

RESEARCH ARTICLE

Cite this: *RSC Med. Chem.*, 2021, 12, 110

Identification of LASSBio-1945 as an inhibitor of SARS-CoV-2 main protease (M^{PRO}) through *in silico* screening supported by molecular docking and a fragment-based pharmacophore model†

Lucas S. Franco, ^{ab} Rodolfo C. Maia ^{bc} and Eliezer J. Barreiro ^{*abc}

In December 2019, an infectious disease was detected in Wuhan, China, caused by a new pathogenic coronavirus, named SARS-CoV-2. It spread very rapidly, and on March 11th of 2020, the outbreak was declared a pandemic by the World Health Organization. Currently, effective treatment options remain limited. SARS-CoV-2 enzyme main protease (M^{PRO}) plays a pivotal role in the viral life cycle, making it a putative drug target. In order to identify suitable hits to develop inhibitors with adequate antiviral properties, we explored the LASSBio Chemical Library employing multiple strategies of virtual screening. A fragment-based pharmacophore model enabled the identification of key interactions involved in the molecular recognition at the catalytic site of M^{PRO}, namely, with amino acid residues His41, His163 and Glu166. Docking-based virtual screening was performed, leading to the identification of LASSBio-1945 (**9**), a new hit of M^{PRO}, presenting an IC₅₀ = 15.97 μM. This compound, an 1,3-benzodioxolyl sulfonamide, represents an interesting starting point for subsequent hit-to-lead optimization steps and, to the best of our knowledge, a new distinct chemotype for M^{PRO} inhibition.

Received 7th August 2020,
Accepted 11th October 2020

DOI: 10.1039/d0md00282h

rsc.li/medchem

Introduction

In December 2019, an infectious disease (COVID-19) was detected in Wuhan, China, caused by a new pathogenic coronavirus (CoV), named SARS-CoV-2. In the past, two other pathogenic CoVs, the severe acute respiratory syndrome coronavirus (SARS-CoV) and the Middle East respiratory syndrome coronavirus (MERS-CoV), both transmitted from animals to humans, have triggered global epidemics in 2003 and 2012, respectively.¹

The new virus has been named SARS-CoV-2 because the RNA genome (GenBank ID: MN908947.3) is about 82% identical to that of SARS-CoV (GenBank ID: NC_004718.3);² both viruses belong to clade b of the genus *Betacoronavirus*.^{3,4}

SARS-CoV-2 spread very rapidly from China to all countries, and on March 11th of 2020, the outbreak was declared a pandemic by the World Health Organization (WHO).⁵

Current disease management is limited to social measures, such as social distancing, travel ban, and lockdown in many regions. Thus, there is an urgent need for the discovery of prevention and treatment strategies for COVID-19. It is acknowledged that the introduction of a vaccine would happen, in the earliest, at the beginning of 2021. While the discovery of a new drug should take even longer, the drug repositioning approach,⁶ *i.e.* the use of an existing drug to treat COVID-19, seems the fastest strategy since these compounds have either regulatory approval as drugs or have already cleared safety studies.^{7,8}

Although several known drugs are currently under investigation,^{7,9} there is a clear need to identify other therapeutic alternatives within the drug repurposing approach. As illustrated by strategies employed to treat HIV nowadays, such as the highly active antiretroviral therapy (HAART), the long-term management of endemic viruses normally includes the use of multiple antiretroviral drugs, interfering with different stages of the virus life-cycle, in an attempt to control infection and the appearance of resistance.^{10,11} Protease inhibition encompasses one of the strategies employed in HIV multi-therapy, and could be successfully mirrored in the treatment of COVID-19. SARS-

^a Programa de Pós-Graduação em Farmacologia e Química Medicinal, Instituto de Ciências Biomédicas, Universidade Federal do Rio de Janeiro, Avenida Carlos Chagas Filho, 373, Ilha do Fundão, 21941-912, Rio de Janeiro, RJ, Brazil

^b Laboratório de Avaliação e Síntese de Substâncias Bioativas (LASSBio®), <http://www.lassbio.icb.ufjf.br>, Instituto de Ciências Biomédicas, CCS, Universidade Federal do Rio de Janeiro, Cidade Universitária, Rio de Janeiro, RJ, Brazil.

E-mail: ejbarreiro@ccsdecania.ufjf.br

^c Instituto Nacional de Ciência e Tecnologia de Fármacos e Medicamentos (INCT-INOFA; <http://www.inct-inofar.ccs.ufjf.br/>), CCS, Universidade Federal do Rio de Janeiro, Cidade Universitária, Rio de Janeiro, RJ, Brazil

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d0md00282h

CoV-2 presents a main protease (M^{PRO}) that plays a pivotal role in mediating viral replication and transcription, making it an attractive drug target.^{12,13} Additionally, it has been demonstrated that the active site of M^{PRO} remains highly conserved across different CoVs, suggesting the possibility of wide-spectrum anti-CoV drug design.^{14,15}

In fact, some approaches towards the identification of antiviral agents acting as M^{PRO} inhibitors have already been reported.^{15,16} In the context of drug repositioning, a combination of SARS-CoV-2 M^{PRO} structure-based virtual screening and high-throughput screening of more than 10 000 compounds (containing approved drugs, clinical trials, and other active substances) identified six compounds as promising M^{PRO} inhibitors (IC_{50} range: 0.67–21.4 μM). Among these identified inhibitors, disulfiram and carmofur are FDA-approved drugs, whereas ebselen, tideglusib, shikonin, TDZD8 and PX-12 are currently under preclinical studies.¹⁶

In another study, a combined strategy employing structure-based drug design, virtual screening and high-throughput screening identified N3, a Michael acceptor inhibitor of M^{PRO} of SARS-CoV and MERS-CoV, as a potential SARS-CoV-2 M^{PRO} inhibitor. Corroborating the high conservation of the active site of M^{PRO} across different CoVs, N3 was shown to form a covalent bond with and to be an irreversible inhibitor of SARS-CoV-2 M^{PRO} .¹⁵

In the same study, ebselen was identified as a SARS-CoV2 M^{PRO} inhibitor after a high-throughput screening approach. Ultimately, in a plaque-reduction assay with simian Vero cells infected with SARS-CoV-2, N3 and ebselen displayed antiviral and cell protection efficacy at EC_{50} values of 16.77 μM and 4.67 μM , respectively,¹⁵ demonstrating their antiviral potential against SARS-CoV-2.

