

Star-shaped Triazine-derivatives: would they crossbind SARS-CoV-2 spike helices?

short title: coronavirus stabilization by Triazine-derivatives

Coll, J*.

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

* Corresponding author

Email: juliocollm@gmail.com (JC)

Julio Coll, orcid: 0000-0001-8496-3493

Abstract

This work describes synthesizable water-soluble Triazine-derivatives computationally crossbinding the S spike helices of Severe Acute Respiratory Syndrome coronavirus (SARS)-CoV-2. The "spring-loaded switch-folding" (S-SLSF) α -helices included in the S homotrimer top-to-bottom cavity and implicated in viral-host membrane fusion were targeted by star-shaped Trihydroxyl-Triphenyl-Triazines (TTT) leads at subnanomolar binding-scores. Exploration of *in silico* leads among millions of molecular candidates, included several similar searches, core-replacement, fragment extensions, or convolutional neural network deep-screening combined with hundreds of water-soluble lead-derivatives identified by manual iterations and commercially available building-blocks for chemical synthesis. The lead-derivatives are briefly discussed for *in vitro* validation and possibilities of fusion inhibition substituting mutations.

Keywords: S spike; synthesizable, water-soluble, star-shaped molecules; Triazine core; prefusion; coronavirus; deep-learning; SARS CoV2; spring-loaded switch-folding

Introduction

The "spring-loaded switch-folding" (SLSF) of wild-type S spikes of SARS-CoV-2 was previously defined as one of the *in silico* predicted targets for fusion inhibition with drugs instead of PP mutations¹. The optimal SLSF-residues targeted for binding were defined from the 960 to the 1010 amino acid sequences located in the upper part of the S2 homotrimer central top-to-bottom cavity of 7-20 Å of diameter². In the wild-type prefusion state, the SLSF-residues and the S-SLSF (SLSF residues at the S trimer) included 3x3 α -helices (Figure 1).

Computational screening of hundred thousands of natural compounds predicted Tinosorb as the top subnanomolar binding lead to SLSF-residues¹. The Tinosorb's preferential binding to SLSF-residues, 3-arms-star-shaped molecular structure³ and crossbinding the inner part of the SLSF-residue α -helices, together with the existence of side and central top-to-bottom surface-accessible cavities to SLSF-residues at the S wild-type prefusion state, suggested that Tinosorb may experimentally crossbind S-SLSF *in vitro* and show binding-dependent biological activities. Tinosorb's crossbinding of the inner space of the S-SLSF 3x3 α -helices could stabilize prefusion states blocking fusion and infection, similarly to what has been previously described for PP mutations of the 986KV residues⁴⁻⁶. However, computational studies could not predict Tinosorb's S-SLSF binding, nor *in vitro* fusion by S-pseudotyped VSV-infectivity could be inhibited by Tinosorb¹. Tinosorb's low solubility in aqueous media together with its relatively large molecular size compared to the narrow surface-accessibility to the S-SLSF binding-pocket, may have contributed to those failures. This work computationally explores Tinosorb-like star-shaped-like molecules among more water-soluble alternatives to best apply possible validation assays. Because whole S trimers may be best targeted to stabilize virions at their prefusion states and/or to interfere with the earliest prefusion steps, the S-SLSF wild-type trimer model (6x8 ID in the RCSB protein data bank) was preferentially selected for this work.

The SLSF-residues contained part of the HR1 heptad-repeat (910 to 988) and part of the CH central helix (986-1033) of the S2 subunit of the S spike of SARS-CoV-2. Each SLSF-residue monomer contains two amino-terminal aligned small α -helices, one elbow-like disorderly bended loop (975-987) and one carboxy-terminal large α -helix. All those 3x3 α -helices are implicated in one spring-loaded mechanism (Figure 1, Bottom). During viral-cellular membrane fusion, the spring-loaded mechanism unfolds/folds by converting the elbow-like disordered loops to α -helices elongating the S-SLSF. After elongation, one coiled-coil bundle involving the newly formed HR1-CH and previous S HR2 α -helices, originates the much larger linear rigid conformation typical of active fusion and postfusion states. Similar spring-loaded mechanisms are common to many other enveloped viruses⁷.⁸. The mutations to prolines (PP mutations) at the tip of the S-SLSF 3x3 α -helices mentioned above, stabilize the loops at their earliest prefusion state blocking viral-host membrane fusion and inhibiting viral infectivity⁴⁻⁶.

Previous and abundant experimental and computational search for anti-coronavirus molecules has been focused on approved drugs (drug repurposing)⁹,¹⁰ to protein targets on either cell hosts or coronavirus particles. For instance, recent experimental screening for *in vitro* inhibitors of coronavirus-infection, identified 90 compounds with EC₅₀ < 96000 nM¹¹ among ~ 12000 drugs from the Repurposing, Focused Rescue and Accelerated MedChem (ReFRAME) bank (<https://reframedb.org>). New enhancers of inhibitors of RNA-dependent RNA

polymerase (RdRp) such as remdesivir (EC₅₀ of 123 nM) or amilmod (11 nM), were experimentally identified. Other coronavirus computational work on their protein targets have been focused to the S2 HR1-HR2 coiled-coil bundle^{12, 13, 14, 15, 16}, the surface interphase of S1 with the ACE2 human receptor¹⁷, and/or the active sites of viral proteases^{10, 18, 19}. Although targeting the coronavirus HR1-HR2 coiled-coils bundle with synthetic peptides reported nanomolar inhibitors of *in vitro* infectivity^{12, 13, 14, 15, 16}, search for more potent ligands among Tinosorb-like star-shaped molecules targeting S-SLSF may be justified by both computational and experimental methods.

Efforts to predict crossbinding possibilities of S-SLSF with ligands rather than with mutations have been explored here by combining computational strategies previously developed by others²⁰. These included searches among similars, core replacement, fragment extensions²¹⁻²⁴, convolutional neural networks (CNN)²⁵, *de novo* manual generation of compounds with drug-like properties²⁶⁻²⁹, filtering for synthetic feasibility, ligand presence in catalogs^{22, 30, 31}, purchasable building-blocks and available chemical synthesis paths.

SLSF-docked leads predicted subnanomolar binding ranges with common Triazine-cores branched by Trihydroxyl-Triphenyls (TTT). Deconstruction of TTT and S-SLSF docking predicted a few more core alternatives (TTX). Further explorations by small fragment extension (F+TTX) and combination of different attachment positions, generated new leads that maintained low their binding-scores. Newly developed CNN models using 2D molecular images²⁵ allowed the screening of larger compound libraries including those for maximal diversity among purchasable chemicals²⁹, and the latest ChEMBL chemical data collection.

Taken together, the results predicted ~ 50 star-shaped leads with lower S-SLSF binding-scores and improved drug-like characteristics compared to the initial Tinosorb molecule. Additional Triazine-derivative leads were then obtained that were water-soluble and can be synthesized from available building-blocks to increase their possibilities for successful *in vitro* validation. Whether some of the Triazine-derivative leads would experimentally crosslink the homotrimer cavity α -helices and inhibit coronavirus fusion, remains to be demonstrated.

Materials and Methods

Molecular characteristics and properties of Tinosorb

Tinosorb (PubMed ID 135487856, ChEMBL 2104956, CAS 187393-00-6), chemically known as bemotrizinol (2,2'-[6-(4-methoxyphenyl)-1,3,5-Triazine-2,4-diyl]bis(5-[(2-ethylhexyloxy)phenol])), has a star-shaped structure with a molecular weight of 627.8 Dalton (Figure S1). Tinosorb adsorbs ultraviolet UV-AB from 280 to 400 nm preventing tissue oxidation by free radicals induced by sun exposure. Due to its low water solubility (0.00033 mg/ml, logP of 10.4), Tinosorb is used in oiled topical creams. Tinosorb's LD₅₀ oral/dermal toxicities are >2g/Kg.

Libraries of ligands used in this work

Tinosorb-similar 1746 molecules were downloaded from PubChem (<https://pubchem.ncbi.nlm.nih.gov/#query=smiles=similarity>). PubChem Trihydroxyl-Triphenyl-Triazine (135616181 ID), and Triazine (9262)-similar searches resulted in 599 and 279689 molecules, respectively. Triazine-similars were downloded to 4346 drug-like molecules <700 Dalton and < 6 logP.

Core-replacement was performed on each of the seeSAR's fragment libraries (pdb and zinc data bases, containing ~20 million of fragments each). Fragment extension screened the seeSAR's library of 100 small fragments enriched with home-designed 10 fragments of 5-7 non-hydrogen atoms.

The SuperNatural II SNII library (http://bioinf-applied.charite.de/supernatural_new/index.php) was splitted in nineteen *.sdf files each containing different molecular weights < 380 Daltons, as described before^{1,32}. Those sdf files were randomly sampled for 10 high-binding-score negative ligands per file to contribute to the design of a training-set to develop 2D deep-learning models. A 0.5 million library of maximized chemotype diversity among the purchasable space²⁹ and the ~2.5 million ChEMBL28 latest library, filtered to drug-like compounds between 250-500 Daltons to ~1.5 million ligands (http://ftp.ebi.ac.uk/pub/databases/chembl/ChEMBLdb/releases/chembl_28/), were screened by the developed DEEPscreen²⁵ CNN T13 model.

The inSolve extremely large libraries provided by BioSolveit (CHEMriya_11bn_2021-05 1.1x10¹⁰ compounds, GalaXi_2.1bn_2020-11 2.1x10⁹, KnowledgeSpace_290tr_2019-05 2.9x10¹⁴, and REALSpace_19bn_2021-04 1.9x10¹⁰) were screened for 10000 compounds with 0.9-1.0 target TTT similarities and 80-90% minimum thresholds.

Commercially-available compounds were searched through the ZINC data base by supplying their smiles (<http://zinc15.docking.org/>) and/or searched in chemical catalogs (Sigma, BLDpharm). Commercially-available building-blocks for chemical synthesis were searched among the data bases developed by Enamine (<https://www.enaminestore.com/search>) and SigmaAldrich (<https://www.aldrichmarketselect.com/>). The duplicates of compounds retrieved from several sources were eliminated by OpenBabel gui 3.1.1. (<https://sourceforge.net/projects/openbabel/postdownload>).

