



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Structure-based drug repositioning over the human TMPRSS2 protease domain: search for chemical probes able to repress SARS-CoV-2 Spike protein cleavages

Natesh Singh^a, Etienne Decroly^b, Abdel-Majid Khatib^{c,d,*}, Bruno O. Villoutreix^{a,*}

^a Univ. Lille, INSERM, Institut Pasteur de Lille, U1177, F-59000 Lille, France

^b Aix Marseille Univ. CNRS, AFMB UMR 7257, Marseille, France

^c Univ. Bordeaux, Allée Geoffroy St Hilaire, 33615 Pessac, France

^d INSERM, LAMC, UMR 1029, Allée Geoffroy St Hilaire, 33615 Pessac, France

ARTICLE INFO

Keywords:

Virtual screening
Drug repurposing
Drug repositioning
Homology modeling
SARS-CoV2
COVID-19
Structural bioinformatics

ABSTRACT

In December 2019, a new coronavirus was identified in the Hubei province of central China and named SARS-CoV-2. This new virus induces COVID-19, a severe respiratory disease with high death rate. A putative target to interfere with the virus is the host transmembrane serine protease family member II (TMPRSS2). This enzyme is critical for the entry of coronaviruses into human cells by cleaving and activating the spike protein (S) of SARS-CoV-2. Repositioning approved, investigational and experimental drugs on the serine protease domain of TMPRSS2 could thus be valuable. There is no experimental structure for TMPRSS2 but it is possible to develop quality structural models for the serine protease domain using comparative modeling strategies as such domains are highly structurally conserved. Beside the TMPRSS2 catalytic site, we predicted on our structural models a main exosite that could be important for the binding of protein partners and/or substrates. To block the catalytic site or the exosite of TMPRSS2 we used structure-based virtual screening computations and two different collections of approved, investigational and experimental drugs. We propose a list of 156 molecules that could bind to the catalytic site and 100 compounds that may interact with the exosite. These small molecules should now be tested in vitro to gain novel insights over the roles of TMPRSS2 or as starting point for the development of second generation analogs.

1. Introduction

A novel coronavirus (SARS-CoV-2) began spreading in December 2019 in the city of Wuhan, China, causing a major outbreak of fatal pneumonia (Zhou et al., 2020). The Covid-19 disease mediated by SARS-CoV-2 is currently detected in most countries around the world and for the time being, there are no approved treatments or vaccines available. Cell entry of coronaviruses relies on the binding of the viral spike (S) proteins to cellular receptors after S protein cleavage by host cell proteases. Like SARS-CoV, SARS-CoV-2 was also found to use the receptor angiotensin-converting enzyme 2 (ACE2) for entry and the serine protease TMPRSS2 for S protein activation (Hoffmann et al., 2020). Interestingly, the S protein of SARS-CoV-2 was found to show a higher cell membrane fusion capacity compared to SARS-CoV (Xia et al., 2020), suggesting that this feature may contribute to the higher contagion and transmissibility of SARS-CoV-2 compared to SARS-CoV. Thereby, interfering between spike (S) and its receptor may

constitute a potential antiviral approach to target Covid-19 disease. The importance of TMPRSS2 in these processes (Iwata-Yoshikawa et al., 2019; Matsuyama et al., 2010; Simmons et al., 2004) suggests that this protein could be a valuable therapeutic target to fight SARS-CoV-2 infection.

TMPRSS2 is a protein that contains several domains: a LDL-receptor like domain, a scavenger receptor cysteine-rich (SRCR) domain and a serine protease domain. It is co-expressed in the lung tissue with the virus receptor angiotensin converting enzyme 2 (Bertram et al., 2012; Heurich et al., 2014). Two main serine protease inhibitors, camostat and nafamostat, have been shown to block the catalytic site of TMPRSS2, highlighting the importance of this human protein as therapeutic target (Hoffmann et al., 2020; Kawase et al., 2012; Yamamoto et al., 2016). Nafamostat inhibited S-mediated membrane fusion of MERS-CoV while camostat appeared to inhibit the entry SARS-CoV or SARS-CoV-2 into human lung cells. While most proteins from SARS-CoV-2 are currently being screened, it would seem beneficial to also

* Corresponding authors.

E-mail address: bruno.villoutreix@inserm.fr (B.O. Villoutreix).

<https://doi.org/10.1016/j.ejps.2020.105495>

Received 9 June 2020; Received in revised form 16 July 2020; Accepted 27 July 2020

Available online 28 July 2020

0928-0987/ © 2020 Elsevier B.V. All rights reserved.

investigate proteins from the host. In the face of this major crisis, a possible strategy is to reposition known drugs on viral or human protein targets. Such compounds should be valuable as possible treatments or to shed light over the mechanisms involved in the disease pathway. Experimental screening (e.g., phenotypic screening) can obviously be used, but the approach is time consuming and may miss some interesting molecules that may not work on cells but that can still be interesting starting points. In front of this emergency crisis, it is thus reasonable to use *in silico* methods to try to identify putative binders and inhibitors acting on a selected target or pathway. Indeed, computational drug repositioning can be of interest to find new indications to approved, investigational or experimental drug compounds (Cereto-Massague et al., 2015; Ekins et al., 2019; Farha and Brown, 2019; Klimenko et al., 2016; Oprea and Overington, 2015; Sam and Athri, 2019; Singh et al., 2020; Stumpfe and Bajorath, 2020; Villoutreix et al., 2013). From the above comments, we here propose to carry out *in silico* drug repositioning on the serine protease domain of TMPRSS2.

Serine protease proteins are endopeptidases that cleave peptide bonds and present with a serine residue in the active site that acts as a nucleophile (Patel, 2017). This family of proteins has many functions in the organism, from metabolism, digestion, blood coagulation, apoptosis, to immunity among others. Based on substrate specificity, serine proteases are ramified into numerous types such as trypsin-like, chymotrypsin-like, etc. Serine proteases are characterized by an active site that is generally made of three highly conserved amino acids, Ser, His and Asp, that form the so-called catalytic triad and work by charge relay network. TMPRSS2 belongs to the trypsin-like type, these enzymes usually cleave peptide bonds at Lys or Arg residues. This so-called P1 residue fits into a negatively charged S1 pocket (Fig. 1). This pocket displays a conserved Asp residue at the bottom and as such tends to prefer a positively charged residue at the P1 position. Numerous molecules (peptides, proteins, small molecules) can naturally block the catalytic site or have been designed specifically to inhibit the catalytic site of various serine proteases. Yet, it is also known that many serine proteases possess exosites. For example, thrombin, a critical serine protease of the blood coagulation system, has two major exosites, exosite I and exosite II (Fig. 1). These exosites can bind negatively charged molecules such as heparin (e.g., thrombin exosite II), protein substrates, protein cofactors and natural or rationally designed inhibitors (e.g., thrombin exosite I) (Huntington, 2014). The binding of molecules to these exosites can for instance inhibit the activity of the enzyme through structural changes that propagate up to the catalytic site (allosteric mechanism) or by blocking the binding of a protein substrate (e.g., some act as protein-protein interaction inhibitors). Finding molecules that do not block the active site but bind to exosite regions can be valuable for many protein families for therapeutic interventions or to probe molecular mechanisms involved in the health and disease states (Nicola et al., 2020; Sperandio et al., 2008; Villoutreix and Miteva, 2016). Thrombin is not the only one serine protease to have exosites and we have for instance identified via structure-based virtual screening and *in vitro* screening small molecules that bind to an exosite of protein C, an anticoagulant serine protease. These compounds interfere with the function of protein C by inhibiting the binding of protein substrates (Sperandio et al., 2014). Considering these observations, we believe that it is here important to not only investigate the catalytic site of TMPRSS2 but also potential exosite(s).

