



Article

A Computational Approach to Explore the Interaction of Semisynthetic Nitrogenous Heterocyclic Compounds with the SARS-CoV-2 Main Protease

Alejandro Llanes ¹, Héctor Cruz ¹, Viet D. Nguyen ², Oleg V. Larionov ^{2,*} and Patricia L. Fernández ^{1,*}

¹ Centro de Biología Celular y Molecular de Enfermedades, Instituto de Investigaciones Científicas y Servicios de Alta Tecnología (INDICASAT AIP), Ciudad del Saber, Panama 0801, Panama; allanes@indicat.org.pa (A.L.); hcruz@indicat.org.pa (H.C.)

² Department of Chemistry, University of Texas at San Antonio, One UTSA Circle, San Antonio, TX 78249, USA; vietnguyen2710@gmail.com

* Correspondence: oleg.larionov@utsa.edu (O.V.L.); pllanes@indicat.org.pa (P.L.F.); Tel.: +1-(210)-458-6050 (O.V.L.); +(507)-517-0700 (P.L.F.)

Abstract: In the context of the ongoing coronavirus disease 2019 (COVID-19) pandemic, numerous attempts have been made to discover new potential antiviral molecules against its causative agent, SARS-CoV-2, many of which focus on its main protease (M^{Pro}). We hereby used two approaches based on molecular docking simulation to explore the interaction of four libraries of semisynthetic nitrogenous heterocyclic compounds with M^{Pro}. Libraries L1 and L2 contain 52 synthetic derivatives of the natural compound 2-propylquinoline, whereas libraries L3 and L4 contain 65 compounds synthesized using the natural compound physostigmine as a precursor. Validation through redocking suggested that the rigid receptor and flexible receptor approaches used for docking were suitable to model the interaction of this type of compounds with the target protein, although the flexible approach seemed to provide a more realistic representation of interactions within the active site. Using empirical energy score thresholds, we selected 58 compounds from the four libraries with the most favorable energy estimates. Globally, favorable estimates were obtained for molecules with two or more substituents, putatively accommodating in three or more subsites within the M^{Pro} active site. Our results pave the way for further experimental evaluation of the selected compounds as potential antiviral agents against SARS-CoV-2.

Keywords: SARS-CoV-2; main protease; semisynthetic compounds; molecular docking simulation



Citation: Llanes, A.; Cruz, H.; Nguyen, V.D.; Larionov, O.V.; Fernández, P.L. A Computational Approach to Explore the Interaction of Semisynthetic Nitrogenous Heterocyclic Compounds with the SARS-CoV-2 Main Protease.

Biomolecules **2021**, *11*, 18. <https://dx.doi.org/10.3390/biom11010018>

Received: 23 November 2020

Accepted: 23 December 2020

Published: 27 December 2020

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Since the beginning of the coronavirus disease 2019 (COVID-19) pandemic, declared on February 2020, many efforts have been made for the identification of potential antivirals against its causative agent, the severe acute respiratory syndrome (SARS)-related coronavirus called SARS-CoV-2. Several studies have focused mainly on impairing the interaction with the human angiotensin-converting enzyme 2 (ACE2) receptor or on inhibiting several viral proteins. The viral proteins mostly used as drugs targets include the RNA-dependent RNA polymerase (RdRp) and two proteases involved in cleavage of viral polyproteins, known as papain-like protease (PL^{Pro}) and 3-chymotrypsin-like protease (3CL^{Pro}) [1]. The latter, also known as the main protease (M^{Pro}), is essential for viral replication and is widely regarded as the most attractive drug target in SARS-related coronaviruses [2].

The native structure of M^{Pro} is a dimer of two identical subunits, each one composed of a single polypeptide chain of ~300 amino acid residues [3–5]. Each subunit can be divided into three well defined domains (I, II and III), with the active site located between domains I and II. Domains II and III are connected by a relatively large linker loop that also flanks the active site. Two residues, His41 and Cys145, which form a catalytic dyad in

the active site, primarily assist the proteolytic activity of M^{Pro}. The M^{Pro} of SARS-related coronaviruses cleaves the viral polyproteins in several positions using core sequences to recognize the cleavage sites. Residues before and after the cleavage site in the substrate are named, respectively, P1 and P1'. Subsequent residues in both directions follow the same pattern in numbering, i.e., P2, P3, P4 and P5 indicate the four residues preceding P1 before the cleavage site. The active site of M^{Pro} can accommodate at least six amino acid residues of the protein substrate, defining highly specific binding subsites often named S5 to S1' after the corresponding residues.

Since the 2003 SARS outbreak, several molecules have been found to inhibit the M^{Pro} of SARS-CoV and SARS-related coronaviruses (see [2] for a comprehensive review). Characterized inhibitors can be divided into generally covalent peptidomimetic inhibitors and small molecule inhibitors, the latter typically considered as non-covalent or reversible. Among the peptidomimetic inhibitors, the best studied is probably the N3 inhibitor, designed to target the SARS-CoV M^{Pro} [6] and further found to inhibit the M^{Pro} of other coronaviruses, including the infectious bronchitis virus (IBV) [4] and SARS-CoV-2 [5]. The N3 inhibitor has groups accommodating in subsites S5 to S1' and forms a covalent bond with the Cys145 residue, all of which ensure a strong binding to the active site and a high inhibitory effect. Several studies have applied strategies based on drug repurposing [7–12] and the discovery of novel drug candidates [5,13–18] to identify inhibitors of M^{Pro} as potential therapeutic agents against SARS-CoV-2. Although particular attention has been given to covalent peptide inhibitors such as N3, authors have also highlighted nitrogenous heterocyclic compounds as potential non-covalent inhibitors of the M^{Pro} of SARS [2,19,20] and SARS-CoV-2 [21]. Based on these promising results, we hereby used a docking approach to computationally explore the interaction of four libraries of semisynthetic *N*-heterocyclic compounds with the SARS-CoV-2 M^{Pro}.

