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: julio.collm@gmail.com (JC)  
Julio Coll, orcid: 0000-0001-9496-3493

Abstract

This work describes synthetizable 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 Trihydroxy-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; synthetizable, 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 muta7ions. 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 diameter 2. In the wild-type pre fusion state, the SLSF-residues and the S-SLSF (SLSF residues at the S trimer) included 3x3 α-helices (Figure 1).

Computational screening of hundreds thousands of natural compounds predicted Tinosorb as the top subnanomolar binding lead to SLSF-residues 1. The Tinosorb's preferential binding to SLSF-residues, 3-arms-star-shaped molecular structure and crossbinding the inner part of the SLSF-α-helices, together with the existence of side and central top-to-bottom surface-accessible cavities to SLSF-residues at the S wild-type pre fusion 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 pre fusion states blocking fusion and infection, similarly to what has been previously described for PP mutations of the S2 HR1-HR2 residues 4,6. However, computational studies could not predict Tinosorb's S-SLSF binding, nor in vitro fusion by S-pseudotyped VSV-infectivity could be inhibited by Tinosorb 1.

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 viron at their pre fusion states and/or to interfere with the earliest pre fusion steps, the S-SLSF wild-type trimer model (fox8 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 viruses 7.  
  
4. The mutations to prolines (PP mutations) at the tip of the S--SLSF 3x3 α-helices mentioned above, stabilize the loops at their earliest pre fusion state blocking viral-host membrane fusion and inhibiting viral infectivity 4,6.

Previous and abundant experimental and computational search for anti-coronavirus molecules has been focused on approved drugs (drug repurposing) 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 nM 1 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 amilorid (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, 14, 15, 16, 18, the surface interphase of S with the ACE2 human receptor 17, and/or the active sites of viral proteases 15, 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 others 20. These included searches among similar, core replacement, fragment extensions 21-24, convolutional neural networks (CNN) 25, de novo manual generation of compounds with drug-like properties 26-29, 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 Trihydroxy-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 images 35 allowed the screening of larger compound libraries including those for maximal diversity among purchasable chemicals 29, 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 135487656, CHEMBL 2104956, CAS 183793-00-6), chemically known as bemoctizol (2,2’-[6-(14-methoxyphenyl)]-1,3,5-Triazine-2,4-diyl)[bis-(2-ethylhexyl)cyclophenol]), has a star-shaped structure with a molecular weight of 527.8 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-like 1740 molecules were downloaded from PubChem https://pubchem.ncbi.nlm.nih.gov/compound?q=555369&term=tinosorb (search for molw: < 400, yl: Triazine) and from PubChem Trihydroxy-Triphenyl-Triazine (135616181 ID), and Tinosorb (9262)-similar searches resulted in 599 and 279689 molecules, respectively. Tinosimb-similar were downsized 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 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 I SNI library (http://biomf-applied.charl.de/supernatural_new/index.php) was split in nineteen *.sdf files each containing different molecular weights < 380 Daltons, as described before. 14. 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 DEEPscreen26 C113 TMD model.

The infinSee extremely large libraries provided by BioSolveIT (CHEMinya_11bn_2021-05_1.1x10^9 compounds), GalaXi_2.1bn_2020-12.1x10^9 KnowledgeSpace_290r_2019-05_2.9x10^10, and REALSpace_18bn_2021-04_1.9x10^11 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 syntheses were searched among the data bases developed by Enamine (https://www.enaminstore.com/search) and SigmaAlrich (https://www.sigmaalrich.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 6x8 S trimeric spike model

