

# Combined Computational NMR and Molecular Docking Scrutiny of Potential Natural SARS-CoV-2 M<sup>pro</sup> Inhibitors

Valentin A. Semenov\* and Leonid B. Krivdin\*



Cite This: *J. Phys. Chem. B* 2022, 126, 2173–2187



Read Online

ACCESS |



Metrics & More

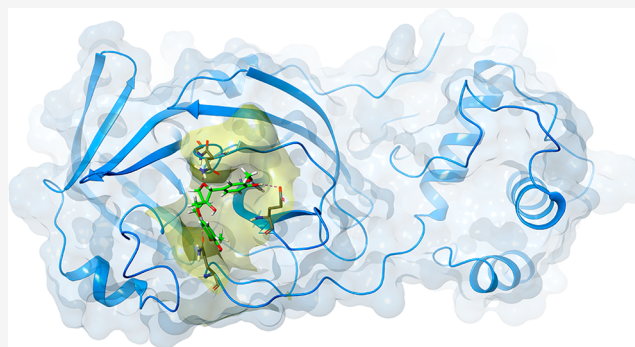


Article Recommendations



Supporting Information

**ABSTRACT:** In continuation of the search for potential drugs that inhibit the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), in this work, a combined approach based on the modeling of NMR chemical shifts and molecular docking is suggested to identify the possible suppressors of the main protease of this virus among a number of natural products of diverse nature. Primarily, with the aid of an artificial neural network, the problem of the reliable determination of the stereochemical structure of a number of studied compounds was solved. Complementary to the main goal of this study, theoretical modeling of NMR spectral parameters made it feasible to perform a number of signal reassignments together with introducing some missing NMR data. Finally, molecular docking formalism was applied to the analysis of several natural products that could be chosen as prospective candidates for the role of potential inhibitors of the main protease. The results of this study are believed to assist in further research aimed at the development of specific drugs based on the natural products against COVID-19.



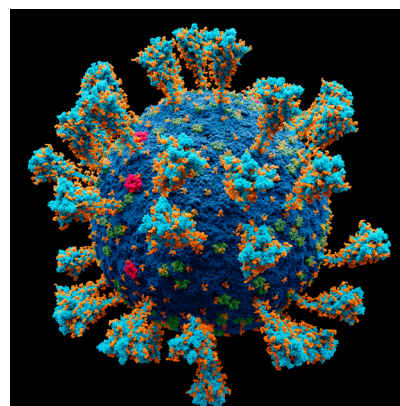
## INTRODUCTION

A highly infectious novel coronavirus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was first identified in the city of Wuhan, Hubei, China. The World Health Organization declared the outbreak of a “Public Health Emergency of International Concern” on January 30, 2020, and a pandemic on March 11, 2020.

Although a detailed understanding of the pathophysiology of SARS-CoV-2 is still not fully gained, the genetic structure of this virus has nevertheless been determined, and several potential targets for the potential drugs have been identified. The virus is known to contain four non-structural proteins: papain-like (PL<sup>pro</sup>) and 3-chymotrypsin-like (3CL<sup>pro</sup> or M<sup>pro</sup>) proteases, RNA polymerase (RDRp), and helicase,<sup>1</sup> as shown in Figure 1.

It was found that only proteases (PL<sup>pro</sup> and M<sup>pro</sup>) are involved in the transcription of this virus.<sup>2</sup> The main protease of SARS-CoV-2 has been shown to have of about 96% sequence similarity to that of SARS-CoV.<sup>3</sup> Therefore, one of the important targets that antiviral molecules can potentially aim at is the main protease M<sup>pro</sup>.

The crystal structure of M<sup>pro</sup> in the complex with the peptide inhibitor N3 was first deposited to the protein databank (PDB ID: 6LU7) in 2020.<sup>4</sup> Since this deposit to date, more than 230 M<sup>pro</sup> structures have been deposited to PDB in combination with various fragments. The M<sup>pro</sup> active site contains catalytic dyad Cys145–His41, located between the two domains of the protease. It is assumed that inhibitors targeting this site should exhibit a broad spectrum of anti-SARS-CoV-2 activity, because