N-heterocyclic compounds are broadly represented among antiviral agents [22]. We therefore have constructed four libraries of *N*-heterocyclic compounds that were systematically selected after consideration of structural diversity, novelty, synthetic accessibility and druglikeness, the latter inferred on the basis of criteria such as compliance with the Lipinski's rule of five (ROF score ≤ 1) [23]. Two broad classes of *N*-heterocycles were chosen based on reported antiviral activities for some representative members of each class. Quinolines have a long history as antiviral agents, and an assortment of structurally novel analogues of 2-alkylquinoline natural products represented by 2-propylquinoline (library L1) (Table S1) were synthesized following a recently reported procedure [24]. A related subclass of compounds, quinoline *N*-oxides (library L2) (Table S1), have recently emerged as promising and understudied derivatives of *N*-heterocycles with expanding applications in medicinal chemistry [25]. These compounds are readily synthetically accessible by oxidation of the parent *N*-heterocycles [26]. Libraries L3 and L4 (Table S2) are based on hexahydropyrrolo[2,3-*b*]indole (HPI) natural products represented by physostigmine, which have been reported to exhibit an array of antiviral activities [27]. Library L3 comprises analogues with the truncated pyrrolidine ring, while library L4 contains structural analogues with the pyrrolidine ring bioisosterically replaced by the partially saturated 1,2-oxazine. Furthermore, 1,2-oxazine systems are encountered in several classes of biologically active natural products, most notably in the trichoderamide series [28,29]. These structurally novel derivatives of HPI natural products can be readily accessed by an inverse electron demand hetero-Diels-Alder reaction of indoles and transient nitrosoalkenes [30]. Our previous studies have indicated that the truncated and 1,2-oxazine-modified derivatives of HPI natural products possess divergent activities, suggesting that both types of analogues should be evaluated to gain insights into their structure–activity relationship [31,32].

2. Materials and Methods

2.1. Preparation of Receptor and Ligand Structures

We used the 1.83 Å crystal structure of SARS-CoV-2 M^{Pro} with Protein Data Bank (PDB) identifier (ID) 5R84 (Fearon et al., unpublished) as the receptor for molecular docking simulations. The structure of the enzyme was prepared for docking by using the DockPrep module of Chimera v. 14 [33]. This step involved the removal of all crystallographic waters and ions, addition of missing hydrogens and side chains and assignment of appropriate charges and protonation states to ionizable amino acids at physiologic conditions (pH = 7.0).

Ligand structures were drawn with ACD/ChemSketch v.2020.1.1 (ACD/Labs, Toronto, ON, Canada) and the corresponding simplified molecular-input line-entry system (SMILES) notations were imported into Chimera for three-dimensional (3D) structure generation. Chimera was additionally used to prepare the molecules for docking with the DockPrep module, as described above for the receptor, and to optimize the geometry of the molecules through an energy minimization step before docking.

2.2. Molecular Docking Simulations

We used two approaches for molecular docking simulations, DOCK v. 6.9 [34], which was used to perform conventional flexible ligand docking, and Smina v. 2019-10-15 [35], a fork of AutoDock Vina v. 1.2.2 [36], which was used to perform a similar flexible ligand docking but with flexibility added to a subset of key amino acid residues in the active site of the receptor. The co-crystallized ligand in PDB 5R84 was used to define the position of the active site and to set box/grid parameters for both programs. Ten amino acid residues predicted to interact with this co-crystallized ligand were set as flexible when performing docking with Smina (using the *-flexres* option of the program), namely Phe140, Asn142, Gly143, Cys145, His164, Glu166, Asn187 and Gln189. Docking energy scores were calculated with the Hawkins Generalized Born/Surface Area (GBSA) scoring function for DOCK and with the custom scoring function for Smina.

3. Results and Discussion

3.1. Validation of Molecular Docking Parameters and Criteria for Selection

In the PDB structure used here as a receptor for molecular docking simulations, the SARS-CoV-2 M^{Pro} was crystallized in complex with ligand 2-cyclohexyl-*N*-(3-pyridyl) acetamide, here called GWS according to its PDB ligand identifier. This molecule is a nitrogenous heterocyclic compound similar in size and chemical nature to many of the molecules considered in this study and therefore, it allowed us to define the position of the active site for docking (Figure 1A). Within the active site of M^{Pro}, GWS extends linearly across the S1 and S2 subsites and forms two hydrogen bonds with residues His163 and Glu166 (Figure 1B; Figure S1A). The molecule also forms several non-covalent interactions with other residues from the active site, including His41, Met49, Phe140, Cys145, His164, Met165, Asn187 and Gln189, the latter two located in the linker loop of domains II and III.

In order to validate our docking parameters, we performed the redocking of the co-crystallized ligand with both docking programs before conducting the actual simulations with the target compounds. Poses with the lowest energy scores obtained through conventional rigid receptor docking with DOCK (−30.3 kcal/mol) and Smina (−6.3 kcal/mol) were relatively similar, with values of root mean square deviation (RMSD) of 2.137 and 1.22 Å, respectively. However, both poses were predicted to form only one of the two hydrogen bonds that GWS forms with residues from the active site. In an attempt to reproduce a more realistic scenario for the docking simulations, we reran Smina with similar parameters but adding flexibility to ten amino acid residues predicted to interact with GWS in the original PDB structure (see Section 2.2), which resulted in a lower, more favorable, energy estimate (−9.7 kcal/mol). The pose obtained with this flexible docking approach was more similar to the original orientation of the compound in the crystal (RMSD of 0.685 Å) and was predicted to form both hydrogen bonds with residues His163 and Glu166, as well as

other non-covalent interactions with other residues in the active site (Figure 1; Figure S1B). Since GWS is a nitrogenous heterocyclic compound, these results suggest that the docking parameters used in these redocking steps are suitable for our target compounds. For further analysis of molecular interaction of these target compounds and M^{pro} , we prioritized the poses obtained with the flexible docking approach with Smina.