The wildtype SARS-CoV-2 trimer (PDB 6x8) 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 1110, as previously reported! (Figure 1) or used for S-SLSF docking (SLSF in the complete trimer structure) using iBioS 10 compounds 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, package (https://pyrx.sourceforge.io/) used multimart on multi-core 67 i7 or 16 processors to speed up docking14. AutoDockVina generates poses with Lamarckian genetic algorithms 28. The corresponding conformation-dependent Gibbs binding-score free-energies (ΔG) for each pose are calculated using semi-empirical data 29. To perform the docking, the *.sdf files required free energy minimization, hydrogen and charges additions to generate *pdbqt files available (OpenBabel). Grids including only the SLSF-residue inner 3x3x3 α-helices were used. Only one pose with the lowest binding-score per docked ligand (* .pdbqt) was retrieved for analysis. To compare with the seeSAR binding-score values, the output ΔG energies in kcal/mol were converted to constant inhibition (Ki) in mM concentrations1. To predict possible seeSAR poses, the vs.10 package (https://www.biosolveit.de/SeeSAR) uses the FLEXx incremental fragment construction method based on software development for computer vision and pattern recognition 27. The corresponding conformation-dependent binding-score Gibbs free-energies (ΔG) for each pose are calculated by the HYDRE scoring function computing both HYdration and DEsolvation values (as calibrated from octanol/water partition data, logP) 30. To reduce false positives, the HYDRE calculations include not only favorable but also unfavorable interactions 6. 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 3C), 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 HYDRE lower-higer 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 (Mobrowser vs3.9-1b1W64bit) 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 files were randomly sampled for 10 binding-score 0s >100 nM from each separate file to generate a representative pool of negative ligands for the training sets (under sampling the majoritant class)31. Other 30 TTT-similarities with binding-scores 0s > 100 nM were added to the negative pool. On the other hand, 48 F+TTX 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-sets as the unique molecular identifiers, using the Dataviewer 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 split 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/padel/descriptor) 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.aiab.si/orange). Their accuracies, however, were low.

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

The DEEPScreen software 36, 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 C113 TMD 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 (http://www.chemaxon.com/calculators/demixp/kljx挽/esolubility). 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 3.9.1b1Win 64bit (MolLogS), Osiris DataWarrior vs5.5.0 (LogS) and Virtual Computational Chemistry Laboratory (http://146.107.217.179/web/algo/ from vcclab.org/labaloggs, aLogS). The estimated consensus LogS water-solubility values in molf were then transformed to mg/l 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 10-100-fold lower binding-scores (higher bindings) for seeSAR compared to AutoDockVina (Figure 1, top, red hexagons versus blue hexagons).

Visualization of the best Tinosorb+SLSF-residues complexes suggested that poses predicted by seeSAR were crossing 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 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 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 1. 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-Triazine). However, most of them were of similar or higher hydrophobicities than Tinosorb (see some selected examples in Table 1) and therefore predicted 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 (Table 1, 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 1998 and Glutamines 1015 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 Y176 and F179 located 4-5 Å in front of the C1s outside S-SLSF (Figure 3E), were also included targets for desolvation (seeSAR, CCP4 and PyMoI 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.