To carry out structure-based virtual drug repositioning over the serine protease domain of TMPRSS2, we followed the workflow depicted in Fig. 2. As there is no experimental structure for TMPRSS2, comparative modeling was used. The search for druggable pockets at the surface of the 3D models obviously suggested to screen the catalytic site and one main possibly important functional exosite. Two different collections of drugs were used and two docking engines were applied on the two selected 3D models of TMPRSS2. Structural analysis taking into account various binding scores obtained with three different scoring

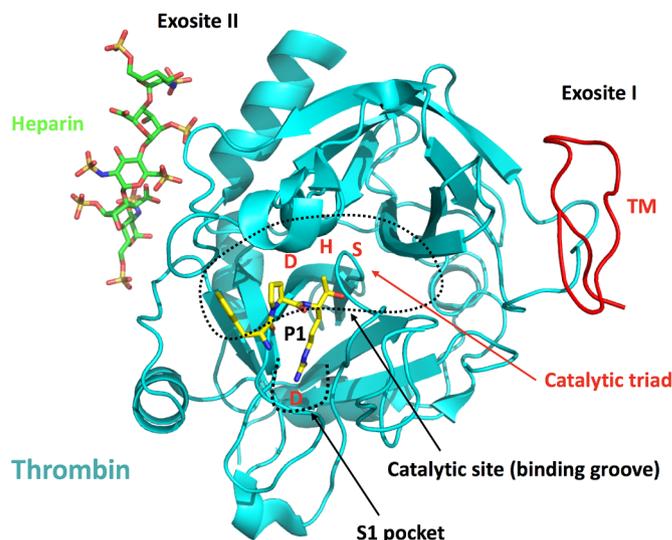


Fig. 1. Overview of a serine protease domain. Thrombin is used here as an example of serine protease. The protein is shown as a cartoon diagram with a view down the active site. The catalytic triad residues are shown and a conserved Asp residue present at the bottom of the S1 pocket for orientation. A small molecule with a positively charged chemical group (the so-called P1 residue) plugs into the S1 pocket. In thrombin, two major exosites are known, exosites I and II. These regions can bind some specific molecules such as heparin or be important for protein-protein interactions (e.g., interactions with protein substrates, protein cofactors and cofactors, designed or naturally occurring inhibitors). The crystal structure of thrombin bound to heparin (octasaccharide chain interacting with residues of the exosite II) and with a small peptide-like inhibitor is shown (PDB file 1XMN (Carter et al., 2005)). A small nonadecapeptide (ribbon in red) belonging to the fight EGF domain of thrombomodulin (TM, an integral membrane protein expressed on the surface of endothelial cells that serves as cofactor for thrombin allowing efficient cleavage and activation of another serine protease named protein C) co-crystallized with thrombin (PDB file 1HLT (Mathews et al., 1994)) was grafted onto the thrombin-heparin structure to highlight the region of exosite I.

functions, consensus scoring results and other rescoring approaches took place to select a list of putative catalytic site inhibitors and a list of compounds that could bind to the exosite of the serine protease domain of TMPRSS2 and perturb the catalytic site and/or interfere with substrate binding or protein partner binding.

2. Methodology

2.1. Homology modeling

The sequence of human TMPRSS2 serine protease domain was obtained from the UniProt database (ID: O15393, residue number 256-489) (UniProt, 2019). Structural templates were searched using the SWISS-MODEL server (Waterhouse et al., 2018) and using BLASTp (Altschul et al., 1997). Multiple sequence to structure alignments were carried out using the MAFFT-DASH server (Rozewicki et al., 2019). Various templates were investigated and BLASTp e-values were analyzed so as to select the best structural template to build the TMPRSS2 model. After appropriate investigations, we built the serine protease domain of TMPRSS2 using as template the experimental structure of human plasma kallikrein co-crystallized with a small molecule catalytic site inhibitor, RCSB Protein Data Bank (PDB) (Burley et al., 2019) file ID: 6O1G (Partridge et al., 2019) (resolution 2.2 Å). The SWISS-MODEL server (Waterhouse et al., 2018) and the MODELLER v9.16 (Sali and Blundell, 1993) package were used to build initial models. One model was built with SWISS-MODEL and 100 models with MODELLER. All of the models generated by MODELLER were assessed on the basis of the normalized Discrete Optimized Protein Energy (z-DOPE) score. The top-

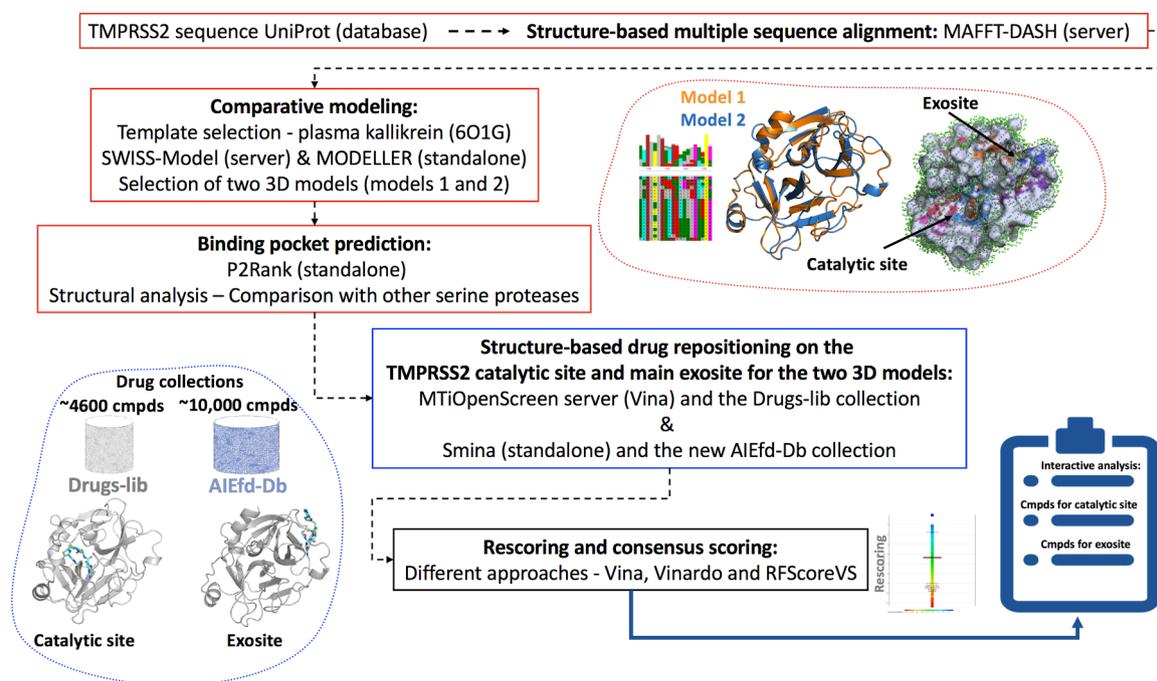


Fig. 2. Structure-based virtual screening workflow. The main approaches used to model the serine protease domain of TMPRSS2 and to screen the catalytic site and main exosite are shown. Two different structural models were selected, two different collections of approved drugs, investigational and experimental compound collections were used and the docking was performed using the MTiOpenScreen server and the standalone application Smina (see the method section).

ranked model showed a Z-DOPE score of -0.55, implying that > 70% of its C α atoms are within 3.5 Å of their accurate positions, thus indicating a reliable predicted structure (Eramian et al., 2008). The protonation state of the titratable residues of the two selected 3D models was investigated with our PCE server (computation of pKa) (Miteva et al., 2005). The quality of the 3D models was also investigated using the PSVS server (Bhattacharya et al., 2008) that implements numerous packages including PROCHECK (Laskowski et al., 1996). Secondary structures were computed via the PDBsum service (Laskowski et al., 2018). Interactive structural analysis was performed with using UCSF Chimera or ChimeraX (Goddard et al., 2018; Pettersen et al., 2004), PyMOL Schrodinger and the ICM-Browser (Molsoft LLC, San Diego, USA).