In this work, we developed a screening strategy to identify M^{PRO} ligands within the LASSBio Chemical Library (LCL),¹⁷ in order to provide a suitable chemical starting point to develop novel inhibitors with adequate antiviral properties. A fragment-based pharmacophore model was created to identify key interactions involved in the molecular recognition at the catalytic site of M^{PRO} by analysing the fragments previously obtained by the Diamond Light Source researchers.¹⁸ Afterwards, molecular docking-based virtual screening on M^{PRO} with the LCL was performed. Next, the information generated by the fragment-based pharmacophore model was employed to assist the analysis of the docking poses, thus refining our final ranking. Additionally, this contribution complements the work done by the Diamond Light Source researchers by focusing the analysis on the non-covalent fragments identified by them and, therefore, it is our hope to collaborate with the worldwide efforts to deliver a strategy that can enhance the success rate of attempts to provide novel therapeutic agents to treat COVID-19.

Results and discussion

Fragment-based pharmacophore model

The fragment-based pharmacophore model (FBPM) was developed based on the fragment-based drug discovery

(FBDD) strategy. The latter is a well-established approach that has been applied to identify fragment-sized hits to provide novel chemical starting points that can be further optimized for the desired molecular target.¹⁹ As fragments have low molecular weights (MWs), they are perfect for probing the most important interactions at the whole active site of a given target, resulting in structural insights about how to increase, fuse or merge these fragments into more potent ligands of higher MW.^{20,21} With this in mind, we figured out that this information could also be used to identify key amino acid residues in the molecular recognition process, allowing the identification of the most important pharmacophore points of interaction. Then, we could use this knowledge to improve our analysis of the poses generated by the docking in the virtual screening stage.

To build the FBPM, we analysed the 22 non-covalent fragments[‡] (detailed in the ESI[†]) co-crystallized with M^{PRO} by the Diamond research group.¹⁸ Considering that the vast majority of LCL compounds were designed as non-covalent ligands, the decision to use only non-covalent fragments was adopted.

Initially, the interactions and crystallographic binding poses of each fragment were individually analysed, aiming to identify if there were patterns of interaction that could highlight specific amino acid residues and/or types of interactions (Table S1, ESI[†]). Curiously, none of the 22 fragments interacts with M^{PRO} through ionic interactions. After mapping out all the interactions, it became clear that residues His41, His163 and Glu166 interacted more frequently than the others. Glu166 was the higher frequency interacting residue, mainly performing hydrogen bonding by means of its backbone and also forming water-mediated interactions, with its side chain carboxylic acid oriented towards the exterior part of the active site (Fig. 1A).

Interestingly, His41 and His163 presented unique and distinct behaviours. His41 was the only residue identified in the active site able to perform π - π stacking interaction with molecular fragments, and it seems to prefer hydrophobic contacts rather than interactions of more polar nature. On the other hand, His163 only interacted through hydrogen bond formation with the fragments, mostly as a hydrogen bond donor. We could also observe that, in most cases, a “convenient” non-classical hydrogen bond with the backbone of the Phe140 residue occurred together with the interaction with His163 (Fig. 1A). This probably happened because of the specific orientation needed by the fragment to interact with His163 which, in turn, projected the electron-deficient hydrogen bonded to the carbon adjacent to the nitrogen to the carbonyl of the Phe140 backbone, favouring this interaction. As mentioned, the interaction with His163 seems to demand a specific orientation inside the M^{PRO} site,

[‡] When this study was performed, there were 22 fragments released. By the time the XRD data of the 23rd fragment was made available, we had the analysis completed.

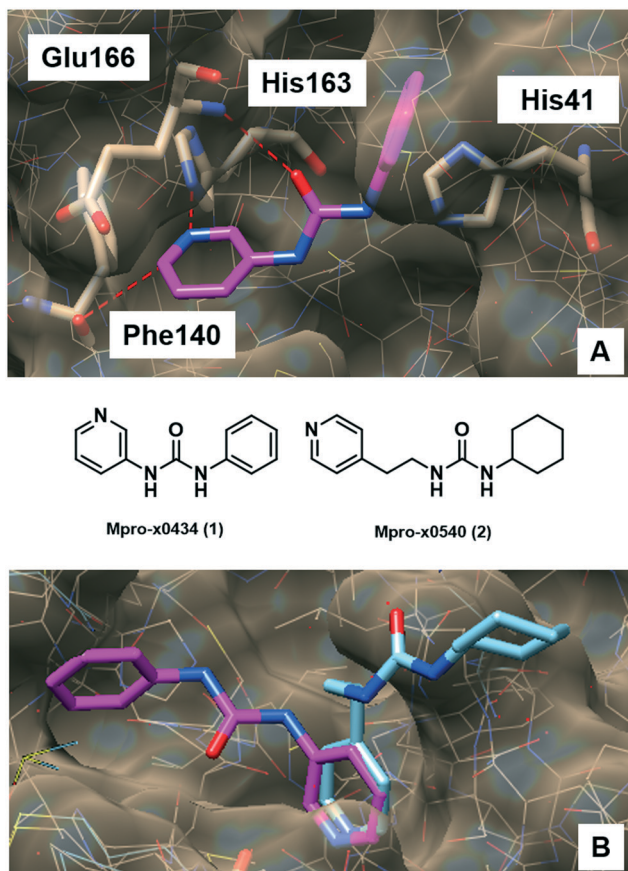


Fig. 1 A) Fragment x0434 (1, magenta) at the M^{PRO} active site and its interactions with His41, Phe140, His163 and Glu166. Hydrogen bonds are shown as red dashed lines. B) Fragments x0434 (1, magenta) and x0540 (2, cyan) superimposed at the M^{PRO} active site. Slight structural changes and differences in the linker size resulted in significantly different binding modes. These images were generated using UCSF Chimera.²²

because even slight structural changes among the fragments that interact with this residue result in different conformational binding orientations (Fig. 1B). Apparently, adopting the conformation that allows for this specific hydrogen bond formation is preferred, indicating that interacting in this fashion with His163 and Phe140 could be a pharmacophoric feature in the molecular recognition process. The way fragment x540 (2) binds corroborates this hypothesis since it orientated its cyclohexyl-urea subunit towards the exterior of the active site in order to adopt the most adequate orientation to allow the pyridine ring to interact properly with His163 instead of trying to occupy other parts of the active site to interact with additional residues at the expense of a not so ideal orientation to interact with His163 (Fig. 1B).