Table 1

<table>
<thead>
<tr>
<th>SLSF-residues lead binding-scores of Tinosorb-similar showing their fragments linked to C1 phenyls</th>
<th>PubChem ID</th>
<th>Binding-score, nM</th>
<th>logP</th>
<th>MW</th>
<th>Smiles</th>
</tr>
</thead>
<tbody>
<tr>
<td>136025237</td>
<td>0.0005</td>
<td>4.4</td>
<td>621.7</td>
<td></td>
<td><img src="image" alt="Smiles" /></td>
</tr>
<tr>
<td>150024905</td>
<td>0.0005</td>
<td>6.2</td>
<td>621.7</td>
<td></td>
<td><img src="image" alt="Smiles" /></td>
</tr>
<tr>
<td>135783913</td>
<td>0.0005</td>
<td>7.9</td>
<td>609.7</td>
<td></td>
<td><img src="image" alt="Smiles" /></td>
</tr>
<tr>
<td>135011728</td>
<td>0.0005</td>
<td>8.2</td>
<td>615.9</td>
<td></td>
<td><img src="image" alt="Smiles" /></td>
</tr>
<tr>
<td>136040444</td>
<td>0.0005</td>
<td>8.3</td>
<td>623.7</td>
<td></td>
<td><img src="image" alt="Smiles" /></td>
</tr>
<tr>
<td>135740105</td>
<td>0.0005</td>
<td>8.7</td>
<td>747.9</td>
<td></td>
<td><img src="image" alt="Smiles" /></td>
</tr>
<tr>
<td>136383973</td>
<td>0.0005</td>
<td>8.7</td>
<td>773.9</td>
<td></td>
<td><img src="image" alt="Smiles" /></td>
</tr>
<tr>
<td>135059048</td>
<td>0.0005</td>
<td>9.0</td>
<td>773.9</td>
<td></td>
<td><img src="image" alt="Smiles" /></td>
</tr>
<tr>
<td>145089938</td>
<td>0.0005</td>
<td>9.1</td>
<td>643.8</td>
<td></td>
<td><img src="image" alt="Smiles" /></td>
</tr>
<tr>
<td>136039929</td>
<td>0.0005</td>
<td>9.6</td>
<td>627.8</td>
<td></td>
<td><img src="image" alt="Smiles" /></td>
</tr>
<tr>
<td>137127599</td>
<td>0.0005</td>
<td>11.6</td>
<td>712.0</td>
<td></td>
<td><img src="image" alt="Smiles" /></td>
</tr>
<tr>
<td>142723658</td>
<td>0.0005</td>
<td>12.0</td>
<td>790.0</td>
<td></td>
<td><img src="image" alt="Smiles" /></td>
</tr>
<tr>
<td>135487756</td>
<td>0.0100</td>
<td>10.4</td>
<td>627.6</td>
<td></td>
<td><img src="image" alt="Smiles" /></td>
</tr>
</tbody>
</table>

Examples of the 17 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 Trihydroxy-Triphenyl bound to the Triazine core (F-TTT, star-shaped molecules with 3-fold symmetry). 1. In Tinosorb’s one of the fragments is chemically different (Figure S1). 135487756, PubChem ID of of Tinosorb (15th 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.

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.

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.

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.
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 Trihydroxy-Triphenyl groups and predicts their resulting binding-scores (Figure S2). The screening of millions of possibilities, identified 4 new cores (TTX chemotypes) with similar SLSF-residues binding-scores than TTT. S-SLSF binding-scores were ~50 fold reduced from $-10^4$ nM with SLSF-residues to $-10^2$ nM with S-SLSF (Figure 4, open and closed red circles, respectively). These new X cores have 1 or 2 Ns rather than 3 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.

Search for new leads using fragment extension

There were 5 different carbons per phenyl for possible fragment attachments since TTT Trihydroxy-Triphenyls were not equivalent due to their titling to each other around their rotate 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 SLSF-residues predicted 10^-1-10^-2 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 generate 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 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 charge inhibition rather than hydrophobicity could partially explain some of the differences in binding-scores when changing targets from SLSF-residues to S-SLSF. For instance, some of the larger or charged fragments that generated SLSF leads, resulted in too high S-SLSF binding-scores ($>10^0$ nM) when attached to the 3 C1 positions, similarly to Tinosorb.

Only one of the S-SLSF lead binding-scores, consisted in a new Trimethoxy-Triphenyl-Pyridine 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.
Screening large libraries by the CNN classifying T13 model

To search for more leads, a newly developed T13 deep-learning CNN model using 2D molecular images was applied to larger libraries (learning rate of the model in Figure S5). A first screened library consisted in 0.5 x 10^10 computationally designed compounds covering a maximum of purchasable chemotypes. Results predicted 105 possible lead candidates. However, their S-SLSF lowest binding-scores were in the high nM ranges. A second screened library consisted in ~2 x 10^8 million compounds downloaded from the last Chembl28 release and downsized to ~1.5 x 10^6 drug-like ligands. Results predicted 8751 possible lead candidates. To downsize such large number of candidates, 3D-docking was made to SLSF- residues. The resulting 34 candidates predicting < 1 nM SLSF-residues binding-scores were docked to S-SLSF identifying 4 new leads. Since one of the leads was previously identified, 3 leads were added to the final lead list (Figure 5, purple stars).

