Would it be possible to stabilize prefusion SARS-COV2 spikes with ligands?

short title: ligand-stabilization of spike coronavirus

Lorenzo, M.M., Blasco, R., and Coll, J*.

Maria M. Lorenzo, orcid: 0000-0001-7588-673X
Rafael Blasco, orcid: 0000-0002-8819-5767
Julio Coll, orcid: 0000-0001-8496-3493

Department of Biotechnology. Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, INIA. Madrid, Spain.

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

Abstract

Fusion to host cells and infection caused by Severe Acute Respiratory Syndrome coronavirus (SARS)-CoV2 was inhibited *in vitro* by PP mutations stabilizing prefusion states of their spike (S) protein native conformation, as reported by several authors. However, the possible stabilization of S by binding-ligands, rather than by mutations, have not been explored, nor it is yet known if it would be possible. In this work, the so called "spring-loaded switch-folding" (SLSF) expanding S amino acid residues 960-1010 was computationally targeted because SLSF surrounded the previously described PP mutations. The SLSF trimeric prefusion conformation consisted in 3x3 α-helices that require a transition to 3 longer α-helices before viral/host membrane fusion, similarly to what occurs in other enveloped viruses. Results of a double computational screening among hundred of thousands of natural compounds for binding to the wild-type isolated SLSF conformer predicted more leads for its trimers than for monomers. Further ranked by the number of SLSF-conformers bound, some of the predicted top-leads may deserve experimental validation. Additional screening among thousands of drugs identified Tinosorb, an star-shaped molecule, as the lowest binding-score lead to SLSF in the low nM range. However, despite its lower binding-score, 3-fold molecular symmetry and fitting the inner part of the SLSF α-helices, we were unable to experimentally show any specific inhibition of S-mediated membrane fusion using an VSV-pseudotyped infectivity assay, nor any virtual binding to S-SLSF using docking to whole native S trimers. Further exploring the star-shaped features may provide new molecular alternatives to cross-bind the α-helices of S-SLSF to hypothetically inhibit coronavirus fusion.

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

Introduction

The surface of infectious SARS-CoV2 (Severe Acute Respiratory Syndrome coronavirus 2), is surrounded by spike (S) glycoprotein trimers forming a corona-like structure. Most of the S trimers in the native virions 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 some 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-CoV2, the S2 subunit (residues 686-1273) contains the fusion peptide (788–806), the amino-terminal heptad repeat 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 specific mutations [10-13]. For instance, mutations to prolines (P) in some of the residues located in HR1-CH 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 other enveloped virions. In wild-type coronavirus, after RBD binding to host receptors, the unfolding of SLS prepares the virion S for viral-host membrane fusion. 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 S 3D-structures were solved using PP mutants (Table S1).

The HR1-CH sequence of SARS-COV2 contains the SLS [10, 14] which maintain 3 α-helices per monomer folded and separated by non-helix residues (spring-loaded) (Figure S1 and 3). Once the SLS folding (SLSF) is unfolded (Figure 1C), refolding inside the S trimer generates three unique longer α-helices (3x3 to 3x1 S-SLSF α-helices transition per trimer). Then HR1-CH-HR2 complexes form a fusion-competent 6-helix core of coiled-coils (one antiparallel complex of 3 internal HR2 + 3 external HR1-CH). A S2' protease-mediated cleavage liberates the internal fusion peptide to be inserted into the host-cell membrane to proceed with viral/host membrane fusion.

Previous successful examples on inhibition of fusion by drugs targeting HRs have been reported in several enveloped viruses, including SARS-COV2 [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, peptides derived from HR2 and binding to HR1 also 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 was 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 searching for more potent binding ligands (i.e., in the low nM range) could be a source of possible prefusion inhibitor candidates for SARS-CoV2. Perhaps targeting SLSF rather than the HR complex core would provide some alternatives. Such possibilities may benefit from preliminary computational predictions.

While most ligands with anti-coronavirus activities 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 SLSF [11-13], the existence of putative binding-ligands in the low nM range were explored here. Such hypothetical ligands may be capable of stabilizing SLSF prefusion conformations, perhaps by non-covalent crosslinking of α-helices or by avoiding their displacement. Whether such ligands binding to SLSFs do exist and could inhibit viral fusion remains to be demonstrated.

The results obtained here predicted that, **i)** top-leads do exist that predict binding-scores to SLSF in the low nM range, including some known drugs, **ii)** The SLSF 6xr8 native trimers were the best targets for such ligands rather than monomers corresponding to any other mutated conformers, and **iii)** Tinosorb, a drug that despite being an star-shaped molecule fitting the inner site of the 3x3 helices of SLSF with the lowest binding-score to SLSF, did not inhibited S-dependent *in vitro* infection, nor showed virtual binding to S-SLSF. Nevertheless, among all these potential SLSF-ligands, Tinosorb-similar 3-fold symmetric structures and/or chemical derivatives smaller in size and more hydrophilic, may offer new opportunities to inhibit coronavirus fusion and infection, and/or, at least, offer new tools to further investigate the prefusion mechanism(s) of coronaviruses.

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). 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 were extracted from the 6xr8 conformer (closed all-down conformer) and used as models. 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 for leads

The BioSolveit seeSAR vs.10 package (https://www.biosolveit.de/SeeSAR/) was used in e7 (8 CPUs) and/or i9 (48 CPUs) desk computers, as previously described [23]. The seeSAR package employs the HYDE scoring function to evaluate HYdration / DEsolvation and to eliminate 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 the internal binding-pockets were used for trimers (Figure S1, yellow shadows). To compute the conformational poses and their corresponding binding-scores, each of the *.sdf files of <9000 ligands, took ~ 2 days of e7 computing. 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 for leads

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 first ffu energy minimized in Open Bable and converted to *.pdbqt files. Further simplifications included setting the SLSF target as rigid (constant covalent lengths and angles) and ligands as flexible (rotatable bonds) and retaining the poses with the lowest ΔG for each *.out.pdbqt for calculations and visualization. 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 °C)[30]. The predicted structures were visualized in PyRx and/or PyMOL.

