Would it be possible to stabilize prefusion SARS-COV-2 spikes with ligands?

short title: ligand-stabilization of spike coronavirus

Coll, J*.

Department of Biotechnology. Instituto Nacional Investigaciones y Tecnologías Agrarias y Alimentarias, INIA. Madrid, Spain.

* Corresponding author
Email: juliocollm@gmail.com (JC)

Abstract

The infection by Severe Acute Respiratory Syndrome coronavirus (SARS)-CoV2 could be inhibited *in vitro* by mutations stabilizing their spike (S) native conformation in prefusion states, as reported by several authors. However, the possible S stabilization by binding-ligands, rather than by mutations, have not been computationally explored, nor it is known if that will be possible. Therefore, to first explore these possibilities, a binding target for predictive programs was focused to where the inhibiting mutations were described in the S coronavirus protein, in particular to the "spring-loaded switch-folding" (SLSF) segment of the S2 subunit, whose prefusion unfolding/refolding is required for viral/host membrane fusion. Similar SLSF prefusion mechanisms have been described in many other enveloped viruses. Results of a double computational screening of hundred of thousands of natural compounds for binding to wild-type SLSF conformer, predicted more leads in the low nM range for trimers than for monomers. Further ranked by the number of bound SLSF-conformers, some of the derived top-leads were predicted that may deserve experimental validation. Additionally, thousands of drugs were also included into the screening, resulting in a few top-lead drugs predicted to bind SLSF targets in the low nM range. All these potentially interacting S-ligands, similar structures and/or chemically improved designs, could be used to experimentally find out whether it will be possible to use them for inhibiting fusion and infection, offer new tools to investigate prefusion mechanism(s) and may contribute to therapeutic purposes.

Keywords: S; S2; conformers; prefusion; coronavirus; computational screening; ligands; SARS CoV-19; spring-loaded switch

Introduction

The surface of infectious SARS-CoV-2, is surrounded by spike (S) glycoprotein trimers forming a corona-like structure. Most of the S trimers are in a host-receptor non-accessible closed prefusion conformation, having their 3 receptor-binding domains (RBD) all-down [1]. Nevertheless, S trimers often displays RBD epitopes targeted by many neutralizing antibodies [2-7], most probably because there may exist RBD spontaneous transitions from closed (down) to exposed (up) receptor-accessible conformations (1, 2 or 3-up) [8]. Upon binding to the host protein receptor, proteolysis separate the S1/S2 subunits, which become non-covalently associated in another prefusion state (Figure 1). In SARS-CoV-2, the S2 subunit (residues 686-1273) contains the fusion peptide (788–806), the amino-terminal HR1 (910 to 988, with non-helix residues at 939-947 and 968-986), the central helix CH (986-1033), the C-terminal HR2 (1162 to 1213), the transmembrane domain (1214–1237) and the cytoplasmic domain (1238–1273) (Figure 1). All these domains participate in several prefusion conformational changes to expose RBDs and trigger viral-host membrane fusion [9, 10].

Infectious coronaviruses can be inactivated by stabilizing their S prefusion all-down conformations by diverse mutations [10-13]. For instance, mutations to prolines (P) in the HR1-CH segment, generated prefusion-stabilized non-infectious MERS [10] and SARS-CoV2 [11-13]. The inhibition of infectivity by double PP mutations is most likely due to the blocking of the "spring-loaded switch" (SLS) unfolding, required for fusion-competent conformations, similarly to those present in many enveloped virions. In wild-type virions, after RBC binding to host receptors, the SLS segment unfolding prepare fusion-competent virions. Although the PP mutants maintained the virion morphology, they were non-infectious. The PP-based strategy may be advantageous for the development of vaccines than other described mutations, since it also increased recombinant S yield and stability. Those two reasons may explain why most of the presently available 3D S structures were solved with PP mutants (Table S1).

The HR1-CH sequence (residues 910-1033) of SARS-COV-2 contains the SLS [10, 14] which maintain 2 α-helices folded but separated by non-helix residues (spring-loaded). Once the SLS folding (SLSF) is unfolded, refolding generates a unique longer α-helix (2 α-helices to 1 α-helix transition) (Figure 1C). Then trimer HR1-HR2 complexes form a fusion-competent 6-helix core of coiled-coils (one antiparallel complex of 3 internal HR2 + 3 external HR1). A S2' protease-mediated cleavage liberates the fusion peptide that now can be inserted into the host-cell membrane for viral/host membrane fusion.

Previous successful examples on fusion inhibition by targeting HRs have been reported in several enveloped viruses, including SARS-COV-2 [15, 16]. Perhaps the best example is Enfurtivide®, an FDA-approved peptide drug blocking HIV infection by inhibiting HR conformational changes in its gp41 membrane protein. In coronaviruses, peptide inhibitors derived from HR2 and binding to HR1 inhibited viral fusion and infectivity. For instance, in SARS-CoV2, the CP-1 peptide showed an inhibition concentration of 19 μM in cell-fusion assays, which could be improved to 0.19 - 0.62 μM by mutations, conjugation to lipids [9, 17] or hydrocarbon-chain stapling [18].

All the above mentioned reports suggests that computationally searching for more potent binding ligands (i.e., in the low nM range) could be a source of possible prefusion inhibitors of SARS-COV-2. Perhaps targeting the SLSF segment rather than the HR complex core would provide some alternatives. Such possibilities may benefit from preliminary computational predictions.

While most anti-coronavirus compounds are being computationally searched among approved drugs targeting the RNA replication complex RdRp core (nsp12), the S1 interface of the RBD / ACE2 host-receptor [19] and/or the viral proteases implicated in viral protein processing [20-22], to our knowledge, there have been no previous reports on computational attempts to search for possible binding ligands targeting SLSFs.Therefore, based on the successful S prefusion stabilization of SARS-CoV2 by mutants on the SLSF segment [11-13], the existence of putative binding-ligands in the low nM range rather than mutations were explored here. Such hypothetical ligands should be capable of stabilizing SLSF prefusion conformations, perhaps by non-covalent crosslinking of α-helices or by avoiding their displacement. Whether such ligands binding SLSFs would exist and be capable of inhibiting viral fusion remains to be demonstrated.

The results obtained predicted that, **i**) top-leads exist predicted to bind SLSF in the low nM range, including some known drugs, and **ii**) The SLSF trimers were the best targets for such binding-ligands rather than monomers. All these potential SLSF-ligands, similar structures and/or chemically altered designs, may be used to experimentally find out whether it will be possible to use them for inhibiting fusion and infection, offer new tools to further investigate the prefusion mechanism(s) of coronaviruses and may contribute to therapeutic purposes.

Materials and Methods

Ligands and tridimensional spring-loaded switch-folding (SLSF) models

To simplify high-throughput screening and facilitate any subsequent practical use of leads, random natural products >380 Daltons and logP>6 were excluded from the initial library. Accordingly, one spatial data file (SNII.sdf) of 325319 compounds from the SuperNatural II (http://bioinf-applied.charite.de/supernatural_new/index.php), was downsized with the DataWarrior program (Osiris DataWarrior vs5.2.1) to 135831 ligands (41.7%). The list was then splitted in subfiles containing ~ 9000 ligands each (16-177, 177-210, 210-231, 231-248, 248-264, 264-280, 280-295, 295-310, 310-320, 320-330, 330-340, 340-350, 350-360, 360-370, 370-380 Daltons).

The S residues expanding amino acids 960 to 1100 were selected as the hypothetical minimal spring-loaded switch-folding (SLSF) segment. To explore possible SLSF conformers, 40 S.pdb structures were downloaded from the RCSB PDB protein data bank (https://www.rcsb.org/) before September of 2020 (Table

S1). The individual 3D 960-1010.pdb files were extracted from the S.pdb files using PyMOL scripts. Structural similarities were then estimated in Å calculated by Root Mean Square Differences (RMSD) of the alpha carbons by superposing the corresponding 3D models in the CCP4 Molecular Graphics program vs2.10.11 (http://www.ccp4.ac.uk/MG). Binding pockets were predicted using the seeSAR vs.10 program (https://www.biosolveit.de/SeeSAR/) (Figure S1).

For high throughput screening, SLSF-trimeric (virion target) and -monomeric (infection target) were used from the 6xr8 conformer (closed all-down conformer) as model. After screening, leads were defined by taking into account the changes in the score profiles (scores values ranked by their relative order among each of the scored compounds) and selecting a minimal of ~1000-3000 or ~1-2 % of the initial downsized library.

