao, L. Cao, . . . Z. Rao. **Structure of the RNA-dependent RNA polymerase from COVID-19 virus**. *Science*. 2020, 368: 779-782. [science.abb7498](https://doi.org/10.1126/science.abb7498) [pii]. <http://dx.doi.org/10.1126/science.abb7498>
- Snijder, E.J., E. Decroly and J. Ziebuhr. **The Nonstructural Proteins Directing Coronavirus RNA Synthesis and Processing**. *Adv Virus Res.* 2016, 96: 59-126. [S0065-3527\(16\)30047-1 \[pii\]](https://doi.org/10.1016/bbs.alvir.2016.08.008). <http://dx.doi.org/10.1016/bbs.alvir.2016.08.008>
- Srinivasan, S., H. Cui, Z. Gao, M. Liu, S. Lu, W. Mkandawire, . . . D. Korkin. **Structural Genomics of SARS-CoV-2 Indicates Evolutionary Conserved Functional Regions of Viral Proteins**. *Viruses*. 2020, 12: [v12040360](https://doi.org/10.3390/v12040360) [pii]. <http://dx.doi.org/10.3390/v12040360>
- Konkolova, E., M. Klima, R. Nencka and E. Boura. **Structural analysis of the putative SARS-CoV-2 primase complex**. *J Struct Biol.* 2020, 211: 107548. [S1047-8477\(20\)30121-0 \[pii\]](https://doi.org/10.1016/j.jsb.2020.107548). <http://dx.doi.org/10.1016/j.jsb.2020.107548>
- Romano, M., A. Ruggiero, F. Squeglia, G. Maga and R. Berisio. **A Structural View of SARS-CoV-2 RNA Replication Machinery: RNA Synthesis, Proofreading and Final Capping**. *Cells*. 2020, 9: [cells9051267](https://doi.org/10.3390/cells9051267) [pii]. <http://dx.doi.org/10.3390/cells9051267>
- Subissi, L., C.C. Posthuma, A. Collet, J.C. Zevenhoven-Dobbe, A.E. Gorbalenya, E. Decroly, . . . I. Imbert. **One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities**. *Proc Natl Acad Sci U S A.* 2014, 111: E3900-9. [1323705111 \[pii\]](https://doi.org/10.1073/pnas.1323705111). <http://dx.doi.org/10.1073/pnas.1323705111>
- Kim, Y., M. Wilamowski, R. Jedrzejczak, N. Maltseva, M. Endres, A. Godzik, . . . A. Joachimiak. **The 1.95 Å 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)*. 2020, DOI: [10.2210/pdb6WQD/pdb](https://doi.org/10.2210/pdb6WQD/pdb)
- Krichel, B., G. Blyapudi, C. Schmidt, C. Blanchet, R. Schubert, L. Brings, . . . C. Uetrecht. **Hallmarks of Alpha- and Betacoronavirus non-structural protein 7+8 complexes**. *bioRxiv*. 2020. [http://dx.doi.org/10.1101/2020.09.30.320762](https://doi.org/10.1101/2020.09.30.320762)
- Peti, W., M.A. Johnson, T. Herrmann, B.W. Neuman, M.J. Buchmeier, M. Nelson, . . . K. Wuthrich. **Structural genomics of the severe acute respiratory syndrome coronavirus: nuclear magnetic resonance structure of the protein nsp7**. *J Virol.* 2005, 79: 12905-13. [79.20.12905 \[pii\]](https://doi.org/10.1128/JVI.79.20.12905-12913.2005). <http://dx.doi.org/10.1128/JVI.79.20.12905-12913.2005>
- Johnson, M.A., A. Chatterjee, B.W. Neuman and K. Wuthrich. **SARS coronavirus unique domain: three-domain molecular architecture in solution and RNA binding**. *J Mol Biol.* 2010, 400: 724-42. [S0022-2836\(10\)00506-1 \[pii\]](https://doi.org/10.1005/506-1). <http://dx.doi.org/10.1016/j.jmb.2010.05.027>
- Gordon, D.E., J. Hiatt, M. Bouhaddou, V.V. Rezelj, S. Ulferts, H. Braberg, . . . N.J. Krogan. **Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms**. *Science*. 2020, 370: [science.abe9403](https://doi.org/10.1126/science.abe9403) [pii]. <http://dx.doi.org/10.1126/science.abe9403>
- Zhai, Y., F. Sun, X. Li, H. Pang, X. Xu, M. Bartlam and Z. Rao. **Insights into SARS-CoV transcription and replication from the structure of the nsp7-nsp8 hexadecamer**. *Nat Struct Mol Biol.* 2005, 12: 980-6. [psmb999 \[pii\]](https://doi.org/10.1038/nsmb999). <http://dx.doi.org/10.1038/nsmb999>