Another relevant observation was that fragments interacting with His41 did not interact with His163 and *vice versa*. The only exception was fragment x0434 (1), which interacted with both His41 and His163 (Fig. 1A). This indicated that there were two subgroups of fragments, and two clear distinct regions inside the active site of M^{PRO} to be

explored during the molecular design. Additionally, this signalled that in order to interact with both His41 and His163 a ligand would need to display an optimal distance between the respective interacting subunits. With this in mind, the fragments were also analysed regarding (i) the presence of subunits able to interact with His41 and His163; (ii) the size of the linker between these subunits, and (iii) the distance between the central atom of the linker and both subunits (Table S2, ESI[†]). This analysis suggested that in order to interact simultaneously with His41 and His163, the fragment needed to have a linker of no more than 3 atoms (approximately 2.50 Å) and the optimal distance between the central atom of the linker and the atom interacting as a hydrogen bond acceptor in the aromatic subunit occupying the His163 site was *ca.* 4.40 Å. The distance of 3.85 Å was the optimal between the central atom of the linker and the centroid of the aromatic subunit interacting with His41, while keeping the adequate orientation to interact with His163 (Table S2, ESI[†]). These findings corroborate why fragments x0434 (1) and x0540 (2) display such diverse binding orientations. Despite the flexible nature of the linker in 2, it is too long (4.46 Å – five atoms) to accommodate both rings inside the active site (Fig. 1B).

All these features resulted in the building of the 3D pharmacophore model based on residues His41, His163 and Glu166 as pharmacophoric points for the establishment of an energetically favoured binding complex (Fig. 2A). By molecular complementarity, a potential ligand would need to present a hydrophobic aromatic subunit to interact with His41 *via* π - π stacking, a second aromatic subunit bearing a

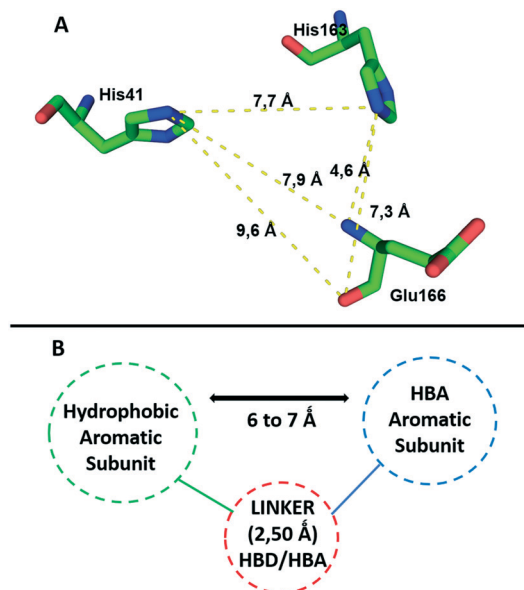


Fig. 2 A) Pharmacophore model based on residues His41, His163 and Glu166 as the key points of interaction required for molecular recognition by M^{PRO}. B) Suggested pharmacophore points, spatial orientation and distances needed for interactions with the key residues by a potential ligand. HBA – hydrogen bond acceptor; HBD – hydrogen bond donor.

hydrogen bond acceptor to interact with His163, and a linker that can interact with Glu166 through H-bonding, and that can also provide a distance in the range of 6–7 Å between the atom interacting as a hydrogen bond acceptor with His163 and the aromatic subunit interacting with His41 (Fig. 2B).

The fragment screening performed by the Diamond Light Source group is part of an international initiative called COVID Moonshot.²³ After the fragments were published, the following actions of the initiative were crowdsourcing molecular designs inspired by these fragments, synthesizing and testing the most promising compounds submitted. Some of these results have already been released in their platform,²⁴ which allowed us to evaluate the hypothesis of our FBPM against some experimental data from non-covalent M^{PRO} inhibitors. These experimental data confirm our suggestion that the His163 residue interacting as a hydrogen bond donor is important for the molecular recognition process and formation of the binding complex. This is illustrated by the activities displayed by compounds TRY-UNI-714a760b-6 (**3**) (IC₅₀ = 24.57 μM) and EDG-MED-0da5ad92-2 (**4**) (IC₅₀ = 53.72 μM). The subtle insertion of a methyl group in position 6 of the 3-pyridine ring was able to promote a 2-fold increase in M^{PRO} inhibition,²⁵ probably due to the induction of a favourable pyridine ring conformation for the interaction with His163 (Fig. 3A). This conformational effect induced by the methyl group is confirmed by other related isomers, where the methyl group was introduced at other positions of the pyridine ring, and all were found to be inactive (Fig. 3B).

The preference of His41 for interacting with more hydrophobic subunits is another feature of the pharmacophore model in agreement with the experimental data released by the Moonshot initiative. It suggests a relationship between the increase in the lipophilic character of the aromatic ring that interacts with His41 and a higher M^{PRO} inhibition (Fig. S2, ESI[†]). It was delightful and reassuring to acknowledge that some of the pharmacophore

model's key interaction hypotheses were supported by the experimental data disclosed so far. The next step of this work consisted of applying the information provided by the pharmacophore model to the analysis of the virtual screening results, which will be better detailed in the discussion of the molecular docking results.

Molecular docking-based virtual screening

The LASSBio Chemical Library (LCL) contains *ca.* 2300 compounds and the library content selection has been driven by medicinal chemistry concepts, with focus on designing compounds with the most adequate lead-like and/or drug-like properties.^{26–28} For instance, approximately 85% of these compounds are compliant with Lipinski's Ro5, and 95% with Veber's rules.^{29,30} Some of the compounds in the LCL have shown *in vivo* activities in one or more animal models, after being administrated orally, which is an indication that they possess overall favourable bioavailability and, hence, adequate pharmacokinetic profiles. From the perspective of a virtual screening campaign, this appropriate profile provides an advantage in accelerating the steps of extensive filtering of huge libraries and, from the perspective of the drug development process, it may provide high quality hits that could progress more effectively into clinical candidates.

Along with rapid development of techniques for determination of biomacromolecules' structures, docking became an important approach of computer-aided drug design.³¹ Although docking is a well-established technique that contributes to the early drug discovery process, one of its flaws is the frequent false positives arising from ranking compounds based on biased scoring functions.^{32,33} Therefore, achieving a successful hit identification rate from a chemical library in a virtual screening campaign is also directly related to the combination of appropriate *in silico* strategies.³⁴ In this context, the integration of both ligand and structure-based methods with visual inspection should be used to help the identification of a subset of compounds from a library to be tested on adequate pharmacological assays.^{35,36} Hence, ligand-based methods such as molecular similarity and substructure searches were implemented to complement molecular docking analysis.³⁷

In this work, the GOLD (Genetic Optimization for Ligand Docking) program of version 5.8.1 (ref. 38) was employed for docking simulations. GOLD has four dimensionless fitness scores named GoldScore,³⁹ ChemScore,⁴⁰ ASP (the Astex Statistical Potential),⁴¹ and ChemPLP.⁴² Even though in each case the scale of the score is different, it is a guide as to how good the pose is, *i.e.* the higher the score, the better the docking result is likely to be. Given the fact that the scoring functions combined had better performance in redocking procedures when compared to each scoring function alone (Table S3, ESI[†]), rank-by-rank consensus scoring was also applied to aid the analysis of docking poses in this study.⁴³