SeeSAR virtual screening

A high-throughput screening of sdf files with a total of 135831 ligands to the 6xr8 conformer and of the resulting leads to 10 SLSF conformers were screened for binding using the BioSolveit seeSAR vs.10 package (https://www.biosolveit.de/SeeSAR/) in e7 desk computers, as previously described [23]. The seeSAR package employs the HYDE scoring function to evaluate HYdration / DEsolvation eliminating unfavorable interactions to reduce false positives [24-26]. To perform the dockings, 2 predicted binding-pockets (Figure S1, colored shadows) were used for monomers while only internal binding-pockets were used for trimers (Figure S1, yellow shadows). To compute the binding-scores and corresponding poses of each of the *.sdf files of <9000 ligands, took ~ 2 days. The program was set to obtain 3-5 binding poses per ligand. For each ligand, the poses with the lower binding-scores were selected for further analysis. The mean nM scores derived from the seeSAR predicted lower / higher boundaries were used for calculations. The predicted structures were visualized in seeSAR and/or PyMOL (https://www.pymol.org/).

AutoDockVina virtual screening

The AutoDockVina program [27] included into the PyRx 0.9.8 package [28] was used to predict Gibbs free-energy (ΔG) as previously described [23, 29]. AutoDockVina algorithm relies on ligand protonation and charge distribution to predict scores [27]. Briefly, the *.sdf files were ffu energy minimized in Open Bable and converted to *.pdbqt files for docking. Further simplifications included setting the SLSF segment as rigid (constant covalent lengths and angles) and the ligands as flexible (rotatable bonds) and retaining the poses with the lowest ΔG of each *.out.pdbqt for calculations and visualization. A high-throughput screening of sdf files with a total of 135831 ligands to the 6xr8 conformer and of the resulting leads with 10 SLSF conformers were screened for binding. To perform the dockings, internal grids to the corresponding molecules were submitted to the program (Figure S1, drawn grids).The output ΔG energies in kcal/mol 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 ºC)[30]. The predicted structures were visualized in PyRx and/or PyMOL.

Heat maps, *in silico* analysis of drug-like properties, commercial availability and known drugs

Top-leads were quantitatively defined by those leads with <50 nM binding-scores which were bound by >4 conformers (voting procedure). To compare leads to select their top-leads, all the conformers and their binding scores were first ordered by ligand name, to obtain lists with one row per ligand / score. The resulting list was then ordered by the 6xr8 binding scores to select those showing binding-scores <50 nM. Top-leads were finally obtained by ordering the first 20-30 leads by those bound to >4 conformers. Conformer bindings defined as positive by using thresholds that resulted in < 35 conformers were 10 µM (seeSAR) or 250 nM (AutoDockVina). Final data were represented as heat maps to facilitate their interpretation.

The water solubility, partition coefficient between *n*-octanol/water (logP), violations of Lipinski's rules, physiological absorption predictions, detoxification by main cytochromes CYP1A2 and CYP3A4 and some other ADME properties were downloaded from the SwissADME web server for the top-leads provided in their SMILES format (http://www.swissadme.ch/). Commercial availability were searched on the ZINC data base for the top-leads provided in their SMILES format (http://zinc15.docking.org/). FDA-approved (https://www.accessdata.fda.gov/scripts/cder/daf/) and NCGC Pharmaceutical Collection (http://tripod.nih.gov/npc/) [31] drugs were retrieved from several different sources, and the duplicates eliminated using OpenBabel (http://openbabel.org/wiki/Windows_GUI, vs 3.3.1).

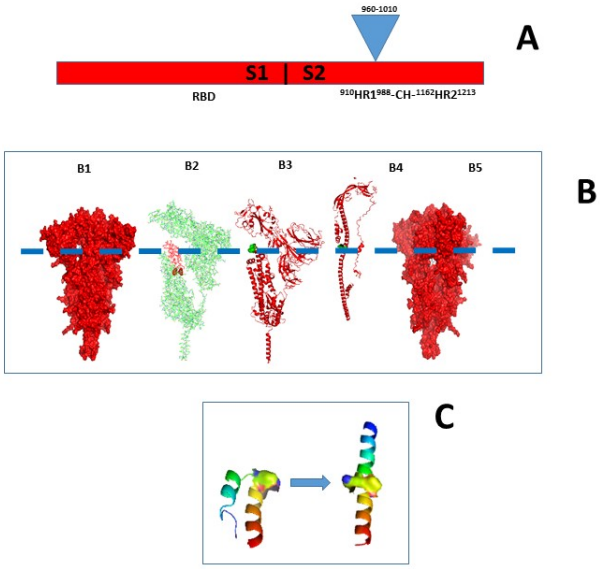

Figure 1. Location of the spring-loaded switch-folding (SLSF) segment of the S spike protein
A) Spike subunits S1 and S2. Blue triangle) Location of the spring-loaded switch-folding (SLSF) 960-1010 segment. B) Position of the exposed SLSF in the trimer. B1) The white spot visualizes the exposed residues 983-985 (983RLD) at the tip of SLSF. B2) The SLSF labeled in red in the down S2 (labeled in green) becomes more accessible after removing S1 (labeled in green). The amino and carboxyl terminus of the SLSF segment were labeled as red spheres. B3) The residues 986/987 (green spheres) mutated to prolines for stabilization, map between the HR1- CH sequences (S2 subunit in red). B4) SLSF longer α-helix in the postfusion state. B5) The white spot visualizes the exposed residues 960-968 (960NTLVKQLSS) of the SLSF segment amino terminus. Hatched blue horyzontal line, location of the tip KV986PP of the SLSF segment. C) Detail of the unfolding-folding helix transition at the tip of SLSF segment (amino-terminal in blue and carboxy-terminal in red). The two mutated prolines labeled in the middle of the SLSF segment .

Results

Definition of the minimal spike S amino acid segment for computational targeting the spring-loaded (SL) switch-folding (SF)

To computationally study the hypothetical binding of ligands to the HR1-CH spring-loaded switch-folding (SLSF) segment with molecules of small molecular size, a preliminary exploration of the HR1-CH was undertaken to locate a minimal target sequence. For that, the 3D structure of the 6xr8 (RCSB ID) conformer, of a wild-type amino acid sequence of a prefusion all RBD-down conformation [32] was used. The 6xr8 conformer was chosen because it may best represent the native virion sequence and structure (Table S1) and therefore may also be one of the best targets for preventive and therapeutic purposes.

The sequence size was expanded around the 2 α-helices around the ...986KV... tip were the inactivating PP mutations were inserted in other conformers [10], using growing grid sizes from a minimal of 9 x 9 x 9 Å centered around that tip. Results concluded that a minimal amino acid segment expanding residues 960-1010, contained most of the significant binding-scores by seeSAR and/or AutoDockVina (Figure 1, B2, B3, C). Additional extensions along any of the 2 α-helices did not discovered any more shorter distances between the 2 α-helices (to be simultaneously bound by short ligands) and/or did not predicted any other binding-pockets (data not shown). Therefore, residues 960-1010 were selected to define the SLSF target for these studies.

Visual inspection predicted that the SLSF segment should be partially accessible through two "windows" visualized when rotating the S trimer compact structure model (Figure 1B, B1 and B5 white spots). The surface-exposed amino acid residues were ~ 983-985 near the tip (983RLD) and ~ 960-968 (960NTLVKQLSS) near the amino end of the SLSF segment.

Number of leads resulting from the high-troughput screening of natural products against the trimer and monomer 6xr8 conformer

High-troughput screening of 135831 random natural compounds to the trimer and monomer 6xr8 conformer were performed in parallel by seeSAR and AutoDockVina.

SeeSAR screening of trimer 6xr8 identified 2948 leads with binding-scores < 200 nM, while that of monomers identified 3045 leads with scores < 10000 nM (Figure S2). The corresponding relative frequency profiles showed that a majority of leads were concentrated in the lowest scores corresponding to peaks at ~ 50 nM in trimers and ~2500 nM in monomers. AutoDockVina screening of trimer 6xr8 identified 1864 leads with binding-scores < 50 nM while monomers identified 1168 leads with scores < 1350 nM (not shown).

Next, to reduce the number of SLSF conformers to facilitate analysis, 40 available SLSF 3D monomer structures were first compared to select 10 representatives. The selected 10 SLSF conformers and the leads predicted as described above were docked by seeSAR and AutoDockVina for top-leads.

Selection of SLSF conformers

Monomer SLSF conformer 3D structures were extracted from the S trimer solved structures available in their pdb files (Table S1).

Structural superposition, showed that most of the SLSF in prefusion (open, intermediate and close states), were similarly folded when compared to the 6xr8 conformer, as suggested by RMSDs < 0.6 Å. Similar conformations were observed among several other SLSF structures [11, 33, 34], despite having neutralizing antibodies bound to the S trimers [2, 4-6], or after further mutating to P [11, 12]. In contrast, RMSD between 0.6 to 2.90 Å were obtained for recent dominant mutants [35, 36], after binding some antibodies [7], after the pH was lowered [37], or after S1-S2 cleavage [38]. Therefore, assuming that similar RMSD in the same amino acid sequences would not significatively change binding-scores, conformers with RMSD < 0.5 Å were excluded from further analysis. Additionally, no superpositions were predicted for postfusion conformers [8, 32] (Table S1).