- Yin, W., C. Mao, X. Luan, D.D. Shen, Q. Shen, H. Su, . . . H.E. Xu. **Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir**. *Science*. 2020. [science.abc1560 \[pii\]](https://doi.org/10.1126/science.abc1560). <http://dx.doi.org/10.1126/science.abc1560>
- Ruan, Z., C. Liu, Y. Guo, Z. He, X. Huang, X. Jia and T. Yang. **SARS-CoV-2 and SARS-CoV: Virtual Screening of Potential inhibitors targeting RNA-dependent RNA polymerase activity (NSP12)**. *J Med Virol.* 2020. [http://dx.doi.org/10.1002/jmv.26222](https://doi.org/10.1002/jmv.26222)
- Kandeel, M. and M. Al-Nazawi. **Virtual screening and repurposing of FDA approved drugs against COVID-19 main protease**. *Life Sci.* 2020, 251: 117627. [S0024-3205\(20\)30375-1 \[pii\]](https://doi.org/10.1016/j.lfs.2020.117627). <http://dx.doi.org/10.1016/j.lfs.2020.117627>
- Tsuji, M. **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.* 2020, 10: 995-1004. [http://dx.doi.org/10.1002/2211-5463.12875](https://doi.org/10.1002/2211-5463.12875)
- Bergman, S.J. **treatment of coronavirus disease 2019 (COVID-19): investigational drugs and other therapies**. <https://emedicine.medscape.com/article/2500116-overview>. 2020:
- Galiano, V., P. Garcia-Valltanen, V. Micol and J.A. Encinar. **Looking for inhibitors of the dengue virus NS5 RNA-dependent RNA-polymerase using a molecular docking approach**. *Drug Des Devel Ther.* 2016, 10: 3163-3181. [http://dx.doi.org/10.2147/DDDT.S117369](https://doi.org/10.2147/DDDT.S117369). [dddt-10.3163 \[pii\]](https://doi.org/10.2147/DDDT.S117369)
- Trott, O. and A.J. Olson. **AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading**. *J Comput Chem.* 2010, 31: 455-61. [http://dx.doi.org/10.1002/jcc.21334](https://doi.org/10.1002/jcc.21334)
- Dallakyan, S. and A.J. Olson. **Small-molecule library screening by docking with PyRx**. *Methods Mol Biol.* 2015, 1263: 243-50. [http://dx.doi.org/10.1007/978-1-4939-2269-7_19](https://doi.org/10.1007/978-1-4939-2269-7_19)
- Bello-Perez, M., A. Falco, B. Novoa, L. Perez and J. Coll. **Hydroxycholesterol binds and enhances the anti-viral activities of zebrafish monomeric c-reactive protein isoforms**. *PLoS One.* 2019, 14: e0201509. [http://dx.doi.org/10.1371/journal.pone.0201509](https://doi.org/10.1371/journal.pone.0201509)
- Bello-Perez, M., R. Medina-Gali, J. Coll and L. Perez. **Viral interference between infectious pancreatic necrosis virus and spring viremia of carp virus in zebrafish**. *Aquaculture.* 2019, 500: 370-377. <https://doi.org/10.1016/j.aquaculture.2018.10.039>
- Shityakov, S. and C. Forster. **In silico predictive model to determine vector-mediated transport properties for the blood-brain barrier choline transporter**. *Adv Appl Bioinform Chem.* 2014, 7: 23-36. [http://dx.doi.org/10.2147/AABC.S63749](https://doi.org/10.2147/AABC.S63749)
- Schneider, N., S. Hindle, G. Lange, R. Klein, J. Albrecht, H. Briem, . . . M. Rarey. **Substantial improvements in large-scale redocking and screening using the novel HYDE scoring function**. *J Comput Aided Mol Des.* 2012, 26: 701-23. [http://dx.doi.org/10.1007/s10822-011-9531-0](https://doi.org/10.1007/s10822-011-9531-0)
- Schneider, N., G. Lange, S. Hindle, R. Klein and M. Rarey. **A consistent description of Hydrogen bond and Dehydration energies in protein-ligand complexes: methods behind the HYDE scoring function**. *J Comput Aided Mol Des.* 2013, 27: 15-29. [http://dx.doi.org/10.1007/s10822-012-9626-2](https://doi.org/10.1007/s10822-012-9626-2)