Conformers top-leads

Top-leads were quantitatively defined by those leads with < 50 nM binding-scores which were also bound by >4 conformers (voting procedure). To compare 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 binding-scores were defined as positive by using thresholds that resulted in < 35 conformers for each target. The resulting thresholds were 10 µM (seeSAR) or 250 nM (AutoDockVina). Final data were represented as qualitative heat maps to facilitate their interpretation.

Drug-like properties, commercial availability and known drugs

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).

Fusion assay by infectivity of S-pseudotyped VSV

Tinosorb (Sigma) was suspended in ethanol-chloroform 1:1 which formed vesicles of different sizes visible at the optical microscope (Figure 9A). Green fluorescent protein (gfp)-expressing, glycoprotein G-deleted Vesicular Stomatitis Virus (VSV) (VSVgfp-ΔG) was pseudotyped with either a codon-optimized SARS-CoV2 spike Sc18 glycoprotein (VSVgfp-ΔG+Sc18) or with its own glycoprotein G as control (VSVgfp-ΔG+G). The Sc18 spike sequence lacking 18 cytosolic amino acids to enhance its incorporation efficiency to the VSV envelop and other conditions were as described before by others[32]. The pseudotyped VSV were titrated by VSV fluorescent forming units (ffu). Infection of ACE2-expressing 293T cells by VSVgfp-ΔG particles pseudotyped with SARS-CoV2 S glycoprotein treated with Tinosorb was used as an assay to measure membrane fusion activity. For this, VSVgfp-ΔG particles were first pre-incubated at 37 °C during 1 h in cell culture medium in the absence of serum with suspended Tinosorb or with anti-RBD rabbit neutralizing antibody (anti-S Ab) purified by antigen-affinity chromatography (GTX135709, Gene Tex). Then, ACE2-293T cell monolayers at 70 % confluence in 24-well plates were incubated with serially diluted VSV+Tinosorb or with anti-RBD Ab preincubated mixtures containing 400 ffu of VSV per well during 1 h in the absence of serum. After 1 day of further incubation, the number of ffu of 3-5 cells were counted with an inverted fluorescent microscope.

Docking of Tinosorb to the S 6xr8 trimer model

Tinosorb was docked to a 3D model of the native 6xr8 conformer [33] of the S trimer of which glycosilations were computationally removed. Docking to whole trimers used the i9 computer. The seeSAR program predicted 36 unique binding-pockets in the whole S trimer with an average of 17 amino acid residues per binding-pocket. The predicted binding-pockets appeared dispersed through the S trimer surface. One of the predicted binding-pockets included 40 amino acids around the central 3x3 α-helices of SLSF (S-SLSF). The seeSAR algorithm calculated the binding-scores for the best 10 poses in free competition among the 36 predicted binding-pockets.

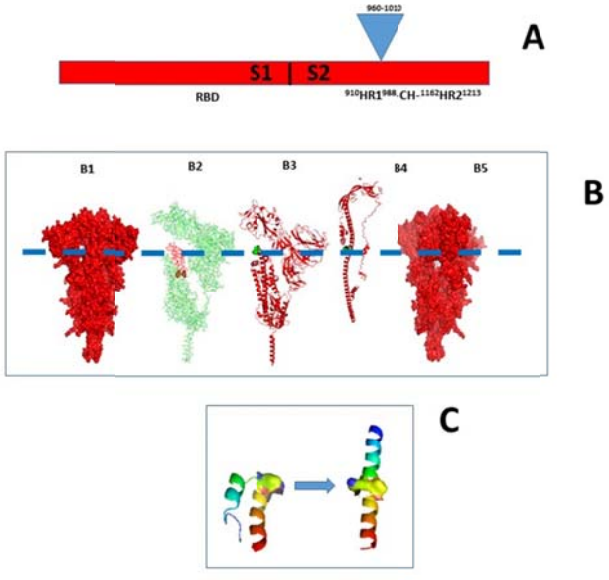

Figure 1
Location of the spring-loaded switch-folding (SLSF) of the S spike protein

A) Scheme of the spike subunits S1 and S2 in the S spike amino acid sequence. **Blue triangle**) Location of the spring-loaded switch-folding (SLSF) expanding residues 960-1010. **B)** Position of the surface-exposed SLSF in the side of the S trimer (S-SLSF). **B1)** The white spot visualizes the surface-exposed residues 983-985 (^{983}RLD) located at the tip of SLSF. **B2)** The SLSF monomer (labeled in red) in the S2 (bottom, labeled in green) becomes more accessible after cleavage of the S1 (top, labeled in green). The amino and carboxyl terminus of SLSF were labeled as red spheres. **B3)** The residues 986/987 (green spheres) mutated to P 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 (^{960}NTLVKQLSS) at the SLSF amino terminus. **Hatched blue horizontal line**, location of the SLSF tip ^{986}KV mutated to PP. **C)** Detail of the monomeric unfolding-folding transition of the 3 helices to 1 helix at the SLSF tip (amino-terminal in blue and carboxy-terminal in red). The mutated PP mapped to the middle of the partial SLSF represented in the scheme (yellow spheres).