Screening extreme larger libraries by infiniSee

To explore larger chemical spaces, the BioSolveIt’s infiniSee program was used to screen for TTT-similars among 4 libraries containing 10^10-10^11 compounds each. However, results only identified compounds with S-SLSF binding-scores >3-4 nM among the best 1000 similar compounds in any of the target similarities or thresholds studied. Additional attempts targeting 100000 compounds were also unsuccessful (not shown).

Binding leads to computationally PP-mutated S trimers

Because preliminary results indicated that some of the leads predicted large differences when docked to other PP-mutated conformers selected from our previous study 1 (data not shown) and their amino acid sequence differences to the wild-type 6x8 trimer were only due to their PP mutations, those mutations were computationally introduced into the 6x8 wild-type trimer. The corresponding D3 trimer remodel derived in the SwissModel web-server from the 6x8PP mutated amino acid sequence was then docked to the leads. The results predicted an estimated 30.9 % of the wild-type 6x8 S-SLSF lead poses altered their binding-scores by the introduction of the PP mutations (Figure 6). In some F-TTX leads those differences were of several orders of magnitude corresponding to poses that did not cross-bind the 3x3 α-helices of S-SLSF (not shown).

Water-soluble Triazine-derivatives with chemical synthesis possibilities

Water-solubilities of the F-TTX leads described at Figure 5, predicted consensus mean values of 0.01 ± 0.04 mg/mL (n=57). Only 6 leads predicted water solubilities >10^4 mg/mL (Figure S6). Such low water-solubility estimations may be insufficient for in vitro validation and/or drug-like purposes. On the other hand, only one of them was commercially available (TTT, PubChem 136616181). Furthermore, for many of them appropriate building-blocks for chemical synthesis were difficult to find (https://askcos.mit.edu/reto/network/; https://www.enaminestore.com/search; https://www.aldichemmarketselect.com/).

To search for new lead-derivatives, chemical spaces for TTT-derivatives with both higher water-solubility estimations and more synthetic possibilities were explored among manual designs. Batches of 5-10 TTT-derived molecules predicting increased logS pH-dependent water-solubility profiles (https://disco.chemaxon.com/calculators/demo/plugins/solubility), were then S-SLSF docked. The resulting leads from TTT-derivatives were selected for the next solubility/docking iteration.

Substitution of the TTT phenyls by piperidine rings (substitution of C4s by N4s and elimination of aromaticity), increased ~100-fold their water-solubility with small variations in their binding-scores (Figure S7). Most probably, the low participation of the phenyl C4s in the binding-score estimations (as shown by visualization of the corresponding docked complexes), could explain these results. After 10 iterations, two new TTT-derivatives were identified with increased water-solubilities (Figure S7). Both C19H30N6O3 and C19H30N6O3 candidates maintained the C3N6 Triamine-Triazine core together with attached Trihydroxyls. The presence of Trihydroxyls was a sine-qua-non requirement for hydrogen bonding as shown by visualization of the predicted docked complexes (not shown).

The Triamine-Triazine core is well known in the literature as Melamine. Melamine and many other highly-reactive Triazine-like compounds (i.e., TriChloro, TriBromo, Cyanuric acid, Trimethoxy, Isocyanuric, Triitol, Tricarboxyl, and/or several combinations among them) are widely available from a high number of commercial sources (Enamine, Sigma, ChemVia, Chemshuttle, Apollo-Scientific). Such Triazine-like compounds have been widely used as the initial cores for a variety of synthetic chemical reactions with the aim to develop related molecules for many different practical applications 2. Therefore, the presence of the binding requirement of Melamine cores increased the synthetic possibilities of any of the lead-derivatives identified.