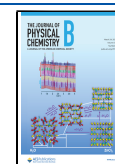


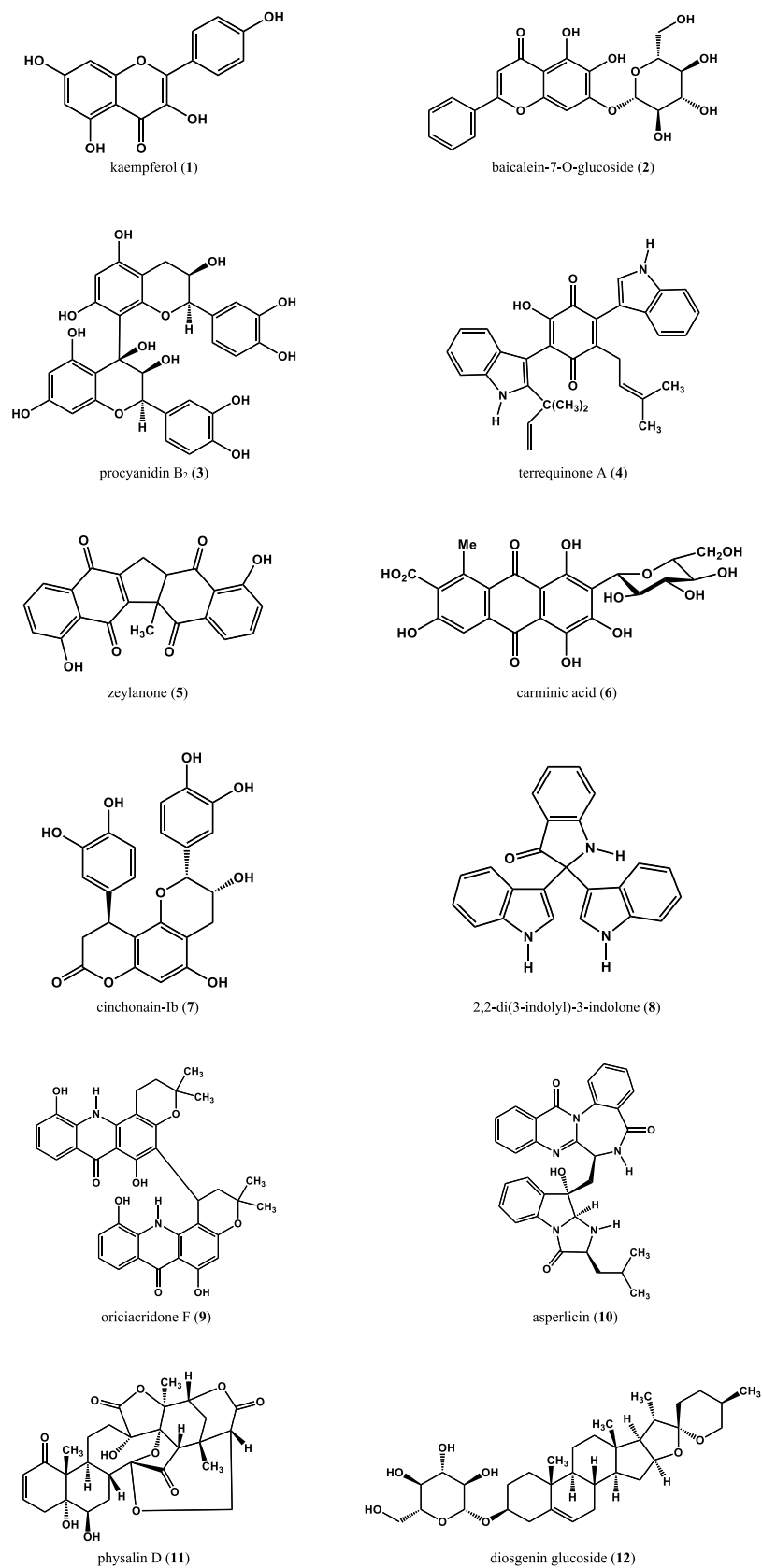
**Figure 1.** Atomic model of the external structure of SARS-CoV-2. Colors: blue—envelope, turquoise—spike glycoprotein, red—envelope proteins, green—membrane proteins, and orange—glycan. Reproduced from Alexey Solodovnikov and Valeria Arkhipova ([https://nl.wikinews.org/wiki/Bestand:Coronavirus\\_SARS-CoV-2.png](https://nl.wikinews.org/wiki/Bestand:Coronavirus_SARS-CoV-2.png)) under the Creative Commons Attribution-NonCommercial 3.0 Unported Licence; see <https://creativecommons.org/licenses/by-nc/3.0/>.