The 9 prefusion and 1 postfusion conformers described in Table S1 were then selected as representative of different SLSF 3D structures. The following conformers were selected, those from prefusion-close all-down (6xlu, 6xm5, 6xey), and prefusion-open or partially open (6vyb, 6xs6, 6zgh, 6zgg, 6xm4) states (Table S1, yellow background and bold lettering). The wild-type ^{986}KV sequence was present in the 6xr8 prefusion and 6xra postfusion, while the rest contained the ^{986}PP mutations. The rest of the amino acid sequences were the same for all conformers (not shown).

Table 1. Top-leads of the seeSAR 6xr8 trimer-leads bound by trimer-conformers

Top-Leads	6xr8	6xlu	6xm5	6xey	6vyb	6xs6	6zgh	6zgg	6xm4	Total
SN00236117	**11.3**									9
SN00333487	**46.3**									9
SN00030711	**1.0**									7
SN00241472	**2.3**									7
SN00339301	**13.4**									7
SN00360448	**4.3**									6
SN00379984	**31.3**									6
SN00350832	**34.5**									6
SN00316933	**40.6**									6
SN00359351	**5.3**									5
SN00037008	**24.8**									5
SN00395077	**33.7**									5
SN00363785	**2.2**									4
SN00272769	**2.6**									4
SN00236633	**3.4**									4
SN00072922	**8.2**									4
SN00071475	**10.2**									4
SN00316223	**12.2**									4
SN00031000	**14.6**									4
SN00334251	**15.7**									4
SN00330379	**18.1**									4
SN00020460	**32.7**									4
SN00306080	**37.4**									4
SN00327581	**37.7**									4
SN00073534	**39.9**									4
SN00024546	**42.7**									4
SN00072921	**44.8**									4
SN00030713	**45.5**									4
SN00334033	**46.6**									4
SN00071389	**47.1**									4
SN00317046	**49.8**									4
Total,%	100	77.4	70.9	35.4	77.4	67.7	41.9	16.1	6.4	

SeeSAR 2948 trimer-leads <50 nM, ranked by the total bound number of conformers with < 10 µM. Red heading, closed-conformers. Blue heading, open-conformers. Dark green, 6xr8 and conformer leads with binding-scores < 50 nM. Ligh green, conformer with binding-scores < 10 µM. Total, number of conformers (vertical) and percentage of top-leads bound by each conformer (down and horyzontal).The data for the 6xra postfusion conformer were not represented. Yellow-background, Top-leads common to Table 1 and 2.

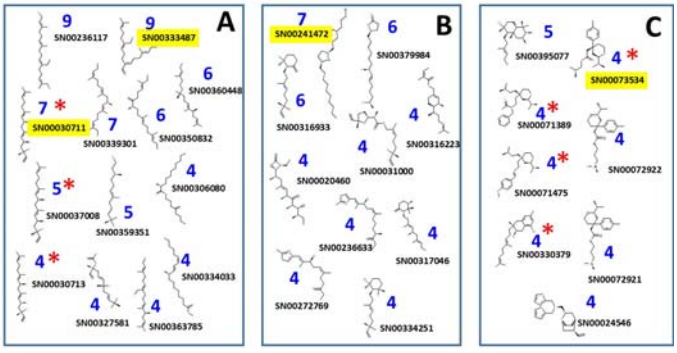

Figure 2. 2D graphic representation of top-leads of Table 1. Top-leads of Table 1. Cluster chemotypes showing none (A), 1 (B) or 2-3 rings (C). Blue numbers, number of trimer-conformers bound. Yellow-background, Top-leads common to trimers (Table 1) and monomers (Table 2). Red stars*, commercially available.

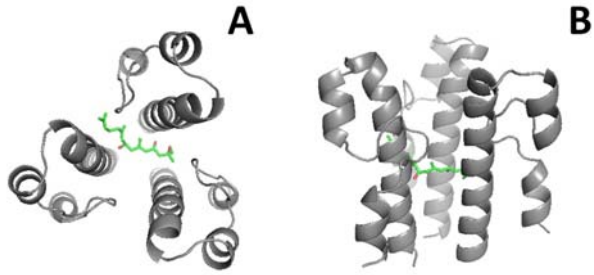

Figure 3. 3D graphic representation of the best pose of the 6xr8 +SN00333487 complex. Most of the top-leads and conformers (Figure 2, Table 1) showed similar best poses to those of SN00333487 and 6xr8 (A and B green lines), as drawn in PyMOL. A, top view. B, side view with the SLSF tip up and their amino and carboxyl ends down. Gray drawings, SLSF trimer structures.

6xr8 leads docked to trimer- and monomer-conformers

The trimer- and monomer-6xr8 leads were docked by seeSAR (Figure S3, A,B,C,D) and by AutoDockVina (Figure S3, E,F,G,H) to the selected trimer and monomer SLSF-conformers to first compare their binding-score profiles.

The seeSAR trimer-leads, were more numerous and lower after docking to trimers (lowest 6xr8 red profile: ~1 nM; Figure S3,A) than to monomers (lowest 6xr8 red profile: ~500 nM; Figure S3, B). Similarly, the seeSAR monomer-leads were more numerous and lower with trimers (lowest 6xr8 red profile: ~1 nM; Figure S3,C) than with monomers (lowest 6xr8 red profile: either with no-binding or ~10 nM; Figure S3, D). Profile variations among conformers were wide in both trimers and monomers (Figure S3, AB and CD, respectively).

The AutoDockVina trimer-leads, were also more numerous and lower after docking to trimers (~5 nM; Figure S3,E) than to monomers (~1000 nM; Figure S3, F). Similarly, the AutoDockVina monomer-leads were more numerous and lower with trimers (~100 nM; Figure S3,G) than with monomers (~2000 nM; Figure S3, H). All these profile variations were small in trimers and monomers (Figure S3, EF and GH, respectively), in contrast to seeSAR's data.

Top-leads were defined as leads with <50 nM binding-scores to the 6xr8 conformer and bound by >4 conformers. Assuming that the lower trimer binding-scores, may be the best predictors for experimental targets, top-leads were identified from seeSAR trimer, seeSAR monomer and AutoDockVina trimer when docked to trimer-conformers (Figure S3, ACE, respectively).

Top-leads of seeSAR 6xr8 trimer-leads docked to trimer-conformers

Top-leads selected among the seeSAR 2948 6xr8 trimer-leads were 31 (Table 1, dark green column). Two top-leads (NS00236117, NS00333487) were bound by 9 and three were bound by 7 conformers.

There was an inverse correlation between the RMSD conformer values relative to 6xr8 and their number of top-leads bound (for instance, 6xey with 2.9 Å and 35.4 % of top-leads bound, 6xm4 with 1.31 Å and 6.4% of top-leads bound, 6zgg with 0.86 Å and 16.1 % of top-leads bound). Additionally, the 6xra conformer with no predicted RMSD, showed only 3.2 % of top-leads bound (not shown). Similarly, more close- (Table 1, red headings) than open- (Table 1, blue headings) conformers bound top-leads.

The seeSAR trimer top-lead chemotypes could be clustered in 3 chemotypes: no rings (A, 41.9%), one ring (B, 32.2%) or 2-3 rings (C, 25.8%) (Figure 2).

Visual inspection of the top-leads complexed with trimer-SLSFs predicted similar interactions with their amino acid neighbors. In the 6xr8 trimer, for example, the neighbors to SN00333487 mapped to the trimer inner part inside the structure binding to each of its 3 monomers (Figure 3, top view) in the middle of the SLSF segment (Figure 3, side view).

Top-leads of seeSAR 6xr8 monomer-leads docked to trimer-conformers

Top-leads selected among the seeSAR 3045 6xr8 monomer-leads were 21 (Table 1, dark green column). Four top-leads (SN00333487, SN00030711, SN00241472, SN00073534) were also among the seeSAR trimer top-leads (Tables 1, 2 and Figures 2, 4, yellow-backgrounds).

There was also an inverse correlation between the higher conformer RMSD values and their lower number of top-leads bound and more close- than open-conformers bound top-leads. The seeSAR monomer top-lead chemotypes were similar to the seeSAR trimer top-leads (Figure 4).

The seeSAR monomer top-leads mapped also to the inner part of their corresponding trimers. SeeSAR monomer top-leads with predicted binding into the middle of the trimer were very rare and with higher binding-scores (data not shown).

Table 2. Top-leads of the seeSAR 6xr8 monomer-leads bound by trimer-conformers

Top-Leads	6xr8	6xlu	6xm5	6xey	5byv	6xs6	6zgh	6zgg	6xm4	Total
SN00249430	5.6									8
SN00241472	7.7									8
SN00400153	17.0									8
SN00278612	28.6									8
SN00359607	3.2									7
SN00300994	4.7									7
SN00282570	7.0									7
SN00254120	19.3									7
SN00335571	24.3									7
SN00316933	32.0									7
SN00307456	42.8									7
SN00030711	1.0									6
SN00272486	4.4									6
SN00333487	14.8									6
SN00312704	26.9									6
SN00362440	43.5									6
SN00356917	3.2									5
SN00334964	25.3									4
SN00073534	39.9									4
SN00046678	47.8									4
SN00400131	49.7									4
Total, %	100	80.9	80.9	52.3	100	80.9	57.1	28.5	33.3	

seeSAR 3045 monomer-leads <50nM, ranked by the total bound number of conformers with < 10μM. Other details as in Table 1.