Linearization of many compounds after eliminating one carbon per piperidine-ring, increased not only their water-solubility but also their synthetic possibilities. After 10 iterations the C19H30N6O3-derivative, resulting after linearization of the C19H30N6O3 molecule was also identified as the other lead-derivative.Manually adding O, OH or NH2 to fragments in many different proportions and combinations, further increased the resulting water-solubilities, but the corresponding binding-scores also increased (not shown).

To explore for more possible fragments to be attached to the two newly identified lead-derivatives, the optimized fragments of the previously identified F-TTX leads (Figure 5) were used as a source to manually generate Triazine piperidine- and linear-derivatives. Therefore, phenyls were substituted by piperidines eliminating the aromatic carbons and substituting the phenyl C4s by N4s. The resulting piperidine rings were then opened for linearization. All possible fragment combinations among the 3 C1 positions were manually generated for new F-C3N6 molecules and their resulting dual-binding-score values eliminated by Open Babel. This process generated 84 F-C3N6-derivatives that were S-SLSF docked resulting in 15 lead-derivatives with binding-scores < 0.6 nM to study their chemical synthesis possibilities. Figure 7 shows the S-SLSF pose profiles docked to the most representative C3N6 lead-derivatives (60 poses per derivative).

As one example among many synthetic possibilities, the 15 new lead-derivatives could be deconstructed in 6 possible candidate ligands by combinations of commercially available building-blocks and cores (Figure 8). Many alternative synthetic paths including different cores and fragments could be followed. As one example, the Ulmann carbon-nitrogen cross-coupling method using the Cyanuric-Tricloride (2,4,6-Trichloro-1,3,5-Triazine) core and R1-NH-R2 fragments, was studied in some detail, since it appears to be one of the most simple to combine with commercially available components 14 (Figure 8).

In such method, substitution of the core chlorine atoms by N- containing fragments may be performed in the presence of one hydrochloride
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 ~0°C, di-substitution at ~40°C and tri-substitution at >120°C, at least in other similar cases. Careful optimization of the amount of each of the fragments, order of fragment addition, temperature/time, solvent and hydrochloride 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/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.sigmaaldrich.com/search). 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 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.

![Graph showing binding-scores of S-SLSF binding-scores of conformational poses after docking lead-derivatives.](image)

**Figure 8**
Scheme of predicted Ulmann 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 (C19N6O3) available from numerous commercial sources and commercially available fragments containing reactive Nitrogen, 4-6 Carbons ± Oxygen. Among many other possibilities, core and fragments could be cross-linked using carbon-nitrogen Ulmann method.


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 TTT (side chain Os bonded to hydroxyl Hs) and QO32 (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.

**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-bound to TTT and QO32 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, TTT and QO32 side-chains implicated in Hydrogen-bonding

Blue sticks, Nitrogen bonds. Red sticks, Oxygen bonds

**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 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 deoxygenating 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.

![Diagram showing the predicted Ulmann crosslinking synthetic paths using Cyanuric-Chloride core and 7 fragments to synthesize 6 lead-derivatives.](image)
Discussion

Despite SLSF-residues docking, Tinosorb failed to S-SLSF docking and did not inhibit S coronavirus-pseudopolyed VSV-infection. 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 Trizaine core with attached Trihydroxyls-Triphenylens (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 TTT and Q<sub>102</sub>. 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 T<sub>102</sub> and F<sub>102</sub>.