Results

Definition of SLSF

To computationally study the hypothetical binding of ligands to the HR1-CH spring-loaded switch-folding (SLSF), the HR1-CH amino acid sequence surrounding the ...[986]KV... tip was explored to define a minimal target size. For that, the 3D structure of the 6xr8 conformer, corresponding to wild-type amino acid sequence in a prefusion all RBD-down conformation [33], was used. The 6xr8 conformer was chosen because it may best represent the native virion sequence and 3D structure (Table S1) and therefore may also be one of the best targets for preventive and therapeutic purposes.

The amino acid sequence size was expanded down from the 2 α-helices around the ...[986]KV... tip where 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 residues 960-1010 with 3 short α-helices contained most of the significant binding-scores by seeSAR and/or AutoDockVina (Figure 1, B2, B3, C and Figure 3A,B). Additional amino acid extensions along the amino and carboxyl-terminal α-helices did not discovered any shorter distances between the α-helices (hypothetically to be simultaneously bound by ligands) and/or did not predicted any other binding-pockets (data not shown). Therefore, residues 960-1010 were selected to define the SLSF sequence as the target for the present docking studies.

Visual inspection of the S trimer compact structure models predicted that the SLSF sequence inside the whole S trimer (S-SLSF) should be partially accessible through two side cavities visualized when rotating the S trimer compact structure model (Figure 1B, B1 and B5 white spots). The corresponding surface-exposed amino acid residues were ~ 983-985 near the tip ([983]RLD) and ~ 960-968 ([960]NTLVKQLSS) near the amino-terminal SLSF sequence. Additionally, an inner cavity extends from top to bottom of the trimer with a variable width of 7-20 Å of diameter including the inner part of the S-SLSF 3x3 α-helices (Figure S1 and 9C).

Lead definition by binding to SLSF 6xr8

High-throughput screening of 135831 natural compounds by docking to the isolated SLSF trimer and monomer models extracted from the 6xr8 native trimer conformer were independently performed by seeSAR and AutoDockVina.

SeeSAR screening to trimers identified 2948 leads with binding-scores < 200 nM, while that to monomers identified 3045 leads with scores < 10000 nM. The corresponding relative frequency profiles showed that a majority of leads were concentrated in the lowest scores corresponding to peaks at ~ 50 nM for trimers and ~2500 nM for monomers (Figure S2).

On the other hand, AutoDockVina screening to trimers identified 1864 leads with binding-scores < 50 nM while monomers identified 1168 leads with scores < 1350 nM (not shown).

Selection of conformers and top-leads

To facilitate conformer top-lead analysis, we try to reduced the number of SLSF possible conformers (Table S1).

3D computational superposition, showed that most of the SLSF in prefusion (open and close states), were similarly folded when compared to the 6xr8 conformer, as suggested by their RMSDs values < 0.6 Å. Similar 3D conformations were observed among several different SLSF structures [11, 34, 35], despite having neutralizing antibodies bound to the S trimers [2, 4-6], or after additional P mutations [11, 12]. In contrast, RMSD between 0.6 to 2.90 Å were obtained for recent dominant mutants [36, 37], after binding other antibodies [7], at lower pHs [38], or after S1-S2 cleavage [39]. Therefore, assuming that similar RMSD in the same amino acid sequences would not highly change their binding-scores, conformers with RMSD < 0.5 Å were excluded from further analysis. No superpositions could be predicted for the postfusion conformer [8, 33] (Table S1). Excluding the 6xr8, a total of 8 prefusion and 1 postfusion conformers among those described in Table S1 were selected as representative of the most different SLSF 3D structures. The prefusion conformers selected were, three from close all-down (6xlu, 6xm5, 6xey), and five from different open states (6vyb, 6xs6, 6zgh, 6zgg, 6xm4) (Table S1, yellow background and bold lettering). The wild-type [986]KV sequence was present only in the 6xr8 prefusion and 6xra postfusion, while the other conformers contained the [986]PP mutations. The rest of the amino acid sequences were identical for all conformers (not shown).

To first compare each of the corresponding conformer binding-profiles with those of the 6xr8, the lead compounds previously defined as described above for the trimer- and monomer-native 6xr8 conformer were docked by seeSAR (Figure S3, A,B,C,D) and by AutoDockVina (Figure S3, E,F,G,H) to the 9 selected SLSF-conformers both in their trimeric and monomeric versions.

Results showed the seeSAR trimer-leads, were more numerous and of lower binding-scores when docked to trimers (Figure S3,A) than to monomers (Figure S3, B). Similarly, the seeSAR monomer-leads were more numerous and of lower binding-scores with trimers (Figure S3,C) than with monomers (Figure S3, D). Profile variations among different conformers were wide in both trimers and monomers (Figure S3, AB and CD, respectively).

The AutoDockVina trimer-leads, were also more numerous and of lower binding-scores when docked to trimers (Figure S3,E) than to monomers (Figure S3, F). Similarly, the AutoDockVina monomer-leads were more numerous and of lower binding-scores with trimers (Figure S3,G) than with monomers (Figure S3, H). All these profile variations were smaller in trimers and monomers (Figure S3, EF and GH, respectively), in contrast to seeSAR's parallel data.

Assuming that the lower binding-scores may be the best to predict experimental binding, the corresponding conformer top-leads were further studied only for the dockings generating the lower binding-scores, in other words to the top-leads obtained from "seeSAR- trimers", "seeSAR-monomers" and "AutoDockVina-trimers" (Figure S3, ACE, respectively). The next sections will study in more detail their corresponding three kinds of top-leads.