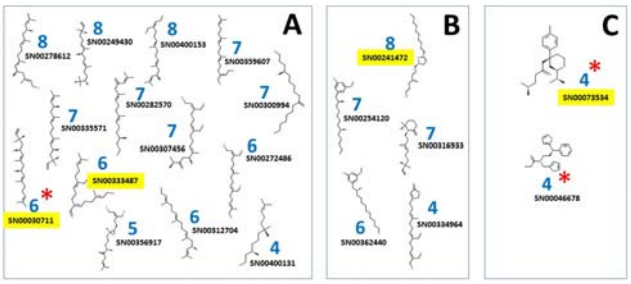

Figure 4. 2D graphic representation of top-leads of Table 2. Top-leads of Table 2. Cluster chemotypes showing none (A), 1 (B) or 2-3 rings (C). Other details as in Figure 2.

Top-leads of AutoDockVina 6xr8 trimer-leads docked to trimer-conformers

Top-leads selected among the 1864 AutoDockVina 6xr8 trimer-leads were 33 (Table 3, dark green column). There were no top-leads common to those predicted by seeSAR trimers and monomers (Table 1 and 2). There was also an inverse correlation between the RMSD and lower number of top-leads bound and more close- than open- conformers bound top-leads.

In contrast to seeSAR trimers and monomers, the AutoDockVina trimer top-leads could be clustered in, 4 different chemotypes: 2 (II, 18.1 %), 3 (III, 42.4 %), 4 (IV, 24.2 %) or 5 (V, 15.1 %) rings (Figure 5).

Visual inspection of the top-leads complexed with trimers predicted similar interactions with their amino acid neighbors and similar to the seeSAR top-leads. Thus, in the 6xr8 trimer, for example, the neighbors to top-leads bound by at least 5 conformers mapped into the inner part inside the trimer (Figure 6, A) and in the middle of the SLSF segment (Figure 6, B).

In silico analysis of drug-like properties of top-leads

The corresponding *in silico* pharmacokinetic parameters, physicochemical and toxicity ADME predicted properties of the top-leads from Tables 1, 2 and 3, showed that they have good characteristics for drug development since most of them were soluble, complied with Lipinski rules and have enough gastrointestinal permeability predictions (Table S2). As expected by the higher number of carbon rings, the top-leads predicted by AutoDockVina (Table 3) were more bulky, worse for gastrointestinal absorption and have more chemical parts known to be toxic or unstable (Brenk alerts) when compared to those predicted by seeSAR (Tables 1 and 2). However, some of the top-leads were predicted to be inhibitors of the most important detoxifying cytochromes P450 (CYP1A2 and CYP3A4), which may raise some physiological problems for drug-like purposes.

Predicted SLSF binding to known drugs

To explore whether there exist any other SLSF-binding compounds among those already FDA-approved or listed as NCGC-pharmaceutical drugs, a maximal number of those were retrieved from different sources. Similar docking screenings to the natural compound library were applied to these libraries but restricted to 6xr8 trimer targets. The 1700 FDA-approved and 7879 NCGC drugs screened by seeSAR predicted only 7 and 22 < 50 nM leads, respectively. Those leads were docked to the rest of trimer-conformers to predict their corresponding top-leads. As expected the profiles of their binding-scores were lower for the 6xr8 conformer and similar to the other profiles obtained with natural products, however their number of leads were ~100-fold lower (Figure S4).

Table 3. Top-leads of the AutoDockVina 6xr8 trimer-leads bound by trimer-conformers

Top-Leads	6xr8	6xlu	6xm5	6xey	6vyb	6xs6	6zgh	6zgg	6xm4	Total
SN00171986	6.1									6
SN00237200	10.1									6
SN00139699	11.9									6
SN00279624	19.8									6
SN00052785	23.5									6
SN00025089	46.1									6
SN00064143	8.5									5
SN00001854	11.9									5
SN00147258	14.1									5
SN00023927	16.7									5
SN00022518	23.5									5
SN00118894	23.5									5
SN00123877	23.5									5
SN00161487	23.5									5
SN00126519	32.9									5
SN00002685	39.0									5
SN00120545	39.0									5
SN00236177	7.2									4
SN00261691	14.1									4
SN00234593	16.7									4
SN00133277	16.7									4
SN00262902	19.8									4
SN00121318	32.9									4
SN00031647	32.9									4
SN00131462	32.9									4
SN00263240	39.0									4
SN00005569	39.0									4
SN00031715	39.0									4
SN00031719	39.0									4
SN00132791	39.0									4
SN00139629	46.1									4
SN00014964	46.1									4
SN00164272	46.1									4
Total, %	100	33.3	63.6	84.8	27.2	39.3	69.6	18.1	21.2	

AutoDockVina 1864 trimer-leads <50 nM ranked by the total bound number of conformers with < 250 nM. Other details as in Table 1.

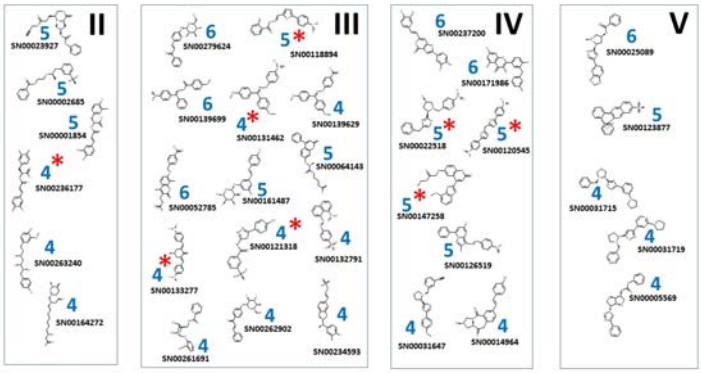

Figure 5 . 2D graphic representation of top-leads of Table 3 . Top-leads of Table 3. Cluster chemotypes showing 2 (II), 3(III), 4 (IV) or 5(V) rings. Other details as in Figure 2.

Figure 6. 3D graphic representation of superposed best poses of the 6xr8 top-leads. Superposed drawings of top-leads bound by >5 conformers (Figure 5, Table 3). Other details as in Figure 3.

There were 7 drug top-leads that bound more than 4 conformers (Table 4), 2 corresponded to FDA-approved drugs and 5 to NCGC drugs (Figure 7). Most of the drug chemotypes predicted to bind at the low nM range to trimer SLSF have 2-fold symmetries except Tinosorb-s which has 3-fold symmetry and have the lowest predicted score for binding to the 6xr8 conformer. The unique 3-fold symmetry structure of Tinosorb s, suggests their fitting to the trimeric inner part of the SLSFs. To note that these drug chemotypes were of higher molecular weights (i.e., 627 Daltons of Tinosorb-s) than any of the natural compounds newly described above (<380 Daltons) and therefore showed lower binding scores when compared to some of the natural compounds.

Visual inspection of 3D models of the best poses of top-leads complexed with trimer-SLSFs predicted similar interactions with their amino acid neighbors. Thus, the amino acid neighbors in the 6xr8 trimer of Tinosorb-s (the lowest predicted binding-score to 6xr8), for example, mapped into the inner part inside the trimer, hypothetically interacting with each of the 3 monomers (Figure 8, A) and in the middle of the SLSF (Figure 8, B). The Tinosorb-s bound to 6xr8 SLSF also confirms its location within the trimer suggested by its 3-fold symmetry. The models were very similar to the best poses of most of the top-leads from the natural compounds described above.

Table 4. Drug top-leads of trimer-seeSAR bound to trimer-conformers

Top-Leads	6xr8	6xlu	6xm5	6xey	6vyb	6xs6	6zgh	6zgg	6xm4	Total
Chaulmosulfone	0.6									9
Latanaprost	13.9									8
Deditonium	28.9									8
Tinosorb-s	0.003									5
Estradiol mustard	0.2									4
Fexofenadine	14.7									4
Laniquidar	40.8									4

The Table shows the 27 seeSAR 6xr8 leads <50nM, ranked by the total bound number of conformers with < 10 μM, obtained from FDA-approved and NCP libraries. One of two independent dockings with similar results were represented. Names in blue, FDA-approved drugs. Names in bold, NCGC drugs. Other details as in Table 1.