**Received:** December 11, 2021

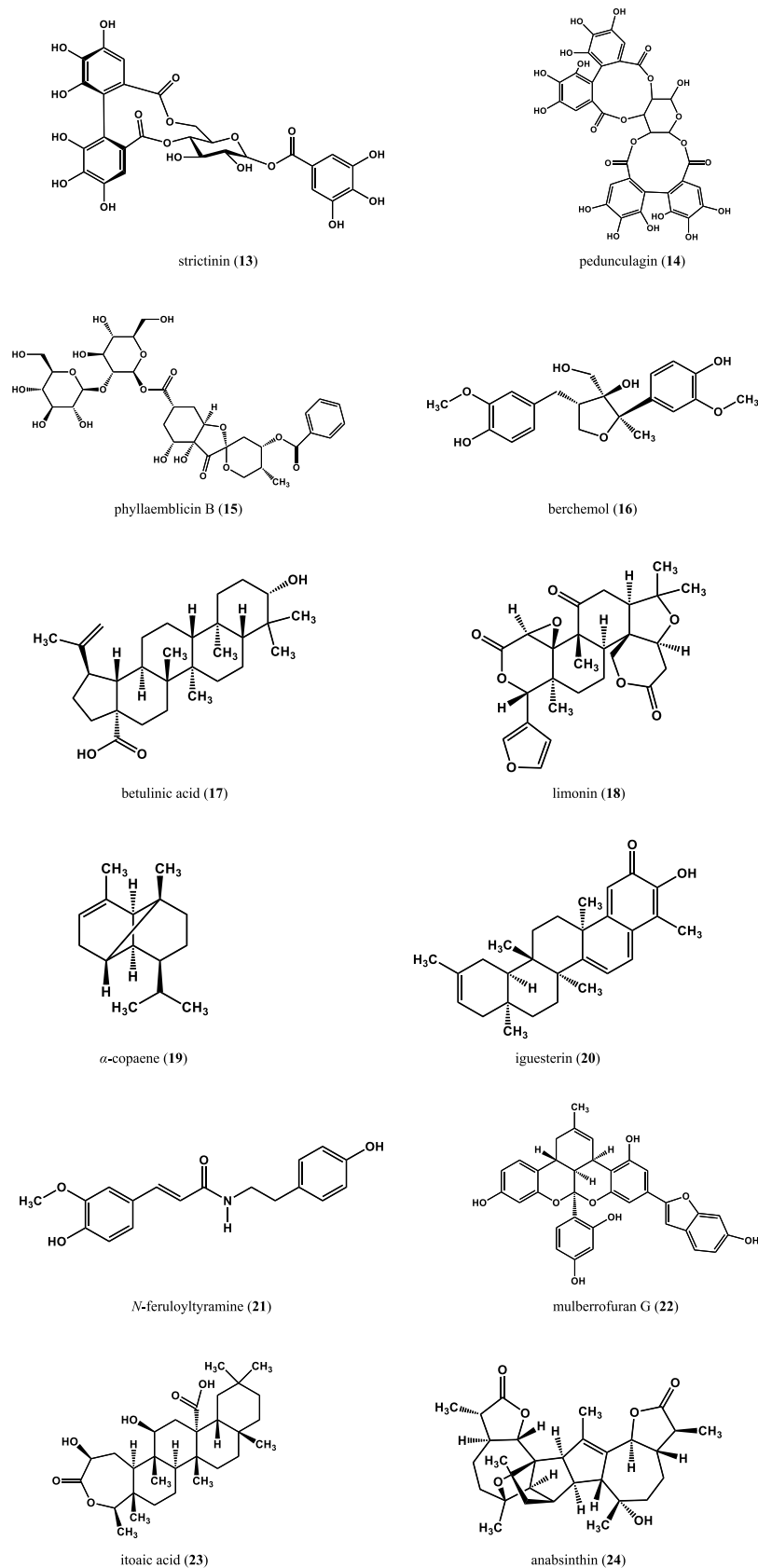
**Revised:** February 25, 2022

**Published:** March 10, 2022



Scheme 1. Trivial Names and Chemical Structures of the Studied Potential Natural SARS-CoV-2 M<sup>PRO</sup> Inhibitors

## Scheme 1. continued



the  $M^{\text{pro}}$  binding site is highly conserved among all types of coronaviruses.<sup>4,5</sup> In the absence of proteases with similar cleavage site specificity, the  $M^{\text{pro}}$  binding site is likely to remain

constant for a long time, because it is less prone to drug resistance mutations.