The developed pharmacophore model (Fig. 2) highlighted three residues mainly responsible for molecular recognition,

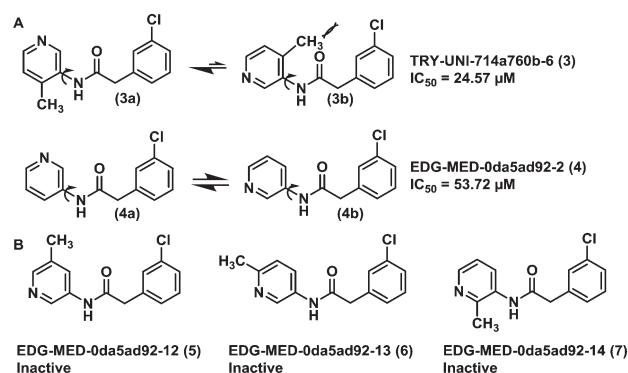


Fig. 3 A) Conformational effect of the methyl group on the pyridine ring orientation of compounds **3** and **4**, demonstrating how the orientation of the nitrogen is key for interaction with His163 and enhances M^{PRO} inhibition. B) Closely related methyl-substituted isomers (**5–7**) that displayed no activity against M^{PRO}, confirming the importance of the conformational effect.

i.e. His41, His163 and Glu166. These amino acid residues are located at two conserved pockets of the M^{PRO} active site, on which His41 forms hydrophobic interactions, His163 acts as a hydrogen bond donor, and Glu166 acts either as a hydrogen bond donor or acceptor. This structural information was used as a criterion to analyse the docking poses of the compounds obtained from the molecular similarity search⁴⁴ using the non-covalent co-crystallized fragments disclosed by the Diamond research group as queries (Scheme S1†). The established similarity coefficient cut-off of 0.4 led to the identification of LASSBio-1962 (**8**)⁴⁵ and LASSBio-1945 (**9**)⁴⁶ (Table 1), which were able to interact with the residues of the pharmacophore model. Additionally, the top 2% ranked LCL compounds were analysed according to the pharmacophore model, and compounds LASSBio-428 (**10**),⁴⁷ and LASSBio-1615 (**11**)⁴⁸ were selected.

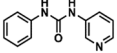
The binding modes of the compounds selected by the molecular similarity search (**8–9**) illustrated the general molecular recognition pattern (Fig. 4) of the identified compounds in Table 1. The compounds occupied the pocket near His163, as depicted by the binding mode of LASSBio-1962 (**8**). In addition, LASSBio-1945 (**9**) was also able to occupy the pocket near His41. LASSBio-1962 (**8**) shares some degree of similarity with x0434 (**1**) and forms a H-bond with His163 using the quinoxaline system. However, probably due to the nature of the bicyclic ring, the distance for H-bonding with this residue is increased for **8** (2.1 Å), when compared to **1** (1.7 Å), which might also reflect the inability of the same linker to adopt a conformation similar to **1** and interact with Glu166 (Fig. 4A). However, the urea linker was able to form a bidentate H-bond with His41, suggesting that it could still bind to the active site. On the other hand, the docking pose of LASSBio-1945 (**9**) (Fig. 4B) suggests that the oxygen atom in the 1,3-benzodioxole ring mimics the hydrogen bond acceptor interaction made by the pyridine nitrogen atom of

1, with the His163 residue. Additionally, the *n*-propyl side chain occupies the so-called S2 hydrophobic pocket, next to His41. X-ray diffraction studies of previously described peptide inhibitors of M^{PRO} show that isobutyl,¹⁵ cyclopropyl,⁴⁹ and cyclohexyl⁵⁰ aliphatic side chains are able to occupy this pocket, supporting the binding mode depicted in Fig. 4B. The hydrogen bonding interaction between Gln189 and the oxygen of the sulphonamide group of LASSBio-1945 (**9**) may also contribute to its molecular recognition.

The binding mode of compounds selected from the top 2% illustrated that the structures which fitted in the His163 and His41 pockets scored higher. LASSBio-428 (**10**) and LASSBio-1615 (**11**) occupied both pockets with aromatic residues, of which **11** interacted with the 2-phenylimidazo[1,2-*a*]pyridine system, and **10** fitted the His163 pocket with its phenyl ring and the His41 pocket with the 3-trifluoromethylaniline subunit. Even though these compounds were able to form H-bonds with amino acid residues in the binding site, none interacted with His163.

Due to the fact that the oxygen atom of the 1,3-benzodioxole moiety acted as a H-bond acceptor, interacting with the pharmacophoric residue His163, as it is depicted in the binding mode of LASSBio-1945 (**9**) (Fig. 4B), we performed further substructure-based searches in the chemical library (Table 2). Firstly, we searched for 1,3-benzodioxole derivatives, leading to the identification of LASSBio-1649 (**13**).⁵¹ Secondly, we defined pyridine and imidazole rings as queries based on pyridine co-crystallized fragments disclosed by Diamond Light Source (*e.g.* **1**, Fig. 1), and on co-crystallized imidazole inhibitors of SARS-CoV M^{PRO}, in which the imidazole ring was able to interact with the conserved histidine residue equivalent to His163.⁵² In an attempt to identify additional chemotypes, a similarity search based on inhibitors of SARS-CoV M^{PRO}, obtained from the

Table 1 Structures, ranking position or similarity coefficient, and nearest neighbour of the identified compounds of the LASSBio Chemical Library through docking-based virtual screening on SARS-CoV-2 M^{PRO}

Compound	Ranking position	Similarity coefficient	Nearest neighbour
 LASSBio-1962 (8)	—	0.52	 Mpro-x0434(1)
 LASSBio-1945 (9)	—	0.42	 Mpro-x0195 (12)
 LASSBio-428 (10)	2	—	—
 LASSBio-1615 (11)	24	—	—