3D wild-type 6xr8 S trimeric spike model

The whole SARS-CoV-2 S trimer 6xr8 vs 4 molecule model (Research Collaboratory for Structural Bioinformatics, RCSB, Protein Data Bank PDB ID) of the wild-type closed, all-down S conformer was either PyMol extracted to isolate the SLSF-residues 960 to 1100, as previously reported¹ (Figure 1) or used for S-SLSF docking (SLSF in the complete trimer structure) using i9 computers with 48 CPUs after eliminating their carbohydrate elements in PyMol (Figure 3).

3D-docking by two algorithms

Two complementary algorithms (AutoDockVina and seeSAR), were employed for tridimensional 3D-docking. The programs differed in both, i) the generation of ligand pose conformations with probable geometric binding to the target binding-pocket (set at 10 poses per ligand) and ii) the quantification of each conformational pose by binding-score calculations. The target S binding-pocket at SLSF-residues or S-SLSF, maintained their covalent lengths/angles while the ligands were flexible, so their rotatable bonds were used by the programs to generate the different pose conformations.

The AutoDockVina included in the PyRx 0.9.8. package³³ (<https://pyrx.sourceforge.io/>) used multithreading on multi-core e7 / i9 computers to speed up docking³⁴. AutoDockVina generates poses with Lamarckian genetic algorithms³⁵. The corresponding conformation-dependent Gibbs binding-score free-energies (ΔG) for each pose are calculated using semi-empirical data³⁶. To perform the docking, the *.sdf files required ffu energy minimization, hydrogen and charges additions to generate *.pdbqt files (PyRx-OpenBabel). Grids including only the SLSF-residue inner 3x3 α -helices were used. Only one pose with the lowest binding-score per docked ligand (*.out.pdbqt) was retained for analysis. To compare with the seeSAR binding-score values, the output ΔG energies in kcal/mol were converted to constant inhibition (Ki) in nM concentrations¹.

To predict possible seeSAR poses, the vs.10 package (<https://www.biosolveit.de/SeeSAR/>) uses the FLEX incremental fragment construction method based on software developed for computer vision and pattern recognition³⁷. The corresponding conformation-dependent binding-score Gibbs free-energies (ΔG) for each pose are calculated by the HYDE scoring function computing both HYdration and DEsolvation values (as calibrated from octanol/water partition data, logP)^{38,39}. To reduce false positives, the HYDE calculations include not only favorable but also unfavorable interactions⁴⁰. To perform the docking, the unique binding-pocket internal to the S-SLSF α -helices or all the 36 binding-pockets (average of 17 amino acids per pocket) predicted by the seeSAR in the whole S trimer (Figure 3,C), were selected. Only one pose with the lowest binding-score per docked ligand was selected for analysis. The binding-scores were expressed as nM means calculated from the HYDE lower-higher estimates (± 100 -fold range).

To facilitate a preliminary interpretation of the abundant docked data, the profiles ordered by binding-scores were first analysed graphically using the Origin program (OriginPro 2015, 64 bit Sr1 b9.2.257, Northampton, MA, USA) and best poses extracted by home-made macros in MS Excel. Binding-score estimations differed <10 % when repeated in different runs (n=3-6). The predicted lead-protein complexes were visualized in PyRx, seeSAR, and/or PyMOL (<https://www.pymol.org/>).

Core-replacement

The Triazine core or each of the Trihydroxyl-Triphenyl fragments attached to Trihydroxyl-Triphenyl-Triazine (TTT) were selected (Figure S2, AB) for core-replacements. To carry out core-replacements, the seeSAR inspirator module was used to screen pdb and zinc seeSAR-provided fragment libraries of ~20 million fragments each. Each iteration of core-replacement selects for the best-similar 10 cores maintaining the rest of the molecule intact while docking. Up to 90 best-fitting new cores were retrieved from each library and their corresponding 3D-binding-scores to SLSF-residues estimated by seeSAR docking. The resulting SLSF-residue leads were then docked to S-SLSF.

Fragment extension

The seeSAR fragment extension feature was applied to identify possible fragments attached to TTT phenyls (Figure S2, C). The sdf file containing 100 small molecular weight fragments provided by seeSAR, enriched with 10 home-made fragments between 1 to 7 non-hydrogen atoms, designed in MarvinSketch 17.1.30.0 (Chemaxon, Oracle Co) was used for this purpose.

To explore each of the C1-C5 carbons, cores were fixed while F+TTT compounds having different fragment extensions at each of the phenyls (RED, GREEN, BLUE) were generated and SLSF docked.

To explore C1 bound fragments at combinations among the 3 positions, additional F+TTT compounds containing 1, 2 or 3 best-binding fragments were manually generated in MolSoft (Molbrowser vs3.9-1bWin64bit) and SLSF-residues docked. The resulting leads were then docked to S-SLSF.

Preparation of training sets for machine learning

Training sets were generated by pooling negative (high binding-scores, 0s) and positive (low binding-scores, 1s) S-SLSF ligands. Our previous library of thousands of 3D-docked compounds separated in 19 molecular weight files¹ were randomly sampled for 10 binding-score 0s >100 nM from each separated file to generate a representative pool of negative ligands for the training sets (under sampling the majoritarian class)⁴¹. Other 30 TTT-similars with binding-scores 0s > 100 nM were added to the negative pool. On the other hand, 48 F+TTT ligands predicting binding-scores 1s < 0.2 nM were included as positives. The final training-set, was obtained in a unique sdf file by pooling the randomized 0s and the 1s in a 4:1 ratio. To provide a common identifier with the target libraries, the *inchikeys* descriptors were calculated and added to the ligands of the training-set as the unique molecular_names, using the DataWarrior program (Osiris, vs 5.5.0. Idorsia Pharmaceuticals Ltd). Possible duplicates were eliminated by OpenBabel. To develop learning models, the final training-set was randomly splitted in 60 % ligands for training, 20 % for validation and 20 % for test.

Learning models using chemical fingerprints as inputs

All chemical fingerprint types were obtained from the PADEL descriptor tool (<http://padel.nus.edu.sg/software/padeldescriptor>) and tested. The resulting PADEL files were splitted into molecular descriptors and algorithm-specific fingerprint bins for comparative tests. Regression prediction models were developed such as AdaBoost, Tree, NeuralNetwork, and others provided by the Orange vs3.27.1 package (<http://www.aillab.si/orange>). Their accuracies, however, were low.

Deep-learning models using convolutional neural networks (CNN) using 2D-molecular images

The DEEPscreen software²⁵, using CNN and molecular 2D-images rather than chemical fingerprints for higher prediction success, was optimized by adjusting their parameters with different training-sets using python and pytorch home-made scripts. To identify new leads on large data bases, the strategy was applied in three steps: i) 3D-docking to obtain enough 0s and 1s ligands to train CNN models, ii) 2D-deep-learning applying the CNN T13 model to downsize large libraries to a lower number of 3D docking-candidates, and iii) 3D-docking of the identified candidates for *in silico* S-SLSF binding validation. The accuracy of the T13 CNN model was 97.4 % on the training-set ligands.

Estimation of water-solubility of lead-derivatives

An estimation of pH-dependent water-solubility was obtained by drawing possible lead-derivatives and studying the resulting alterations of their pH-dependent logS profiles on the ChemAxon web-server by manual trial-and-error approximations (<https://disco.chemaxon.com/calculators/demo/plugins/solubility/>). To increase the accuracy of water-solubility values, consensus logS for selected candidates were then calculated by averaging the prediction values obtained from molSoft Browser vs. 3.9-1b/Win 64bit (MolLogS), Osiris DataWarrior vs5.5.0 (cLogS) and Virtual Computational Chemistry Laboratory (<http://146.107.217.178/web/aloggs/> from vcclab.org/lab/aloggs, aLogS). The estimated consensus LogS water-solubility values in mol/l were then transformed to mg/ml for each new compound and their mean and standard deviations calculated (n=3).

Results

Docking of Tinosorb-similars to SLSF-residues

Tinosorb-similars were docked to SLSF-residues by AutoDockVina and seeSAR. Results predicted 10²-10⁵-fold lower binding-scores (higher bindings) for seeSAR compared to AutoDockVina (Figure 1 top, **red hexagons** versus **blue hexagons**).

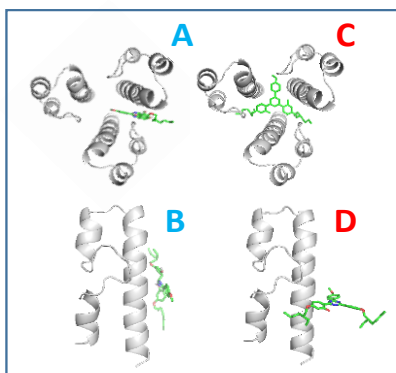
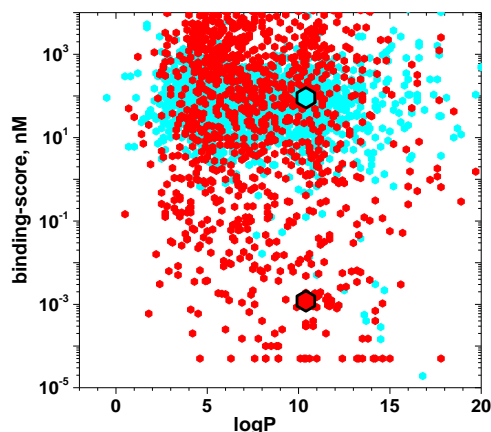


Figure 1

SLSF-residues binding-scores of Tinosorb-similars and their logP (UP) and drawings of complexes of SLSF-residues+Tinosorb by AutoDockVina (A,B) and seeSAR (C,D) (DOWN)

UP) **Blue hexagons**, Tinosorb-similars docked by AutoDockVina.

Black-edged blue hexagon, Tinosorb docked by AutoDockVina

Red hexagons, Tinosorb-similars docked by seeSAR.

Black-edged red hexagon, Tinosorb docked by seeSAR.

DOWN) **Green**, best Tinosorb docked pose.

Gray, SLSF 960-1010 residue ribbons

A, Top-view of trimeric SLSF-residues 3x3 α -helices complexed to Tinosorb by AutoDockVina.

B, Side-view of one SLSF-residues complexed to Tinosorb by AutoDockVina. Two monomers removed for clarity.

C, Top-view of trimeric SLSF-residues 3x3 α -helices complexed to Tinosorb by seeSAR.

D, Side-view of one SLSF-residues complexed to Tinosorb by seeSAR. Two monomers removed for clarity.