- Reau, M., F. Langenfeld, J.F. Zagury and N. Montes. **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.* 2018, 32: 231-238. [http://dx.doi.org/10.1007/s10822-017-0063-0](https://doi.org/10.1007/s10822-017-0063-0)
- Alam, S. and F. Khan. **Virtual screening, Docking, ADMET and System Pharmacology studies on Garcinia caged Xanthone derivatives for Anticancer activity**. *Sci Rep.* 2018, 8: 5524. [http://dx.doi.org/10.1038/s41598-018-23768-7](https://doi.org/10.1038/s41598-018-23768-7) [pii]
- Huang, R., N. Southall, Y. Wang, A. Yasgar, P. Shinn, A. Jadhav, . . . C.P. Austin. **The NCGC pharmaceutical collection: a comprehensive resource of clinically approved drugs enabling repurposing and chemical genomics**. *Sci Transl Med.* 2011, 3: 80ps16. [380ps16 \[pii\]](https://doi.org/10.1126/scitranslmed.3001862). <http://dx.doi.org/10.1126/scitranslmed.3001862>
- Wilamowski, M., Y. Kim, R. Jedrzejczak, N. Maltseva, M. Endres, A. Godzik, . . . A. Joachimiak. **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)*. 2020, DOI: [10.2210/pdb6WQ/pdb](https://doi.org/10.2210/pdb6WQ/pdb)
- Konkolova, E., M. Klima and E. Boura. **Crystal structure of the nsp7-nsp8 complex of SARS-CoV-2**. *RCSB protein Data Bank (to be published)*. 2020, DOI: [10.2210/pdb6YHU/pdb](https://doi.org/10.2210/pdb6YHU/pdb)
- Peng, Q., R. Peng, B. Yuan, J. Zhao, M. Wang, X. Wang, . . . Y. Shi. **Structural and Biochemical Characterization of the nsp12-nsp7-nsp8 Core Polymerase Complex from SARS-CoV-2**. *Cell Rep.* 2020: 107774. [S2211-1247\(20\)30754-3 \[pii\]](https://doi.org/10.1016/j.celrep.2020.107774). <http://dx.doi.org/10.1016/j.celrep.2020.107774>
- Kirchdoerfer, R.N. and A.B. Ward. **Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors**. *Nat Commun.* 2019, 10: 2342. [http://dx.doi.org/10.1038/s41467-019-10280-3](https://doi.org/10.1038/s41467-019-10280-3) [pii]
- Wang, Q., J. Wu, H. Wang, Y. Gao, Q. Liu, A. Mu, . . . Z. Rao. **Structural Basis for RNA Replication by the SARS-CoV-2 Polymerase**. *Cell.* 2020. [S0092-8674\(20\)30629-2 \[pii\]](https://doi.org/10.1016/j.cell.2020.05.034). <http://dx.doi.org/10.1016/j.cell.2020.05.034>
- Gomez-Carballa, A., X. Bello, J. Pardo-Seco, F. Martín-Ortega and A. Salas. **The impact of super-spreaders in COVID-19: mapping genome variation worldwide**. *bioRxiv* 2020. <https://doi.org/10.1101/2020.05.19.097410>
- Maia, E.H.B., L.R. Medaglia, A.M. da Silva and A.G. Taranto. **Molecular Architect: A User-Friendly Workflow for Virtual Screening**. *ACS Omega.* 2020, 5: 6628-6640. [http://dx.doi.org/10.1021/acsomega.9b04403](https://doi.org/10.1021/acsomega.9b04403)
- Plewczynski, D., M. Lazniewski, R. Augustyniak and K. Ginalska. **Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database**. *J Comput Chem.* 2011, 32: 742-55. [http://dx.doi.org/10.1002/jcc.21643](https://doi.org/10.1002/jcc.21643)
- Smith, R.D., J.B. Dunbar, Jr., P.M. Ung, E.X. Esposito, C.Y. Yang, S. Wang and H.A. Carlson. **CSAR benchmark exercise of 2010: combined evaluation across all submitted scoring functions**. *J Chem Inf Model.* 2011, 51: 2115-31. [http://dx.doi.org/10.1021/ci200269g](https://doi.org/10.1021/ci200269g)
- Atkovska, K., S.A. Samsonov, M. Paszkowski-Rogacz and M.T. Pisabarro. **Multipose binding in molecular docking**. *Int J Mol Sci.* 2014, 15: 2622-45. [ijms15022622 \[pii\]](https://doi.org/10.3390/ijms15022622). <https://doi.org/10.3390/ijms15022622>