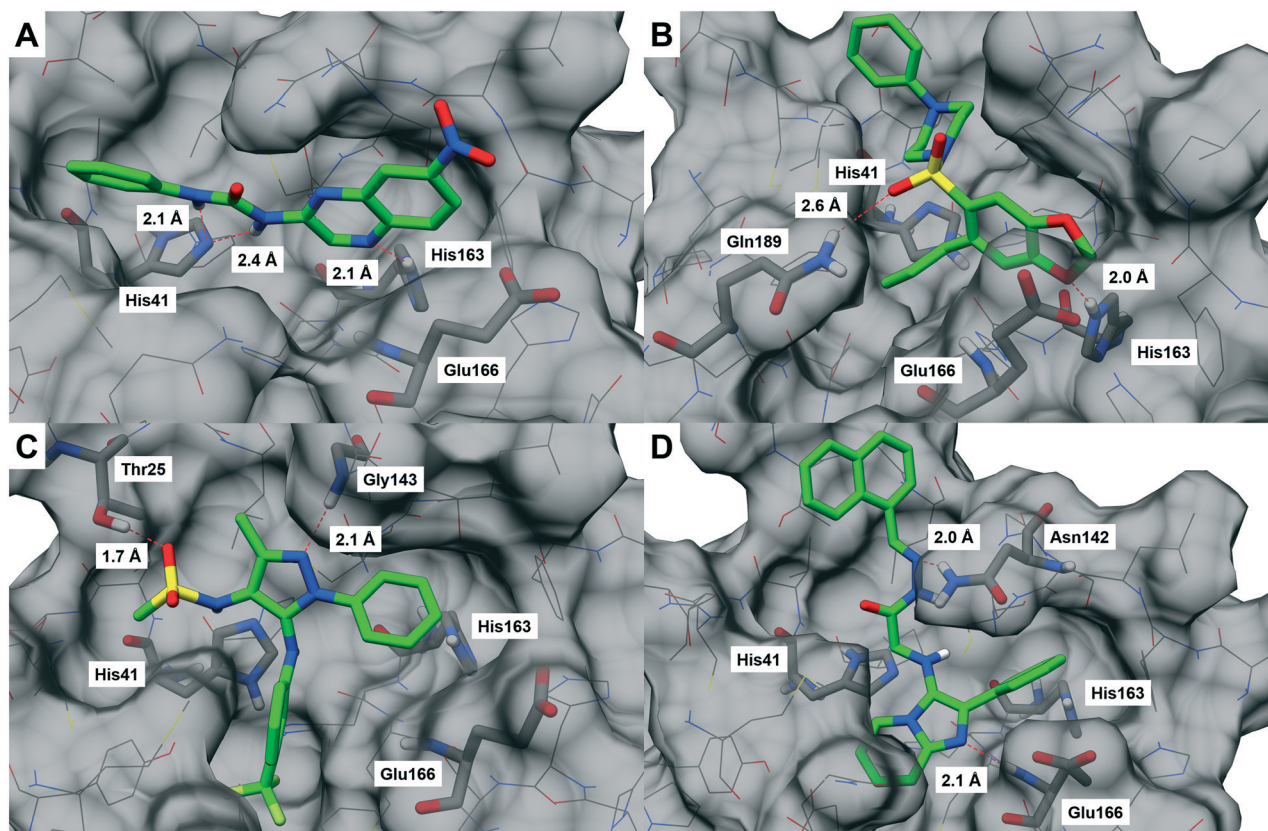


Fig. 4 Docking poses of (A) LASSBio-1962 (**8**),⁴⁵ (B) LASSBio-1945 (**9**),⁴⁶ (C) LASSBio-428 (**10**),⁴⁷ and (D) LASSBio-1615 (**11**).⁴⁸ The structure of SARS-CoV-2 M^{PRO} is shown as lines and surface, with pharmacophore model residues and additional predicted interacting residues shown as sticks (carbon atoms in grey). Docked compounds are shown as sticks (carbon atoms in green). Hydrogen bonds are shown as red dashed lines. These images were generated using UCSF Chimera.²²

Table 2 Structures and query substructure or nearest neighbour of the identified compounds of the LASSBio Chemical Library through docking-based virtual screening on SARS-CoV-2 M^{PRO}

Compound	Query	Compound	Query or nearest neighbour
 LASSBio-1649 (13)		 LASSBio-1600 (15)	
 LASSBio-1891 (14)		 LASSBio-1652 (16)	 (17)

scientific literature and patents available on the Integrity database, was also conducted (Table 2). After visual analysis based on the pharmacophore model (Fig. S3–S6, ESI[†]), LASSBio-1649 (**13**),⁵¹ LASSBio-1891 (**14**),⁵³ and LASSBio-1600 (**15**)⁵⁴ were selected as a result of searches using 1,3-benzodioxole, pyridine and imidazole subunits as queries, respectively. LASSBio-1652 (**16**)⁵¹ was selected from the similarity search which resulted in compound **17** (IC₅₀ = 17 μM; M^{PRO} SARS-CoV)⁵⁵ as the nearest neighbour with a Tanimoto coefficient of 0.51 (Table 2).

All these compounds belong to the class of *N*-acylhydrazones (NAHs), and this is probably due to the nature of the LASSBio Chemical Library. In fact, the NAH functional group is the major chemotype studied in LASSBio⁵⁶ and, consequently, it is present in approximately 50% of the LCL compounds. In the particular example of 1,3-benzodioxole, which was also selected in the similarity search (LASSBio-1652, **16**), it is an outcome of the continuous studies based on the use of safrole, an abundant Brazilian natural product and the principal chemical constituent of sassafras oil (*Ocotea pretiosa*).⁵⁷ In conclusion,

these additional searches allowed the identification of another four compounds (13–16) for investigation of potential M^{PRO} inhibition.

Main protease (M^{PRO}) inhibitory activity

The selected compounds in the virtual screening steps were assayed regarding their M^{PRO} inhibition of enzymatic activity by employing the RapidFire High-Throughput Mass Spectrometry assay.⁵⁸ Initially, these compounds were tested at 50 μM concentration and LASSBio-1945 (9), which presented a percentage of inhibition greater than 50%, had its full dose–response curve determined (Table 3).

As demonstrated in Table 3, this iteration of hit discovery allowed us to identify, among the eight compounds evaluated, compound LASSBio-1945 (9)⁴⁶ as a M^{PRO} inhibitor in the micromolar range (IC₅₀ = 15.97 μM), which means a success rate of 12.5%.

While it was observed that the 1,3-benzodioxole subunit of LASSBio-1945 (9) could mimic the interaction pattern of the 3-amino-pyridinyl subunit of other M^{PRO} inhibitors (Fig. 4B), it was intriguing to note that compounds LASSBio-1649 (13) and LASSBio-1652 (16), also bearing the 1,3-benzodioxole subunit, were inactive. One of the reasons might be the size of the linker of both compounds. The distances between the carbonyl and the imine carbon were 2.9 Å and 3.6 Å for LASSBio-1649 (15) and LASSBio-1652 (18), respectively.