Table 1

PubChem ID	Binding-score, nM	logP	MW	Smiles	Trihydroxyl-Triphenyl – Triazine (TTT)
136025237	0.00005	4.4	621.7	<chem>CCC(CO-C1</chem>	
155024065	0.00005	6.2	621.7	<chem>CCCCCOC1=CC(=C(C=C1</chem>	
135783913	0.00005	7.9	609.7	<chem>C1CCCC(C1</chem>	
135611720	0.00005	8.2	615.8	<chem>CCC(C)CO-C1</chem>	
136044044	0.00005	8.3	623.7	<chem>C1CCCC(C1</chem>	
135740105	0.00005	8.7	747.9	<chem>CCCCCOC(CO-C1</chem>	
136383973	0.00005	8.7	773.9	<chem>CC(C)CCCCCOC(=O)C(C)O-C1</chem>	
136058049	0.00005	9.0	773.9	<chem>CCCCCCCCCOC(=O)C(C)O-C1</chem>	
149408938	0.00005	9.1	643.8	<chem>CCCCCOC(CO-C1</chem>	
136030929	0.00005	9.6	627.8	<chem>CCC-C1</chem>	
137127598	0.00005	11.6	712.0	<chem>CCCCC(C)CO-C1</chem>	
142723568	0.00005	12.0	790.0	<chem>CCCCCCCCCOC-C1</chem>	
135487856	0.00100	10.4	627.8	<chem>*CCCCC(CO)CO-C1</chem>	

Examples of the 57 leads and their fragments obtained by docking Tinosorb-similars predicting lower binding-scores than Tinosorb to SLSF-residues ordered by their logP. The same fragments were attached to the C1 carbons at each of the Trihydroxyl-Triphenyls bound to the Triazine core (F+TTT, star-shaped molecules with 3-fold symmetries). *, In Tinosorb's one of the fragments is chemically different (Figure S1). 135487856, PubMed ID of Tinosorb (58th of the leads).

Binding-score, mean of the seeSAR's estimations in nM.

LogP, logarithm of the proportion between water and octanol solubilities.

MW, molecular weight in Daltons.

Smiles, Simplified Molecular Input Line Entry Systems.

Visualization of the best Tinosorb+SLSF-residues complexes suggested that poses predicted by seeSAR were crossbinding the inner sides of the 3x3 α -helix residues (Figure 1 Down **C,D**), while those generated by AutoDockVina predicted interactions with only two α -helices (Figure 1 Down **A,B**). These observations may explain the binding-score differences between the two programs. Together, these data suggested that to explore for new candidates targeting the binding-site of the S-SLSF 3x3 α -helices, the seeSAR program was the best option. Therefore, the seeSAR program was chosen for the rest of this work.

By seeSAR docking, 57 Tinosorb-similars reduced their binding-scores (leads) compared to Tinosorb (Figure 1, up). Most of the Tinosorb-similar leads contained a central Triazine core (N Nitrogens at the 1,3 and 5 positions of Triazine C=N rings), with Trihydroxyl-Triphenyl groups attached to its C Carbons (TTT). Similar chemical structures have been described in the chemical-synthesis literature as star-shaped molecules ³.

Additional structural variations among the Tinosorb-similar leads predicted different fragments attached to the **C1s** of the Triphenyl groups (Table 1). Many of them maintained 3-fold symmetrical star-shaped molecules (F+TTT, Fragment-Trihydroxyl-Triphenyl-Triazines). However, most of them were of similar or higher hydrophobicities than Tinosorb (see some selected examples in Table 1) and therefore predicting reduced drug-like possibilities. For instance, although 57 Tinosorb-similars predicted lower binding-scores than Tinosorb (Figure 1 up, **red large hexagon**), only one showed a minimal lower logP corresponding to moderately higher water-solubility (Table1, 136025237 ID).

Docking of TTT-similars to S-SLSF

Results of docking to TTT-similars predicted several leads with lower S-SLSF binding-scores compared to SLSF-residues (Figure 2). These results confirmed the chemical structural features than for Tinosorb-similars and suggested the implication of other amino acid residues.

Although predicting ~10-fold higher binding-score than Tinosorb (Figure 2), TTT had properties improving its drug-likeness such as lower molecular weight, higher water-solubility and 3-fold symmetry. Therefore, TTT was selected as the main initial structure for further improvements. On the other hand, there were also two other leads showing other cores and/or with less hydroxyls which suggested there may exist other chemotype leads.

Visual inspection of the TTT+S-SLSF complex confirmed docking to the inner binding-pocket of the 3x3 α -helices at the upper part of the top-to-bottom central cavity of the S trimer (Figure 3, A,B,D). Such binding-pocket was fitting TTT despite the existence of 36 competitors (Figure 3C, different colors). In contrast, Tinosorb was unable to predict any docking to S-SLSF ¹. Detailed exploration of the TTT+S-SLSF complex (Figure 3E) predicted Threonines T⁹⁹⁸ and Glutamines Q¹⁰⁰² as the main S-SLSF residues forming hydrogen bonds with TTT Trihydroxyls. The rest of the TTT atoms, including those of the Triazine cores, the Triphenyls and/or their attached fragments (in the case of TTT-similar leads) contributed to the final binding-score only by favoring hydrophobic desolvation around many of the amino acid residues between the S-SLSF 994 and the 1006 positions. Hydrophobic binding residues Y⁷⁵⁶ and F⁷⁵⁹ located 4-5 Å in front of the **C1s** outside S-SLSF (Figure 3E), were also included targets for desolvation (seeSAR, CCP4 and PyMol visualizations), confirming the participation of such residues in the S-SLSF bindings in contrast to SLSF-residues. The reduced participation of Triazine C=N atoms in binding, suggested that the Triazine scaffolds were shaping and holding together the "star-shaped" molecule.

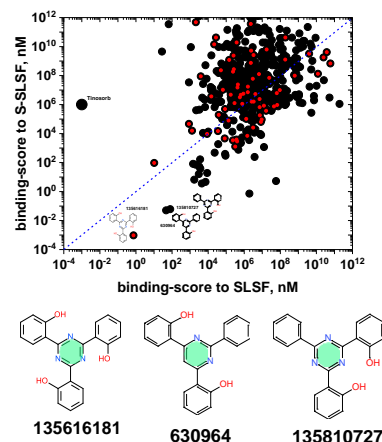


Figure 2

SLSF-residues and S-SLSF binding-scores of TTT-similars

Black circles, asymmetrical TTT-similars. **Red circles**, 3-fold symmetrical TTT-similars. **Black numbers and down chemical structures**, PubChem IDs of some of the leads. **Blue hatched line**, equal SLSF-residues and S-SLSF binding-scores.

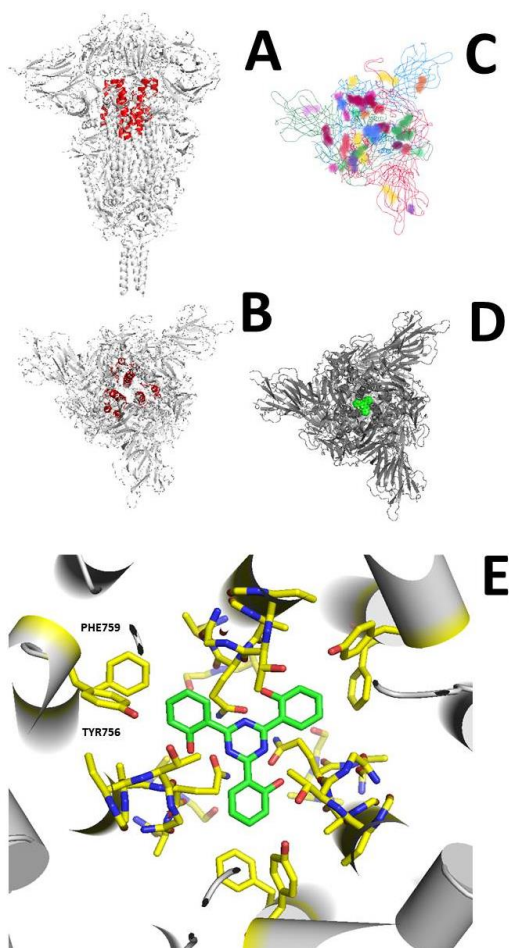


Figure 3

Mapping S-SLSF residues at the S trimer (A,B), S-SLSF binding-pockets predicted by seeSAR (C) and S-SLSF+ TTT complexes (D,E)

A, side-view of the wild-type S 6x8 trimer and location of the 3x3 α -helices of S-SLSF (red). Gray lines, rest of amino acid residues of the S trimer
B, top-view of the wild-type S 6x8 trimer and central location of its 3x3 α -helices of S-SLSF showing the central homotrimer cavity of 7-20 Å wide (red). Gray lines, rest of the amino acid residues of the S trimer
C, top-view of the 36 binding-pockets predicted by seeSAR in the exposed surfaces of the wild-type S 6x8 trimer in several background colors. Red, green, blue lines, rest of amino acid residues belonging to each of the S monomers in a different color line.
D, top-view of the S-SLSF docked to TTT. Green, TTT docked to the wild-type S 6x8 trimer in competition with the other 36 binding-pockets schematized in C. Gray lines, amino acids residues of the S trimer
E, top-view of the side chains of S-SLSF docked to TTT. Side-chain carbons corresponding to S-SLSF-residues 994-1006 and Y⁷⁵⁶ and F⁷⁵⁹ located at 4-5 Å complexed with TTT.
 Gray ribbons, sliced ribbon cylinders of the 3x3 α -helices of S-SLSF, the rest omitted for clarity.
 Green sticks, TTT carbon bonds.
 Yellow sticks, side chains of S-SLSF-residues 994-1006 plus outsider residues Y⁷⁵⁶ and F⁷⁵⁹
 Blue sticks, nitrogen bonds.
 Red sticks, Oxygen bonds

Search for new cores by Triazine-core replacement

Since 1,3,5 Triazine was not unique to the leads, explorations for alternative cores were attempted among Triazine-similars. However, although yielding a large number of 4346 drug-like molecules, none of them predicted S-SLSF lead binding-scores.