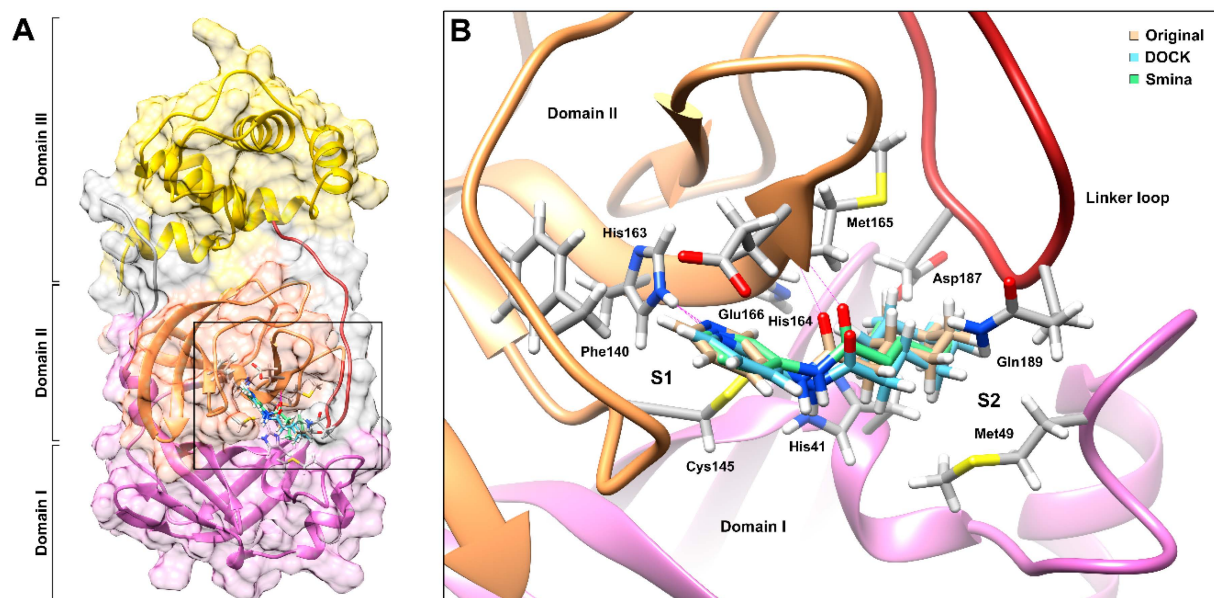


Figure 1. Redocking of the ligand co-crystallized with chain A of SARS-CoV-2 main protease (M^{pro}) in Protein Data Bank (PDB) 5R84. (A) Location of the active site relative to the domains I, II and III of M^{pro} , respectively, colored purple, orange and yellow. Linker loop connecting domains II and III is colored dark red. (B) Comparison of docking poses obtained with DOCK and Smina (flexible docking approach) with the original orientation of the co-crystallized ligand. Hydrogen bonds with residues His163 and Glu166 are represented with thin purple lines.

After performing docking simulations for the target compounds with the selected parameters, all molecules could fit within the M^{pro} active site, although individual poses varied significantly among members of the four libraries. Molecular docking simulations resulted in DOCK's GBSA energy scores of -21.2 to -39.4 kcal/mol (Figure 2A) and Smina affinity scores of -6.7 to -13.4 kcal/mol (Figure 2B). Predicted energy estimates and the lowest energy poses for all the compounds and their natural precursors are shown in Table S3. The difference in scale between these two types of energy scores is a result of the two programs using different algorithms and therefore these estimates cannot be directly compared. Although the word affinity is used in the definition of Vina and Smina energy scores, these values are not related to the affinity of the molecules to the receptor, such as the dissociation (K_d) or inhibition (K_i) constants. Taking into account the inherent differences between the two types of energy scores, we subsequently treated them independently and used the values of the lower quartiles as soft cut-offs to further select the compounds with the most favorable energy estimates (Figure 2C). These values should not be considered as thresholds to predict the binding of compounds to the target protein. These cut-off values were used to prioritize compounds for further analysis of their molecular interactions within the M^{pro} active site.

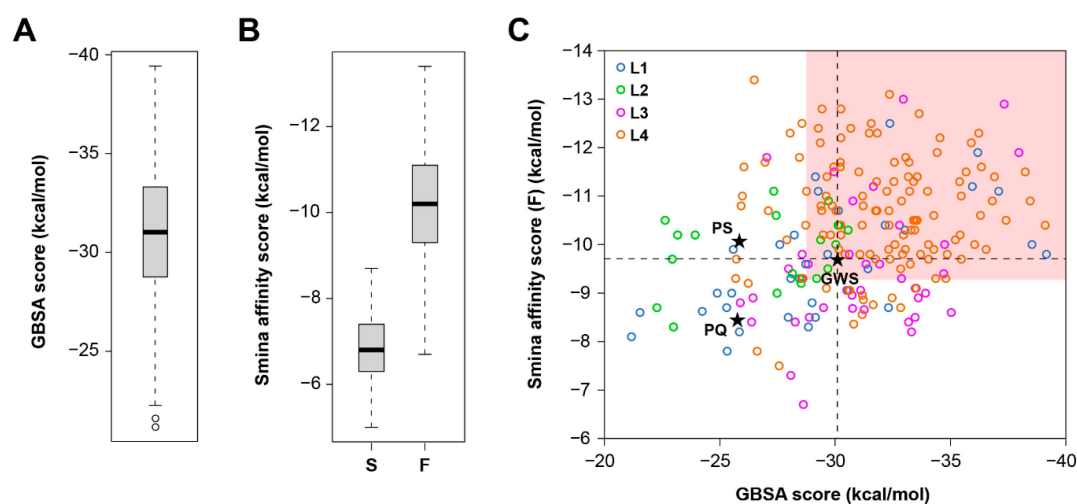


Figure 2. Values of docking energy estimates obtained for the target compounds. (A) Box-and-whisker plot for the DOCK energy scores. (B) Box-and-whisker plots for the Smina affinity scores obtained with standard (S) rigid receptor docking and the flexible receptor (F) approach. (C) Distribution of values among the four libraries of compounds evaluated (L1, L2, L3 and L4), with compounds meeting the selection criteria clustering in the area highlighted in pink. Energy scores obtained for the co-crystallized ligand (GWS) are indicated with dashed lines. PQ: 2-propylquinoline, PS: physostigmine.

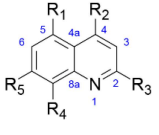
3.2. Quinoline and Quinoline N-Oxide Derivatives

We evaluated 52 compounds belonging to libraries L1 and L2 including quinoline (compounds 1–25) and quinoline *N*-oxide (compounds 26–52) derivatives, respectively. Molecular docking simulations resulted in 11 compounds from each of the libraries L1 (Figure 3A) and L2 (Figure 3B) meeting the criteria defined for selection. Energy scores computed for the 2-propylquinoline natural precursor, -8.5 kcal/mol (DOCK) and -25.8 kcal/mol (Smina), were less favorable than those estimated for the vast majority of compounds in these two libraries. In the binding poses predicted for this precursor, which were similar for both programs, the molecule was oriented so that the quinoline nucleus accommodates between the S3 and S4 subsites and the propyl chain protrudes towards the S2 site, leaving the S1 and S1' subsites empty. Binding poses for compounds 1 and 2 were similar to that of the precursor, but the presence of an additional phenyl group in compounds 3–7 appears to promote a more favorable orientation of the quinoline nucleus in the active site, with the best energy scores estimated for compounds 4, 5 and 6. A similar orientation was predicted for compounds 9 and 15, which have an additional cyclopentane ring at position C2 of the quinoline nucleus with one or two additional Cl substituents; as well as for compound 17, which is similar to compound 15 but has a 4-chlorophenyl group at C2 instead of the cyclopentane ring.