As already discussed in the pharmacophore model section, the linker should not have more than 3 atoms (*ca.* 2.50 Å) in order to be able to interact simultaneously with His41 and His163. Probably, NAH derivatives 13 and 16 behave in the same way as fragment x0540 (2) (Fig. 1B), *i.e.* in order to accommodate the 1,3-benzodioxole moiety to interact with His163, they adopt a binding pose where the aromatic subunit supposed to interact in the His41 pocket is oriented towards another position within the active site, resulting in the loss of that additional, relevant interaction. This size of the linker issue could also be one of the factors contributing to the inactivity of compounds LASSBio-1891 (14) and LASSBio-1600 (15), since they also have the NAH linker, despite 15 being able to interact with His41 *via* hydrophobic

Table 3 IC₅₀ values and percentage of inhibition at 50 μM concentration for M^{PRO} enzymatic activity of the selected compounds in the RapidFire High-Throughput Mass Spectrometry assay

Compound	IC ₅₀ (μM)	Inhibition at 50 μM (%)
LASSBio-428 (10)	N.D.	−10.63
LASSBio-1600 (15)	N.D.	−16.29
LASSBio-1615 (11)	N.D.	7.53
LASSBio-1649 (13)	N.D.	−6.66
LASSBio-1652 (16)	N.D.	−12.83
LASSBio-1891 (14)	N.D.	−6.32
LASSBio-1945 (9)	15.97	68.26
LASSBio-1962 (8)	N.D.	0.77

N.D. – not determined.

interactions and with His163 acting as a hydrogen bond acceptor (Fig. S5, ESI†). The results of LASSBio-1962 (8) support the hypothesis that the interactions with both His163 and His41 are important for activity. Even though the docking pose suggests that 8 was able to interact with His163, LASSBio-1962 (8) was not able to form aromatic interactions with His41. In the cases of LASSBio-428 (10) and LASSBio-1615 (11), the absence of the H-bond interaction with His163 may be the reason for the inactivity.

In summary, these results uncover a potential bioisosteric relationship between the 1,3-benzodioxole moiety present in LASSBio-1945 (9) and the 3-amino-pyridinyl moiety displayed by several fragments, such as 1, TRY-UNI-714a760b-6 (3) (IC₅₀ = 24.57 μM) and EDG-MED-0da5ad92-2 (4) (IC₅₀ = 53.72 μM). The experimental results suggest that the interactions with the residues His41 and His163 of the pharmacophore model are essential for molecular recognition. In addition, the interaction of LASSBio-1945 (9) with Gln189 suggests that perhaps the interaction with Glu166 is not essential if the linker interacts *via* H-bonding with another close residue. The superimposition of LASSBio-1945 (9) with fragment x0434 (1) illustrates how the *n*-propyl chain occupies the pocket next to His41, and a possible non-classical hydrogen bond formed by the methylene group of 1,3-benzodioxole, which might be equivalent to the one pointed out for x0434 (1) (Fig. 5). Moreover, LASSBio-1945 (9) is a M^{PRO} inhibitor with a unique structural pattern and a suitable candidate for subsequent optimization steps through medicinal chemistry-oriented SAR cycles.

Experimental

Fragment-based pharmacophore model

The fragments were analysed using the fragment 3D viewer in the Fragalysis online platform.⁵⁹

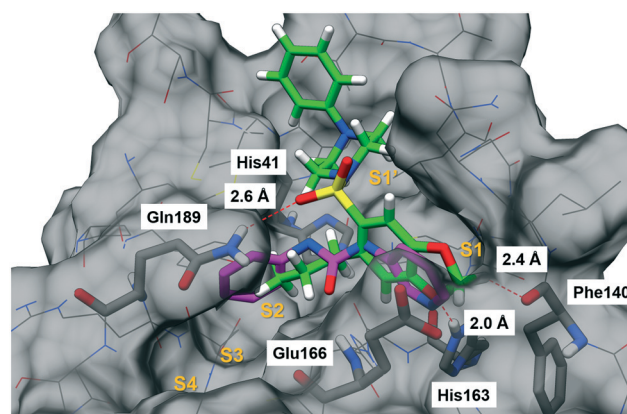


Fig. 5 Superimposition of the docking pose of LASSBio-1945 (9, green, non-polar hydrogen atoms visible)⁴⁶ and x0434 (1, magenta). The structure of SARS-CoV-2 M^{PRO} is shown as lines and surface, with pharmacophore model residues and additional predicted interacting residues shown as sticks (carbon atoms in grey). Putative interactions are shown as red dashed lines. Sub-pockets S1–S4 are indicated in orange. This image was generated using UCSF Chimera.²²

Redocking

The redocking study of the co-crystallized fragments was performed using all GOLD 5.8.1 scoring functions and, in conjunction, the GOLD scoring functions had a better performance than considering each function separately. The docking poses presenting values of root mean square deviation (RMSD) ≤ 1.6 were considered validated (Table S3, ESI†). The RMSD cut-off is the mean value of the crystal structures' resolution.

Virtual screening

The protein structure was prepared using GOLD 5.8.1, by removing water and cofactors, and for protonation with default settings. For the docking procedure, the M^{PRO} PDB structure 5RF7 (chain A) was used. The binding site was defined in the prepared structure, selecting residues within the 10 Å radius of the sulphur atom of Cys145. Visual inspection of the superimposed crystal structures allowed the identification of amino acid residues whose side chains adopt different conformations to allow the molecular recognition of the co-crystallized fragments (Fig. S7, ESI†). Based on this analysis, the side chains of binding site residues His41, Met49, Asn142, Cys145, Met165, Glu166 and Gln189 were set with free flexibility. The GOLD genetic algorithm (GA) was set to virtual screening mode and 10 GA runs were performed for each ligand. The highest scored docking poses (ChemPLP) were analysed using Hermes, and the distances and angles were measured using UCSF Chimera.²²

Molecular similarity and substructure searches

Molecular similarity searches were performed in the KNIME Analytics Platform⁶⁰ using the RDKit⁶¹ and CDK⁶² extension packages. Substructure searches were performed in the ChemInventory⁶³ server.

Main protease (M^{PRO}) inhibition assay

The M^{PRO} inhibition activity was measured employing the RapidFire High-Throughput Mass Spectrometry assay.⁵⁸

Conclusions

In this work, with the aim to identify suitable chemical starting points in the LASSBio Chemical Library to develop inhibitors with adequate antiviral properties, we employed multiple strategies of virtual screening. First, we created a fragment-based pharmacophore model to identify key interactions involved in the molecular recognition at the catalytic site of M^{PRO}. After performing docking-based virtual screening, the combination of pharmacophore-based analysis, substructure and similarity searches, and consensus scoring allowed the selection of eight compounds for the enzymatic assay, which led to the identification of LASSBio-1945 (9) as a M^{PRO} inhibitor (IC₅₀ = 15.97 μM). This hit (9) represents an interesting starting point for subsequent hit-to-lead optimization steps and, to the best of our knowledge, it is a distinct chemical structure pattern

for M^{PRO} inhibition, from what have been published in the literature so far. Further SAR studies and analysis will be carried out with this hit (9) in order to increase its potency and arrive at a promising lead candidate. Additionally, we hope to complement the work done by the Diamond Light Source researchers by providing additional insights into chemical subunits useful for the application of medicinal chemistry strategies, and to collaborate with the worldwide efforts to deliver novel therapeutic agents to treat COVID-19.