To further explore for Triazine-core alternatives, the seeSAR core-replacement feature was then employed. This methodology substituted Triazine for any core by different molecules while maintaining the rest of the Trihydroxyl-Triphenyl groups and predicts their resulting binding-scores (Figure S2). The screening of millions of possibilities, identified 4 new cores (TTX chemotypes) with similar S-SLSF-residues binding-scores than TTT. S-SLSF binding-scores were ~50 fold reduced from $\sim 10^{-2}$ nM with S-SLSF-residues to $\sim 10^{-4}$ nM with S-SLSF (Figure 4, open and closed red circles, respectively). These new X cores have 1 or 2 Ns (N1, N14, N35, N34). N1 and N35 were present in both pdb and zinc libraries (Figure 4).

These results suggested that other cores also predict molecules with subnanomolar lead binding to S-SLSF.

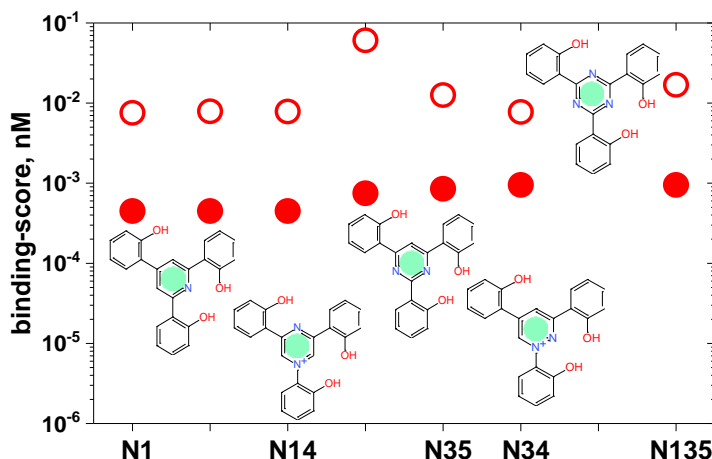


Figure 4

S-SLSF-residues and S-SLSF Triazine-core replacement binding-scores

The N135 (TTT) Triazine core was replaced by X-cores (TTX) using the core-replacement feature of seeSAR which screens the pdb / zinc fragment-libraries of tenths of millions molecules each (Figure S2). Y-axis, binding-scores of the newly generated core-replaced TTX molecules. X-axis, leads with cores labeled by N followed by their position in the 6 atom cores. The numbers corresponding to the pdb and zinc fragment IDs were as follows N1 (3qx502P1H51, zinc263631), N14 (zinc1593398), N35 (3bhh5CP1B600F, zinc8300484), N34 (zinc1564326) and N135 (PubChem 135616181).
 Open red circles, S-SLSF-residues binding-scores.
 Solid red circles, S-SLSF binding-scores.

Search for new leads using fragment extension

There were 5 different carbons per phenyl for possible fragment attachments since TTT Trihydroxyl-Triphenyls were not equivalent due to their tilting to each other around their rotatable C4 to C Triazine bonds (Figure S1). To study whether or not there were any differences among the Triphenyls, those were labeled as RED, GREEN and BLUE (RGB), using their default colors in seeSAR. Each of them were independently studied as possible targets for fragment extension (Figure S2). Comparative docking to S-SLSF-residues predicted 10^3 - 10^6 -fold lower binding-scores to the fragments attached to C1 (Figure 5), confirming previous observations made on Tinosorb, Tinosorb-similars, TTT and TTT-similars. Therefore, the C1 positions were targeted for additional searches for possible fragment extensions.

A first search for other possible C1-fragments among PubMed similars only found fragments for N1, N35 and N135 cores, but none of those displayed optimal drug-like characteristics (i.e., high hydrophilicities). On the other hand, the "build evolutionary library" option of DataWarrior did not generated alternatives for any of the N1, N14, N35, N34 and N135 cores (not shown).

The fragment extension feature of seeSAR was then applied individually to each of the 3 C1s RGB positions of each of the 5 cores. A fragment library using 100 fragment extensions (provided by seeSAR) enriched with 10 home-made small size fragments, were tested for fragment extension. The resulting TTX molecules docked to S-SLSF predicted that most, but not all, the present in leads were bound to each of the three RGB C1 positions with similar binding-scores (Figure S3, red bar means with low standard deviations among the 3 positions).

Those fragments/positions predicting binding-scores to < 0.2 nM (Table S1) were then combined to construct the corresponding F+TTX complete molecules for S-SLSF docking. The results identified dozens of new leads (Figure 5), predicting both symmetric and asymmetric fragment structures. Compared to Tinosorb, the new lead logPs were reduced from 10.4 to ~3-6, increasing their water-solubility. Steric or charged inhibition rather than hydrophobicity could partially explain some of the differences in binding-scores when changing targets from S-SLSF-residues to S-SLSF. For instance, some of the larger or charged fragments that generated S-SLSF leads, resulted in too high S-SLSF binding-scores ($> 10^6$ nM) when attached to the 3 C1 positions, similarly to Tinosorb¹.

Only one of the S-SLSF lead binding-scores, consisted in a new Trimethoxyl-Triphenyl-Pyrimidine chemotype molecule. However, further search did not identified any lower binding-score among 510-similars, nor fragment extension attempts to its N35 core could identify any lower or similar S-SLSF binding-scores (not shown).

Together, the fragment extension F+TTXs leads showed a S-SLSF first docking group on the ~ 0.001 nM binding-score range and a second group on the ~ 0.1 nM range (Figure 5). Up to 43.7 % of the F+TTX leads were found in PubMed while the rest corresponded to newly described molecules identified by fragment extension.

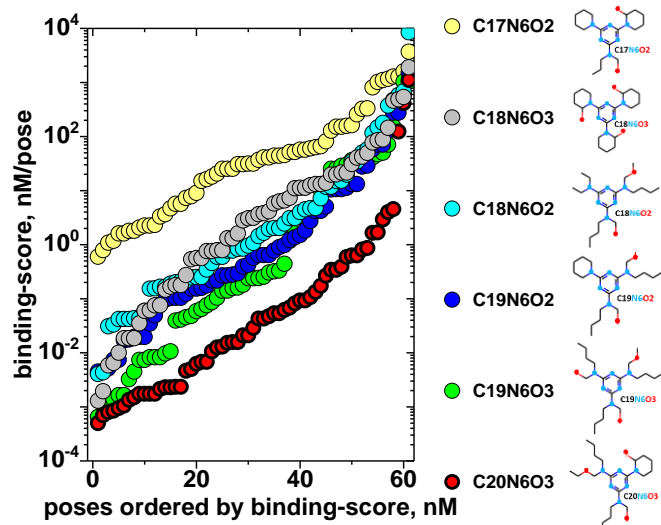


Figure 7

Profiles of S-SLSF binding-scores of conformational poses after docking lead-derivatives

To visualize their binding characteristics, 60 conformational poses per lead-derivative were generated and their corresponding S-SLSF binding-scores compared.

acceptor to block the liberated corrosive hydrochloride acid molecules (i.e., sodium carbonate, bicarbonate or hydroxide). The preparation of mono-, di- and tri-substituted cores of Cyanuric Chloride may be performed in a unique pot controlled step-by-step by increasing temperatures. This may be possible because generally, mono-substitution is predicted to occur at $\sim 0^\circ\text{C}$, di-substitution at $\sim 40^\circ\text{C}$ and tri-substitution at $> 120^\circ\text{C}$, at least in other similar cases⁴². Careful optimization of the amount of each of the fragments, order of fragment addition, temperature/time, solvent and hydrochlorine acceptor would be required, would be required to synthesize each of the desired lead-derivatives under appropriated safe conditions. Although wet-lab experimentation is required to successfully carry out such chemical synthesis, most paths mentioned above could be mimicked by some reaction-predictive programs (i.e., "Suggest Synthesis Route/of Structure using SPAYA.AI", DataWarrior). Although some of the identified fragments could be also synthesized before attachment to the core, the R1-NH-R2 fragments shown in Figure 8 were actually available at the building-block collections of Enamine (<https://www.enaminestore.com/search>) and/or Sigma (<https://www.aldrichmarketselect.com/>). Theoretically, with the 7 fragments detected at the building-block databases mentioned above, 6 lead-derivatives could be synthesized (Figure 8). Most probably the easiest lead-derivative to synthesize would be C18N6O3 because has the same fragments attached at each of its 3 possible positions.

The computational visualization of the TTT initial molecule and their lead-derivatives generated after substituting its aromatic carbons, opening the

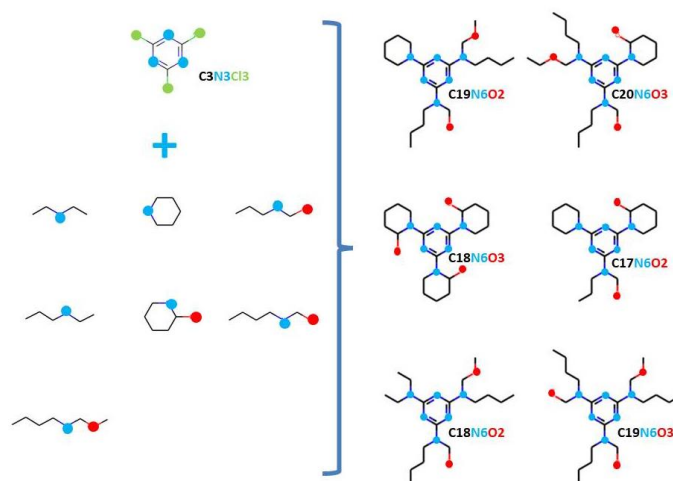


Figure 8

Scheme of predicted Ullmann crosslinking synthetic paths using a Cyanuric-Chloride core and 7 fragments to synthesize 6 lead-derivatives

The proposed example starts with the highly reactive Cyanuric Chloride core (C3N3Cl3) available from numerous commercial sources and commercially available fragments containing reactive Nitrogen, 4-6 Carbons \pm Oxygen. Among many other possibilities, core and fragments could be cross-linked using carbon-nitrogen Ullmann method. Blue circles, Nitrogen atoms. Red circles, Oxygen Atoms. Green circles, Chloride atoms. ---, Carbon bonds. Hydrogens have been omitted to increase clarity.

piperidine rings and attaching some of the above mentioned carbon/oxygen fragments, predicted similar S-SLSF docked complexes. Despite lead-dependent small alterations, their central location in between residues T⁹⁹⁸ (side chain Os bonded to hydroxyl Hs) and Q¹⁰⁰² (side chain Hs bonded to hydroxyl Os) forming two hydrogen-bonds per hydroxyl perpendicular to the homotrimer cavity were conserved. Figure 9 shows some comparative details of the relative positions in the TTT (C21N3O3) and its lead-derivatives C18N6O3 and C19N6O3.