Compounds 20 and 21 have a 4-fluorophenyl group linked through an ether bond to C4 that accommodates around the S1 and S2 subsites, possibly contributing to more favorable energy estimates, but still leaving the area around the catalytic dyad and the S1' subsite empty. Similar orientations were predicted for compounds 23, 24 and 25, which also have relatively large substituents attached to C4 through a secondary amine bond and containing a tertiary amine moiety. The only compound of library L1 that was predicted to fill the active site in a way resembling that of inhibitor N3 was compound 22, which is also the only compound in the library with two stereoisomers (designated 22_R and 22_S). Although energy scores estimated for this compound were not the lowest in the library, this is the most attractive candidate for experimental evaluation among the quinoline derivatives, since both stereoisomers were concurrently predicted to fill all the five subsites in the active site (Figure 4). Both stereoisomers were predicted to form a hydrogen bond with His164, a residue that is sterically close to the His41 and Cys145 residues forming the catalytic dyad. The 22_R stereoisomer was also predicted to form an additional hydrogen

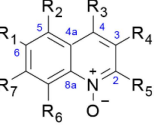
bond with residue Asn142. Compounds 22–25 were constructed to mimic antimalarial drugs chloroquine (22) and amodiaquine (compounds 23–25), both of which were recently shown to have notable antiviral activities against SARS-CoV-2 [37–39]. However, clinical effectiveness of hydroxychloroquine in the treatment of COVID-19 patients has been recently questioned [40].

A



	R1	R2	R3	R4	R5	DOCK (kcal/mol)	Smina (kcal/mol)
4	-			-	-	-28.2	-10.2
5	-			-	-	-27.6	-10.0
6	-			-	-	-30.1	-10.7
15		-				-31.4	-9.5
17		-				-29.1	-11.4
20				-		-33.0	-10.3
21				-		-32.4	-12.5
22	-			-		-39.1 (S)	-10.0 (R)
23	-			-		-37.1	-11.1
24	-			-		-36.2	-11.9
25	-			-		-35.9	-11.2

B



	R1	R2	R3	R4	R5	R6	R7	DOCK (kcal/mol)	Smina (kcal/mol)
27	-	-		-		-		-32.2	-10.4
30	-			-	-	-		-29.3	-11.1
34	-	-	-	-	-		-	-29.4	-10.1
40		-	-	-	-		-	-30.0	-10.5
42		-	-	-	-		-	-29.2	-9.3
43	-	-	-		-		-	-29.7	-9.5
44	-	-	-		-		-	-30.2	-10.4
48	-	-	-	-			-	-29.7	-10.9
49	-	-	-	-	-		-	-30.2	-10.4
50		-	-	-	-		-	-30.1	-10.4
52		-	-	-	-		-	-30.6	-10.3

Figure 3. Compounds from libraries L1 (A) and L2 (B) with the most favorable energy scores. The compounds were selected according to the cut-off values defined in Section 3.1. For compound 22, position of the stereocenter is indicated with “*” and the most favorable, lowest, energy score between the two stereoisomers is shown for each program.

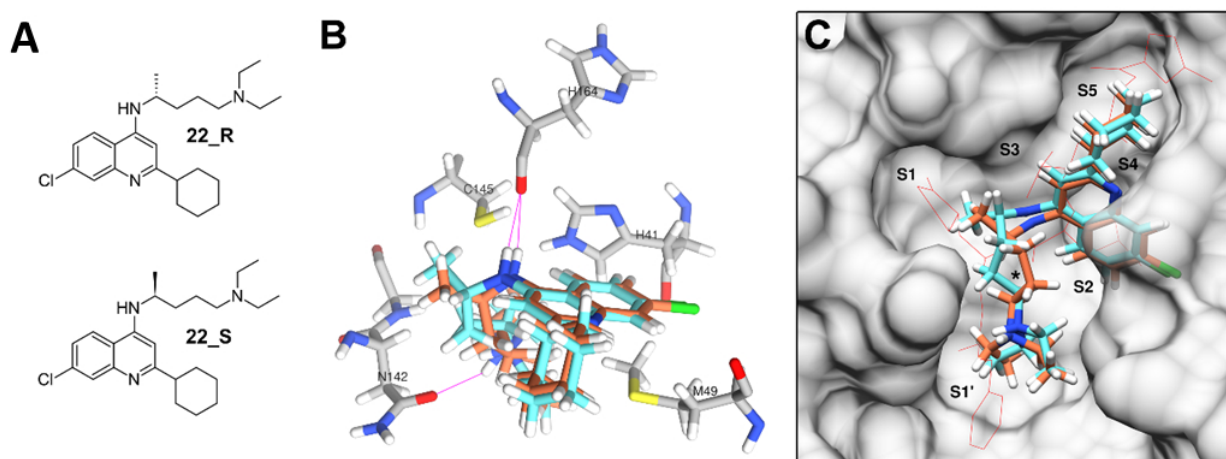


Figure 4. Docking poses predicted for compound 22. (A) Structure of stereoisomers 22_R and 22_S. (B) Orientation predicted for stereoisomers 22_R (cyan) and 22_S (orange) in the active site of MPTO. Hydrogen bonds are represented by violet lines. (C) Extent of the orientation predicted for both stereoisomers in the active site, compared with the orientation of the N3 inhibitor (red wire). Approximate position of the His41-Cys145 catalytic dyad is indicated with “*”.

Globally, quinoline *N*-oxide derivatives from library L2 received relatively lower energy scores when compared to the quinoline derivatives in library L1. These compounds tended to accommodate in the relatively large cavity formed in the active site around subsites S2, S3, S4 and S5, in some cases forming a hydrogen bond between the charged oxygen atom at position N1 of the quinoline nucleus and the Gln189 residue in the linker loop. Of the compounds selected from this family, only compounds 48, 49, 50 and 52, containing trifluoromethyl groups as substituents, were concurrently predicted by both programs to cover a wider region of the active site, involving the S1 and S1' subsites. Docking poses for compounds 48 and 49 are very similar to the orientation of GWS in the crystal, with the molecules extending almost linearly along the S1 and S2 subsites and, in the case of compound 48, forming a hydrogen bond with residue Glu166. In compounds 50 and 52, the relatively bulky phenyl ring with two trifluoromethyl substituents was predicted to accommodate around the S1' and S4 subsites, respectively, causing the quinoline nucleus to position very close to residue Cys145.

3.3. Derivatives of Hexahydropyrrolo[2,3-*b*]indole (HPI) Natural Products

We also evaluated 65 compounds from two additional libraries of HPI derivatives, L3 and L4, containing the truncated pyrrolidine ring (compounds 53–90) and the tetrahydro-1,2-oxazino indole (TOI) nucleus (compounds 91–117), respectively. Molecular docking simulations resulted in 14 compounds from library L3 (Figure 5) and 22 compounds from library L4 (Figure 6) meeting the criteria for selection. The energy score estimated with DOCK for physostigmine (−26.5 kcal/mol), which is the natural precursor of these libraries, was less favorable than those computed for most of the compounds in both libraries. However, the Smina score computed for this precursor (−10.1 kcal/mol) was more favorable than that computed by DOCK, when referring to the individual thresholds selected for both programs. In the lowest-energy poses predicted by these two programs, which were also roughly consistent, the molecule extends across the S1 and S2 sites.