2.2. Approved, investigational and experimental drug collections

In order to identify advanced molecules acting on TMPRSS2, two different collections of approved, investigational and experimental drugs were used: Drugs-lib and AIEfd-Db. We have reported recently the Drugs-lib collection (Lagarde et al., 2018). Briefly, about 22,000 approved and investigational drugs (and a few experimental compounds) were obtained from DrugBank (Wishart et al., 2018), DrugCentral (Ursu et al., 2019), SuperDrug2 (Siramshetty et al., 2018), and ChEMBL (Mendez et al., 2019). Molecules were curated, duplicates were removed and filtering was performed to reject compounds with documented toxicophores using FAF-Drug4 (Lagorce et al., 2017). Only molecules with less than 20 rotatable bonds and a MW below 1000 Da were kept (i.e., docking accuracy decreases significantly when molecules are too flexible/large, thus we kept molecules within the mentioned thresholds). Further, in Drugs-lib, only molecules that could be found in chemical vendor catalogs were kept, leading to about 4600 unique molecules. The protonation state was predicted using ChemAxon (<https://chemaxon.com/>) and the 3D structures were built using Corina (<https://www.mn-am.com/>).

For the present study, we generated a newer and different collection of small molecules that we call AIEfd-Db (Approved, Investigational and Experimental compounds for docking database) using the following protocol: about 30,000 non-unique approved, experimental and

investigational compounds were downloaded from the last release of DrugBank (Wishart et al., 2018), DrugCentral (Ursu et al., 2019), e-Drug3d (Dougnet, 2018), SuperDrug2 (Siramshetty et al., 2018), SWEETLEAD (Novick et al., 2013) and ChEMBL (Mendez et al., 2019). Some drugs accepted by the FDA at the end of 2019 and early 2020 were added manually. Some of the molecules present in these databases have been withdrawn from the market but some compounds are now in clinical trials and thus belong to the investigational set. In addition, over 17,500 molecules were extracted from Wikipedia Chemical Structures using utilities implemented in the DataWarrior package (Sander et al., 2015). Molecules were also filtered using our FAF-Drug4 server (Lagorce et al., 2017) but only to remove compounds with inorganic atoms. In this electronic library, drug compounds with putative toxicophores were only flagged and kept in the collection. Also, by contrast to the Drugs-lib collection, molecules that could not be found in major commercial vendor catalogs (Lagarde et al., 2018) were kept as they might be available from some companies not included in our initial search. Molecules were curated, salts and duplicates were removed as for the Drugs-lib collection. Manual inspection took place, further guided by considering a drug-likeness score and several others computed physicochemical properties such as logP. In addition, as for the Drugs-lib collection only molecules that could be docked with a reasonable chance of success were kept (e.g., molecules with less than 20 rotatable bonds and with a MW below 1000 Da). The selected molecules were generated in 3D and protonated using Surfex tools (Jain et al., 2019).

2.3. Structure-based virtual screening

Prediction of likely binding pockets for small molecules was performed with the P2Rank (Krivak and Hoksza, 2018), a machine learning-based tool that proposes putative ligand binding sites using as input a protein 3D structure. Analysis of the prediction together with comparisons with other serine proteases indicated, obviously, that the catalytic site was interesting to screen but that a major exosite, next to the catalytic site, could also be valuable to investigate by docking. The MTiOpenScreen web-server with the Drugs-lib collection

(Lagarde et al., 2019; Lagarde et al., 2018) was applied on the two structural models of TMPRSS2 on the catalytic site and the main predicted exosite. This server implements AutoDock Vina for structure-based virtual screening and scoring tool (Forli et al., 2016). Similar computations were carried out with the standalone tool Smina (Koes et al., 2013), a fork of AutoDock Vina, that implements different types of scoring function including Vinardo, an approach that was shown to outperform the Vina scoring function on several targets (Quiroga and Villarreal, 2016). In this case, the AIEfd-Db collection described above was used. For all computations, the three best energy poses were kept initially. Rescoring of all the poses (the ones obtained from the MTiOpenScreen server and from Smina on the two structural models and two binding pockets) was carried out with Smina, using the same protocol for all molecules, and applying the implemented Vina (hybrid scoring function, empirical + knowledge-based) (Trott and Olson, 2010) and Vinardo (physics-based) (Quiroga and Villarreal, 2016) scoring functions so as to have all the molecules evaluated with the same parameters. In addition, rescoring was also carried out using the random forest scoring function (RF-Score-VS-v2) (Wojcikowski et al., 2017) so as to gain additional information about the molecules with yet a different type of scoring function (Ain et al., 2015).

There are different solutions to perform consensus scoring (reviewed in (Feher, 2006)). With this approach, the predicted binding affinities or scores of each compound for a binding pocket are predicted by using more than one scoring method. In this study, a consensus scoring approach using the 'rank-by-rank' method (Wang and Wang, 2001) was applied to evaluate the hits obtained from high throughput docking. All the candidates were ranked by the average ranks predicted by all the scoring functions. This strategy uses relative ranks rather than absolute binding affinities for ranking. Since the compound docking scores obtained from Vina, Vinardo, and RF-Score-VS programs are of different natures, data were normalized to bring all the scores in a common scale ranging from 0 to 1 using Eqs. 1 and 2.

$$\begin{aligned} & \text{(For positive scores); Normalized Score} \\ & = \frac{\text{Docking Score} - \text{Docking Score}_{\min}}{\text{Docking Score}_{\max} - \text{Docking Score}_{\min}} \end{aligned} \quad (1)$$

$$\begin{aligned} & \text{(For negative scores); Normalized Score} \\ & = 1 - \frac{\text{Docking Score} - \text{Docking Score}_{\min}}{\text{Docking Score}_{\max} - \text{Docking Score}_{\min}} \end{aligned} \quad (2)$$

The hits were then ranked based on the normalized docking scores from the three scoring functions (or two scoring functions). Finally, the results were combined by averaging the rank of each molecule obtained from the individual scoring function (Eq. 3). The compounds were ranked from best to worst based on their consensus rank.

$$\text{Consensus Rank} = \frac{\text{Vina}_{\text{rank}} + \text{Vinardo}_{\text{rank}} + \text{RF-Score} - \text{VS}_{\text{rank}}}{3} \quad (3)$$

In addition to scores and consensus scoring, an interactive analysis of the docked molecules was also performed to guide further the selection of molecules. Indeed, scoring or ranking the compounds is a weak point in structure-based virtual screening. This interactive analysis took into consideration ligand energy strains (i.e., ligands with favorable binding scores but high internal energy due, for instance, to unfavorable torsion angles). Also, some large molecules eventually making meaningless contacts with the target can have a high score but the removal of such compounds via computational means is difficult (Zhu et al., 2013). Thus we analyzed the docked pose interactively while looking not only at the scores but also at the scores divided by the number of non-hydrogen atoms (Pan et al., 2003) or the scores divided by the molecular complexity index as computed with DataWarrior (Sander et al., 2015). There are many ways of computing diversity

(Mendez-Lucio and Medina-Franco, 2017), here, the approach cut molecules in different types of fragments (each fragment has 7 bonds unless the molecule has less than 14 bonds in total) and counts the number of diverse fragments. This number tends to grow exponentially with the size of the molecule. The fragment count was then normalized by the method by taking the logarithm of the fragment count divided by the size of the compound.