Table 1
Conformer top-leads corresponding to seeSAR trimer-leads

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

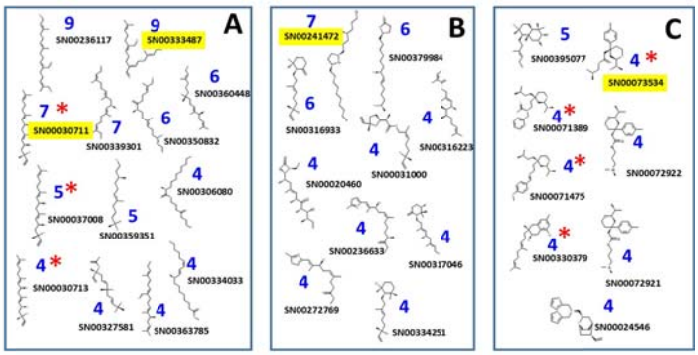

Figure 2
2D representation of conformer top-leads of Table 1
Clustered chemotypes of top-leads from Table 1 showing none (**A**), 1 (**B**) or 2-3 rings (**C**).
Blue numbers, number of trimer-conformers bound.
Yellow-background, Top-leads common to trimers (Table 1) and monomers (Table 2).
Red stars*,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

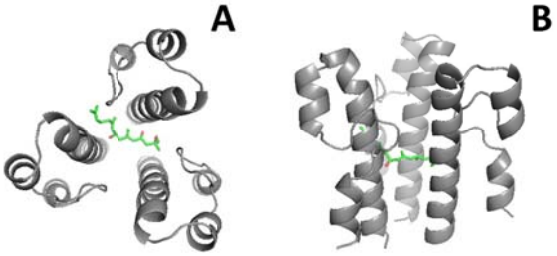

Figure 3
Representation of the best pose of SN00333487 complexed with SLSF 3x3 α-helices of 6xr8
Most of the top-leads and conformers (Figure 2, Table 1) showed similar best poses to those of SN00333487 and 6xr8 (A and B green lines), as drawn in PyMOL. **A,** SLSF top view. **B**, side view with the SLSF tip up and their amino and carboxyl ends down. **Gray drawings**, SLSF trimer amino acid ribbon structures. Note the 3 α-helices separated by 2 loops per monomer.

Conformer top-leads of seeSAR trimer-leads

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

There was an inverse correlation between the RMSD conformer values relative to 6xr8 and the number of top-leads bound. Additionally, the 6xra conformer with no predicted RMSD, predicted binding to only 3.2 % of top-leads (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 SLSF (Figure 3, side view).

Conformer top-leads of seeSAR monomer-leads

Top-leads selected among the seeSAR 3045 6xr8 monomer-leads were 21 (Table 1, dark green column). Four of these top-leads were also among the above described 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. Also more close- than open-conformers bound top-leads. The seeSAR monomer top-lead chemotypes were similar to the seeSAR trimer top-leads (Figure 4).

Most of the seeSAR monomer top-leads mapped 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 corresponding to seeSAR monomer-leads

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 < 50 nM, ranked by the total number of conformers docked with binding-scores < 10 μM. Other details as in Table 1.

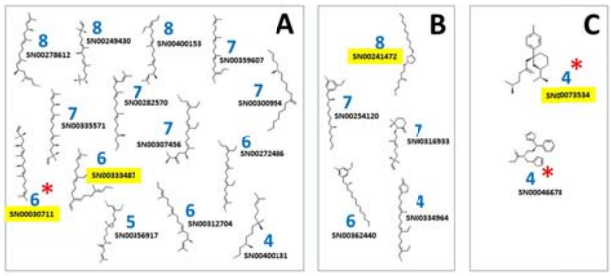

Figure 4
2D representation of conformer top-leads of Table 2.
Clustered chemotypes of top-leads of Table 2 showing none (**A**), 1 (**B**) or 2-3 rings (**C**). Other details as in Figure 2.

Conformer top-leads of AutoDockVina trimer-leads

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 SLSF (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 most of them have good characteristics for drug development since many 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 some of the most important detoxifying cytochromes P450 (CYP1A2 and CYP3A4), which may raise some physiological concerns if used for drug-like purposes.

Table 3
Top-leads corresponding to AutoDockVina trimer-leads

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 number of conformers docked with binding-scores < 250 nM. Other details as in Table 1.

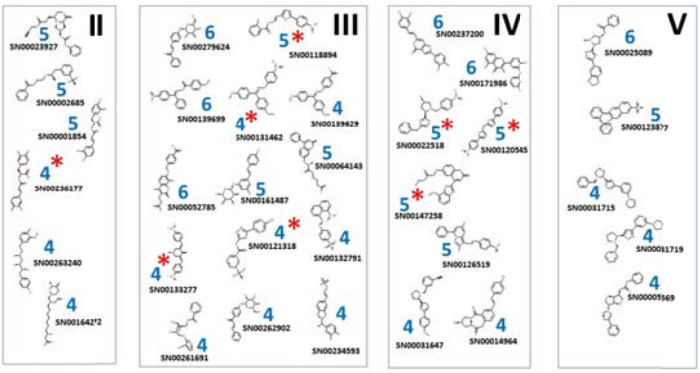

Figure 5
2D representation of conformer top-leads of Table 3
Clustered chemotypes of the top-leads of Table 3 showing 2 (**II**), 3(**III**), 4 (**IV**) or 5(**V**) rings. Other details as in Figure 2.