Binding poses predicted for several compounds belonging to library L3 followed the same orientation predicted for physostigmine, although the presence of one or two additional hydrophobic and bulky substituents at positions N1 or C3 of the indole nucleus typically resulted in lower, more favorable, energy estimates. For instance, compounds 53 and 69, which, respectively, have a methyl and an isopropyl group in position N1, received relatively unfavorable scores. However, compound 63, which differs from compound 53 in the presence of a 2,5-dimethylbenzyl group in N1, received more favorable energy scores. In the binding poses for most of these compounds, the additional bulky substituents

appear to accommodate around the S2/S4 subsites or the S1/S1' subsites. No molecule from this library was concurrently predicted by both programs to occupy more than three subsites in the active site. Additionally, due to the general lack of donor or acceptor groups in compounds from this library, almost none of them were predicted to form hydrogen bonds with residues from the active site. However, some of the energy scores computed for these molecules were similar or better than those estimated for compounds from the other libraries forming one or more hydrogen bonds. This suggests that other non-covalent interactions, including hydrophobic interactions, may have the largest contribution to energy scores estimated for compounds from this library, particularly in those with highly hydrophobic or aromatic substituents. Compound 64 was the only molecule in the library that was predicted to form a hydrogen bond with a residue from the active site (Gln189), in this case due to the presence of a carbonyl group at C3.

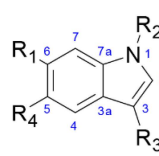






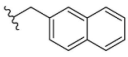

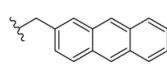













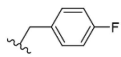











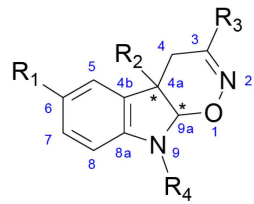
	R1	R2	R3	R4	DOCK (kcal/mol)	Smina (kcal/mol)
						
62	-			-	-31.9	-9.6
63	-			-	-32.9	-9.3
64	-			-	-31.3	-6.6
65	-			-	-31.6	-11.2
66	-			-	-32.9	-13.0
68	-			-	-29.9	-11.5
71	-			-	-30.6	-9.8
75				-	-28.8	-9.6
76				-	-34.7	-10.0
77				-	-30.8	-10.9
78	-			-	-32.8	-10.4
84	-				-37.3	-12.9
87	-				-38.0	-11.9
90	-				-34.7	-9.4

Figure 5. Compounds from L3 library with the most favorable energy scores. The compounds were selected according to the cut-off values defined in Section 3.1.















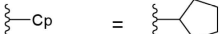




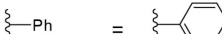









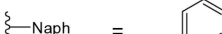










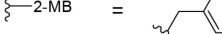







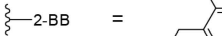




































	R1	R2	R3	R4	DOCK (kcal/mol)	Smina (kcal/mol)
	-				-34.4 (RS)	-11.3 (SS)
	-				-34.0 (RR)	-10.8 (SS)
	-				-37.4 (SR)	-10.5 (SR)
	-				-31.2 (SS)	-12.8 (RR)
 = 					-32.5 (RS)	-13.4 (SR)
 = 					-31.8 (RR)	-12.3 (RR)
 = 					-33.2 (SR)	-11.1 (SR)
 = 					-31.7 (SS)	-12.5 (RR)
					-34.0 (SR)	-11.7 (SS)
					-35.9 (RS)	-12.7 (SR)
 = 					-36.2 (RS)	-13.1 (SR)
					-34.3 (RS)	-11.8 (SS)
 = 					-32.6 (SR)	-10.4 (RS)
					-36.3 (SR)	-10.6 (SR)
					-39.1 (SR)	-10.4 (SR)
					-33.0 (SS)	-12.4 (SR)
					-33.2 (RS)	-12.4 (SR)
					-33.2 (RS)	-12.4 (SR)
	-				-33.4 (SR)	-11.7 (SS)
	-				-36.0 (SS)	-11.1 (RS)
	-				-36.5 (SR)	-10.8 (RS)
					-33.4 (SS)	-12.3 (SR)

Figure 6. Compounds from L4 library with the most favorable energy scores. The compounds were selected according to the cut-off values defined in Section 3.1. All the selected compounds have at least one stereoisomer meeting the selection criteria. Position of stereocenters in the tetrahydro-1,2-oxazino indole (TOI) nucleus is indicated with “*”. Only the value for the stereoisomer with the most favorable, lowest, energy score is shown for each compound and program. Cp: cyclopentyl; Ph: phenyl; Bn: benzyl; Naph: naphthyl; 2-MB: 2-methylbenzyl; 2-BB: 2-bromobenzyl; 2,5-DMP: 2,5-dimethylphenyl.

On the other hand, the compounds containing the tetrahydro-1,2-oxazino indole (TOI) nucleus from library L4 received the most favorable energy scores among all the four libraries examined here. Unlike those from all other libraries, compounds from this library have four stereoisomers due to the presence of two stereocenters, located at positions 4a and 9a of the TOI nucleus. Here, the four stereoisomers were designated with suffixes “_RR” (4aR,9aR), “_RS” (4aR,9aS), “_SR” (4aS,9aR) and “_SS” (4aS,9aS). Each of these stereoisomers was submitted separately to the docking simulations. Poses predicted for the four stereoisomers were notably different for almost all compounds in the library. However, stereoisomers with the most favorable energy scores in the library, including 97_RR, 104_RR, and 110_SR, tended to adopt a similar orientation within the active site, despite having substituents of different chemical nature (Figure 7). In this conserved orientation, the TOI

nucleus occupied a relatively central position in the active site, very close the catalytic dyad, while the substituents in positions 3, 4a and 9 accommodate in the S4, S2 and S1 subsites, respectively. This orientation within the active site also appears to be favored by the formation of a hydrogen bond between the oxygen atom of the oxazine ring of the TOI nucleus and residue Glu166. All these findings suggest that these three stereoisomers and/or the corresponding racemates are the most attractive candidates for experimental evaluation among the compounds in this library.