For Tinosorb, nor TTT-similar subnanomolar leads were previously detected among natural compounds, most probably due to the molecular weight cut-off used. Nevertheless, the TTT positive leads and the not-binding negative natural ligands previously identified, could be pooled to generate appropriately 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 any TTT-similars of much larger libraries (10<sup>5</sup>-10<sup>4</sup> compounds), did not contribute any new leads. In the future, further explorations including more star-shaped Trizaine-derivative chemical spaces may be tried but libraries enriched in Trizaine-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 phenomena, some problems with the second type of sequence file conversion of Trizaine rings by OpenBabel for AutoDockVina docking, may result in their higher binding-score predictions. In contrast, all the Trizaine 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 rodenticides or prediction of interactions between graphene and detoxifying enzymes. 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 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 TTT and Q<sub>102</sub> amino acids. There were none or few alternative fragments which could replace the contribution of the Trihydroxylic groups, confirming the importance of their hydrogen-bonds. The small distance-range, for hydrogen-bond docking, may be amplified by the contribution of the S-SLSF 3 α-helices, like those described for PP mutans,<sup>4</sup> could increase their corresponding lead binding-scores. On the other hand, the rest of the lead atoms, only contributed to the binding-scores 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 Trizaine-core and the F-XTX fragments were selected. The Trizaine-core was chosen because it has more synthetizable possibilities than any other core of the identified cores. The F-XTX 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, Pipeline - and with the derivatives were manually generated by iteration of combinations between Trizaine-cores and F-XTX 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 hydrophobic, 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, the nucleation of S pseudo-pseudopolyed VSV-infection (as we described before) 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 who can access the inner S-SLSF binding-pocket target. The target may be reached i) at the compact wild-type pre-fusion state, ii) at the pre-fusion states, or iii) during infection at trimer biosynthesis. Most of the leads docked to the S-SLSF binding-pocket of the wild-type pre-fusion state, visually appeared tightly "sandwiched" between Throneines (Ts) + Tyrosines (Ys) and below Glutamates (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 A side, respectively) or trilateral (~7-20 A wide) cavities. At this respect, the 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 6x8 conformer. Such possible lead alternatives require further computational studies and experimentation. Focusing on their possible computational aspects, the preliminary docking of the 1C24-remodeled TTT x-ray structure showed unfavorable increases in their lead binding-scores. Those could be due to the ~2-4 A displacements detected on their YF937 Oxygen-triangle distances while those of T<sub>102</sub> and Q<sub>102</sub> 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 structures.<sup>5</sup> Furthermore, unfavorable higher binding-scores were obtained also with the 6x8 3- sms of the 6x8 conformer which displayed Oxygen-triangle distances among Qs ~ 3 Å longer than those of 6x8 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 only 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 state may be available for more accurate computational predictions including molecular dynamic studies to best mimic the conformational variations during infection 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 work also with the preliminary molecular work we have carried out. 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 6x8 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-ethylhexyloxyethoxy-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 seSAR'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 tested 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, C9 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 extensions leads having binding-scores < 0.2 nM. Mean ± 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