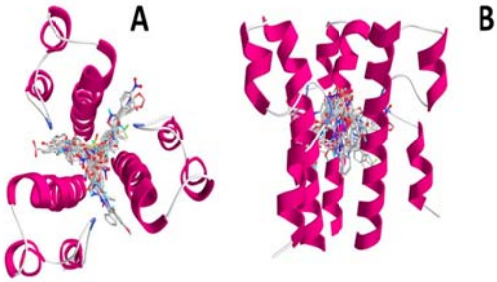

Figure 6
3D graphic representation of superposed best poses of the 6xr8 top-leads.

Superposed drawings of top-leads bound by >5 conformers (Figure 5, Table 3). Other details as in Figure 3.

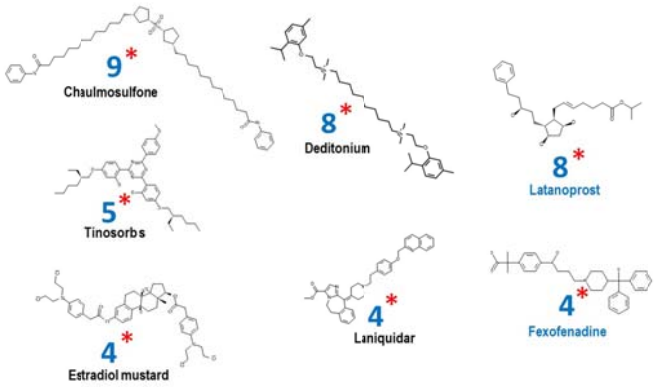

Figure 7
2D representation of drug top-leads of Table 4 .
Names in blue corresponded to FDA-approved drugs while the rest were from NCGC drugs. Other details as in Figure 2.

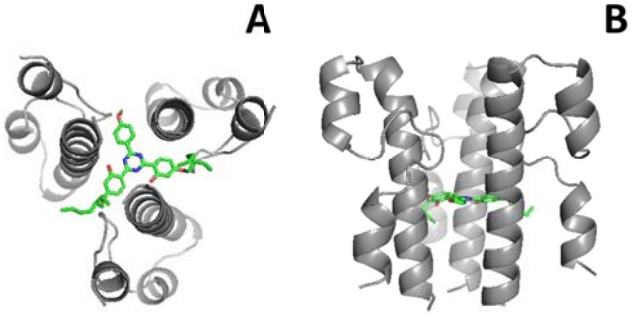

Figure 8
3D representation of the complex of the 6xr8 SLSF and the best pose of bound star-shaped Tinosorb.
Representative drawing of some top-leads and conformers (Figure 7, Table 4). Other details as in Figure 3.

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 the 6xr8 trimer native target. 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 SLSF trimer-conformers to predict their corresponding top-leads. As expected the profiles of their binding-scores were lower for the 6xr8 trimer conformer and similar to the other profiles obtained with natural products (Figure S4).

There were 7 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 which has a unique 3-fold symmetry forming a star-shaped molecule (branched compounds having several chains linked to a central core). Tinosorb displayed the lowest binding-score to the 6xr8 trimer at the low nM range. The unique 3-fold symmetry structure of Tinosorb, suggests their fitting to the inner part of the 3x3 α-helices of the S-SLSF trimers. To note that these drug chemotypes were of higher molecular weights (i.e., 627 Daltons of Tinosorb) than any of the natural compounds described above (<380 Daltons) and therefore showed lower binding-scores when compared to some of the natural compounds.

Visual inspection of the 3D models predicted for the best poses of drug top-leads complexed with SLSF trimers showed similar interactions with their amino acid neighbors. The amino acid neighbors in the 6xr8 trimer of Tinosorb-SLSF complexes, for example, mapped into the inner part inside the trimer, hypothetically interacting with the 3 monomers (Figure 8, A), in the middle of the SLSF (Figure 8, B) and confirming its location within the trimer as suggested above by its 3-fold symmetry. Other drug lead 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 corresponding to trimer-seeSAR leads

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	0.003									5
Estradiol mustard	0.2									4
Fexofenadine	14.7									4
Laniquidar	40.8									4

The Table shows the seeSAR 6xr8 leads <50 nM, ranked by the total number of conformers bound with < 10 μM binding-scores, 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 black**, NCGC drugs. Other details as in Table 1.

Fusion inhibition assay of Tinosorb

The Tinosorb binding-score to SLSF in the nM range, the existence of side, top and bottom axis windows on the native S trimer and the Tinosorb star-shaped structure fitting the inner part of the SLSF trimer α-helices (Figure 8), strongly suggested a possible biological activity. Therefore, Tinosorb was assayed in cells to investigate a possible interference with SARS-CoV2 S fusion activity. One of the first challenges for these assays was Tinosorb high hydrophobicity (logP ~ 10.4). Among the several solvents that were tested, Ethanol/Chloroform 1:1 or DMSO offered partial solubilities. Therefore, Tinosorb was suspended in these solvents and assayed at the nM-μM range at maximum concentrations to try to increase Tinosorb availability in the aqueous cell culture medium despite its low water solubility. Complete solubility was not obtained because vesicles, micelles or droplets of different sizes < 200 μm could be observed in sonicated 10 mM Tinosorb suspensions at the optical microscope (Figure 9A, up insert). To assay for the effect on S glycoprotein activity, S-pseudotyped VSV particles were preincubated with Tinosorb suspensions before using them to infect susceptible cell monolayers. Results showed that S-mediated infectivity was not significantly inhibited by Tinosorb at the range of concentrations explored with respect to controls made in the absence of Tinosorb. In contrast, neutralizing anti-RBD antibody completely inhibited infectivity (>90 %) in parallel assays (Figure 9A).