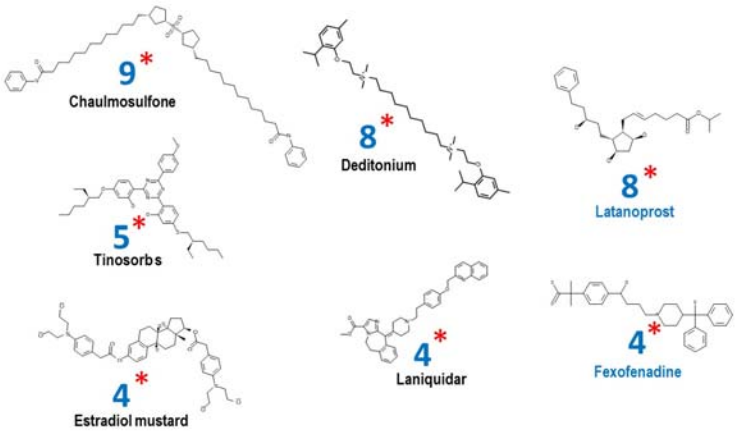

Figure 7 . 2D graphic representation of drug top-leads of Table 4 . Names in blue corresponded to FDA-approved drugs while the rest were from NCGC drugs. Other details as in Figure 2.

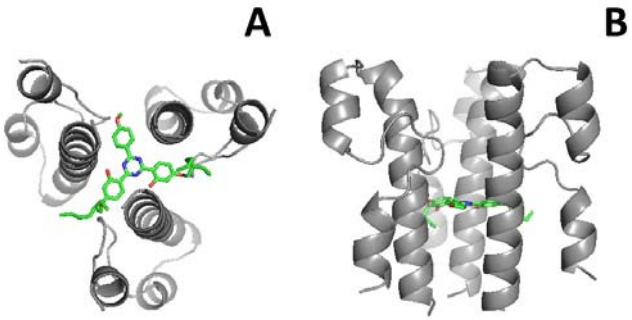

Figure 8. 3D graphic representation of the complex of the 6xr8 SLSF and the best pose of Tinosorb-s. Representative drawing of top-leads and conformers (Figure 7, Table 4). Other details as in Figure 3.

Discussion

After exploring the trimer and monomer conformations of SLSF (S residues 960-1010) of SARS-CoV2 by computational ligand screening, dozens of leads and top-leads were predicted to bind in the low nM range to a high number of conformers from thousands of natural compounds with <380 Dalton and < 6 logP. Surprisingly, leads and top-leads identified here were targeted to SLSF residues far from their tips, where PP mutations affecting fusion and infectivity had been previously located. Nevertheless, the predicted top-leads may be used to experimentally find out whether it will be possible to inhibiting fusion and infection.

The SLSF segment was chosen as a computational target for ligands because it had not been explored before with such purpose. The SLSF segment contains an spring-loaded switch folding that when unfolded-refolded converts two α-helices to a larger α-helix required for fusion. The double screening (seeSAR and AutoDockVina) approach used here consisted in a combination of high-throughput screening of a native trimer and monomer SLSF conformer to predict leads, followed by docking to several SLSF conformers. After estimating that the variations in conformation which may had been introduced by using PP mutants did not cause conformational changes at the SLSF segment (low RMSD values), top-leads were defined by taking into account the number of conformers bound (vote approach). Rather than relying solely in lead binding-scores, these additional criteria was expecting to increase their chances to predict experimental success. This conformer approach may also be looked as a simplified alternative to molecular dynamic procedures, which although will best mimic some of the experimental situations, it uses prohibitive computational costs for screening a large number of ligands. Additionally, two different docking algorithms were used because of previously reported experimental failures, due to limitations of actual molecular scoring programs [39-41]. Nevertheless, because of the absence of experimental data to evaluate the relative physiological importance of each conformer, it should be recognized that all these assumptions are only some among many other possibilities. An example of the above considerations, is the lack of coincidence among the top-lead chemotypes predicted by seeSAR and AutoDockVina. These data not only may reflect the different algorithms these programs use but also the molecular diversity of possible solutions to find compounds which bind the inner part of the SLSF segment.

One of the challenges for experimental prediction is how accessible is the SLSF segment in the native infectious viral particle. The partial accessibility of SLSF predicted by modeling the native closed all-down S trimers, suggests that SLSF may be reached even when inside the trimeric highly compacted prefusion S conformations. Theoretically, experimental binding could be possible for some of the top-leads with the highest affinities and small and elongated-structures. Furthermore, the presence of partially open S structures in some prefusion states described by several authors, suggests that the S1-S2 interactions may be under spontaneous continuous displacements, thus theoretically increasing the accessibility to SLSF. Therefore, possibilities may be high for high-affinity and small ligands to get access to bind SLSF at the prefusion states. However, experimental binding and whether binding top-leads would inhibit fusion, remains to be investigated. In this respect, experimental difficulties may also arise by top-leads in the low nM binding-score (high affinity) range. Since to stabilize the S prefusion state any of the studied drugs must predict strong ligand affinities to SLSF to lock the spring-loaded mechanism, such high affinities may also recognize similar combinations of amino acids in other proteins, generating unexpected undesirable side-effects.

No top-leads could be found for the post-fusion state, making the top-lead binding possibilities less likely once SLSF reaches that conformation. Although, there is a requirement for a trimer-dependent inner binding-pocket for the 6xr8-dependent top-lead bindings, alternatively, it would be possible to high-throughput screen for postfusion targets using the 6xra rather than the 6xr8 conformer in future work. Similarly, other possibilities may also be explored such as targeting the surface of the SLSF or the binding interfaces with other S domains or using other SLSF conformers for the initial high-throughput screening.

The conformer-dependent wide variation in binding-scores, despite having the same amino acid sequences and similar 3D solved structures (low RMSDs), was remarkable. At this respect, the conservation of S amino acid sequences among SARS-CoV2 isolates is high since only 9 amino acid substitutions, most of them conservative, were found among 61 SARS-CoV2 S sequences (Global Initiative on Sharing All Influenza Data database, https://www.gisaid.org/) [1]. Therefore, most of the observed binding-score differences among SLSF conformers are most probably due to small differences on 3D structures. Among those sequence variations, the D614G mutation that has become dominant as the pandemic proceeds, has been implicated in increasing the spread of the virus by favoring S up-conformations and by changing tissue tropisms [35, 36]. The SLSF corresponding to the D614G mutation caused an 0.6 Å RMSD change, however 67.7 % of the top-leads still were predicted to bind to it, therefore minimizing possible scapes of the mutant to those top-leads.

A few FDA and NCPG drugs were also predicted to bind SLSF in the low nM range. Although all those drugs were of higher molecular weights than those investigated in this work, they may also deserve further experimental investigations. For instance, Tinosorb-s, a compound used to prevent damage in the skin by ultraviolet light, could be a candidate to stabilize prefusion conformations, however, because of their physiological properties, it may be only suitable to disinfect surfaces rather than for therapeutic purposes.

Among many other possibilities, the lists of leads/top-leads and not-binding ligands identified here may be used as several choices for training sets for deep learning approaches. Deep learning approaches may serve to screen larger libraries of millions of compounds based solely on their chemical attributes rather than on the labor-intensive docking. Such approaches may be used as alternatives for a first step of the first high-throughput computer-intensive screening looking for downsized leads, thus leaving the more traditional docking to predict top-leads for experimental validation. Finally, *in vitro* assays testing binding to recombinant S, inhibition of fusion and possible blocking of *in vitro* coronavirus cellular infection, could be employed to experimentally validate some of the proposed predictions.

Supporting information

Table S1. 3D similarities among SLSF trimer-conformers

PDB S code	RMSD Å	Characteristics	S state	Ref
6xr8	0.00	RBD 3down	prefusion closed	32
6xra	none	helix+helix	postfusion	
6vxx	0.37	RBD 3down	prefusion	33
6vyb	0.40	RBD 1up	prefusion open	
6x2a	0.46	PP+ RDB 1up	prefusion intermediate	11
6x2b	0.51	PP+ RBD 2up	prefusion intermediate	
6x2c	0.42	PP+ RBD 3down	prefusion closed	
6x29	0.44	PP+C mutant	prefusion closed	
6wpt	0.41	+ NAb S309	prefusion open	2
6wps	0.41	+ NAb S309	prefusion closed	
6x6p	0.41	RBD 3down	prefusion closed	34
6xcn	0.53	+ NAb C105 #2	prefusion open	3
6xcm	0.55	+ NAb C105 #1	prefusion open	
6xkl	0.51	PPPPPP	prefusion	12
6vsb	0.50	PP	prefusion intermediate	
6xm3	0.52	RBD 1up pH5.5 #1	prefusion closed	37
6xm4	1.31	RBD 1up pH5.5 #2	prefusion open	
6xlu	0.46	RBD 3down pH4	prefusion closed	
6xm0	0.51	consensus at pH5.5	prefusion closed	
6xm5	0.53	RBD 3down at pH5	prefusion closed	
6xs6	0.60	CoV2-D614G	prefusion open	35
7c2l	0.50	CoV2- D614G+Ab4A8	prefusion open	
7byr	0.53	+ BD23 NAb	prefusion	4
6z43	0.51	+ NAb	prefusion open	5
6z97	0.41	+ MAb CR3022	prefusion open	
6zdh	0.42	+serum Ab	prefusion closed	6
6zox	0.44	S-R/x2	prefusion closed	13
6zoy	0.41	S-R/PP/x1	prefusion closed	
6zoz	0.43	S-R/PP/x1	prefusion locked	
6zp0	0.43	S-R	prefusion closed	
6zp1	0.43	S-R/PP	prefusion closed	
6zp2	0.30	S-R/PP	prefusion locked	
6zge	0.48	uncleaved	prefusion closed	38
6zgi	0.53	S1-S2 cleaved	prefusion closed	
6zgh	0.78	S1-S2 cleaved	prefusion intermediate	
6zgg	0.86	S1-S2-cleaved	prefusion open	
6xey	2.90	+anti-RBD NMAb2-4	prefusion closed	7
6zow	0.45	higher resolution hr	prefusion	8
6zp5	none	hr. helix-helix	postfusion closed	
6zp7	none	hr. helix-helix	postfusion open	