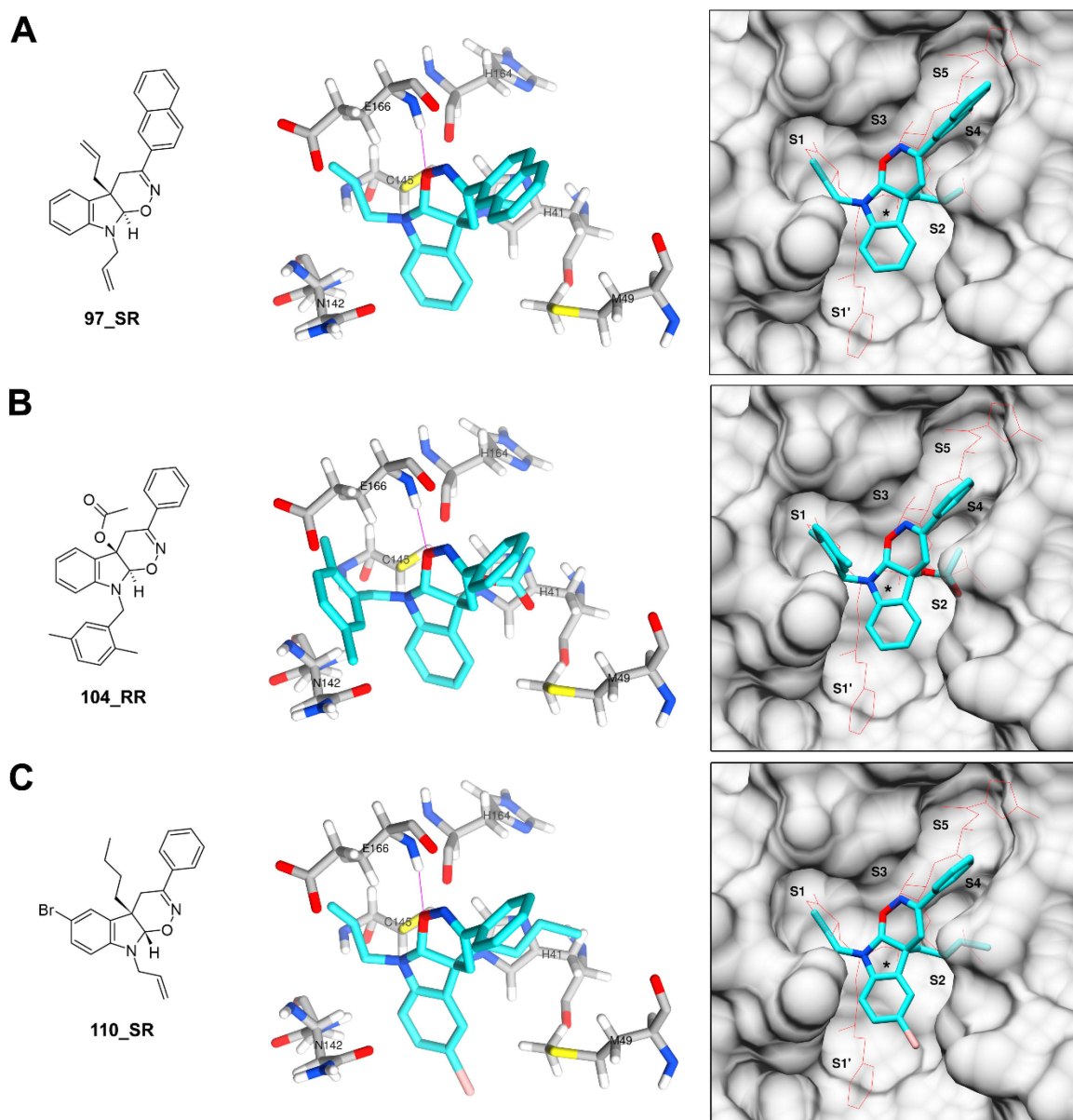


Figure 7. Docking poses predicted for stereoisomers 97_RR (A), 104_RR (B) and 110_SR (C). These stereoisomers are colored in cyan with hydrogen bonds represented by violet lines. The orientation of the N3 inhibitor in surface representations is indicated by a red wire and the approximate position of the His41-Cys145 catalytic dyad is indicated with “*”.

Several other stereoisomers with favorable energy scores followed a similar orientation, with relatively small changes in the local position of the TOI nucleus and its substituents but occupying the same subsites. These include 105_RR, 106_RS, 107_RR, 108_RR, 111_SR, 112_SR, 116_SR and three stereoisomers of compound 117 (117_RR, 117_RS and 117_SR). Finding regularities that can serve as hypotheses about binding of these compounds to the active site is difficult due to the diverse nature of their substituents, but the

presence of a hydrophobic or bulky group at position C3 again seems to be relevant. The Br substituent at position C6 also appears to favor positioning of the TOI nucleus, since in almost all the compounds with this substitution there is at least one stereoisomer in which the halogen atom orients towards the S1' subsite. Conversely, in all the stereoisomers of compound 116, which lack this substituent at C6 but have a 4-bromophenyl substituent at C3, the TOI ring accommodates so that the 4-bromophenyl group is oriented towards S1'. In compound 117, which have an I instead of a Br at position C6, the three stereoisomers mentioned above have the halogen atom oriented towards S1'.

4. Conclusions

This study proposes a methodology for rapid prediction of binding of *N*-heterocycles to the M^{Pro} of SARS-related coronaviruses. We used two molecular docking simulation approaches to explore the interaction of four libraries of *N*-heterocyclic compounds with the SARS-CoV-2 M^{Pro}. Globally, we observed no significant correlation between the results obtained through the rigid receptor approach (DOCK) and the flexible receptor approach (Smina), although both approaches tended to agree better in molecules with favorable energy estimates. Docking of co-crystallized ligand and careful inspection of molecular interactions of target compounds suggested that the flexible approach, which adds flexibility to 10 key amino acid residues of the M^{Pro} active site, was a better model to predict the orientations and interactions within the active site. Using empirical energy score thresholds, we selected 58 compounds from the four libraries to further discuss their predicted interaction with M^{Pro}. Most favorable predictions were obtained for relatively large molecules from libraries L1 and L4, containing substituents capable of accommodating in three or more subsites within the M^{Pro} active site. The presence of two or more large, relatively bulky substituents is common in previously reported inhibitors of the SARS-CoV M^{Pro}, including peptidomimetic inhibitors containing Michael acceptors and other non-peptidic molecules [2].

Due to the relevance of M^{Pro} in viral replication, this protein has been considered one of the most attractive targets for the identification of new antiviral drugs against SARS-related coronaviruses, including SARS-CoV-2 [1,2]. However, despite all efforts thus far, no M^{Pro} inhibitor has advanced to clinical trials. The identification of new molecules with the potential to inhibit viral replication and consequently the infection, continues to be urgent. Herein, we have identified several compounds predicted to interact with M^{Pro}. Further studies are necessary to evaluate, *in vitro*, the effect of those compounds on SARS-CoV-2 replication. The potential advance of new molecules in preclinical and clinical programs offers hope for the development of effective therapies that help control the spread of the virus. This work adds to the efforts of the scientific community in the fight against SARS-CoV-2.

Supplementary Materials: The following are available online at <https://www.mdpi.com/2218-273X/11/1/18/s1>, Figure S1: Non-covalent interactions among the GWS ligand and amino acid residues from the active site of SARS-CoV-2 M^{Pro}, Table S1: Compounds from libraries L1 and L2, Table S2: Compounds from libraries L3 and L4, Table S3: Docking energy scores and lowest-energy poses for the compounds considered in this study.