Virtual binding of Tinosorb to S-SLSF (SLSF inside the S trimer model) rather than to SLSF was then estimated (Figure 9B, shows all possible binding pockets predicted by seeSAR). However, none of the 10 poses of Tinosorb predicted any binding-scores < 500 μM (not shown). Furthermore, visual inspection of the resulting Tinosorb-S complexes located the best binding poses outside the inner SLSF α-helices (Figure 9C, **green spheres**). These results suggest that in the fusion assay, Tinosorb did not even reached the inner S-SLSF, even if it could be 100 % solubilized in water. On the other hand, these results suggest a new model to computationally search for alternative ligands targeting S-SLSF in competition with the 36 binding-pocket possibilities at the S trimer molecule. Although more costly in terms of computation, this model may best mimic the whole virus situation and therefore help to make more accurate ligand predictions.

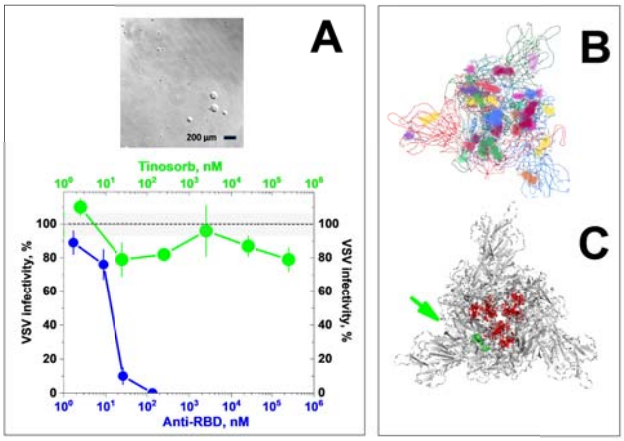

Figure 9
Tinosorb assays of inhibition of fusion (A) and binding to the S 6xr8 trimer model (B,C)

A) Top insert, microscopic aspect of Tinosorb sonicated in ethanol:chloroform (1:1). **Bottom graph,** serial dilutions of Tinosorb suspensions pre-incubated for 1 h with VSVgfp-ΔG+Sc18 and then added to 293T-ACE2 cell monolayers and incubated for 1 day before counting ffu. **Hatched horizontal line**, VSVgfp-ΔG+G and VSVgfp-ΔG+Sc18 infectivity range (93-107 %, shadowed background). **Green circles,** VSVgfp-ΔG+Sc18 + Tinosorb **Blue circles,** VSVgfp-ΔG+Sc18 + anti-RBD antibody. The percentages of VSV infectivity were calculated by the formula, ffu +Tinosorb or +anti-RBD antibody / ffu without any additives. Mean and standard deviations from triplicates are represented.

B) Top view of the 36 binding-pockets represented in different colors, averaging 17 amino acids each as predicted by seeSAR in the S 6xr8 trimer model. The central **blue** binding-pocket includes 40 amino acids around the inner 3x3 α-helices of S-SLSF.

C) Top view of Tinosorb best conformation (pose) bound to the S 6xr8 trimer (binding-score > 500 µM, n=10). **Red lines,** SLSF 3x3 helices. **Green spheres,** bound Tinosorb. **Gray amino acids**, rest of the S amino acids in the trimer.

Discussion

After exploring 9 trimer and monomer prefusion conformations of SLSF (S residues 960-1010 per monomer) of SARS-CoV2 by computational ligand screening, top-leads from hundreds of thousands of natural compounds with <380 Dalton and < 6 logP were identified with binding-scores in the nM range to a high number of conformers. Surprisingly, most leads were targeted to SLSF amino acid residues other than those located at their tips, where the most used PP stabilizing mutations affecting fusion and infectivity had been previously described.

The double screening (seeSAR and AutoDockVina) approach used here consisted in a combination of a first high-throughput screening of a native trimer and monomer SLSF conformer (6xr8) to define leads, followed by docking to several selected SLSF conformers obtained from PP mutant models to define top-leads. After estimating that the variations in conformation which may had been introduced by using PP mutants did not cause excessive conformational changes at SLSF (low RMSD values), top-leads were defined by taking into account the number of conformers each of them bound within a threshold binding-score (vote approach). Rather than relying solely in lead binding-scores, these additional criteria chose because it was expected to increase their chances to predict experimental success may not be accurate. 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 [40-42]. 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 a few 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 of these programs use but also the molecular diversity of possible solutions to find compounds which may bind to SLSF.

One of the challenges to predict experimental success in the present case, is how accessible is the SLSF target sequence in the native infectious viral particle (S-SLSF) compared to our isolated models (SLSF). The partial accessibility of S-SLSF predicted by modeling the native closed all-down S trimers, suggests that S-SLSF may be reached even when inside the highly compacted prefusion S trimeric conformations. For instance, we could detect both side and top-to-bottom axis accessible cavities of 7-20 Å as estimated in ribbon/sphere representations in PyMol whole S trimer models. Those cavities although small will be enough to allow for surface exposure or penetration of the smaller molecules of some of the top-leads predicted here with the lowest binding-scores. Furthermore, the presence of partially open S prefusion structures (1, 2 or 3 RBD-up states) described by several authors, suggests that the S1-S2 interactions may be under spontaneous continuous changes, thus theoretically increasing the accessibility to S-SLSF. Therefore, possibilities may be high for low binding-score small ligands to get access and to bind S-SLSF at some of the prefusion states. A complete computational analysis of all these possibilities must be performed before we can reach firm predictions. In this respect, experimental difficulties may also arise by top-leads in the low nM binding-score range that while required to lock the spring-loaded mechanism, may also recognize similar combinations of amino acid targets in other proteins, generating unexpected undesirable side-effects.