For compounds expected to bind in the exosite cavity, as this region of the protease is more solvent-exposed than the catalytic site pocket and that molecules binding to superficial pockets tend to have different physico-chemical properties than compounds binding to deep cavity (Trisciuzzi et al., 2019), we decided to flag the docked compounds using a random forest classifier built using as a training a collection of solvent-exposed (about 1,000 molecules) and buried (about 1,000 molecules) ligands (Trisciuzzi et al., 2019). These molecules were reported previously and extracted from about 15,000 co-crystal structures downloaded from the PDBbind database (Liu et al., 2015). The random forest classifier reported previously (Trisciuzzi et al., 2019) was applied to all the molecules docked into the TMPRSS2 exosite so as to annotate the molecules as putative "solvent-friendly or superficial" binders (i.e., molecules that have molecular descriptors that make them belong to the class of co-crystallized compounds that we found in more solvent-exposed pockets).

3. Result

3.1. Homology modeling

The serine protease domain is well-conserved and considering a sequence identity of 44% between the serine protease domain of TMPRSS2 and the plasma kallikrein structural template, and the lack of major insertion or deletion, accurate 3D models for TMPRSS2 can be built (Suppl. Fig 1). We decided to keep two best models (Fig. 3a and 3b) with small variations in the orientation of some side chains or in the backbone atoms of some loops in the catalytic site and exosite area so as to have a small conformational ensemble for the subsequent structure-based virtual screening computations. This is usually considered valuable as a way to indirectly take into account receptor flexibility during docking computations. By contrast, using too many many receptor structures (often above 3 structures) tend to increase the noise and limit the chance of finding bioactive compounds (Rueda et al., 2010). The overall stereochemical quality of the two selected models is seen in the Ramachandran plot (suppl. Fig. 2) as most residues are in favorable and allowed regions of the plot. In addition, the root-mean-square deviation (rmsd) between the backbone atoms and the experimental template kallikrein structure is very low (around 0.39Å), indicating further that the amino acids of TMPRSS2 can easily accommodate in the selected template (e.g., the rmsd among experimental structures in the serine protease family is often between 0.5 to 1 Å as the fold is very conserved). The catalytic site region of TMPRSS2 is similar to the one of plasma kallikrein while several aromatic residues appear solvent-exposed in the area of the main exosite (see below). Such regions are generally important for protein-protein interactions. Overall, the structural analysis suggests that the two selected homology models are of good quality and can be used for virtual screening studies.

3.2. Approved, investigational and experimental drug collections

To identify putative small molecules (approved drugs, FDA approved or approved in other countries than the US, investigational or experimental compounds) we used two different collections. The Drugs-lib collection has been described previously and contains about 4600 unique molecules (Lagarde et al., 2018). The protonation state and 3D structures for this collection were predicted using ChemAxon and Corina, respectively. The newly developed collection, AIEfd-Db, contains about 10,000 unique compounds (7864 approved in some

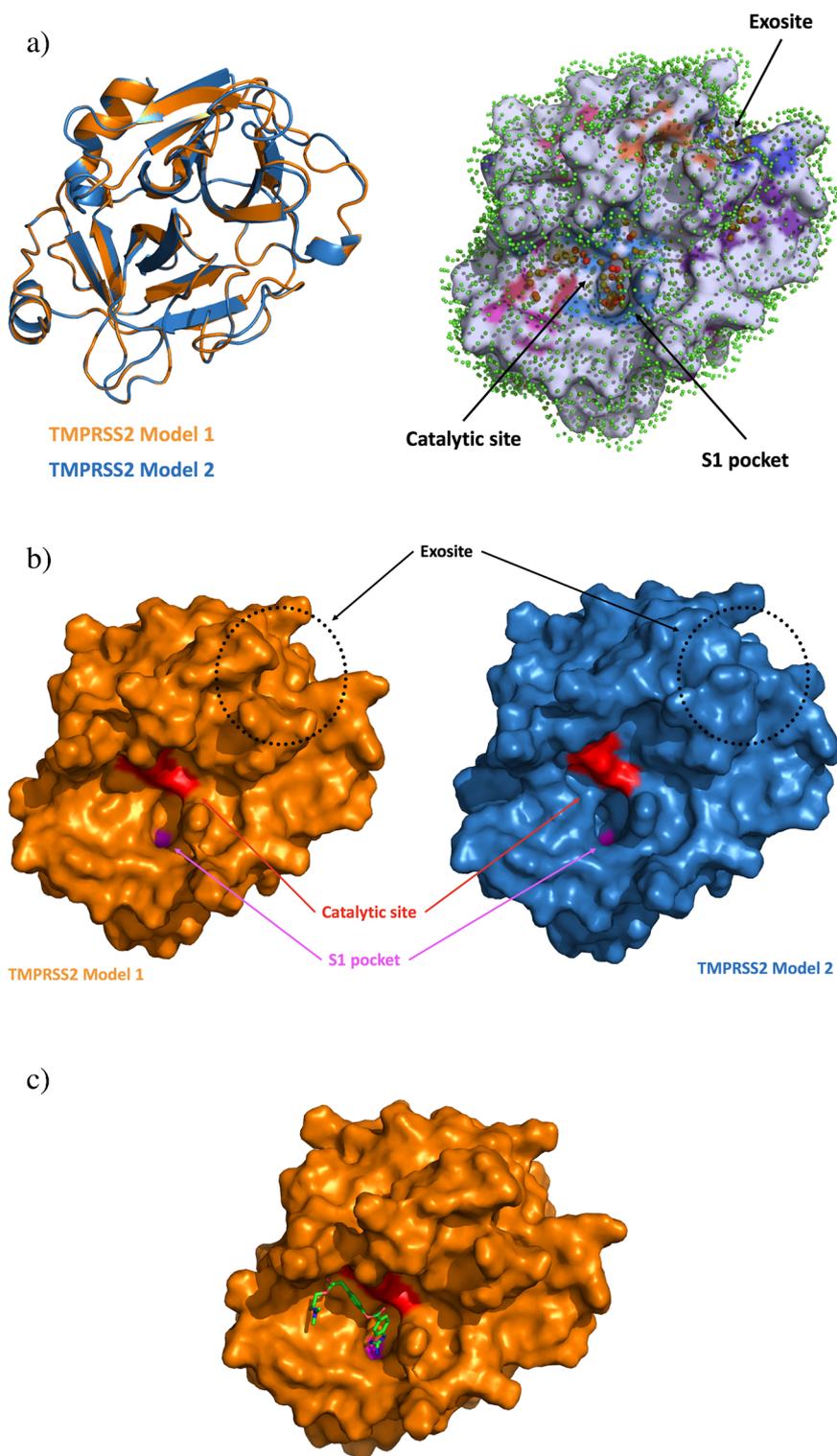


Fig. 3. TMPRSS2 serine protease homology models

Fig. 3a: Two selected models for TMPRSS2. Two best models were selected and are shown with a view down the catalytic site. They differ slightly in some loop regions and in the orientation of some side chains in the catalytic site. These two structures were selected to mimic receptor flexibility during the structure-based screening step. Predictions of binding pockets were performed. A major binding pocket was obviously found and corresponds to the catalytic site. Another critical binding pocket was predicted; it is here simply called exosite. This binding pocket, likely involved in protein-protein interaction, is somewhat similar to thrombin exosite I.

Fig. 3b: Molecular surface of the two structural models. The molecular surface for the two selected homology models of TMPRSS2 is shown with a view down the catalytic site. The exosite is labeled and so are the catalytic site and S1 binding pockets. Small differences in the loops and in the orientation of some side chains are visible in the areas of the exosite and of the catalytic pocket. This should help to explore better the protein during the subsequent docking computations.