Abbreviations

CoV	Coronavirus
FBDD	Fragment-based drug discovery
FBPM	Fragment-based pharmacophore model
FDA	Food and Drug Administration
GOLD	Genetic optimization for ligand docking
HAART	Highly active antiretroviral therapy
HIV	Human immunodeficiency virus
M ^{PRO}	Main protease
MERS-CoV	Middle East respiratory syndrome coronavirus
MW	Molecular weight
SARS-CoV	Severe acute respiratory syndrome coronavirus
WHO	World Health Organization

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors thank the Brazilian funding agencies for the financial support involved in this work: Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq (INCT-INOVAR Grant #465.249/2014-0; Grant #381138/2019-4); Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro – FAPERJ (Grant #E-26/010.000090/2018). The authors also thank the COVID Moonshot initiative (<https://postera.ai/covid>) for performing the main protease (M^{PRO}) inhibition activity assay of the compounds of this work.

Notes and references

- 1 C. I. Paules, H. D. Marston and A. S. Fauci, *JAMA*, 2020, **323**, 707.
- 2 J. S. Morse, T. Lalonde, S. Xu and W. R. Liu, *ChemBioChem*, 2020, **21**, 730–738.
- 3 F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu, Z.-W. Tao, J.-H. Tian, Y.-Y. Pei, M.-L. Yuan, Y.-L. Zhang, F.-H. Dai, Y. Liu, Q.-M. Wang, J.-J. Zheng, L. Xu, E. C. Holmes and Y.-Z. Zhang, *Nature*, 2020, **579**, 265–269.
- 4 P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W. Zhang, H.-R. Si, Y. Zhu, B. Li, C.-L. Huang, H.-D. Chen, J. Chen, Y. Luo, H. Guo, R.-D. Jiang, M.-Q. Liu, Y. Chen, X.-R. Shen, X. Wang, X.-S. Zheng, K. Zhao, Q.-J. Chen, F. Deng, L.-L. Liu, B. Yan, F.-X. Zhan, Y.-Y. Wang, G.-F. Xiao and Z.-L. Shi, *Nature*, 2020, **579**, 270–273.

- 5 World Health Organization, *WHO Coronavirus Disease (COVID-19) Dashboard*, <https://covid19.who.int/> (accessed 3 July 2020).
- 6 T. Pillaiyar, S. Meenakshisundaram, M. Manickam and M. Sankaranarayanan, *Eur. J. Med. Chem.*, 2020, **195**, 112275.
- 7 G. Li and E. De Clercq, *Nat. Rev. Drug Discovery*, 2020, **19**, 149–150.
- 8 D. L. McKee, A. Sternberg, U. Stange, S. Laufer and C. Naujokat, *Pharmacol. Res.*, 2020, **157**, 104859.
- 9 C. Gil, T. Ginex, I. Maestro, V. Nozal, L. Barrado-Gil, M. Á. Cuesta-Geijo, J. Urquiza, D. Ramírez, C. Alonso, N. E. Campillo and A. Martínez, *J. Med. Chem.*, 2020, DOI: 10.1021/acs.jmedchem.0c00606.
- 10 R. D. Moore and R. E. Chaisson, *AIDS*, 1999, **13**, 1933–1942.
- 11 R. W. Eisinger, C. W. Dieffenbach and A. S. Fauci, *JAMA*, 2019, **321**, 451.
- 12 R. Hilgenfeld, *FEBS J.*, 2014, **281**, 4085–4096.
- 13 T. Pillaiyar, M. Manickam, V. Namasivayam, Y. Hayashi and S.-H. Jung, *J. Med. Chem.*, 2016, **59**, 6595–6628.
- 14 F. Wang, C. Chen, W. Tan, K. Yang and H. Yang, *Sci. Rep.*, 2016, **6**, 22677.
- 15 Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, Y. Duan, J. Yu, L. Wang, K. Yang, F. Liu, R. Jiang, X. Yang, T. You, X. Liu, X. Yang, F. Bai, H. Liu, X. Liu, L. W. Guddat, W. Xu, G. Xiao, C. Qin, Z. Shi, H. Jiang, Z. Rao and H. Yang, *Nature*, 2020, **582**, 289–293.
- 16 S. A. Amin and T. Jha, *Eur. J. Med. Chem.*, 2020, **201**, 112559.
- 17 *LASSBio Chemical Library*, <http://www.lassbiochemicallibrary.com>, (accessed 10 September 2020).
- 18 A. Douangamath, D. Fearon, P. Gehrtz, T. Krojer, P. Lukacik, C. D. Owen, E. Resnick, C. Strain-Damerell, A. Aimon, P. Ábrányi-Balogh, J. Brandão-Neto, A. Carbery, G. Davison, A. Dias, T. D. Downes, L. Dunnett, M. Fairhead, J. D. Firth, S. P. Jones, A. Keely, G. M. Keserü, H. F. Klein, M. P. Martin, M. E. M. Noble, P. O'Brien, A. Powell, R. Reddi, R. Skyner, M. Snee, M. J. Waring, C. Wild, N. London, F. von Delft and M. A. Walsh, *Nat. Commun.*, 2020, **11**, 5047.
- 19 P. N. Mortenson, V. Berdini and M. O'Reilly, *Methods Enzymol.*, 2014, **548**, 69–92.
- 20 D. A. Erlanson, S. W. Fesik, R. E. Hubbard, W. Jahnke and H. Jhoti, *Nat. Rev. Drug Discovery*, 2016, **15**, 605–619.
- 21 M. Rachman, D. Bajusz, A. Hetényi, A. Scarpino, B. Meró, A. Egyed, L. Buday, X. Barril and G. M. Keserü, *RSC Med. Chem.*, 2020, **11**, 552–558.
- 22 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, 2004, **25**, 1605–1612.
- 23 J. Chodera, A. A. Lee, N. London and F. von Delft, *Nat. Chem.*, 2020, **12**, 581.
- 24 *Post Era*, COVID Moonshot, <https://postera.ai/covid> (accessed 03 August 2020).
- 25 E. J. Barreiro, A. E. Kümmerle and C. A. M. Fraga, *Chem. Rev.*, 2011, **111**, 5215–5246.
- 26 N. M. Colodette, L. S. Franco, R. C. Maia, H. H. Fokoue, C. M. R. Sant'Anna and E. J. Barreiro, *J. Comput.-Aided Mol. Des.*, 2020, **34**, 1091–1103.
- 27 S. J. Teague, A. M. Davis, P. D. Leeson and T. Oprea, *Angew. Chem., Int. Ed.*, 1999, **38**, 3743–3748.
- 28 A. Ajay, W. P. Walters and M. A. Murcko, *J. Med. Chem.*, 1998, **41**, 3314–3324.
- 29 C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, *Adv. Drug Delivery Rev.*, 1997, **23**, 3–26.
- 30 D. F. Veber, S. R. Johnson, H.-Y. Cheng, B. R. Smith, K. W. Ward and K. D. Kopple, *J. Med. Chem.*, 2002, **45**, 2615–2623.
- 31 D. E. Clark, *Expert Opin. Drug Discovery*, 2008, **3**, 841–851.
- 32 Y.-C. Chen, *Trends Pharmacol. Sci.*, 2015, **36**, 78–95.
- 33 D. B. Kitchen, H. Decornez, J. R. Furr and J. Bajorath, *Nat. Rev. Drug Discovery*, 2004, **3**, 935–949.
- 34 P. Ripphausen, B. Nisius, L. Peltason and J. Bajorath, *J. Med. Chem.*, 2010, **53**, 8461–8467.
- 35 J. G. B. Pedreira, L. S. Franco and E. J. Barreiro, *Curr. Top. Med. Chem.*, 2019, **19**, 1679–1693.
- 36 J. Bajorath, *Nat. Rev. Drug Discovery*, 2002, **1**, 882–894.
- 37 G. Maggiora, M. Vogt, D. Stumpfe and J. Bajorath, *J. Med. Chem.*, 2014, **57**, 3186–3204.
- 38 *Bases de Estruturas Cristalinas – CAPES*, <https://bdec.dotlib.com.br/>, (accessed 30 June 2020).
- 39 G. Jones, P. Willett, R. C. Glen, A. R. Leach and R. Taylor, *J. Mol. Biol.*, 1997, **267**, 727–748.
- 40 M. D. Eldridge, C. W. Murray, T. R. Auton, G. V. Paolini and R. P. Mee, *J. Comput.-Aided Mol. Des.*, 1997, **11**, 425–445.
- 41 W. T. M. Mooij and M. L. Verdonk, *Proteins: Struct., Funct., Bioinf.*, 2005, **61**, 272–287.
- 42 O. Korb, T. Stützle and T. E. Exner, *J. Chem. Inf. Model.*, 2009, **49**, 84–96.
- 43 R. Wang and S. Wang, *J. Chem. Inf. Comput. Sci.*, 2001, **41**, 1422–1426.
- 44 D. J. Rogers and T. T. Tanimoto, *Science*, 1960, **132**, 1115–1118.
- 45 D. N. do Amaral, J. Lategahn, H. H. Fokoue, E. M. B. da Silva, C. M. R. Sant'Anna, D. Rauh, E. J. Barreiro, S. Laufer and L. M. Lima, *Sci. Rep.*, 2019, **9**, 14.
- 46 C. G. e Sá, *PhD Thesis*, Universidade Federal do Rio de Janeiro, Brazil, 2017.
- 47 M. P. Veloso, *PhD Thesis*, Universidade Federal do Rio de Janeiro, Brazil, 2000.
- 48 R. B. Lacerda, N. M. Sales, L. L. da Silva, R. Tesch, A. L. P. Miranda, E. J. Barreiro, P. D. Fernandes and C. A. M. Fraga, *PLoS One*, 2014, **9**, e91660.
- 49 L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox and R. Hilgenfeld, *Science*, 2020, eabb3405.
- 50 W. Dai, B. Zhang, H. Su, J. Li, Y. Zhao, X. Xie, Z. Jin, F. Liu, C. Li, Y. Li, F. Bai, H. Wang, X. Cheng, X. Cen, S. Hu, X. Yang, J. Wang, X. Liu, G. Xiao, H. Jiang, Z. Rao, L.-K. Zhang, Y. Xu, H. Yang and H. Liu, *Science*, 2020, eabb4489.
- 51 T. A. Soares, *MSc. Dissertation*, Universidade Federal do Rio de Janeiro, Brazil, 2011.
- 52 Y. Shimamoto, Y. Hattori, K. Kobayashi, K. Teruya, A. Sanjoh, A. Nakagawa, E. Yamashita and K. Akaji, *Bioorg. Med. Chem.*, 2015, **23**, 876–890.