Drug-like properties of the S-SLSF leads and lead-derivatives

The corresponding *in silico* pharmacokinetic parameters, physicochemical and toxicity ADME predicted to the leads (Table S2) and to lead-derivatives (Table S3) showed that while most of the leads were classified as only moderately soluble (mean logP = 4 ± 0.7), the lead-derivatives were ~ 100 -fold more soluble (mean LogP = 2.1 ± 0.6). Furthermore, while most of the leads predicted inhibition of many cytochrome P-450 detoxifying enzymes (Table S2), the lead-derivatives were more favorable (Table S3). Advantages of lead-derivatives respect to leads were also apparent when comparing synthetic difficulties calculated by fragmentation (Table S2) with predicted synthetic paths (Figure 8). Therefore, in comparison with the leads, the lead-derivatives not only predicted improved water-solubility and synthetic possibilities but may also improve other drug-like characteristics.

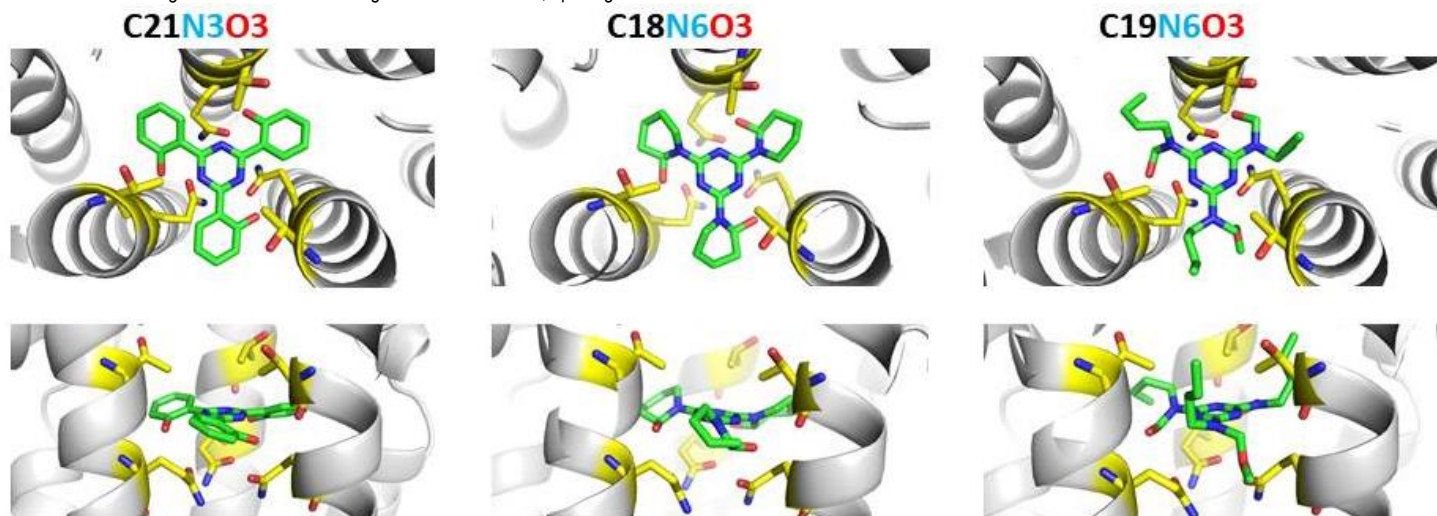


Figure 9

Top (up) and bottom (down) views of TTT and representative TTT- derivatives complexed with S-SLSF

The hydroxyls of TTT (C21N3O3) and TTT-derivatives (C18N6O3, C19N6O3) were Hydrogen-bonded to T⁹⁹⁸ and Q¹⁰⁰² S-SLSF side-chains as visualized in PyMol.

Gray partial ribbons, partial locations of the S-SLSF 3x3 α -helices, the rest omitted for clarity.

Green sticks, TTT and TTT-derivative carbon bonds. Yellow sticks, T⁹⁹⁸ and Q¹⁰⁰² side-chains implicated in Hydrogen-bonding

Blue sticks, Nitrogen bonds.

Red sticks, Oxygen bonds

Discussion

Despite SLSF-residues docking, Tinosorb failed to S-SLSF docking and did not inhibit S coronavirus-pseudotyped VSV-infection¹. Whether other Tinosorb-like star-shaped molecules do exist with improved drug-like properties at subnanomolar binding ranges have been computationally explored here.

First results were encouraging since ~ 50 more SLSF-residues leads could be identified among Tinosorb-similars. However, all except one, predicted higher logP hydrophobicities than Tinosorb, reducing their drug-like possibilities. Nevertheless, most of them showed a common minimal star-shaped molecular structure consisting of one Triazine core with attached Trihydroxyls-Triphenyls (TTT). Computational TTT core-replacement and fragment extension, identified dozens of star-shaped molecular lead alternatives including those containing hydrophilic short fragments (F) linked to their Triphenyl C1 carbons (F+TTT). Most important, ~10-fold lower S-SLSF binding-scores compared to SLSF-residues were predicted. Trihydroxyls formed hydrogen-bonds to S-SLSF T⁹⁹⁸ and Q¹⁰⁰². Reduction of S-SLSF binding-scores, could be due to additional desolvation interactions with hydrophobic residues outside the S-SLSF, such as those detected with Y⁷⁵⁶ and F⁷⁵⁹.

Nor Tinosorb, nor TTT-similar subnanomolar leads were previously detected among natural compounds, most probably due to the molecular weight cut-off used¹. Nevertheless, the TTT positive leads and the not-binding negative natural ligands previously identified, could be pooled to generate appropriated training-sets to optimize CNN models. Thus, the resulting trained-sets were used to develop a T13 deep-learning model to search for additional leads in large libraries using 2D molecular images. The T13 model successfully identified new 2D molecular candidates among ~ 2.5 million compounds in a short time compared to 3D docking, however, only 3 new star-shaped leads were detected. Further screening for TTT-similars of much larger libraries (10⁹-10¹⁴ compounds), did not contribute any new leads. In the future, further explorations including more star-shaped Triazine-derivative chemical spaces may be tried but libraries enriched in Triazine-derivatives should be first generated. Up to now, the star-shaped lead molecules were detected using work-intensive core-replacement, fragment-extension and/or manual iteration efforts. Most probably, automation by new algorithms based on similar search methods could contribute to discover any new star-shaped leads. Because trimer α -helices similar to S-SLSF may be participating in other important biological interactions, such algorithms could be of a more general interest.

Only seeSAR predicted star-shaped structures docking to the inner part of the 3x3 S-SLSF-residues at subnanomolar ranges. Among other possibilities, some problems with the sdf to pdbqt file conversion of Triazine rings by OpenBabel for AutoDockVina docking, may result in their higher binding-score predictions. In contrast, all the Triazine chemical features, including their correct placements of double bonds, were conserved during seeSAR docking. Nevertheless, AutoDockVina performed better dockings than seeSAR in other examples, such as screening for new rodenticides⁴³ or prediction of interactions between graphene and detoxifying enzymes⁴⁴. It seems likely that the best docking algorithms may depend on each particular ligand/target system.

Visualization of the S-SLSF docked complexes predicted only slightly different lead-dependent interactions. For most leads, the main contributions to their estimated binding-scores were made by hydrogen-bonds between their Trihydroxyls and the three S-SLSF T⁹⁹⁸ and Q¹⁰⁰² amino acids. There were none or few alternative fragments which could replace the contribution of the Trihydroxyl groups, confirming the importance of their hydrogen-bonds. The small distance-requirement for hydrogen-bonds may also explain why any short displacement, amplified by the contribution of the S-SLSF 3 α -helices, like those described for PP mutants⁴⁻⁶, could increase their corresponding lead binding-scores. On the other hand, the rest of the lead atoms, only contributed to the binding-score values by favoring desolvation (displacement of water molecules to allow for hydrophobic interactions). Such desolvations included amino acid atom targets both inside and outside the S-SLSF. Although weak for each pair of individual atoms, they made important desolvation contributions when added together for each of the lead binding-score computations. All fragment alternatives tested that contained charged atoms predicted unfavorable interactions, increased their corresponding binding-scores, and suggested that electrostatic charges were not implicated in the formation of S-SLSF+ lead complexes.

To explore for any more possible lead-derivatives with improved drug-like properties, the Triazine-core and the F-TTX fragments were selected. The Triazine-core was chosen because it has more synthesizable possibilities than any other of the identified cores. The F-TTX fragments were chosen because they were the result of S-SLSF docking optimization after exploring millions of alternatives. Their piperidine- and linear-derivatives were generated because these modifications of the TTT molecular structure increased their water-solubility maintaining their subnanomolar binding-scores. Manual iterations were employed because appropriated open-source programs including predictions of water-solubility, synthetic possibilities and building-block availabilities could not be found. Therefore, Piperidine- and their linear-derivatives were manually generated by iteration of combinations between Triazine-cores and F-TTX fragments. Those

generated candidates predicting subnanomolar S-SLSF binding-scores, were selected to study both their synthetic possibilities and their corresponding building-block availabilities. Although many other possible solutions do exist since a more comprehensive study has not been carried out, the manual strategy identified a few lead-derivatives which combined all the above mentioned practical requirements.

Compared to Tinosorb or TTT, the lead and lead-derivatives (leads for short) were more hydrophilic, of smaller size and maintained subnanomolar binding-scores. The leads have also a high proportion of rotatable bonds which may improve their penetration possibilities to reach their S-SLSF optimal binding-site by going through the side-chains surrounding the homotrimer cavity. However, some of these leads may still be difficult to dissolve in water for efficient delivery for *in vitro* assays or may be inhibitors of detoxifying cytochromes for *in vivo* drug-like purposes. Therefore, these aspects should be also considered when predicting or interpreting future validations. Experimental assays such as solid-phase binding to recombinant S spikes, cell culture inhibition of S pseudotyped VSV fusion (as we described before¹) and/or possible blocking of cellular infection, could be employed to *in vitro* validate some of the proposed lead biological activities against coronaviruses. Considering that molecules with 3-fold symmetries may be the most favorable for chemical synthesis, the corresponding leads may be preferred for initial validations.

The main challenge for experimental success of subnanomolar leads remains on how or when to access the inner S-SLSF binding-pocket target. The target may be reached i) at the compact wild-type prefusion state, ii) at the open prefusion states, or iii) during infection at trimer biosynthesis.