In addition to the known drugs, to date several synthetic antiviral compounds are at the stage of their preclinical and clinical trials. At the same time, a possible negative modulation of the main protease of SARS-CoV-2, which could lead to new developments in therapeutic drugs, was demonstrated by several natural products, including a number of alkaloids, flavonoids, and terpenoids with high affinity to M<sup>pro</sup>.<sup>6–11</sup> Moreover, several theoretical approaches based on the calculations of free energy and molecular dynamics, as well as virtual screening of exposed ligand series, are useful in identifying the inhibitory activity of M<sup>pro</sup>.<sup>12–24</sup>

It is well known that the theoretical calculation of NMR chemical shifts provides a powerful tool in structural elucidation of organic molecules,<sup>25–29</sup> such as natural products, carbohydrates, and biochemical species, including those with potential anti-SARS-CoV-2 activity. In continuation of our recent computational NMR survey of parent strychnine<sup>30</sup> with a related family of *Strychnos* alkaloids<sup>31,32</sup> and their much larger derivatives<sup>33,34</sup> together with a number of antimalarial plant and marine alkaloids,<sup>35</sup> in this paper we investigated a series of natural products of diverse chemical origins providing potential SARS-CoV-2 M<sup>pro</sup> inhibitory activities. Here, we report on the combination of the predictive power of computational NMR and molecular docking to the study of the compounds, which were shown<sup>36</sup> to provide inhibitory activities against the main protease: flavonoids (1–3), quinones (4–6), coumarin (7), alkaloids (8–10), steroids (11, 12), tannins/ellagitannins (13–15), lignan (16), and other terpenes and terpenoids (17–24); see Scheme 1.

## ■ COMPUTATIONAL DETAILS

**Geometry Optimization.** Geometry optimization of 1–24 was performed with GAUSSIAN 09<sup>37</sup> code at the M06-2X/cc-pVTZ level (for nitrogen and oxygen atoms, the extended aug-cc-pVTZ basis set with diffuse functions was used to take into account the effect of the diffused lone pairs). Evaluated structures were identified as minima on their potential energy surfaces. Solvent effects were accounted for within the integral equation formalism polarizable continuum model (IEF-PCM).<sup>38,39</sup> Cartesian coordinates of all optimized structures 1–24 are given in the Supporting Information.

**Calculation of Chemical Shifts.** All calculations of <sup>1</sup>H and <sup>13</sup>C NMR isotropic magnetic shielding constants (the latter being converted into a chemical shift scale) were carried out at the gauge including atomic orbital (GIAO) DFT level for a liquid phase by implying the DALTON package.<sup>40</sup> The well-known GIAO method is based on the approach that introduces the local gauge origins to define the vector potential of the external magnetic field. In our calculations, we used the mPW1PW91/6-31G(d,p) computational scheme. In this scheme, the one-parameter hybrid functional mPW1PW91 (which is based on the Perdew–Wang exchange), as modified by Adamo and Barone, combined with PW91 correlation<sup>41</sup> was employed throughout in combination with Pople's double-zeta basis set.<sup>42</sup>

Calculated proton and carbon isotropic magnetic shielding constants were converted into <sup>1</sup>H and <sup>13</sup>C NMR chemical shifts as recommended by the International Union of Pure and Applied Chemistry (IUPAC).<sup>43</sup> To take into account the systematic error of calculated chemical shifts, we established correlations between their calculated and experimental values. These correlations were used further to find out the linear correlation equations of the  $y = ax + b$  type. The slope  $a$  and

intercept  $b$  were then used for recalculating theoretical chemical shifts into experimental  $\delta$ -scale by using the equation  $\delta_{\text{recalc}} = (\delta_{\text{calc}} - b)/a$ . The mean absolute errors (MAE) and the corrected mean absolute errors (CMAE) were evaluated in all particular cases.