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

The conformer-dependent wide variation in binding-scores, despite having the same amino acid sequences except the PP mutations 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 their 3D structures. Among those sequence variations, the D614G mutation which became dominant as the pandemic proceeded, has been implicated in increasing the spread of the virus by favoring S up-conformations and by changing tissue tropisms [36, 37]. 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, at least to SLSF.

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, a star-shaped compound [43] actually been used to prevent ultraviolet light damage in the skin, was a candidate to stabilize prefusion conformations. However, because of its high hydrophobicity, it may be only suitable to disinfect surfaces rather than for therapeutic purposes. Furthermore, despite Tinosorb's low nM binding-score, fitting the inside of SLSF trimeric 3x3 α-helices and S-SLSF possible accessibility through side-top-bottom cavities in whole S trimers, we failed to demonstrate inhibition of viral-cellular fusion by pre-incubating the coronavirus with Tinosorb, at least within the variables used for our assay. Furthermore, we could not demonstrate any virtual low binding-score of Tinosorb to S-SLSF. It is unlikely that the inhibition failure could be due to different S conformers for modeling and pseudotyping, because of the small sequence variations of SLSF (Table S1). Although, small sequence variations outside the S-SLSF could sequester Tinosorb, those are also unlikely because of its high binding-scores at the >500 µM range obtained to S compared to the nM ranges when bound to SLSF. Due to Tinosorb's high hydrophobicity, aggregation may still occur after adding any solvent-solubilized mixture into the aqueous cell culture media contributing to reduce Tinosorb free concentration and explaining its failure to inhibit fusion. Although hydrophilic cyclodextrins could had been added as carriers to reduce the hydrophobicity of Tinosorb, the resulting increase in molecular size of such cyclodextrin-Tinosorb complexes would contribute to more difficulties to reach the inner S-SLSF. Perhaps the most likely explanations of the failures of Tinosorb to inhibit fusion and to predict binding to S-SLSF may be due to its hydrophobicity and/or steric hindrance, respectively. Although the many glycans surrounding the S trimer in the 3D models should not interfere in the penetration to the internal S-SLSF (none of them appear to be on the way in the models), the ~20 Å small molecular size of the trimer cavities to S-SLSF may be still problematic for large molecules like Tinosorb that consist of a ~17 Å side triangle. Therefore, the most probable alternatives for Tinosorb-similar-star-shaped molecules to reach S-SLSF, may be among smaller and more hydrophilic derivatives. As an alternative, absence of activity might be explained if the proposed binding pocket had become inaccessible after the S homotrimer was synthesized and assembled by other reasons. In this hypothesis, addition of the drug during infection and biosynthesis of the S glycoprotein could provide a plausible approach that may deserve further experimentation.

Among many other possibilities, the predicted leads/top-leads and not-binding ligands identified here may be used to develop training sets for deep-learning approaches. In the present context, deep-learning may serve to screen larger libraries of millions of compounds based solely on their chemical properties rather than on computer-intensive docking. Other lead alternatives may be identified also by deconstructing the Tinosorb molecular characteristics. For instance, by identification of the minimal size molecule required for the lower binding-scores. More chemotype possibilities may still lay ahead on alternative ligands with lower molecular sizes and binding-scores, with more probabilities to inhibit coronavirus fusion.

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	33
6xra	**none**	**helix+helix**	postfusion	
6vxx	0.37	RBD 3down	prefusion	34
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	35
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	38
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	36
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	39
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 isolated sequences were obtained by extracting them from 3D S spike trimeric protein models of SARS-CoV2 (SARS, **S**evere **A**cute **R**espiratory **S**yndrome) downloaded from the **R**esearch **C**ollaboratory for **S**tructural **B**ioinformatics (RCSB) **P**rotein **D**ata **B**ank. Conformer amino acid sequences 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 residues 383 and 985. **x2**, stabilizer disulphide link between residues 413 and 987.
Bold & yellow, conformers selected for docking studies.
Red PDB, closed conformer representatives.
Blue PDB, open conformer representatives.
Gray PDB, postfusion conformer representative.

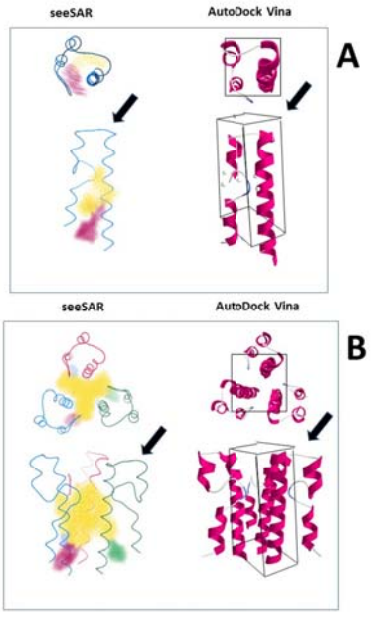

Figure S1
Scheme of the monomer (A) and trimer (B) SLSF 3D models for docking by seeSAR and AutoDockVina

seeSAR predicted a number of binding-pockets around the 3x3 α-helices (colored shadows at the left).
AutoDockVina works on submitted grids around the inner parts of the 3x3 α-helices (rectangles at the right).
A, monomers
B, trimers
Top A and B, top views.
Bottom A and B, side views.
Black arrows, locations of the SLSF tips where the prefusion-stabilizing PP mutations were introduced by other authors.