Author Contributions: Conceptualization, O.V.L. and P.L.F.; methodology, A.L., H.C. and V.D.N.; formal analysis, A.L., H.C., O.V.L. and P.L.F.; writing—original draft preparation, A.L., O.V.L. and P.L.F.; writing—review and editing, A.L., O.V.L. and P.L.F. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Sistema Nacional de Investigación, Panama (SNI-142-2018) and Welch Foundation (AX-1788).

Data Availability Statement: Data is contained within the article or supplementary material.

Acknowledgments: Financial support by Sistema Nacional de Investigación (SNI), Panama and the Welch Foundation (AX-1788) is gratefully acknowledged. We thank Jean-Didier Maréchal from the Autonomous University of Barcelona for his help with the flexible docking methodology.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. Li, G.; De Clercq, E. Therapeutic options for the 2019 novel coronavirus (2019-nCoV). *Nat. Rev. Drug Discov.* **2020**, *19*, 149–150. [[CrossRef](#)] [[PubMed](#)]
2. Pillaiyar, T.; Manickam, M.; Namasivayam, V.; Hayashi, Y.; Jung, S.-H. An Overview of Severe Acute Respiratory Syndrome-Coronavirus (SARS-CoV) 3CL Protease Inhibitors: Peptidomimetics and Small Molecule Chemotherapy. *J. Med. Chem.* **2016**, *59*, 6595–6628. [[CrossRef](#)] [[PubMed](#)]
3. Anand, K.; Ziebuhr, J.; Wadhwani, P.; Mesters, J.R.; Hilgenfeld, R. Coronavirus main proteinase (3CLpro) structure: Basis for design of anti-SARS drugs. *Science* **2003**, *300*, 1763–1767. [[CrossRef](#)] [[PubMed](#)]
4. Xue, X.; Yu, H.; Yang, H.; Xue, F.; Wu, Z.; Shen, W.; Li, J.; Zhou, Z.; Ding, Y.; Zhao, Q.; et al. Structures of two coronavirus main proteases: Implications for substrate binding and antiviral drug design. *J. Virol.* **2008**, *82*, 2515–2527. [[CrossRef](#)] [[PubMed](#)]
5. Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C.; et al. Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* **2020**, *582*, 289–293. [[CrossRef](#)] [[PubMed](#)]
6. Yang, H.; Xie, W.; Xue, X.; Yang, K.; Ma, J.; Liang, W.; Zhao, Q.; Zhou, Z.; Pei, D.; Ziebuhr, J.; et al. Design of wide-spectrum inhibitors targeting coronavirus main proteases. *PLoS Biol.* **2005**, *3*, e324.
7. Wang, J. Fast Identification of Possible Drug Treatment of Coronavirus Disease-19 (COVID-19) through Computational Drug Repurposing Study. *J. Chem. Inf. Model.* **2020**, *60*, 3277–3286. [[CrossRef](#)]
8. Chtita, S.; Belhassan, A.; Aouidate, A.; Belaidi, S.; Bouachrine, M.; Lakhlifi, T. Discovery of Potent SARS-CoV-2 Inhibitors from Approved Antiviral Drugs via Docking Screening. *Comb. Chem. High Throughput Screen.* **2020**, *23*. [[CrossRef](#)]
9. Peele, K.A.; Potla Durthi, C.; Srihansa, T.; Krupanidhi, S.; Ayyagari, V.S.; Babu, D.J.; Indira, M.; Reddy, A.R.; Venkateswarulu, T.C. Molecular docking and dynamic simulations for antiviral compounds against SARS-CoV-2: A computational study. *Inform. Med. Unlocked* **2020**, *19*, 100345. [[CrossRef](#)]
10. Mittal, L.; Kumari, A.; Srivastava, M.; Singh, M.; Asthana, S. Identification of potential molecules against COVID-19 main protease through structure-guided virtual screening approach. *J. Biomol. Struct. Dyn.* **2020**. [[CrossRef](#)]
11. Kandeel, M.; Al-Nazawi, M. Virtual screening and repurposing of FDA approved drugs against COVID-19 main protease. *Life Sci.* **2020**, *251*, 117627. [[CrossRef](#)] [[PubMed](#)]
12. Fischer, A.; Sellner, M.; Neranjan, S.; Smieško, M.; Lill, M.A. Potential Inhibitors for Novel Coronavirus Protease Identified by Virtual Screening of 606 Million Compounds. *Int. J. Mol. Sci.* **2020**, *21*, 3626. [[CrossRef](#)] [[PubMed](#)]
13. Tsuji, M. Potential anti-SARS-CoV-2 drug candidates identified through virtual screening of the ChEMBL database for compounds that target the main coronavirus protease. *FEBS Open Bio* **2020**, *10*, 995–1004. [[CrossRef](#)] [[PubMed](#)]
14. Mirza, M.U.; Froeyen, M. Structural elucidation of SARS-CoV-2 vital proteins: Computational methods reveal potential drug candidates against main protease, Nsp12 polymerase and Nsp13 helicase. *J. Pharm. Anal.* **2020**, *10*, 320–328. [[CrossRef](#)] [[PubMed](#)]
15. Liu, S.; Zheng, Q.; Wang, Z. Potential covalent drugs targeting the main protease of the SARS-CoV-2 coronavirus. *Bioinformatics* **2020**, *36*, 3295–3298. [[CrossRef](#)] [[PubMed](#)]
16. Grassmann, A.A.; Kremer, F.S.; Dos Santos, J.C.; Souza, J.D.; Pinto, L.D.S.; McBride, A.J.A. Discovery of Novel Leptospirosis Vaccine Candidates Using Reverse and Structural Vaccinology. *Front. Immunol.* **2017**, *8*, 463. [[CrossRef](#)]
17. Zhang, H.; Saravanan, K.M.; Yang, Y.; Hossain, M.T.; Li, J.; Ren, X.; Pan, Y.; Wei, Y. Deep Learning Based Drug Screening for Novel Coronavirus 2019-nCoV. *Interdiscip. Sci. Comput. Life Sci.* **2020**, *12*, 368–376. [[CrossRef](#)]
18. Berry, M.; Fielding, B.C.; Gamielien, J. Potential Broad Spectrum Inhibitors of the Coronavirus 3CLpro: A Virtual Screening and Structure-Based Drug Design Study. *Viruses* **2015**, *7*, 6642–6660. [[CrossRef](#)]
19. Lu, I.-L.; Mahindroo, N.; Liang, P.-H.; Peng, Y.-H.; Kuo, C.-J.; Tsai, K.-C.; Hsieh, H.-P.; Chao, Y.-S.; Wu, S.-Y. Structure-based drug design and structural biology study of novel nonpeptide inhibitors of severe acute respiratory syndrome coronavirus main protease. *J. Med. Chem.* **2006**, *49*, 5154–5161. [[CrossRef](#)]
20. Blanchard, J.E.; Elowe, N.H.; Huitema, C.; Fortin, P.D.; Cechetto, J.D.; Eltis, L.D.; Brown, E.D. High-throughput screening identifies inhibitors of the SARS coronavirus main proteinase. *Chem. Biol.* **2004**, *11*, 1445–1453. [[CrossRef](#)]
21. Ton, A.-T.; Gentile, F.; Hsing, M.; Ban, F.; Cherkasov, A. Rapid Identification of Potential Inhibitors of SARS-CoV-2 Main Protease by Deep Docking of 1.3 Billion Compounds. *Mol. Inform.* **2020**, *39*, e2000028. [[CrossRef](#)] [[PubMed](#)]
22. Vitaku, E.; Smith, D.T.; Njardarson, J.T. Analysis of the structural diversity, substitution patterns, and frequency of nitrogen heterocycles among U.S. FDA approved pharmaceuticals. *J. Med. Chem.* **2014**, *57*, 10257–10274. [[CrossRef](#)] [[PubMed](#)]
23. Petit, J.; Meurice, N.; Kaiser, C.; Maggiora, G. Softening the Rule of Five—where to draw the line? *Bioorg. Med. Chem.* **2012**, *20*, 5343–5351. [[CrossRef](#)] [[PubMed](#)]
24. Larionov, O.V.; Stephens, D.; Mfuh, A.; Chavez, G. Direct, catalytic, and regioselective synthesis of 2-alkyl-, aryl-, and alkenyl-substituted N-Heterocycles from n-oxides. *Org. Lett.* **2014**, *16*, 864–867. [[CrossRef](#)]
25. Mfuh, A.M.; Larionov, O. V Heterocyclic N-Oxides - An Emerging Class of Therapeutic Agents. *Curr. Med. Chem.* **2015**, *22*, 2819–2857. [[CrossRef](#)]