All the SLSF were obtained from 3D S spike trimeric protein models of SARS-CoV2 (SARS, Severe Acute Respiratory Syndrome) downloaded from the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank. Conformers contained the ^{986}PP mutations except the native 6xr8 and 6xra which contained the ^{986}KV wild-type sequence. Structural similarity relative to 6xr8 was expressed in RMSD Å (http://www.ccp4.ac.uk/MG). RMSD, Root Mean Square Differences. S state, prefusion, spike structures before viral/host membrane fusion. postfusion, spike structures after viral/host membrane fusion. Closed, RBDs down. Open, 1,2 or 3 RBDs up (receptor-accessible). Locked, a more dense closed conformation [13]. S-R, the S1/S2 cleavage site was replaced by arginine (R). x1, stabilizer disulphide link between 383 and 985. x2, stabilizer disulphide link between 413 and 987. Bold & yellow, conformers selected for further studies. Red PDB, closed conformation representatives. Blue PDB, open conformation representatives. Gray PDB, postfusion representative.

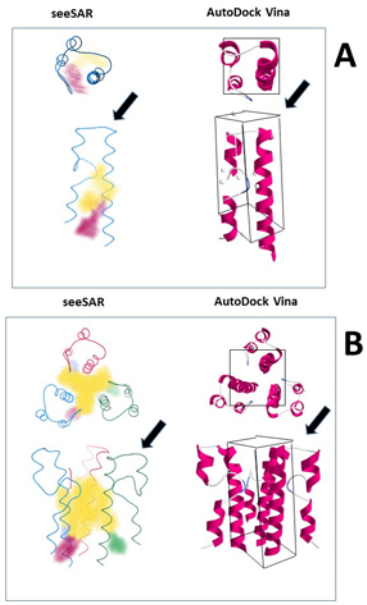

Figure S1. Scheme of the SLSF monomer (A) and trimer (B) 3D models.
SeeSAR predicted a number of binding-pockets (colored shadows at the left). AutoDockVina works on submitted internal grids (rectangles at the right). A and B up, top views. A and B bottom, side views. Black arrows, location of the SLSF tips on side views.

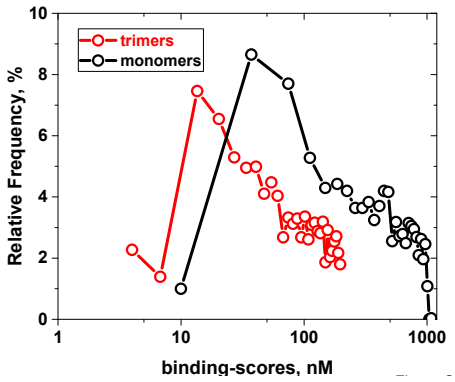

Figure S2. Comparison of relative frequencies of lead binding-scores < 1000 nM from the 6xr8 trimer and monomer. The screenings for binding of natural ligands to the 6xr8 conformer was carried out by seeSAR. The distribution of lead frequencies were calculated and represented in Origin.

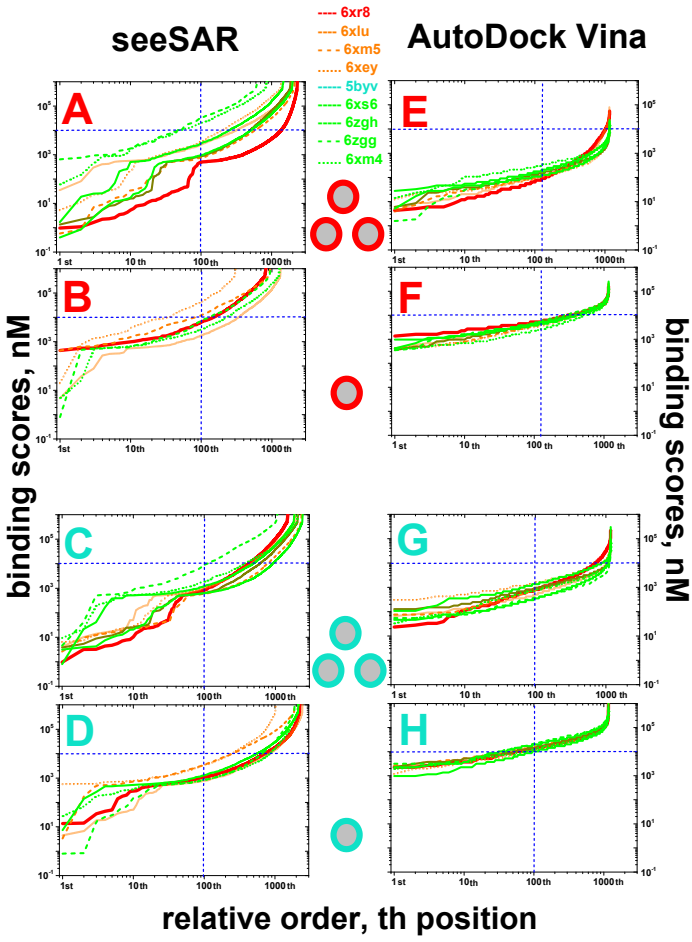

Figure S3. Binding of trimer and monomer 6xr8 leads to conformers. Each set of nM binding-scores were ordered from lower to higher and the first 1000 represented.The 6xr8 trimer (red A,B,E,F) and monomer (green C.D,G,H) leads were docked to trimers (3 central circles) and monomers (one central circle) by seeSAR (A,B,C,D) or by AutoDockVina (E,F,G,H). Red-edged grey circles, trimer 6xr8 leads (bound to 3 trimer or 1 monomer SLSFs). Green-edged grey circles, monomer 6xr8 leads (bound to 3 trimer or 1 monomer SLSFs). Red line, 6xr8 conformer. Orange line, 6xlu. Orange dashed-line, 6xm5. Orange dot-line, 6xey. Dark green, 6vyb. Green lines, 6xs6 and 6zgh. Green dash-line, 6zgg. Green dot-line, 6xma. A) seeSAR trimer 6xr8 2948 leads docked to trimer-conformers. B) seeSAR trimer 6xr8 2948 leads docked to monomer-conformers. C) seeSAR monomer 6xr8 3045 leads docked to trimer-conformers. D) seeSAR monomer 6xr8 3045 leads docked to monomer-conformers. E) AutoDockVina trimer 6xr8 1864 leads docked to trimer-conformers. F) AutoDockVina trimer 6xr8 1864 leads docked to monomer-conformers. G) AutoDockVina monomer 6xr8 1168 leads docked to trimer-conformers. H) AutoDockVina monomer 6xr8 1168 leads docked to monomer-conformers.

Table S2. Drug-like characteristics of top-leads predicted by the SwissADME web server

Top-leads Table 1	TPSA	LIPK	LogP	Sol	GIA	1A2 inh	3A4 inh	PAINS	Brenk
SN00236117	60.7	0	4.1	S	High	Yes	Yes	0	1
SN00333487	57.5	0	4.1	S	High	No	Yes	0	3
SN00030711	60.7	0	4.0	S	High	No	No	0	1
SN00241472	49.7	0	5.1	M	High	No	No	0	0
SN00339301	60.7	0	4.1	S	High	Yes	Yes	0	1
SN00316933	20.2	1	5.3	M	High	No	No	0	1
SN00350832	60.7	0	4.1	S	High	Yes	Yes	0	1
SN00360448	60.7	0	4.1	S	High	Yes	No	0	1
SN00379984	66.8	0	3.5	S	High	No	No	0	1
SN00037008	60.7	0	3.9	S	High	No	No	0	1
SN00359351	40.5	0	3.7	S	High	Yes	No	0	1
SN00395077	80.9	0	2.7	S	High	No	No	0	1
SN00020460	63.6	0	4.2	M	High	No	Yes	0	2
SN00024546	65.4	0	2.2	M	High	No	No	0	1
SN00030713	60.7	0	4.0	S	High	No	No	0	1
SN00031000	55.8	0	3.9	S	High	No	Yes	0	1
SN00071389	29.5	0	5.1	M	High	No	Yes	0	0
SN00071475	30.8	0	5.4	M	High	No	No	0	1
SN00072921	42.8	0	2.7	M	High	No	No	0	0
SN00072922	42.8	0	2.7	M	High	No	No	0	0
SN00073534	38.3	0	4.2	M	High	No	No	0	0
SN00236633	78.4	0	4.7	M	High	Yes	Yes	0	1
SN00272769	78.4	0	3.8	S	High	No	No	0	1
SN00306080	63.6	0	4.5	M	High	Yes	Yes	0	0
SN00316223	46.5	0	4.2	M	High	No	Yes	0	2
SN00317046	60.7	0	3.6	S	High	No	No	0	1
SN00327581	66.8	0	2.8	S	High	No	No	0	1
SN00330379	49.7	0	4.8	M	High	No	Yes	0	1
SN00334033	46.5	0	4.9	M	High	Yes	Yes	0	1
SN00334251	40.5	0	4.6	M	High	No	No	0	1
SN00363785	60.7	0	4.0	S	High	Yes	Yes	0	1