Fig. 3c: Camostat docked into the catalytic site of TMPRSS2. Camostat is a known inhibitor of TMPRSS2. It was docked in a proper orientation with both MTiOpenScreen (that performs docking with AutoDock Vina) and Smina. The positively charged group of camostat plugs into the S1 pocket and makes favorable interactions with a conserved Asp residue located at the bottom of the pocket.

countries and investigational, 2073 experimental) and was protonated and generated in 3D with the Surflex tools. As such, by using two different collections (although 4600 molecules are present in both collections) prepared differently (3D generation and predicted protonation states), we expect to identify more putative hit compounds as compared to using only one input collection.

3.3. Structure-based virtual screening

While the catalytic site is an obvious zone to perform virtual

screening computations, we were also interested in screening other regions of the protein. Predictions of alternative binding pockets was performed with P2Rank. Only one key exosite was found, it is located nearby the catalytic site, in a region that is often used in the serine protease family to bind substrates, protein cofactors or inhibitors (Fig. 3a). The volumes of the other predicted sites were small and most likely unable to bind drug-like compounds (i.e., they might bind small fragments). Docking computations were thus performed on the catalytic site and on the key exosite of both TMPRSS2 models with MTiOpenScreen (Drugs-lib collection) and Smina (AIEfd-Db collection). For

each compound, three best poses were initially kept.

In order to investigate our virtual screening results, it was interesting to first analyze independently the scores obtained with the three different scoring functions (Vina scores, Vinardo scores, RFScoreVs) as well as the docked poses obtained with Autodock Vina and Smina on the two models and on the catalytic site and exosite and then investigate consensus scoring and related strategies.

We started by analyzing the results obtained for the catalytic site. For this region, it is easier to analyze the data as two serine protease inhibitors (approved or investigational-experimental), nafamostat (Yamamoto et al., 2016) and camostat (Hoffmann et al., 2020) are known. These compounds can be used to assess the poses and the predicted scores. Camostat mesylate (MW = 399, rotatable bonds = 9), presently in clinical trials for Covid-19 treatment, is well-positioned in the binding groove of our 3D models (Fig. 3c). The compound positively charged group points toward the highly conserved negatively charged Asp at the bottom of the S1 pocket and fill the pocket in a similar manner as numerous small molecules co-crystallized in the catalytic site of serine proteases. Yet, this molecule was ranked at position 1052 using Vina rescoring, at position 361 via rescoring with Vinardo and at position 185 when using RF-Score-VS. Definitely, some larger molecules that are absolutely not compatible with a binding into the TMPRSS2 catalytic site were creating noise. Analysis of nafamostat (MW = 349, rotatable bonds = 5) showed that it fits also very well in the binding pocket (data not shown), it was ranked at position 130 after rescoring with the Vina scoring function implemented in Smina, while it was at position 53 with Vinardo and position 41 with RF-Score-VS.

Consensus scoring can in some situations help selecting likely bioactive molecules but not always (Masters et al., 2020). For camostat and nafamostat, consensus scoring improved the ranking but still, some very unlikely catalytic site binders such as some approved or experimental antibiotics (e.g., apramycin approved for animal use, MW = 540, rotatable bonds = 6) had better scores and could be found in the top 20-100 best compounds depending on the scoring functions or on the consensus scoring results. Clearly, these scores for antibiotics should be due to a scoring artifact. The ranking of these antibiotics or of some other, relatively large, molecules that do not appear after structural analysis to be likely catalytic site binders such as tirilazad (MW = 625, rotatable bonds = 6; an experimental molecule proposed to be used to treat acute ischaemic stroke that worked in animal models but failed in humans (van der Worp et al., 2002)) was problematic and expected (i.e., scoring is a weak point in structure-based screening). Indeed, most of these molecules were well-ranked by each independent scoring function while their chemistry does not seem compatible with the catalytic site. The strategy of dividing the binding scores by the number of non-hydrogen atoms (Pan et al., 2003) to disqualify larger compounds was not efficient in our study as this led to a large number of small molecules, equally unlikely to bind to the catalytic site, ranked among the top 200-300 scores. With this calibration of the scores (dividing by the number of non-hydrogen atoms), the known active compounds, camostat and nafamostat, were then found at position 2399 and 571, respectively, using for instance Vinardo.

However, we observed that some highly unlikely catalytic site binders were not much larger or much more flexible than some serine protease inhibitors, but they had a high computed complexity index. It is known that, for instance, antibiotic molecules are usually chemically complex (Bottcher, 2016). As such, we divided our scores with the molecular complexity index computed by DataWarrior and found that most of the very complex structures, containing fused rings and several sugar moieties, were then ranked toward the bottom of the list while camostat was at position 106 and nafamostat at position 57 (e.g., with Vinardo re-scoring). On the other hand, a molecule like tirilazad which was initially ranked at position 47, for instance with Vinardo, was now found at position 1403 while apramycin, found for instance at position 115 with Vinardo, was then at position 1529.

While we did not reject compounds based on this empirical

calibration of the scores, we used this information during our interactive analysis. Thus, all the docked poses on the two structural models were interactively analyzed in the light of the various scores, consensus scores and calibrated scores by the molecular complexity. Further, internal ligand energy values (e.g., that can reflect high energy strains in the docked ligands and thus unlikely candidate) as computed in Smina were also considered during the interactive analysis. The different investigations led to different lists of compounds that were all merged into a single list, resulting in the selection of 156 compounds that could be inhibitor of the catalytic site of TMPRSS2 (Suppl. Fig. 3 and supplement Table 1). In this list, among the top 20 molecules, 8 molecules are known to inhibit serine proteases.

For the exosite, we could not calibrate the poses or the scores as we have no reference compounds. Yet, consideration of the scores, consensus scoring, ligand internal energies, non-covalent interactions and the results of our random forest classifier expected to flag superficial binders versus compounds that should prefer to bind to buried pockets helped in the rational selection of 100 compounds (Suppl Fig 4 and Suppl Table 2). Some selected molecules were found after our structural analysis to be protein-protein interaction inhibitors (e.g., venetoclax (Yap et al., 2017)) or peptidomimetics (e.g., anamorelin (Currow et al., 2018)) and were also flagged as putative superficial binder with our random forest classifier.

4. Discussion

Numerous strategies can be used to carry out in silico drug re-positioning (Cereto-Massague et al., 2015; Ekins et al., 2019; Farha and Brown, 2019; Klimenko et al., 2016; Oprea and Overington, 2015; Sam and Athri, 2019; Singh et al., 2020; Stumpfe and Bajorath, 2020; Villoutreix et al., 2013). The small molecules identified via such approaches can be assessed experimentally and if valuable be used in clinical trials or to probe physiopathological mechanisms or else used as a starting point to design second generation analogs. With regard to finding small molecules that could block the activity of TMPRSS2, structure-based approaches appear to be well-suited. We used the protocol summarized in Fig. 2 to search for putative approved, investigational or experimental drugs that could interfere with the activity of TMPRSS2. Indeed, as the serine protease domain is highly conserved in 3D, it was possible to build reasonable models for this region of TMPRSS2 (Fig. 3; Suppl Fig. 1 and 2). On these predicted 3D models, the catalytic site was found to be well-conserved as compared to the plasma kallikrein X-ray template while we could predict a binding pocket possibly involved in protein-protein interaction that we name here exosite (Fig. 3). Such a region could have some similarities with the exosite 1 of thrombin in terms of interaction with protein partners and substrates. As the 3D models were built using a template that had a co-crystallized ligand, the situation should be favorable for docking computations in the catalytic site as it is known that homology models can be used for screening and that better results are observed when docking into holo structures as compared to docking in apo conformations (Cavasotto, 2011; Phatak et al., 2009; Rognan, 2017). Docking into well-defined binding pocket allows in general to identify bioactive compounds (Willems et al., 2020) within the top 50-1000 scores, however, docking molecules into more solvent-exposed exosites, is more challenging (Bienstock, 2012; Kruger et al., 2012; Nero et al., 2014; Perot et al., 2010; Trisciuzzi et al., 2019; Villoutreix et al., 2014). This is due to the fact that more exposed binding pockets involved in protein-protein interactions are more flat and are composed of several small cavities and because some plasticity is often present in such regions. To indirectly consider flexibility, we carried out our docking experiments on two different models that vary in the orientation of some side chains in the catalytic site area and of some loops in the main exosite region (Fig. 3).