<table>
<thead>
<tr>
<th>core</th>
<th>C1-fragment</th>
<th>RED</th>
<th>GREEN</th>
<th>BLUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1</td>
<td>N1+</td>
<td>100</td>
<td>0.11</td>
<td>0.01</td>
</tr>
<tr>
<td>N1+</td>
<td>C[R]</td>
<td>100</td>
<td>0.11</td>
<td>0.01</td>
</tr>
<tr>
<td>N1+</td>
<td>CCC[R]</td>
<td>0.05</td>
<td>0.03</td>
<td>0.01</td>
</tr>
<tr>
<td>N1+</td>
<td>CCCCC[R]</td>
<td>0.07</td>
<td>0.03</td>
<td>0.01</td>
</tr>
<tr>
<td>N1+</td>
<td>CCCCCC[R]</td>
<td>0.12</td>
<td>0.06</td>
<td>0.03</td>
</tr>
<tr>
<td>N1+</td>
<td>CCCCCCRR</td>
<td>0.09</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>N1+</td>
<td>CCCCCCRRR</td>
<td>0.05</td>
<td>0.01</td>
<td>0.03</td>
</tr>
<tr>
<td>N1+</td>
<td>CCCCCCCRR</td>
<td>0.14</td>
<td>0.01</td>
<td>0.03</td>
</tr>
<tr>
<td>N1+</td>
<td>CCCCCCRRR</td>
<td>0.12</td>
<td>0.07</td>
<td>0.01</td>
</tr>
<tr>
<td>N1+</td>
<td>CCCCCCCRRR</td>
<td>260.74</td>
<td>0.04</td>
<td>0.07</td>
</tr>
<tr>
<td>N1+</td>
<td>CCCCCCRRRR</td>
<td>0.12</td>
<td>0.08</td>
<td>0.08</td>
</tr>
<tr>
<td>N1+</td>
<td>CCCCCCRRRRR</td>
<td>0.91</td>
<td>0.03</td>
<td>0.09</td>
</tr>
<tr>
<td>N35</td>
<td>N35+</td>
<td>[R]</td>
<td>0.41</td>
<td>0.07</td>
</tr>
<tr>
<td>N35+</td>
<td>C[R]</td>
<td>0.12</td>
<td>0.06</td>
<td>0.03</td>
</tr>
<tr>
<td>N35+</td>
<td>CCC[R]</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>N35+</td>
<td>CCCCC[R]</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>N35+</td>
<td>CCCCCC[R]</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>N35+</td>
<td>CCCCCC[R]</td>
<td>0.11</td>
<td>0.01</td>
<td>0.05</td>
</tr>
<tr>
<td>N35+</td>
<td>CCCCCC[R]</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>N34+</td>
<td>[R]</td>
<td>33.70</td>
<td>0.12</td>
<td>0.06</td>
</tr>
<tr>
<td>N34+</td>
<td>C[R]</td>
<td>1.34</td>
<td>1.34</td>
<td>0.17</td>
</tr>
<tr>
<td>N135</td>
<td>N135+</td>
<td>[R]</td>
<td>0.38</td>
<td>0.40</td>
</tr>
<tr>
<td>N135+</td>
<td>C[R]</td>
<td>0.12</td>
<td>0.14</td>
<td>0.10</td>
</tr>
<tr>
<td>N135+</td>
<td>CCC[R]</td>
<td>0.16</td>
<td>0.09</td>
<td>0.05</td>
</tr>
<tr>
<td>N135+</td>
<td>CCCCC[R]</td>
<td>0.06</td>
<td>0.02</td>
<td>0.04</td>
</tr>
<tr>
<td>N135+</td>
<td>CCCCCC[R]</td>
<td>0.04</td>
<td>0.05</td>
<td>0.07</td>
</tr>
<tr>
<td>N135+</td>
<td>CCCCCC[R]</td>
<td>0.07</td>
<td>0.07</td>
<td>0.01</td>
</tr>
<tr>
<td>N135+</td>
<td>CCCCCC[R]</td>
<td>0.01</td>
<td>0.22</td>
<td>0.02</td>
</tr>
<tr>
<td>N135+</td>
<td>CCCCCCRR</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
</tr>
</tbody>
</table>

The F-TTX structures were generated by the seeSAR fragment extension of hydroxy-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., N1+1). Those fragments represented by individual [R] corresponded to the initial core-replacement molecules of Figure 4. The represented fragment binding leads to SLSF-residues as defined in those F-TTX 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-TTX molecules to be tested by SLSF docking (see results in Figure S5). The minimal fragments were rejected by the fragment extension program, but their binding-scores included in the final F-TTX 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 CNN13 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 T3T-similar + 162 SN1 as negatives (0s). To train the T33 model, the training set of 240 classified compounds was randomized and split in 60 % for training, 20 % for validation and 20 % for test. Loss, mean differences between the T33 model predictions and the 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).

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

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