Most of the leads docked to the S-SLSF binding-pocket of the wild-type prefusion state, visually appeared tightly "sandwiched" between above Threonines (Ts) + Tyrosines (Ys) and below Glutamines (Qs) + phenylalanines (Fs). Such locations suggested difficult penetration throughout the side-chains of either top-to-bottom (Ts, Qs, Ys and Fs Oxygen atoms form triangles of ~ 9, 4, 17 and 17 Å side, respectively) or trilateral (~7-20 Å wide) cavities. At this respect, lead flexibility by having more rotatable C-C bonds may favor their possibilities for penetration compared to the more rigid alternatives containing aromatic rings.

Other alternatives for penetration may include targeting the more relaxed open (receptor binding domain, RBD-up) S conformers rather than the compact wild-type 6xr8 conformer. Such possible lead alternatives require further computational studies and experimentation. Focusing on their possible computational aspects, the preliminary docking to PP-mutated SwissModel-remodeled 6xr8 conformer showed unfavorable increases in their lead binding-scores. Those could be due to the ~ 2-4 Å displacements detected on their YF⁹⁸⁷ Oxygen-triangle distances while those of T⁹⁹⁸ and Q¹⁰⁰² remained similar (± 0.8 Å), as suggested by PyMol visualization (not shown) and confirming the low propensity for dynamic conformational changes of the wild-type homotrimer cavity previously described in crystalized structures². Furthermore, unfavorable higher binding-scores were obtained also with the 6xs6 3-open PP-mutated conformer which displayed Oxygen-triangle distances among Qs ~ 3 Å longer than those of 6xr8, while its Oxygen-triangle Ts distances were conserved (not shown). Both preliminary observations suggest that the lead bindings may have very strict S-SLSF 3x3 α -helix structural length requirements making it probable that S open conformations could require other lead chemotypes to target their slightly displaced 3x3 α -helices. Perhaps in the future, 3D structures of wild-type S in their open states may be available to allow for more accurate docking predictions, including molecular dynamic studies to best mimic the conformational variations during all fusion steps.

Other opportunities to reach the S-SLSF binding-pocket may be offered by targeting viral infection steps rather than prefusion viral particles as discussed above. For instance, targeting the S biosynthesis at the formation of dimers/trimers from monomers during the coronavirus infection. Any of these many possibilities need further computational mimicking and experimental work.

We may conclude that there are star-shaped leads and lead-derivatives with predicted subnanomolar binding-scores to S-SLSF which are highly specific for the wild-type 6xr8 conformers present in the surface of wild-type coronaviruses. Such leads have probabilities to inhibit coronavirus fusion, however, it is not known whether their accession to the target binding-site located at the inner part of the S-SLSF α -helices would be possible. In other words, despite their predicted subnanomolar binding-scores, small molecular sizes, high water-solubilities, synthesizability and expected accessibility possibilities to their target, it is not yet clear if any of the leads/lead-derivatives described here will successfully bind to S-SLSF, block fusion and inhibit coronavirus infectivity. Because some computational data indicated that the smaller the size, the better, any other leads with simpler chemical structures should be also explored including any possibility for covalent cross-linking the α -helices implicated taking advantage of the high specificity of the star-shaped leads identified.

Supporting information

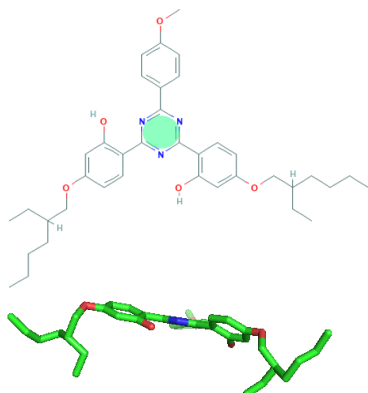


Figure S1

Molecular 2D and 3D molecular structures of Tinosorb

- A. 2D representation of Tinosorb (bis-ethylhexyloxyphenyl-methoxyphenyl-Triazine)
 B. 3D representation of Tinosorb best conformational pose when docked to SLSF

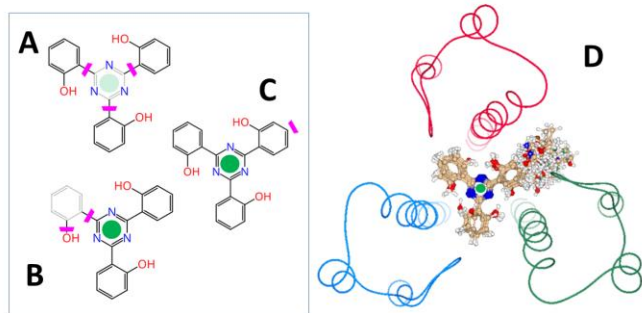


Figure S2

TTT core- / phenyl-replacements and hydroxyl-phenyl C1 fragment extensions

Core- and phenyl-replacements (A,B) were screened among ~40 millions of fragments of seeSAR's supplied libraries. Hydroxyl-phenyl C1 fragment extensions (C) were screened among the 100 low molecular weight fragment library provided by seeSAR and enriched with 10 home-made fragments. Each of the resulting new F-TTT molecules were then SLSF docked.

- A, core-replacements in **gray** and the rest of the maintained structures in **black**, separated by the **pink** insertion locations.
 B, phenyl-replacements in **gray** and the rest of the maintained structures in **black**, separated by the **pink** insertion locations.
 C, hydroxyl-phenyl C1 extension separated by the **pink** insertion location.
 D, example of merged fragments extended in position C1 at the **RED** hydroxyl-phenyl.

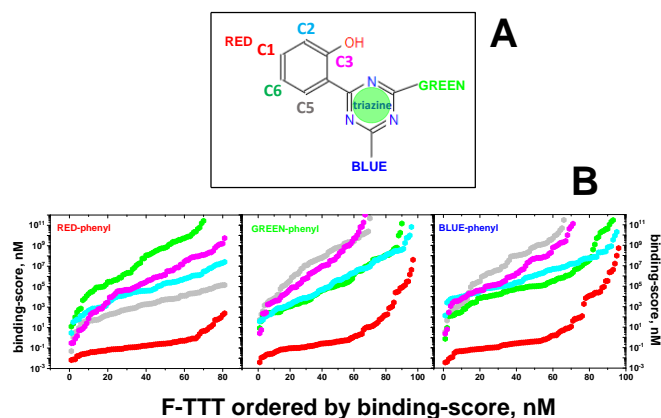


Figure S3

SLSF-docking of fragments extended to different C positions of TTT

- One hundred fragments provided by seeSAR's were linked to each of the TTT 1-6 Carbon phenyl positions (except position 4). **RED, GREEN** and **BLUE** hydroxyl-phenyls were independently extended and docked to SLSF.
 A, scheme of the TTT molecular structure, with only the **RED** hydroxyl-phenyl drawn (**GREEN** and **BLUE** hydroxyl-phenyls were omitted in the TTT structure for clarity).
 B, Binding-score profiles to SLSF-residues of the resulting F+TTT complexes.
Red hexagons, C1 phenyl carbons.
Blue hexagons, C2 phenyl carbons.
Purple hexagons, C3 phenyl carbons..
 Gray hexagons, C5 phenyl carbons..
Green hexagons, C6 phenyl carbons.

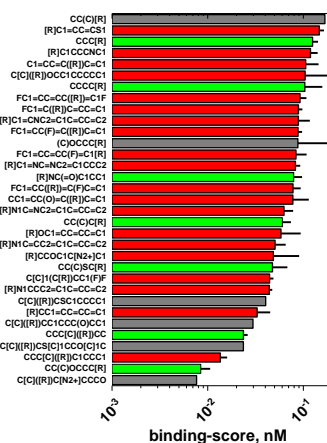


Figure S4

RGB lead binding-scores to SLSF-residues of the C1 fragment extension

The fragments represented in their smiles formula were those selected among the seeSAR's fragment extension leads having binding-scores < 0.2 nM. Mean \pm sd were calculated from the binding-scores of the 3 RGB hydroxyl-phenyl groups at each of the F+TTT complexes.

Y-axis, fragment smiles formula where [R] indicates the covalent bond to C1.

Gray bars, bound only to one of the RGB. **Red bars**, fragments with > 7 non-hydrogen atoms. **Green bars**, fragments with < 7 non-hydrogen atoms.

Table S1
SLSF binding-scores of cores + fragment extensions

			phenyl		
			RED	GREEN	BLUE
core		C1-fragment	nM	nM	nM
N1	N1+	[R]			
	N1+1	C[R]			
	N1+2	CC[R]	0.14	0.11	0.11
	N1+3	CCC[R]	0.05	0.03	0.05
	N1+4	CCCC[R]	0.07	0.03	0.05
	N1+5	CC(C)SC[R]	0.02	0.01	0.01
	N1+6	CC(C)C[R]	0.02	0.03	0.01
	N1+7	CC(C)OCCC[R]	0.01	0.02	0.07
	N1+8	(C)OCCC[R]	0.00	0.23	0.01
	N1+9	CS[R]	0.09	0.12	0.08
	N1+10	CC(O)[R]	0.13	1.30	0.09
	N1+11	COCN[R]	0.03	0.09	7.63
N1+12	CCC(C)([R])CC	0.11	0.01	0.01	
N14	N14+	[R]			
	N14+1	C[R]			
	N14+2	CC[R]	0.11	0.07	0.10
	N14+3	CCC[R]	0.06	0.03	0.02
	N14+4	CCCC[R]	0.09	0.02	0.02
	N14+5	CC(C)SC[R]	0.05	0.01	0.04
	N14+6	CC(C)C[R]	0.14	0.01	0.03
	N14+9	CS[R]	0.12	0.07	0.10
	N14+13	CC(C)[R]	260.74	0.04	0.07
	N14+14	COC([R])=O	0.12	0.08	0.08
	N14+15	C1COCCC1[R]	0.91	0.03	0.09
	N35	N35+0	[R]		
N35+1		C[R]			
N35+3		CCC[R]	0.41	0.07	0.08
N35+4		CCCC[R]	0.12	0.06	0.03
N35+5		CC(C)SC[R]	0.01	0.01	0.01
N35+6		CC(C)C[R]	0.05	0.05	0.05
N35+7		CC(C)OCCC[R]	0.00	0.00	0.01
N35+8		(C)OCCC[R]	0.01	0.01	0.05
N35+12		CCC(C)([R])CC	0.01	0.01	0.01
N34	N34+	[R]			
	N34+1	C[R]			
	N34+13	CC(C)[R]	33.70	0.12	0.06
	N34+16	CO[R]	1.34	1.34	0.17
N135	N135+	[R]			
	N135+1	C[R]	0.38	0.40	0.45
	N135+3	CCC[R]	0.12	0.14	0.10
	N135+4	CCCC[R]	0.16	0.09	0.05
	N135+5	CC(C)SC[R]	0.06	0.02	0.04
	N135+6	CC(C)C[R]	0.04	0.05	0.07
	N135+7	CC(C)OCCC[R]	0.01	0.01	0.01
	N135+8	(C)OCCC[R]	0.01	0.22	0.02
	N135+12	CCC(C)([R])CC	0.02	0.02	0.02