**Preparation of Protein for Docking and Grid Generation.** The crystal structure of the main protease (having resolution 2.16 Å,  $R$ -value free 0.24,  $R$ -value work 0.20) in a complex with a peptide-like inhibitor N3 was obtained from the Protein Data Bank<sup>44</sup> (PDB ID: 6LU7<sup>4</sup>). The 6LU7 protein contains two chains, A and B, which form a homodimer. Chain A was used for macromolecule preparation with AutoDockTools.<sup>45</sup> The co-factor and water molecules were removed, and hydrogen atoms were added. The partial atomic charges were predicted via the Gasteiger–Marsili approach<sup>46</sup> after putting hydrogens in all the required atoms of the protein.

The receptor grid maps were calculated with AutoGrid 4.2,<sup>45</sup> mapping the receptor interaction energies using the ligand atom types as probes. The grid of 60 × 60 × 60 Å with 0.375 Å spacing was centered on the coordinates of the ligand originally present in the 6LU7 crystal.

**Molecular Docking Simulations.** The molecular docking using the AutoDock 4.2 software<sup>45</sup> was employed to rapidly determine the ligand-binding pose and affinity to SARS-CoV-2 M<sup>pro</sup>. The energetically minimized structures of all studied compounds, which were obtained as a result of optimization in the first stage of this study, were used to simulate docking as ligand molecules. A Lamarckian genetic algorithm<sup>47</sup> was used to generate docking simulations with an initial population size of 300 random individuals and 50 independent runs. Initial orientation and position of the ligands were randomly set.

The best docking modes of all compounds were selected from their conformations based on the binding energy values as well as on significant non-valence interactions observed with receptors. The docking analyses were performed by using Schrödinger Maestro 11.5,<sup>48</sup> MOE 2014,<sup>49</sup> and AutoDockTools.<sup>45</sup>

**Validation of the Docking Protocol.** The docking procedure was validated by re-docking the peptide-like inhibitor N3 extracted from a crystallographic structure (6LU7) into the receptor at the same binding pocket. These results indicated that the applied protocol was reliable because a good overlap was achieved between the ligand pose and the X-ray pose with a binding affinity of −9.0 kcal/mol and a root mean square deviation (RMSD) reference value of 1.90 Å.

**Identification of Possible Stereochemical Misassignments.** All studied compounds were checked for possible inconsistencies with their originally established structures. For this purpose, the artificial neural network (ANN) pattern recognition methodology developed by Sarotti<sup>50</sup> was used throughout. Calculated shielding constants were compared with the experimental chemical shifts by using the two-layer feed-forward ANN. Within the framework of the formalism of this neural network, three different sets of statistical descriptors were used as input vectors for the input layer, namely, correlation coefficient  $R^2$ , slope  $a$ , and intercept  $b$  (see the section "Calculation of Chemical Shifts"). The MAE, standard deviation, maximum error, and their corresponding corrected descriptors were also used for the statistical treatment of the results. Thus, for each reference standard used in the calculation of NMR chemical shifts, a total of nine statistical parameters were evaluated. As a result, two categories were used as the

patterns for the output layer, namely, correct structures and incorrect structures.

## RESULTS AND DISCUSSION

**Molecular Objects.** Herein, we will shortly comment on the isolation and general biological activity of natural products 1–24 selected for this study.

Kaempferol (1) together with other strawberry flavonoids isolated from *Fragaria ananassa* Duch was found to effectively suppress degranulation in rat basophilic leukemia cells.<sup>51</sup> It was established that the suppression of the Ag-stimulated degranulation by kaempferol was mainly due to the suppression of the Ca<sup>2+</sup> elevation together with the spleen tyrosine kinase activation.

Baicalein-7-*O*-glucoside (2) has been isolated from the seeds of *Oroxylum indicum* and identified by the high-speed counter-current chromatography by Chen, Games, and Jones<sup>52</sup> and later by Yuan and coauthors.<sup>53</sup> *O. indicum* is a traditional herbal medicine, possessing analgesic, antitussive, and anti-inflammatory activity. Active ingredients characterized from the seeds and leaves of this plant are several flavonoids involving baicalein-7-*O*-glucoside, which were shown to reduce the total cholesterol level and have detoxification and chemo-preventative effects.