Top-leads Table 2	TPSA	LIPK	LogP	Sol	GIA	1A2 inh	3A4 inh	PAINS	Brenk
SN00249430	80.9	0	3.3	M	High	No	No	0	1
SN00241472	49.7	0	5.1	P	High	No	No	0	0
SN00400153	40.5	1	5.0	P	High	Yes	Yes	0	1
SN00278612	60.7	0	4.1	M	High	Yes	No	0	1
SN00359607	40.5	0	4.8	M	High	Yes	Yes	0	1
SN00300994	37.3	0	5.0	P	High	Yes	Yes	0	0
SN00282570	37.3	0	5.0	P	High	Yes	Yes	0	2
SN00254120	69.9	0	4.1	P	High	Yes	No	0	0
SN00335571	60.7	0	4.0	M	High	No	No	0	1
SN00316933	20.2	1	5.3	P	High	No	No	0	1
SN00307456	80.9	0	3.6	M	High	Yes	No	0	1
SN00030711	60.7	0	4.0	M	High	No	No	0	1
SN00272486	80.9	0	3.3	M	High	No	Yes	0	1
SN00333487	57.5	0	4.1	M	High	No	Yes	0	3
SN00312704	40.5	0	4.8	P	High	Yes	Yes	0	1
SN00362440	49.7	0	5.1	P	High	Yes	No	0	0
SN00356917	70.1	0	3.4	M	High	No	No	0	2
SN00334964	87.0	0	2.9	S	High	No	No	0	1
SN00073534	38.3	0	4.2	M	High	No	No	0	0
SN00046678	46.6	0	3.7	M	High	No	Yes	0	0
SN00400131	40.5	0	3.6	M	High	No	No	0	1

Top-leads Table 3	TPSA	LIPK	LogP	Sol	GIA	1A2 inh	3A4 inh	PAINS	Brenk
SN00171986	131.4	0	2.7	M	Low	Yes	No	0	0
SN00237200	124.3	0	3.0	M	High	Yes	No	1	1
SN00139699	96.5	0	3.4	M	High	No	No	0	2
SN00279624	116.5	0	1.1	S	High	No	No	0	1
SN00052785	119.7	0	2.7	M	High	Yes	No	0	1
SN00025089	90.9	0	1.4	M	High	No	Yes	0	0
SN00064143	108.7	0	2.6	S	High	No	Yes	0	1
SN00001854	147.4	0	1.3	S	Low	No	No	1	2
SN00147258	88.1	0	3.6	M	High	Yes	Yes	0	1
SN00023927	103.8	0	-0.3	S	High	No	No	0	1
SN00022518	108.2	0	2.1	M	High	No	Yes	0	1
SN00118894	120.7	0	2.4	M	High	Yes	No	0	2
SN00123877	91.4	0	3.3	M	High	Yes	No	1	3
SN00161487	139.8	1	0.6	S	High	No	No	0	1
SN00126519	143.8	0	2.4	M	Low	No	No	1	2
SN00002685	42.9	0	3.6	M	High	Yes	No	0	0
SN00120545	117.9	0	3.8	M	Low	Yes	No	0	1
SN00236177	144.5	0	1.6	S	Low	No	No	1	2
SN00261691	125.7	0	0.9	S	High	No	No	0	0
SN00234593	154.0	0	1.0	S	Low	No	No	1	4
SN00133277	108.7	0	3.5	M	Low	No	Yes	1	2
SN00262902	116.5	0	1.0	S	High	No	No	0	1
SN00121318	68.0	0	4.2	M	High	Yes	No	0	0
SN00031647	92.1	0	1.5	S	High	No	Yes	0	0
SN00131462	88.7	0	3.7	M	High	No	No	0	2
SN00263240	110.4	0	2.2	S	High	No	No	0	0
SN00005569	100.4	0	1.5	S	High	No	Yes	0	0
SN00031715	60.7	0	2.0	M	High	Yes	Yes	0	0
SN00031719	73.6	0	1.4	S	High	No	Yes	0	0
SN00132791	136.1	0	2.9	M	Low	No	No	1	3
SN00139629	83.0	0	3.4	M	High	No	No	0	1
SN00014964	89.9	0	1.7	S	High	No	No	0	1
SN00164272	114.3	0	0.9	S	High	No	No	0	1

The corresponding 2D structures to the SuperNatural II SN numbers can be consulted at Figure 2,4 and 5 or at http://bioinf-applied.charite.de/supernatural_new/index.php. Sol, solubilities in water classified in general classes. LIPK, number of violations of Lipinski rules that would make the ligand less likely to be an orally administrable drug if >5. It 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. TPSA, estimates of the amount of topological polar molecular surface area, lowest values facilitate permeation of cell membranes (best to be <90 Å^2). 1A2, 3A4, inhibition of the main detoxyfying cytochromes P450. GIA, prediction of gastro-intestinal adsorption. PAIN, Pan Assay Interference Structures (PAINS), alerting of the number of chemical fragments that return false positive signals in virtual binding. Brenk, alerting of the number of chemical moieties known to be toxic and/or unstable. Green, favorable. Yellowish, moderate. Reddish, unfavorable.

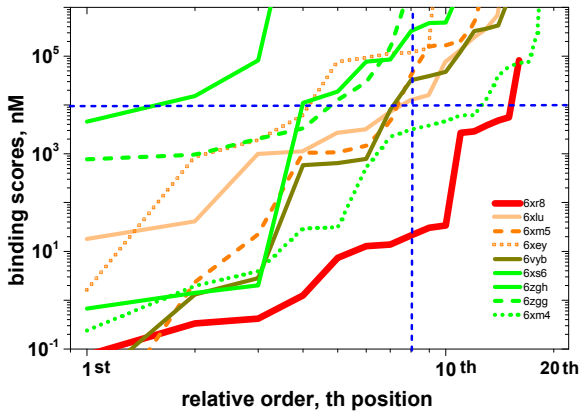

Figure S4. Binding of drug seeSAR trimer 6xr8 leads to conformers. Each set of nM binding-scores were ordered from lower to higher and the first 27 represented. Other details as in Figure S3.

Funding
The work was carried out without any external financial contribution

Competing interests
The author declares no competing interests

Authors' contributions
JC designed, performed and analyzed the dockings, and drafted the manuscript.

Acknowledgements
Thanks are due to Dr. Alberto Villena from the University of Leon (Spain) and to Dr. Ignacio Garcia from the Hospital Gomez Ulla (Madrid, Spain) for their help with the bibliography and data, to Dr. Rafael Blasco at INIA for his original ideas and discussions and to Dr.Jose Antonio Encinar from the IBMC-UMH, Elche (Spain) for providing the initial SuperNatural II sdf file of random natural compounds.