It is well-known that scoring is a weak point in structure-based virtual screening (Rognan, 2017). Different strategies have been used,

such as rescoring with different types of scoring functions, consensus scoring or using more CPU-demanding free energy computations, but in all cases, it turns out that the approaches are target-dependent and that it is difficult to predict which strategy is going to perform best unless one has numerous true active and inactive compounds to calibrate the methods or to select the best docking-scoring protocols.

Here we used different approaches to select the compounds including rescoring, consensus scoring, consideration of ligand-internal energies, modification of the scores by taking into account the complexity of the molecules and the use of a random forest classifier to try to distinguish molecules that like to bind in more solvent exposed binding pockets. All the data were considered while performing interactive analysis of the docked poses in the two different models with two different methods, namely Autodock Vina as implemented in the MTiOpenScreen server that used the Drugs-lib collection and Smina, a standalone application, that used the AIEfd-Db collection. After extensive investigations, we propose a list of 156 molecules that could bind to the catalytic site of TMPRSS2 (Suppl Fig 3 and Suppl Table 1) and 100 compounds that could fit into the main serine protease domain exosite (Suppl Fig 4 and Suppl Table 2). In fact, considering potential scoring errors due to positioning, lack of appropriate treatment of flexibility or the potential role of water molecules, we believe that it is more appropriate to present a more extended list of compounds than the top 5-10 best scoring molecules. While we have no reference compounds to judge the quality of docking into the exosite region, we are confident that some molecules predicted to inhibit the catalytic site are likely to be true binders. Indeed, during our selection of the molecules, the names and the potential therapeutic activities and intended targets of the molecules were not considered initially but for camostat and nafamostat. Of interest, we could find in the list of putative catalytic site inhibitors several known serine protease inhibitors. Yet, proposing serine protease inhibitors to inhibit a serine protease is not sufficient in our hands to probe the function of TMPRSS2 and this is why many other types of compounds are reported in the Suppl Table 1. Further, in the list of molecules proposed to bind to the TMPRSS2 exosite (Suppl Table 2), some molecules are known protein-protein interaction inhibitors or peptidomimetics. The suggested exosite binders appear to be reasonable starting points to probe this region of TMPRSS2 as it should bind a linear peptide, either from a substrate or a protein partner.

5. Conclusion

In silico drug repositioning results could be of interest to assist the design of a treatment for SARS-CoV-2 infection. We provide two lists of molecules, one for the catalytic site of TMPRSS2 and one for a main binding exosite expected to be part of a protein-protein interaction region. The lists of molecules that we propose contain different types of approved drugs, experimental and investigational compounds, with favorable predicted binding scores and/or interesting non-covalent interactions within the binding site regions. Some of these molecules should be valuable to probe the molecular functions of TMPRSS2 for research purposes, some might be of interest as potential treatments while others could be used as starting points to develop novel molecules, starting from advanced compounds for which numerous information are available (e.g., pharmacokinetics, adverse drug reactions in human or animal models). While at present we can not test experimentally these molecules, we believe that our two lists of molecules can be valuable to other research groups.

Author statement

Project conceived: BV and AMK; Project workflow design: NS and BV. Collection of small molecule: BV; Homology modelling: NS and BV; Virtual screening and structural analysis: NS and BV; Supplementary material: NS and BV; Manuscript draft: AMK, NS and BV; Manuscript

writing: NS, ED, AMK, BV. All authors proofread the final manuscript.

Declaration of Competing Interest

The authors claim that the researchers in this study have no conflict of interest.

Acknowledgment

The Inserm Institute, I-Site Université Lille Nord Europe and Région Hauts-de-France, program STaRS (to BV).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.ejps.2020.105495](https://doi.org/10.1016/j.ejps.2020.105495).

Appendix

Supplementary materials

References

- Ain, Q.U., Aleksandrova, A., Roessler, F.D., Ballester, P.J., 2015. Machine-learning scoring functions to improve structure-based binding affinity prediction and virtual screening. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 5, 405–424.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402.
- Bertram, S., Heurich, A., Lavender, H., Gierer, S., Danisch, S., Perin, P., Lucas, J.M., Nelson, P.S., Pohlmann, S., Soilleux, E.J., 2012. Influenza and SARS-coronavirus activating proteases TMPRSS2 and HAT are expressed at multiple sites in human respiratory and gastrointestinal tracts. *PLoS One* 7, e35876.
- Bhattacharya, A., Wunderlich, Z., Monleon, D., Tejero, R., Montelione, G.T., 2008. Assessing model accuracy using the homology modeling automatically software. *Proteins* 70, 105–118.
- Bienstock, R.J., 2012. Computational drug design targeting protein-protein interactions. *Curr. Pharm. Des.* 18, 1240–1254.
- Botcher, T., 2016. An additive definition of molecular complexity. *J. Chem. Inf. Model.* 56, 462–470.
- Burley, S.K., Berman, H.M., Bhikadiya, C., Bi, C., Chen, L., Di Costanzo, L., Christie, C., Dalenberg, K., Duarte, J.M., Dutta, S., Feng, Z., Ghosh, S., Goodsell, D.S., Green, R.K., Guranovic, V., Guzenko, D., Hudson, B.P., Kalro, T., Liang, Y., Lowe, R., Namkoong, H., Peisach, E., Periskova, I., Prlc, A., Randle, C., Rose, A., Rose, P., Sala, R., Sekharan, M., Shao, C., Tan, L., Tao, Y.P., Valasatava, Y., Voigt, M., Westbrook, J., Woo, J., Yang, H., Young, J., Zhuravleva, M., Zardecki, C., 2019. RCSB protein data bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Res.* 47, D464–D474.
- Carter, W.J., Cama, E., Huntington, J.A., 2005. Crystal structure of thrombin bound to heparin. *J. Biol. Chem.* 280, 2745–2749.
- Cavasotto, C.N., 2011. Homology models in docking and high-throughput docking. *Curr. Top. Med. Chem.* 11, 1528–1534.
- Cereto-Massague, A., Ojeda, M.J., Valls, C., Mulero, M., Pujadas, G., Garcia-Valls, S., 2015. Tools for in silico target fishing. *Methods* 71, 98–103.
- Currow, D.C., Maddocks, M., Cella, D., Muscaritoli, M., 2018. Efficacy of anamorelin, a novel non-peptide ghrelin analogue, in patients with advanced non-small cell lung cancer (NSCLC) and Cachexia-review and expert opinion. *Int. J. Mol. Sci.* 19.
- Douquet, D., 2018. Data sets representative of the structures and experimental properties of FDA-approved drugs. *ACS Med. Chem. Lett.* 9, 204–209.
- Ekins, S., Puhl, A.C., Zorn, K.M., Lane, T.R., Russo, D.P., Klein, J.J., Hickey, A.J., Clark, A.M., 2019. Exploiting machine learning for end-to-end drug discovery and development. *Nat. Mater.* 18, 435–441.
- Eramian, D., Eswar, N., Shen, M.Y., Sali, A., 2008. How well can the accuracy of comparative protein structure models be predicted? *Protein Sci.* 17, 1881–1893.
- Farha, M.A., Brown, E.D., 2019. Drug repurposing for antimicrobial discovery. *Nat. Microbiol.* 4, 565–577.
- Feher, M., 2006. Consensus scoring for protein-ligand interactions. *Drug Discov. Today* 11, 421–428.
- Forli, S., Huey, R., Pique, M.E., Sanner, M.F., Goodsell, D.S., Olson, A.J., 2016. Computational protein-ligand docking and virtual drug screening with the AutoDock suite. *Nat. Protoc.* 11, 905–919.
- Goddard, T.D., Huang, C.C., Meng, E.C., Pettersen, E.F., Couch, G.S., Morris, J.H., Ferrin, T.E., 2018. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.* 27, 14–25.
- Heurich, A., Hofmann-Winkler, H., Gierer, S., Liepold, T., Jahn, O., Pohlmann, S., 2014. TMPRSS2 and ADAM17 cleave ACE2 differentially and only proteolysis by TMPRSS2 augments entry driven by the severe acute respiratory syndrome coronavirus spike