The F+TTT structures were generated by the seeSAR fragment extension of hydroxyl-phenyls **C1s** in the 5 different cores of Figure 4. The resulting structures were labeled by an N followed by their N positions in the core as in Figure 4. The selected fragments were arbitrarily numbered from +1 to +16 (i.e., N34+13). Those fragments represented by individual [R] corresponded to the initial core-replacement molecules of Figure 4. The represented fragment binding leads to SLSF-residues were defined as those F+TTT structures with predicting binding-scores < 0.2 nM. The tabulated fragments were then computationally drawn as been bound to their corresponding **C1** positions on the final F+TTT molecules to be tested by S-SLSF docking (final results in Figure 5). The minimal fragments C[R] were rejected by the fragment extension program, but their binding-scores included in the final F+TTT reconstructed molecules for S-SLSF docking.
Numbers in red, fragments with a high nM binding-score having low probability of fitting to the SLSF-residues and/or S-SLSF.

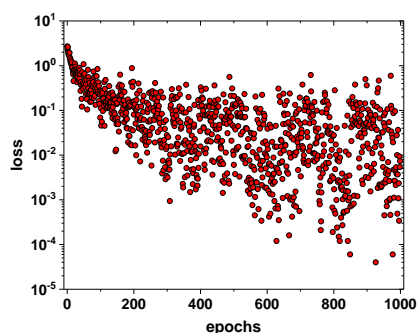


Figure S5

Learning curve of the DEEPScreen CNN T13 model developed for high-throughput screening of binding candidates among those of large libraries

The training-set contained 48 F+TTX ligands classified as positives (1s) (Figure 5) and randomized/size-selected 30 TTT-similars + 162 SNII as negatives (0s). To train the T13 model, the training-set of 240 classified compounds was randomized and splitted in 60 % for training, 20 % for validation and 20 % for test. **Loss**, mean differences between the T13 model predictions and their true classifications. **Epochs**, number of forward/backward CNN iterations through the training-set.

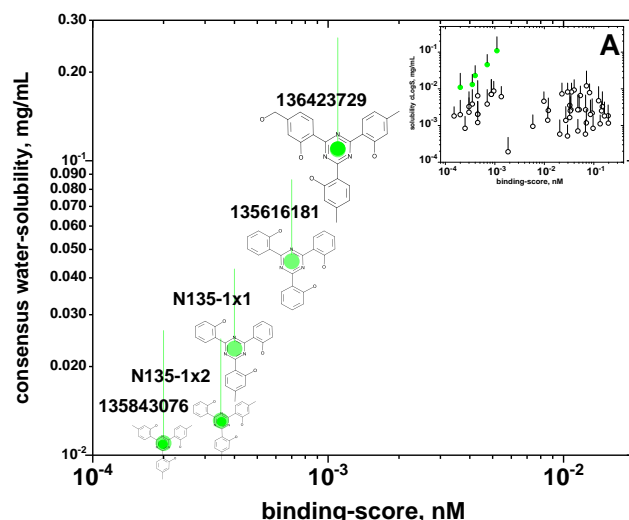


Figure S6

Predicted higher water-solubilities vs binding-scores of the leads

The leads from Figure 5 showing the higher mean water-solubilities in mg/ml were represented together with their corresponding formulas. **A insert**, water-solubilities vs binding-scores of all the leads from Figure 5. **Green circles and vertical lines**, mean and standard deviation water-solubility (n=3)

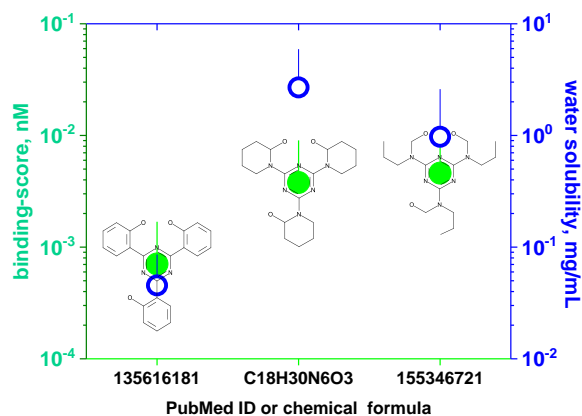


Figure S7

Lead-derivatives with higher water-solubilities and their corresponding binding-scores

Lead-derivatives theoretically drawn from TTT (PubMed ID, 135616181) predicting the most soluble pH-dependent logS profiles (<https://disco.chemaxon.com/calculators/demo/plugins/solubility>), were S-SLSF docked. Only the lead-derivatives with the lowest binding-scores were selected for the next solubility/docking iteration. After 10 iterations, two new chemotypes were identified with increased water-solubilities maintaining low binding-scores (C18H30N6O3 and 155346721 or C15H30N6O3). They contained a C3N6 Triamine-Triazine core with the attached Trihydroxyls (-COOH) required for hydrogen binding to the T⁹⁸⁸ and Q¹⁰⁰² residues of the S-SLSF 3x3 α-helices. **Green circles, and vertical upper lines**, mean and standard deviation of binding-scores expressed in nM (n=3). **Open blue circles and vertical upper lines**, mean and standard deviations of water-solubility expressed in mg/ml (n=3).

Table S2
Lead drug-like characteristics predicted by the SwissADME web-server

ID or Name	MW	LogP	sol	GIA	1A2	2C19	2C9	2D6	3A4	LipK	PAINS	SynAcc
141325665	0.0001	398.5	4.5									3.1
135843076	0.0002	399.4	4.2									2.7
N14+1x2	0.0002	385.4	3.0									3.2
N135+1x3	0.0002	399.4	4.2									2.7
N1+1x2	0.0002	383.4	4.6									3.0
N14+1x1	0.0003	371.4	2.6									3.1
N35+1x2	0.0003	384.4	4.2									3.0
N135+1x2	0.0003	385.4	3.9									2.6
N35+1x1	0.0003	370.4	3.9									2.9
N135+1x1	0.0004	371.4	3.5									2.5
137100803	0.0004	355.4	4.0									2.8
N14	0.0004	357.4	2.3									3.0
N1	0.0004	355.4	4.0									2.8
N1+1x1	0.0004	369.4	4.4									2.9
N135	0.0007	357.4	3.2									2.4
N35+1x1	0.0007	370.4	3.9									2.7
141325666	0.0008	356.4	3.7									2.8
N35	0.0008	355.4	3.7									2.8
13655883	0.0009	357.4	3.2									2.4
N34	0.0009	357.4	2.7									3.0
136423729	0.0011	415.4	3.5									2.9
135960896	0.0013	413.5	4.6									2.9
chembl220560	0.0018	444.4	5.5									3.0
N1+2x1	0.0058	383.4	4.7									2.9
N34+1x1	0.0098	371.4	3.1									3.2
N35+4x1	0.0117	412.5	4.9									3.2
13606876	0.0120	383.4	4.6									2.7
chembl4098217	0.0205	373.8	5.0									2.8
135839733	0.0227	355.4	4.0									2.5
N14+3x1	0.0271	399.5	3.3									3.3
153499063	0.0294	323.4	4.8									2.6
N34+16x1	0.0295	387.4	2.8									3.2
N35+1x3	0.0325	398.5	4.6									3.1
136627870	0.0327	427.5	4.9									3.1
135800943	0.0335	340.4	3.9									2.7
136159616	0.0344	339.4	4.4									2.8
N34+16x1	0.0354	387.4	2.7									3.2
136058071	0.0400	355.4	4.0									2.5
135981233	0.0474	339.4	4.4									2.7
N1+4x1	0.0476	411.5	5.4									3.2
630964	0.0505	340.4	4.0									2.7
136052741	0.0535	401.5	4.4									3.9
13762037	0.0678	373.8	5.0									2.8
chembl1096434	0.0681	339.4	4.4									2.7
135458191	0.0706	353.4	4.7									2.8
136052737	0.0710	401.5	4.1									3.9
N34+16x2	0.0820	417.4	2.7									3.4
N14+1x2	0.0855	385.4	3.0									3.2
N14+2x1	0.0926	385.4	3.0									3.1
N1+1x2	0.0965	383.4	4.7									3.0
N135+4x1	0.1250	413.5	4.6									2.9
N1+8x1	0.1343	427.5	4.7									3.2
N34+1x2	0.1437	385.4	3.4									3.3
N14+1x1	0.1491	371.4	2.7									3.1
136432921	0.1641	423.4	5.0									2.8
N35+8x1	0.1948	428.5	4.2									3.2
N1+8x1	0.1969	427.5	4.7									3.2

Lead numbers, PubMed IDs
Lead chembl number, ChEMBL IDs
Lead N numbers, core (Figure 4) + fragment number (Table S1) x 1.2 or 3 fragments.
LIPK, number of violations of Lipinski rules that would make the ligand less likely to be an orally administrable drug if >5.
LogP, consensus value of multiple predictions of lipophilicity.
1A2, 2C19, 2C9, 2D6, 3A4, in green the leads predicted to inhibit the main detoxifying cytochromes P450.
GIA, in green predictions of high gastro-intestinal adsorption.
PAIN, Pain Assay Interference Structures (PAINS), alerting of the number of chemical fragments that return false positives in virtual binding.
Green, favorable
Light green, moderately favorable.
Empty in white backgrounds, unfavorable.
SynAcc, synthetic difficulties calculated by fragmentation of the leads and ranged from 1-10 (the highest the most difficult to synthesize)

Table S3
Drug-like characteristics of lead-derivatives predicted by the SwissADME web-server

Formula	Cons. LogP	Cons. Sol mg/mL	GI	1A2	C19	2C9	2D6	3A4
C20N6O3	2.5	0.03						
C19N6O3	2.4	0.02						
C18N6O3	1.1	13.38						
C19N6O2	2.5	0.01						
C18N6O2	2.7	0.02						
C17N6O2	1.6	0.69						

Cons., SwissADME consensus LogP and Sol water-solubility.