- 53 S. Herrmann, T. Schübel, F. N. Costa, M. L. C. Barbosa, F. F. Ferreira, T. L. M. F. Dias, M. V. Araújo, M. S. Alexandre-Moreira, L. M. Lima, S. Laufer and E. J. Barreiro, *An. Acad. Bras. Cienc.*, 2018, **90**, 1073–1088.
- 54 T. F. Silva, W. Bispo Júnior, M. S. Alexandre-Moreira, F. N. Costa, C. Monteiro, F. Furlan Ferreira, R. C. R. Barroso, F. Noël, R. T. Sudo, G. Zapata-Sudo, L. M. Lima and E. Barreiro, *Molecules*, 2015, **20**, 3067–3088.
- 55 W. Liu, Z. Lou, Z. Rao, B. Sun, C. Yang, T. Zhou and G. Niu, CN103159666A, 2011.
- 56 S. Thota, D. A. Rodrigues, P. de S. M. Pinheiro, L. M. Lima, C. A. M. Fraga and E. J. Barreiro, *Bioorg. Med. Chem. Lett.*, 2018, **28**, 2797–2806.
- 57 E. J. Barreiro and C. A. M. Fraga, *Quim. Nova*, 1999, **22**, 744–759.
- 58 M. Leveridge, L. Collier, C. Edge, P. Hardwicke, B. Leavens, S. Ratcliffe, M. Rees, L. P. Stasi, A. Nadin and A. D. Reith, *J. Biomol. Screening*, 2016, **21**, 145–155.
- 59 *Fragalysis*, Fragment 3D viewer, <https://fragalysis.diamond.ac.uk/> (accessed 14 April 2020).
- 60 M. R. Berthold, N. Cebren, F. Dill, T. R. Gabriel, T. Kötter, T. Meinel, P. Ohl, C. Sieb, K. Thiel and B. Wiswedel, in *Data Analysis, Machine Learning and Applications*, ed. C. Preisach, H. Burkhardt, L. Schmidt-Thieme and R. Decker, Springer, Albert-Ludwigs-Universität Freiburg, 2008, pp. 319–326.
- 61 *RDKit: Open-source cheminformatics*, <http://www.rdkit.org>, (accessed 3 August 2020).
- 62 C. Steinbeck, Y. Han, S. Kuhn, O. Horlacher, E. Luttmann and E. Willighagen, *J. Chem. Inf. Comput. Sci.*, 2003, **43**, 493–500.
- 63 *ChemInventory*, <https://www.cheminventory.net/>, (accessed 3 August 2020).