Procyanidin B<sub>2</sub> (3), extracted from hawthorn, belongs to the well-known proanthocyanidins, which are widely distributed in nature, providing a range of therapeutic effects and often considered as the active ingredients of herbal medicines. Khan, Haslam, and Williamson<sup>54</sup> established its structure as a proanthocyanidin dimer and performed a detailed conformational study by applying a heteronuclear NMR correlation.

In a study to discover potential anticancer agents from rhizosphere fungi of Sonoran desert plants, a new metabolite, terrefuranone (4), was isolated from *As. terries*, occurring in the rhizosphere of *Ambrosia ambrosoides*.<sup>55</sup> The structure of 4 was elucidated mainly by means of <sup>1</sup>H and <sup>13</sup>C NMR techniques involving one- and two-dimensional COSY, ROESY, NOESY, and HMBC together with UV, IR, and MS including APCIMS and HRFABMS methods. Terrequinone A was found to exhibit moderate selective cytotoxicity against cancer cell lines as compared with the normal fibroblast cells.

Zeylanone (5) was obtained from the crude extracts of *Drimys maritima* bark, which is traditionally used in Indonesia for the treatment of rheumatic diseases. This natural product displays a range of biological effects, such as ichthyotoxicity and antibacterial activity. The structure of zeylanone was most completely elucidated by Gu and coauthors<sup>56</sup> by means of diverse NMR experiments involving, in particular, COSY, ROESY, HSQC, and HMBC.

The structure of carminic acid (6), a degradation product of the well-known coloring matter of cochineal and the color pigment of the Middle and South American shield louse *Coccus cacti coccinelliferi*, was established by chemical means as early as in 1964 by Overeem and van der Kerk.<sup>57</sup> It was reinvestigated much later by Schmitt et al.,<sup>58</sup> who performed a detailed NMR investigation of this interesting natural product by applying a variety of one- and two-dimensional NMR techniques.

Cinchonain-Ib (7) was isolated by Nonaka and Nishioka<sup>59</sup> as a result of the investigation of the relatively low molecular weight phenolic compounds in the bark of *Cinchona succirubra*. A direct coupling of caffeic acid with epicatechin leading to the formation of the chinchonine alkaloids has also been documented in that study. The structure and stereochemistry of cinchonain-Ib were

established based on the physicochemical methods available at that period of time.

2,2-Di(3-indolyl)-3-indolone (8) was originally isolated from a *Vibrio* sp. separated from the marine sponge *Hyrtios altum*.<sup>60</sup> This molecule has also been very recently produced from the clone based on the functional screening of the metagenome library generated from the marine sponge *Discodermia calyx*.<sup>61</sup> Obtained in that study, <sup>1</sup>H and <sup>13</sup>C NMR spectra and COSY, HMBC, and HMQC correlations together with the high-resolution electrospray ionization mass spectrometry (ESIMS) enabled unambiguous identification of its structure.

The dichloromethane/methyl alcohol extract of the stem bark of *Oriciopsis glaberrima* ENGL. (Rutaceae) afforded four new acridone alkaloids of the oriciacridone family including oriciacridone F (9).<sup>62</sup> All structures were established based on the MS, 1D, and 2D NMR experiments. Acridone F showed potent activity against  $\alpha$ -glucosidase together with moderate free radical scavenging activity against 1,1-diphenyl-2-picrylhydrazyl (known as DPPH).

Asperlicin (10), a competitive cholecystokinin antagonist produced from *Aspergillus alliaceus*, was originally isolated by Chang et al.,<sup>63</sup> and later, its structure was revised by Sun, Byard, and Cooper<sup>64</sup> by means of mainly DEPT, HMQC, and HMBC experiments to yield the reestablished structure.

Physalin D (11), extracted from the fractions of *Physalis angulata* L., was shown to play a relevant role in the antimycobacterial activity.<sup>65</sup> Structural elucidation of physalins D performed in that study was based on detailed <sup>1</sup>H and <sup>13</sup>C NMR spectral analyses with the aid of 2D-correlation spectroscopy (COSY, HSQC, and HMBC). The assignment of its <sup>13</sup>C NMR chemical shifts was reported in that study for the first time.

Diosgenin glucoside (12), known also as diosgenin-3-*O*- $\beta$ -D-glucopyranoside, was extracted from the natural steroidal saponin mixture, which are widely distributed in plants, such as *Paridis polyphylla*.<sup>66</sup> Diosgenin glucoside has been known to improve cardiovascular function, anti-platelet aggregation, and antitumor and anti-diabetic activities. The structure of that compound was deduced from mainly <sup>1</sup>H and <sup>13</sup>C NMR data.