- protein. *J. Virol.* 88, 1293–1307.
- Hoffmann, M., Kleine-Weber, H., Schroeder, S., Kruger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.H., Nitsche, A., Muller, M.A., Drosten, C., Pohlmann, S., 2020. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell*.
- Huntington, J.A., 2014. Natural inhibitors of thrombin. *Thromb. Haemost.* 111, 583–589.
- Iwata-Yoshikawa, N., Okamura, T., Shimizu, Y., Hasegawa, H., Takeda, M., Nagata, N., 2019. TMPRSS2 contributes to virus spread and immunopathology in the airways of murine models after coronavirus infection. *J. Virol.* 93.
- Jain, A.N., Cleves, A.E., Gao, Q., Wang, X., Liu, Y., Sherer, E.C., Reibarkh, M.Y., 2019. Complex macrocycle exploration: parallel, heuristic, and constraint-based conformer generation using ForceGen. *J. Comput. Aided Mol. Des.* 33, 531–558.
- Kawase, M., Shirato, K., van der Hoek, L., Taguchi, F., Matsuyama, S., 2012. Simultaneous treatment of human bronchial epithelial cells with serine and cysteine protease inhibitors prevents severe acute respiratory syndrome coronavirus entry. *J. Virol.* 86, 6537–6545.
- Klimenko, K., Marcou, G., Horvath, D., Varnek, A., 2016. Chemical space mapping and structure-activity analysis of the ChEMBL antiviral compound set. *J. Chem. Inf. Model.* 56, 1438–1454.
- Koes, D.R., Baumgartner, M.P., Camacho, C.J., 2013. Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. *J. Chem. Inf. Model.* 53, 1893–1904.
- Krivak, R., Hoksza, D., 2018. P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. *J. Cheminform.* 10, 39.
- Kruger, D.M., Jessen, G., Gohlke, H., 2012. How good are state-of-the-art docking tools in predicting ligand binding modes in protein-protein interfaces? *J. Chem. Inf. Model.* 52, 2807–2811.
- Lagarde, N., Goldwasser, E., Pencheva, T., Jereva, D., Pajeva, I., Rey, J., Tuffery, P., Villoutreix, B.O., Miteva, M.A., 2019. A free web-based protocol to assist structure-based virtual screening experiments. *Int. J. Mol. Sci.* 20.
- Lagarde, N., Rey, J., Gyulkhandanyan, A., Tuffery, P., Miteva, M.A., Villoutreix, B.O., 2018. Online structure-based screening of purchasable approved drugs and natural compounds: retrospective examples of drug repositioning on cancer targets. *Oncotarget* 9, 32346–32361.
- Lagorce, D., Bouslama, L., Becot, J., Miteva, M.A., Villoutreix, B.O., 2017. FAF-Drugs4: free ADME-tox filtering computations for chemical biology and early stages drug discovery. *Bioinformatics* 33, 3658–3660.
- Laskowski, R.A., Jablonska, J., Pravda, L., Varekova, R.S., Thornton, J.M., 2018. PDBsum: structural summaries of PDB entries. *Protein Sci.* 27, 129–134.
- Laskowski, R.A., Rullmann, J.A., MacArthur, M.W., Kaptein, R., Thornton, J.M., 1996. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* 8, 477–486.
- Liu, Z., Li, Y., Han, L., Li, J., Liu, J., Zhao, Z., Nie, W., Liu, Y., Wang, R., 2015. PDB-wide collection of binding data: current status of the PDBbind database. *Bioinformatics* 31, 405–412.
- Masters, L., Eagon, S., Heying, M., 2020. Evaluation of consensus scoring methods for AutoDock Vina, smina and idock. *J. Mol. Graph. Model.* 96, 107532.
- Mathews, II, Padmanabhan, K.P., Tulinsky, A., Sadler, J.E., 1994. Structure of a non-adecepeptide of the fifth EGF domain of thrombomodulin complexed with thrombin. *Biochemistry* 33, 13547–13552.
- Matsuyama, S., Nagata, N., Shirato, K., Kawase, M., Takeda, M., Taguchi, F., 2010. Efficient activation of the severe acute respiratory syndrome coronavirus spike protein by the transmembrane protease TMPRSS2. *J. Virol.* 84, 12658–12664.
- Mendez, D., Gaulton, A., Bento, A.P., Chambers, J., De Veij, M., Félix, E., Magarinos, M.P., Mosquera, J.F., Mutowo, P., Nowotka, M., Gordillo-Maranon, M., Hunter, F., Junco, L., Mugumbate, G., Rodriguez-Lopez, M., Atkinson, F., Bosc, N., Radoux, C.J., Segura-Cabrera, A., Hersey, A., Leach, A.R., 2019. ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res.* 47, D930–D940.
- Mendez-Lucio, O., Medina-Franco, J.L., 2017. The many roles of molecular complexity in drug discovery. *Drug Discov. Today* 22, 120–126.
- Miteva, M.A., Tuffery, P., Villoutreix, B.O., 2005. PCE: web tools to compute protein continuum electrostatics. *Nucleic Acids Res.* 33, W372–W375.
- Nero, T.L., Morton, C.J., Holien, J.K., Wielens, J., Parker, M.W., 2014. Oncogenic protein interfaces: small molecules, big challenges. *Nat. Rev. Cancer* 14, 248–262.
- Nicola, G., Kufareva, I., Ilatovskiy, A.V., Abagyan, R., 2020. Druggable exosites of the human kino-pocketome. *J. Comput. Aided Mol. Des.* 34, 219–230.
- Novick, P.A., Ortiz, O.F., Poelman, J., Abdulhay, A.Y., Pande, V.S., 2013. SWEETLEAD: an in silico database of approved drugs, regulated chemicals, and herbal isolates for computer-aided drug discovery. *PLoS One* 8, e79568.
- Oprea, T.I., Overington, J.P., 2015. Computational and practical aspects of drug repositioning. *Assay Drug Dev. Technol.* 13, 299–306.
- Pan, Y., Huang, N., Cho, S., MacKerell Jr., A.D., 2003. Consideration of molecular weight during compound selection in virtual target-based database screening. *J. Chem. Inf. Comput. Sci.* 43, 267–272.
- Partridge, J.R., Choy, R.M., Silva-Garcia, A., Yu, C., Li, Z., Sham, H., Metcalf, B., 2019. Structures of full-length plasma kallikrein bound to highly specific inhibitors describe a new mode of targeted inhibition. *J. Struct. Biol.* 206, 170–182.
- Patel, S., 2017. A critical review on serine protease: Key immune manipulator and pathology mediator. *Allergol. Immunopathol. (Madr.)* 45, 579–591.
- Perot, S., Sperandio, O., Miteva, M.A., Camproux, A.C., Villoutreix, B.O., 2010. Druggable pockets and binding site centric chemical space: a paradigm shift in drug discovery. *Drug Discov. Today* 15, 656–667.
- Petersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612.
- Phatak, S.S., Stephan, C.C., Cavasotto, C.N., 2009. High-throughput and in silico screenings in drug discovery. *Expert Opin. Drug Discov.* 4, 947–959.
- Quiroga, R., Villarreal, M.A., 2016. Vinaro: a scoring function based on AutoDock Vina improves scoring, docking, and virtual screening. *PLoS One* 11, e0155183.