References

[1] Wrapp, D., N. Wang, K.S. Corbett, J.A. Goldsmith, C.L. Hsieh, O. Abiona, . . . J.S. McLellan.Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation *Science.* 2020, 367: 1260-1263.*science.abb2507 [pii], http://dx.doi.org/10.1126/science.abb2507*

[2] Pinto, D., Y.J. Park, M. Beltramello, A.C. Walls, M.A. Tortorici, S. Bianchi, . . . D. Corti.Structural and functional analysis of a potent sarbecovirus neutralizing antibody *bioRxiv.* 2020.*http://dx.doi.org/10.1101/2020.04.07.023903*

[3] Barnes, C.O., A.P. West, Jr., K.E. Huey-Tubman, M.A.G. Hoffmann, N.G. Sharaf, P.R. Hoffman, . . . P.J. Bjorkman.Structures of human antibodies bound to SARS-CoV-2 spike reveal common epitopes and recurrent features of antibodies *bioRxiv.* 2020.*http://dx.doi.org/10.1101/2020.05.28.121533*

[4] Cao, Y., B. Su, X. Guo, W. Sun, Y. Deng, L. Bao, . . . X.S. Xie.Potent Neutralizing Antibodies against SARS-CoV-2 Identified by High-Throughput Single-Cell Sequencing of Convalescent Patients' B Cells *Cell.* 2020, 182: 73-84 e16.*http://dx.doi.org/10.1016/j.cell.2020.05.025, S0092-8674(20)30620-6 [pii]*

[5] Huo, J., Y. Zhao, J. Ren, D. Zhou, H.M.E. Duyvesteyn, H.M. Ginn, . . . D.I. Stuart.Neutralization of SARS-CoV-2 by Destruction of the Prefusion Spike *Cell Host Microbe.* 2020.*S1931-3128(20)30351-6 [pii], http://dx.doi.org/10.1016/j.chom.2020.06.010*

[6] Zhou, D., H.M.E. Duyvesteyn, C.P. Chen, C.G. Huang, T.H. Chen, S.R. Shih, . . . K.A. Huang.Structural basis for the neutralization of SARS-CoV-2 by an antibody from a convalescent patient *Nat Struct Mol Biol.* 2020.*http://dx.doi.org/10.1038/s41594-020-0480-y, 10.1038/s41594-020-0480-y [pii]*

[7] Liu, L., P. Wang, M.S. Nair, J. Yu, M. Rapp, Q. Wang, . . . D.D. Ho.Potent neutralizing antibodies directed to multiple epitopes on SARS-CoV-2 spike *Nature.* 2020.*http://dx.doi.org/10.1038/s41586-020-2571-7, 10.1038/s41586-020-2571-7 [pii]*

[8] Melero, R., C.O.S. Sorzano, B. Foster, J.L. Vilas, M. Martinez, R. Marabini, . . . J.M. Carazo.Continuous flexibility analysis of SARS-CoV-2 Spike prefusion structures *bioRxiv.* 2020.*http://dx.doi.org/10.1101/2020.07.08.191072*

[9] Xia, S., Y. Zhu, M. Liu, Q. Lan, W. Xu, Y. Wu, . . . L. Lu.Fusion mechanism of 2019-nCoV and fusion inhibitors targeting HR1 domain in spike protein *Cell Mol Immunol.* 2020, 17: 765-767.*http://dx.doi.org/10.1038/s41423-020-0374-2, 10.1038/s41423-020-0374-2 [pii]*

[10] Pallesen, J., N. Wang, K.S. Corbett, D. Wrapp, R.N. Kirchdoerfer, H.L. Turner, . . . J.S. McLellan.Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen *Proc Natl Acad Sci U S A.* 2017, 114: E7348-E7357.*1707304114 [pii], http://dx.doi.org/10.1073/pnas.1707304114*

[11] Henderson, R., R.J. Edwards, K. Mansouri, K. Janowska, V. Stalls, S. Gobeil, . . . P. Acharya.Controlling the SARS-CoV-2 Spike Glycoprotein Conformation *bioRxiv.* 2020.*http://dx.doi.org/10.1101/2020.05.18.102087*

[12] Hsieh, C.L., J.A. Goldsmith, J.M. Schaub, A.M. DiVenere, H.C. Kuo, K. Javanmardi, . . . J.S. McLellan.Structure-based design of prefusion-stabilized SARS-CoV-2 spikes *Science.* 2020.*science.abd0826 [pii], http://dx.doi.org/10.1126/science.abd0826*

[13] Xiong, X., K. Qu, K.A. Ciazynska, M. Hosmillo, A.P. Carter, S. Ebrahimi, . . . J.A.G. Briggs.A thermostable, closed SARS-CoV-2 spike protein trimer *Nat Struct Mol Biol.* 2020.*http://dx.doi.org/10.1038/s41594-020-0478-5, 10.1038/s41594-020-0478-5 [pii]*

[14] Carr, C.M. and P.S. Kim.A spring-loaded mechanism for the conformational change of influenza hemagglutinin *Cell.* 1993, 73: 823-832.*https://doi.org/10.1016/0092-8674(93)90260-W*

[15] Cannalire, R., I. Stefanelli, C. Cerchia, A.R. Beccari, S. Pelliccia and V. Summa.SARS-CoV-2 Entry Inhibitors: Small Molecules and Peptides Targeting Virus or Host Cells *Int J Mol Sci.* 2020, 21.*ijms21165707 [pii], http://dx.doi.org/10.3390/ijms21165707*

[16] Tang, T., M. Bidon, J.A. Jaimes, G.R. Whittaker and S. Daniel.Coronavirus membrane fusion mechanism offers a potential target for antiviral development *Antiviral Res.* 2020, 178: 104792.*S0166-3542(20)30206-0 [pii], http://dx.doi.org/10.1016/j.antiviral.2020.104792*

[17] Xia, S., L. Yan, W. Xu, A.S. Agrawal, A. Algaissi, C.K. Tseng, . . . L. Lu.A pan-coronavirus fusion inhibitor targeting the HR1 domain of human coronavirus spike *Sci Adv.* 2019, 5: eaav4580.*http://dx.doi.org/10.1126/sciadv.aav4580, aav4580 [pii]*

[18] Wang, C., S. Xia, P. Zhang, T. Zhang, W. Wang, Y. Tian, . . . K. Liu.Discovery of Hydrocarbon-Stapled Short alpha-Helical Peptides as Promising Middle East Respiratory Syndrome Coronavirus (MERS-CoV) Fusion Inhibitors *J Med Chem.* 2018, 61: 2018-2026.*http://dx.doi.org/10.1021/acs.jmedchem.7b01732*

[19] Wu, C., Y. Liu, Y. Yang, P. Zhang, W. Zhong, Y. Wang, . . . H. Li.Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods *Acta Pharm Sin B.* 2020.*http://dx.doi.org/10.1016/j.apsb.2020.02.008, S2211-3835(20)30299-9 [pii]*

[20] 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*

[21] 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], http://dx.doi.org/10.1016/j.lfs.2020.117627*

[22] 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*

[23] Blasco, R. and J.M. Coll.In silico screening for natural ligands to non-structural nsp7 conformers of SARS coronaviruses *ChemRxiv.* 2020.*https://doi.org/10.26434/chemrxiv.12952115.v1*

[24] 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*

[25] 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*

[26] Reau, M., F. Langenfeld, J.F. Zagury and M. 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, 10.1007/s10822-017-0063-0 [pii]*

[27] 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*

[28] 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*

[29] 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*

[30] 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*

[31] 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.*3/80/80ps16 [pii], http://dx.doi.org/10.1126/scitranslmed.3001862*

[32] Cai, Y., J. Zhang, T. Xiao, H. Peng, S.M. Sterling, R.M. Walsh, Jr., . . . B. Chen.Distinct conformational states of SARS-CoV-2 spike protein *Science.* 2020.*science.abd4251 [pii], http://dx.doi.org/10.1126/science.abd4251*

[33] Walls, A.C., Y.J. Park, M.A. Tortorici, A. Wall, A.T. McGuire and D. Veesler.Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein *Cell.* 2020.*S0092-8674(20)30262-2 [pii], http://dx.doi.org/10.1016/j.cell.2020.02.058*

[34] Herrera, N.G., N.C. Morano, A. Celikgil, G.I. Georgiev, R.J. Malonis, J.H. Lee, . . . S.C. Almo.Characterization of the SARS-CoV-2 S Protein: Biophysical, Biochemical, Structural, and Antigenic Analysis *bioRxiv.* 2020.*http://dx.doi.org/10.1101/2020.06.14.150607*

[35] Yurkovetskiy, L., K.E. Pascal, C. Tompkins-Tinch, T. Nyalile, Y. Wang, A. Baum, . . . J. Luban.SARS-CoV-2 Spike protein variant D614G increases infectivity and retains sensitivity to antibodies that target the receptor binding domain *bioRxiv.* 2020.*http://dx.doi.org/10.1101/2020.07.04.187757*

[36] Plante, J.A., Y. Liu, J. Liu, H. Xia, B.A. Johnson, K.G. Lokugamage, . . . P.Y. Shi.Spike mutation D614G alters SARS-CoV-2 fitness *Nature.* 2020.*http://dx.doi.org/10.1038/s41586-020-2895-3, 10.1038/s41586-020-2895-3 [pii]*

[37] Zhou, T., Y. Tsybovsky, A.S. Olia, J. Gorman, M.A. Rapp, G. Cerutti, . . . P.D. Kwong.A pH-dependent switch mediates conformational masking of SARS-CoV-2 spike *bioRxiv.* 2020.*http://dx.doi.org/10.1101/2020.07.04.187989*

[38] Wrobel, A.G., D.J. Benton, P. Xu, C. Roustan, S.R. Martin, P.B. Rosenthal, . . . S.J. Gamblin.SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects *Nat Struct Mol Biol.* 2020, 27: 763-767.*http://dx.doi.org/10.1038/s41594-020-0468-7, 10.1038/s41594-020-0468-7 [pii]*

[39] 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*

[40] Plewczynski, D., M. Lazniewski, R. Augustyniak and K. Ginalski.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*

[41] 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/ci200269q*