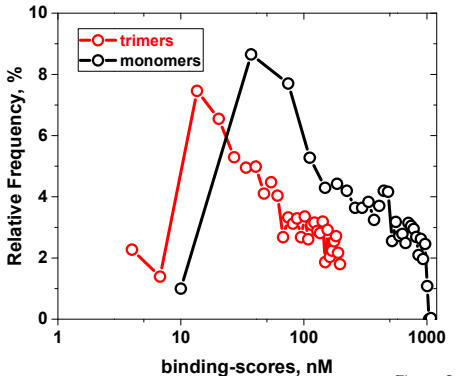

Figure S2
Comparison of relative frequencies of seeSAR lead binding-scores < 1000 nM from the 6xr8 trimer and monomer.

The binding screenings of natural compounds to the SLSF 6xr8 conformer were carried out with seeSAR. The distribution of lead frequencies were calculated and represented in Origin.

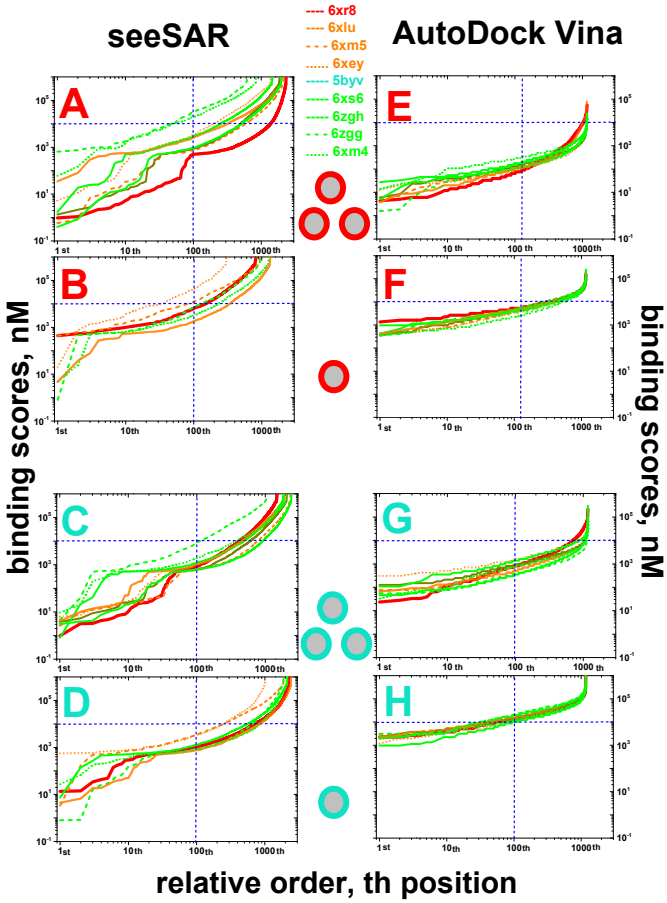

Figure S3
Binding of trimer and monomer 6xr8 leads to conformers.
Each set of nM binding-scores were ordered from lower to higher and the first 1000 represented. The 6xr8 trimer (**red A,B,E,F**) and monomer (**green C.D,G,H**) leads were docked to other conformer trimers (3 central circles) and monomers (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 properties of top-leads predicted by SwissADME

Top-leads of 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 of 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 of 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.

TPSA, estimates of the amount of topological polar molecular surface area, lowest values facilitate permeation of cell membranes (best to be <90 Å²).

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.

Sol, solubilities in water classified in general classes

GIA, prediction of gastro-intestinal adsorption.

1A2, 3A4, inhibition of the main detoxyfying cytochromes P450.

PAINS, 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.

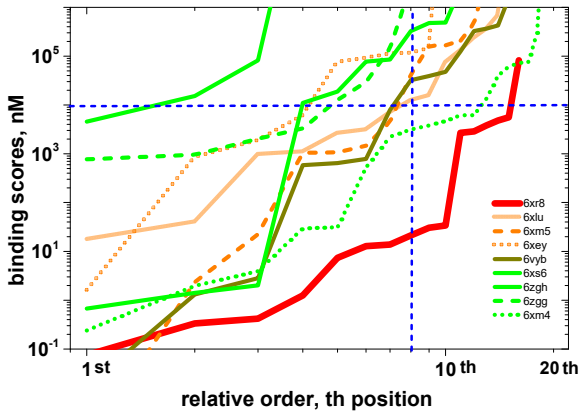

Figure S4
Binding-profiles of drug leads to conformers
Each set of nM binding-scores per conformer were ordered from lower to higher and the first 27th represented. Other details as in Figure S3.

Funding

The *in vitro* experimentation was supported by grant COV20-0901 from Instituto de Salud Carlos III. The computational work was carried out without any external financial contribution.

Competing interests

The authors declare no competing interests

Authors' contributions

MS and RB designed and performed the fusion assays. 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, 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.v2: https://doi.org/10.26434/chemrxiv.12952115.v2.

[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. 10.1007/s10822-017-0063-0 [pii] http://dx.doi.org/10.1007/s10822-017-0063-0.

[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]Johnson, M.C., T.D. Lyddon, R. Suarez, B. Salcedo, M. LePique, M. Graham, . . . D.G. Ritter. Optimized Pseudotyping Conditions for the SARS-COV-2 Spike Glycoprotein. *J Virol.* 2020, 94: JVI.01062-20 [pii], http://dx.doi.org/10.1128/JVI.01062-20.

[33]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.

[34]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.

[35]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.

[36]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.

[37]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].

[38]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.

[39]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].

[40]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.

[41]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.

[42]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.

[43]Diab, H.M., A.M. Abdelmoniem, M.R. Shaaban, I.A. Abdelhamid and A.H.M. Elwahy. An overview on synthetic strategies for the construction of star-shaped molecules. *Royal Society Chemistry Advances.* 2019, 9: 16606-16682. http://dx.doi.org/10.1039/c9ra02749a.