- Rognan, D., 2017. The impact of in silico screening in the discovery of novel and safer drug candidates. *Pharmacol. Ther.* 175, 47–66.
- Rozewicki, J., Li, S., Amada, K.M., Standley, D.M., Katoh, K., 2019. MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic Acids Res.* 47, W5–W10.
- Rueda, M., Bottegoni, G., Abagyan, R., 2010. Recipes for the selection of experimental protein conformations for virtual screening. *J. Chem. Inf. Model.* 50, 186–193.
- Sali, A., Blundell, T.L., 1993. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234, 779–815.
- Sam, E., Athri, P., 2019. Web-based drug repurposing tools: a survey. *Brief. Bioinform.* 20, 299–316.
- Sander, T., Freyss, J., von Korff, M., Rufener, C., 2015. DataWarrior: an open-source program for chemistry aware data visualization and analysis. *J. Chem. Inf. Model.* 55, 460–473.
- Simmons, G., Reeves, J.D., Rennekamp, A.J., Amberg, S.M., Piefer, A.J., Bates, P., 2004. Characterization of severe acute respiratory syndrome-associated coronavirus (SARS-CoV) spike glycoprotein-mediated viral entry. *Proc. Natl. Acad. Sci. U. S. A.* 101, 4240–4245.
- Singh, N., Chaput, L., Villoutreix, B.O., 2020. Virtual screening web servers: designing chemical probes and drug candidates in the cyberspace. *Brief Bioinform.*
- Siramshetty, V.B., Eckert, O.A., Gohlke, B.O., Goede, A., Chen, Q., Devarakonda, P., Preissner, S., Preissner, R., 2018. SuperDRUG2: a one stop resource for approved/ marketed drugs. *Nucleic Acids Res.* 46, D1137–D1143.
- Sperandio, O., Miteva, M.A., Segers, K., Nicolaes, G.A., Villoutreix, B.O., 2008. Screening outside the catalytic site: inhibition of macromolecular inter-actions through structure-based virtual ligand screening experiments. *Open Biochem. J.* 2, 29–37.
- Sperandio, O., Wildhagen, K.C., Schrijver, R., Wielders, S., Villoutreix, B.O., Nicolaes, G.A., 2014. Identification of novel small molecule inhibitors of activated protein C. *Thromb. Res.* 133, 1105–1114.
- Stumpfe, D., Bajorath, J., 2020. Current trends, overlooked issues, and unmet challenges in virtual screening. *J. Chem. Inf. Model.*
- Trisciuzzi, D., Nicolotti, O., Miteva, M.A., Villoutreix, B.O., 2019. Analysis of solvent-exposed and buried co-crystallized ligands: a case study to support the design of novel protein-protein interaction inhibitors. *Drug Discov. Today* 24, 551–559.
- Trott, O., Olson, A.J., 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31, 455–461.
- UniProt, C., 2019. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 47, D506–D515.
- Ursu, O., Holmes, J., Bologa, C.G., Yang, J.J., Mathias, S.L., Stathias, V., Nguyen, D.T., Schurer, S., Oprea, T., 2019. DrugCentral 2018: an update. *Nucleic Acids Res.* 47, D963–D970.
- van der Worp, H.B., Kappelle, L.J., Algra, A., Bar, P.R., Orgogozo, J.M., Ringelstein, E.B., Bath, P.M., van Gijn, J., Investigators, T., Investigators, T.I., 2002. The effect of tirilazad mesylate on infarct volume of patients with acute ischaemic stroke. *Neurology* 58, 133–135.
- Villoutreix, B.O., Kuenemann, M.A., Poyet, J.L., Bruzzoni-Giovanelli, H., Labbe, C., Lagorce, D., Sperandio, O., Miteva, M.A., 2014. Drug-like protein-protein interaction modulators: challenges and opportunities for drug discovery and chemical biology. *Mol. Inf.* 33, 414–437.
- Villoutreix, B.O., Lagorce, D., Labbe, C.M., Sperandio, O., Miteva, M.A., 2013. One hundred thousand mouse clicks down the road: selected online resources supporting drug discovery collected over a decade. *Drug Discov. Today* 18, 1081–1089.
- Villoutreix, B.O., Miteva, M.A., 2016. Discoidin domains as emerging therapeutic targets. *Trends Pharmacol. Sci.* 37, 641–659.
- Wang, R., Wang, S., 2001. How does consensus scoring work for virtual library screening? An idealized computer experiment. *J. Chem. Inf. Comput. Sci.* 41, 1422–1426.
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P.,empfer, C., Bordoli, L., Lepore, R., Schwede, T., 2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 46, W296–W303.
- Willems, H., De Cesco, S., Svensson, F., 2020. Computational chemistry on a budget - supporting drug discovery with limited resources. *J. Med. Chem.*
- Wishart, D.S., Feunang, Y.D., Guo, A.C., Lo, E.J., Marcu, A., Grant, J.R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., Maciejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, D., Pong, A., Knox, C., Wilson, M., 2018. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 46, D1074–D1082.
- Wojcikowski, M., Ballester, P.J., Siedlecki, P., 2017. Performance of machine-learning scoring functions in structure-based virtual screening. *Sci. Rep.* 7, 46710.
- Xia, S., Liu, M., Wang, C., Xu, W., Lan, Q., Feng, S., Qi, F., Bao, L., Du, L., Liu, S., Qin, C., Sun, F., Shi, Z., Zhu, Y., Jiang, S., Lu, L., 2020. Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion. *Cell Res.* 30, 343–355.
- Yamamoto, M., Matsuyama, S., Li, X., Takeda, M., Kawaguchi, Y., Inoue, J.I., Matsuda, Z., 2016. Identification of nafamostat as a potent inhibitor of middle east respiratory syndrome coronavirus S protein-mediated membrane fusion using the split-protein

- based cell-cell fusion assay. *Antimicrob. Agents Chemother.* 60, 6532–6539.
- Yap, J.L., Chen, L., Lanning, M.E., Fletcher, S., 2017. Expanding the cancer arsenal with targeted therapies: disarmament of the antiapoptotic Bcl-2 proteins by small molecules. *J. Med. Chem.* 60, 821–838.
- Zhou, P., Yang, X.L., Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B., Huang, C.L., Chen, H.D., Chen, J., Luo, Y., Guo, H., Jiang, R.D., Liu, M.Q., Chen, Y., Shen, X.R., Wang, X., Zheng, X.S., Zhao, K., Chen, Q.J., Deng, F., Liu, L.L., Yan, B., Zhan, F.X., Wang, Y.Y., Xiao, G.F., Shi, Z.L., 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579, 270–273.
- Zhu, T., Cao, S., Su, P.C., Patel, R., Shah, D., Chokshi, H.B., Szukala, R., Johnson, M.E., Hevener, K.E., 2013. Hit identification and optimization in virtual screening: practical recommendations based on a critical literature analysis. *J. Med. Chem.* 56, 6560–6572.