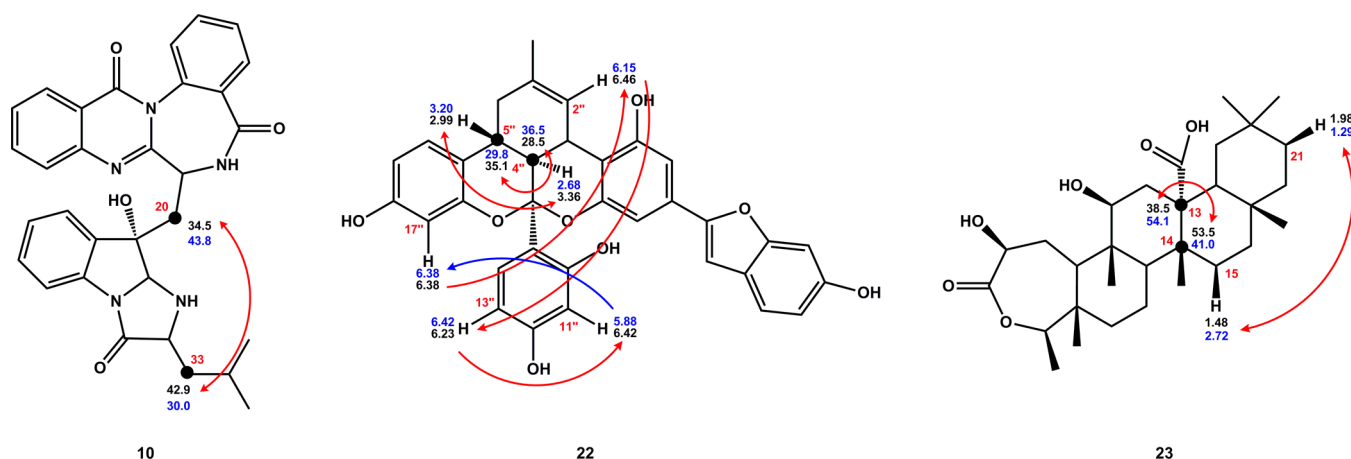
Strictinin (13) was extracted from the leaves of *Camellia irrawadiensis* and was found to show antiallergic effect and hair growth promotion.<sup>67</sup> It was identified as 1-*O*-galloyl-4,6-*O*-(*S*)-hexahydroxydiphenoyl- $\beta$ -D-glucose (13) by a direct comparison of its spectral data (NMR, MS, and circular dichroism) and specific rotation with those of an authentic sample, which was isolated from the green tea extract derived from *C. sinensis*.

Pedunculagin (14), an active inhibitor against carbonic anhydrase,<sup>68</sup> was isolated much earlier from the pericarps of *Punica granatum* L. Its stereochemical structure was established by Tanaka, Nonaka, and Nishioka<sup>69</sup> by means of an extensive NMR study.

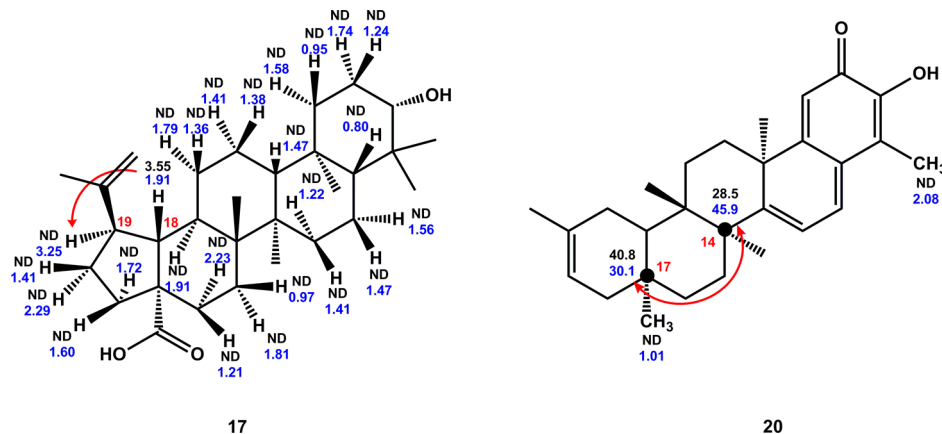
Phyllaemblicin B (15) was isolated from the roots of *Phyllanthus emblica* along with 15 tannins and related compounds,<sup>70</sup> the latter known for their anti-inflammatory and antipyretic effects in many local traditional medicinal systems, such as the Chinese herbal medicine, Tibetan medicine, and Ayurvedic medicine. The structure of phyllaemblicin B was established by means of NMR and other spectral and chemical methods.