26. Larionov, O.V.; Stephens, D.; Mfuh, A.M.; Arman, H.D.; Naumova, A.S.; Chavez, G.; Skenderi, B. Insights into the mechanistic and synthetic aspects of the Mo/P-catalyzed oxidation of N-heterocycles. *Org. Biomol. Chem.* **2014**, *12*, 3026–3036. [[CrossRef](#)]
27. Ruiz-Sanchis, P.; Savina, S.A.; Albericio, F.; Álvarez, M. Structure, bioactivity and synthesis of natural products with hexahydropyrrolo[2,3-b]indole. *Chemistry* **2011**, *17*, 1388–1408. [[CrossRef](#)]
28. Mfuh, A.M.; Zhang, Y.; Stephens, D.E.; Vo, A.X.T.; Arman, H.D.; Larionov, O.V. Concise Total Synthesis of Trichodermamides A, B, and C Enabled by an Efficient Construction of the 1,2-Oxazadecaline Core. *J. Am. Chem. Soc.* **2015**, *137*, 8050–8053. [[CrossRef](#)]
29. Jans, P.E.; Mfuh, A.M.; Arman, H.D.; Shaffer, C.V.; Larionov, O.V.; Mooberry, S.L. Cytotoxicity and Mechanism of Action of the Marine-Derived Fungal Metabolite Trichodermamide B and Synthetic Analogues. *J. Nat. Prod.* **2017**, *80*, 676–683. [[CrossRef](#)]
30. Zhang, Y.; Stephens, D.; Hernandez, G.; Mendoza, R.; Larionov, O.V. Catalytic diastereo- and enantioselective annulations between transient nitrosoalkenes and indoles. *Chemistry* **2012**, *18*, 16612–16615. [[CrossRef](#)]
31. Herrera, L.; Stephens, D.E.; D'Avila, A.; George, K.G.; Arman, H.; Zhang, Y.; Perry, G.; Leonart, R.; Larionov, O.V.; Fernández, P.L. Insights into the structural patterns of the antileishmanial activity of bi- and tricyclic N-heterocycles. *Org. Biomol. Chem.* **2016**, *14*, 7053–7060. [[CrossRef](#)] [[PubMed](#)]
32. Doens, D.; Valdés-Tresanco, M.E.; Vasquez, V.; Carreira, M.B.; De La Guardia, Y.; Stephens, D.E.; Nguyen, V.D.; Nguyen, V.T.; Gu, J.; Hegde, M.L.; et al. Hexahydropyrrolo[2,3-b]indole Compounds as Potential Therapeutics for Alzheimer's Disease. *ACS Chem. Neurosci.* **2019**, *10*, 4250–4263. [[CrossRef](#)] [[PubMed](#)]
33. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera—A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612. [[CrossRef](#)] [[PubMed](#)]
34. Allen, W.J.; Balius, T.E.; Mukherjee, S.; Brozell, S.R.; Moustakas, D.T.; Lang, P.T.; Case, D.A.; Kuntz, I.D.; Rizzo, R.C. DOCK 6: Impact of new features and current docking performance. *J. Comput. Chem.* **2015**, *36*, 1132–1156. [[CrossRef](#)] [[PubMed](#)]
35. Koes, D.R.; Baumgartner, M.P.; Camacho, C.J. Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. *J. Chem. Inf. Model.* **2013**, *53*, 1893–1904. [[CrossRef](#)] [[PubMed](#)]
36. Trott, O.; Olson, A.J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461. [[CrossRef](#)]
37. Gendrot, M.; Andreani, J.; Boxberger, M.; Jardot, P.; Fonta, I.; Le Bideau, M.; Dufлот, I.; Mosnier, J.; Rolland, C.; Bogreau, H.; et al. Antimalarial drugs inhibit the replication of SARS-CoV-2: An in vitro evaluation. *Travel Med. Infect. Dis.* **2020**, *37*, 101873. [[CrossRef](#)]
38. Ianevski, A.; Yao, R.; Fenstad, M.H.; Biza, S.; Zusinaite, E.; Reisberg, T.; Lysvand, H.; Løseth, K.; Landsem, V.M.; Malmring, J.F.; et al. Potential antiviral options against SARS-CoV-2 infection. *Viruses* **2020**, *12*, 642. [[CrossRef](#)]
39. Devaux, C.A.; Rolain, J.-M.; Colson, P.; Raoult, D. New insights on the antiviral effects of chloroquine against coronavirus: What to expect for COVID-19? *Int. J. Antimicrob. Agents* **2020**, *55*, 105938. [[CrossRef](#)]
40. WHO Solidarity Trial Consortium; Pan, H.; Peto, R.; Henao-Restrepo, A.-M.; Preziosi, M.-P.; Sathiyamoorthy, V.; Abdool Karim, Q.; Alejandria, M.M.; Hernández García, C.; Kieny, M.-P.; et al. Repurposed Antiviral Drugs for Covid-19 - Interim WHO Solidarity Trial Results. *N. Engl. J. Med.* **2020**. [[CrossRef](#)]