Other details as in Table S2

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 deep-screen, and drafted the manuscript.

Acknowledgements

Thanks are specially due to Dr. Judd Duncan of Awridian Ltd at United Kingdom by kindly providing the 500k synthetic chemotype-maximized purchasable library, to Dr. Tunca Dogan of the University of Ankara at Turkey for his help to understand his simple and powerful DEEPScreen program, to Dr. Markus Gastreich of BioSolveIT GmbH at Germany for his zoom-help with the inspirator features of seeSAR. Thanks are also due to Dr. Alberto Villena from the University of Leon (Spain), Luis Maestre (telecommunication engineer) from Madrid and to Dr. Ignacio Garcia from the Hospital Gomez Ulla of Madrid (Spain) for their help with the bibliography, to Dra.Maria M.Lorenzo for her preliminary tests and to Dr. Rafael Blasco at INIA (Madrid, Spain) for his original ideas and discussions.

References

- ¹Lorenzo, M.M., R. Blasco and J.M. Coll. Would it be possible to stabilize prefusion SARS-CoV-2 spikes with ligands? *ChemRxiv*. 2021: <http://dx.doi.org/10.26434/chemrxiv.13453919.v2>.
- ²Kalathiya, U., M. Padariya, M. Mayordomo, M. Lisowska, J. Nicholson, A. Singh, . . . J.A. Alfaro. Highly Conserved Homotrimer Cavity Formed by the SARS-CoV-2 Spike Glycoprotein: A Novel Binding Site. *J Clin Med*. 2020, 9: <http://dx.doi.org/10.3390/jcm9051473>.
- ³Diab, 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>.
- ⁴Henderson, 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>.
- ⁵Hsieh, 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>.
- ⁶Xiong, 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].
- ⁷Carr, C.M. and P.S. Kim. A spring-loaded mechanism for the conformational change of influenza hemagglutinin. *Cell*. 1993, 73: 823-832. [http://dx.doi.org/10.1016/0092-8674\(93\)90260-W](http://dx.doi.org/10.1016/0092-8674(93)90260-W).
- ⁸Pallesen, J., N. Wang, K.S. Corbett, D. Wrapp, R.N. Kirchoerfer, 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>.
- ⁹Kandeel, M., A.H.M. Abdelrahman, K. Oh-Hashi, A. Ibrahim, K.N. Venugopala, M.A. Morsy and M.A.A. Ibrahim. Repurposing of FDA-approved antivirals, antibiotics, anthelmintics, antioxidants, and cell protectives against SARS-CoV-2 papain-like protease. *J Biomol Struct Dyn*. 2020: 1-8. <http://dx.doi.org/10.1080/07391102.2020.1784291>.
- ¹⁰Kandeel, M. and M. Al-Nazawi. Virtual screening and repurposing of FDA approved drugs against COVID-19 main protease. *Life Sci*. 2020, 251: 117627. <http://dx.doi.org/10.1016/j.lfs.2020.117627>.
- ¹¹Bakowski, M.A., N. Beutler, K.C. Wolff, M.G. Kirkpatrick, E. Chen, T.-T.H. Nguyen, . . . T.F. Rogers. Drug repurposing screens identify chemical entities for the development of COVID-19 interventions. *Nature Communications*. 2021, 12: 3309-3323. <http://dx.doi.org/10.1038/s41467-021-23328-0>.
- ¹²Xia, 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].
- ¹³Xia, 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].
- ¹⁴Wang, 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>.
- ¹⁵Cannalire, 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>.
- ¹⁶Tang, 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>.
- ¹⁷Wu, 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].
- ¹⁸Ruan, 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>.
- ¹⁹Tsuji, 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>.
- ²⁰de Souza Neto, L.R., J.T. Moreira-Filho, B.J. Neves, R. Maidana, A.C.R. Guimaraes, N. Furnham, . . . F.P. Silva, Jr. In silico Strategies to Support Fragment-to-Lead Optimization in Drug Discovery. *Front Chem*. 2020, 8: 93. <http://dx.doi.org/10.3389/fchem.2020.00093>.
- ²¹Liu, T., M. Naderi, C. Alvin, S. Mukhopadhyay and M. Brylinski. Break Down in Order To Build Up: Decomposing Small Molecules for Fragment-Based Drug Design with eMolFrag. *J Chem Inf Model*. 2017, 57: 627-631. <http://dx.doi.org/10.1021/acs.jcim.6b00596>.
- ²²Patel, H., W.D. Ihlenfeldt, P.N. Judson, Y.S. Moroz, Y. Pevzner, M.L. Peach, . . . M.C. Nicklaus. SAVI, in silico generation of billions of easily synthesizable compounds through expert-system type rules. *Sci Data*. 2020, 7: 384. <http://dx.doi.org/10.1038/s41597-020-00727-4>.
- ²³Polishchuk, P. CREM: chemically reasonable mutations framework for structure generation. *J Cheminform*. 2020, 12: 28. <http://dx.doi.org/10.1186/s13321-020-00431-w>.
- ²⁴Jahnke, W., D.A. Erlanson, I.J.P. de Esch, C.N. Johnson, P.N. Mortenson, Y. Ochi and T. Urushima. Fragment-to-Lead Medicinal Chemistry Publications in 2019. *J Med Chem*. 2020, 63: 15494-15507. <http://dx.doi.org/10.1021/acs.jmedchem.0c01608>.
- ²⁵Rifaiglu, A.S., E. Nalbat, V. Atalay, M.J. Martin, R. Cetin-Atalay and T. Dogan. DEEPScreen: high performance drug-target interaction prediction with convolutional neural networks using 2-D structural compound representations. *Chem Sci*. 2020, 11: 2531-2557. <http://dx.doi.org/10.1039/c9sc03414e>.
- ²⁶Yang, X., J. Zhang, K. Yoshizoe, K. Terayama and K. Tsuda. ChemTS: an efficient python library for de novo molecular generation. *Sci Technol Adv Mater*. 2017, 18: 972-976. <http://dx.doi.org/10.1080/14686996.2017.1401424>. 1401424 [pii].
- ²⁷Spiegel, J.O. and J.D. Durrant. AutoGrow4: an open-source genetic algorithm for de novo drug design and lead optimization. *J Cheminform*. 2020, 12: 25. <http://dx.doi.org/10.1186/s13321-020-00429-4>.
- ²⁸Chevillat, F., S. Stotani, A. Karawajczyk, S. Hristeva, E. Pardon, J. Steyaert, . . . P. Kolb. Interrogating dense ligand chemical space with a forward-synthetic library. *Proc Natl Acad Sci U S A*. 2019, 116: 11496-11501. 1818718116 [pii]. <http://dx.doi.org/10.1073/pnas.1818718116>.
- ²⁹Volochnyuk, D.M., S.V. Ryabukhin, Y.S. Moroz, O. Savych, A. Chuprina, D. Horvath, . . . D.B. Judd. Evolution of commercially available compounds for HTS. *Drug Discov Today*. 2019, 24: 390-402. S1359-6446(18)30242-3 [pii]. <http://dx.doi.org/10.1016/j.drudis.2018.10.016>.
- ³⁰Polishchuk, P. Control of Synthetic Feasibility of Compounds Generated with CREM. *J Chem Inf Model*. 2020, 60: 6074-6080. <http://dx.doi.org/10.1021/acs.jcim.0c07092>.
- ³¹Yang, T., Z. Li, Y. Chen, D. Feng, G. Wang, Z. Fu, . . . M. Zheng. DrugSpaceX: a large screenable and synthetically tractable database extending drug space. *Nucleic Acids Res*. 2021, 49: D1170-D1178. 5940503 [pii]. <http://dx.doi.org/10.1093/nar/gkaa920>.
- ³²Blasco, R. and J.M. Coll. In silico screening for natural ligands to non-structural nsp7 conformers of SARS coronaviruses. *ChemRxiv*. 2020: <http://dx.doi.org/10.26434/chemrxiv.12952115.v2>.
- ³³Dallakyan, 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.
- ³⁴Trott, 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>.
- ³⁵Morris, G.M., R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell and A.J. Olson. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem*. 2009, 30: 2785-91. <http://dx.doi.org/10.1002/jcc.21256>.
- ³⁶Huey, R., G.M. Morris, A.J. Olson and D.S. Goodsell. A semiempirical free energy force field with charge-based desolvation. *J Comput Chem*. 2007, 28: 1145-52. <http://dx.doi.org/10.1002/jcc.20634>.
- ³⁷Rarey, M., B. Kramer, T. Lengauer and G. Klebe. A fast flexible docking method using an incremental construction algorithm. *J Mol Biol*. 1996, 261: 470-89. S0022-2836(96)90477-5 [pii]. <http://dx.doi.org/10.1006/jmbi.1996.0477>.
- ³⁸Schneider, 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>.
- ³⁹Schneider, 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>.
- ⁴⁰Reau, 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>.
- ⁴¹Garcia, S. and F. Herrera. Evolutionary undersampling for classification with imbalanced datasets: proposals and taxonomy. *Evol Comput*. 2009, 17: 275-306. <http://dx.doi.org/10.1162/evco.2009.17.3.275>.
- ⁴²Kotha, S., M. Meshram, N.R. Panguluri, V.R. Shah, S. Todeti and M.E. Shirbhate. Synthetic Approaches to Star-Shaped Molecules with 1,3,5-Trisubstituted Aromatic Cores. *Chem Asian J*. 2019, 14: 1356-1403. <http://dx.doi.org/10.1002/asia.201801912>.
- ⁴³Bermejo-Nogales, A., J.M. Navas and J.M. Coll. Computational ligands to VKORC1s and CYPs. Could they predict new anticoagulant rodenticides? *BioRxiv*. 2021: <http://dx.doi.org/10.1101/2021.01.22.426921>.
- ⁴⁴Connolly, M.C., J.M. Navas and J. Coll. Prediction of nanographene binding-scores to trout cellular receptors and cytochromes. *bioRxiv*. 2021: <http://dx.doi.org/10.1101/2021.02.20.432107>.