Berchemol (16), a representative of the epoxy lignans family, was extracted from the seeds of *Centaurea cyanus* and identified by means of UV, ESIMS, high-resolution MS, <sup>1</sup>H and <sup>13</sup>C NMR, and a series of 2D NMR experiments. Unambiguous and complete assignment of its <sup>1</sup>H and <sup>13</sup>C NMR signals was

**Scheme 2. Reassignments (Red Arrows) of <sup>1</sup>H and <sup>13</sup>C NMR Chemical Shifts (Calculated and Measured Values Are Given Accordingly, in Blue and Black) of Asperlicin (10), Mulberrofuran G (22), and Iguesterin (20) Based on the Performed Calculations; Enumeration of Atoms (Given in Red) Is Taken from the Original Publications**



**Scheme 3. Reassignments (Red Arrows) and Predictions of <sup>1</sup>H and <sup>13</sup>C NMR Chemical Shifts (Calculated and Measured Values Are Given Accordingly, in Blue and Black) of Betulinic Acid (17) and Itoaic Acid (23) Based on the Performed Calculations; Enumeration of Atoms Is Taken from the Original Publications**



performed later on the basis of DEPT, COSY, TOCSY, HMQC, HMBC, and NOESY experiments.<sup>71</sup>

Betulinic acid (17) was isolated earlier from the leaves of *Nerium Oleander*<sup>72</sup> and most recently from the stems of *Combretum laxum*,<sup>73</sup> the former showing strongest activities against *Candida albicans* and *Cryptococcus neoformans* with a minimum inhibitory concentration of 100 μg/mL.

Limonin (18) was isolated from the chloroform layer of the bark of *Phellodendron amurense* and was found to provide little cytotoxicity against the human cancer cell lines.<sup>74</sup> The structure of limonin was identified by a comparison of its spectral data (UV, IR, MS, <sup>1</sup>H, and <sup>13</sup>C NMR) with those reported in the literature.

Copaene (19), a potent attractant for male Mediterranean fruit flies, *Ceratitidis capitata*, was found to be a minor component in the oils of various plant species, including its hosts such as orange, guava, and mango.<sup>75</sup> It was also shown in that study that copaene affected virgin females, provoking “pseudomale” courtship behavior in the short-range bioassay. Mating occurred exclusively on the artificial leaves treated with copaene, suggesting that the compound potentially served as a chemical cue to facilitate orientation of flies to the “rendezvous” site. Historically, copaene was first synthesized by Wenkert, Bookser, and Arrhenius.<sup>76</sup>

Iguesterin (20), which belongs to the quinone-methide triterpene family, was isolated from *Tritergium regelii*<sup>77</sup> and provided potent inhibitory activities against SARS. The worldwide outbreak of SARS was caused by the infection with novel SARS-CoV-2 3CL<sup>pro</sup>,<sup>78</sup> which mediated the proteolytic processing of replicase polypeptides into functional proteins. Iguesterin was found to be an effective drug against SARS coronavirus.

*N*-Feruloyltyramine (21) was isolated and characterized by Park<sup>79</sup> as the P-selectin expression suppressor from garlic (*Allium sativum*), a medicinal and culinary plant reported to have several positive health effects on cardiovascular diseases, particularly via the suppression of platelet activation. It was shown that *N*-feruloyltyramine was able to effectively suppress P-selectin expression on platelets. Potential effects of *N*-feruloyltyramine on the cyclooxygenase enzymes were found to be of major importance.

Mulberrofuran G (22) was isolated from the ethyl acetate extract of the root bark of cultivated mulberry tree *Morus lhou* Koidz.<sup>80</sup> The structure of mulberrofuran G was established based on the spectral and chemical evidence. It was found that intravenous injection of mulberrofuran G caused a marked depression effect in rabbits.

Scheme 4. Additional Assignments of  $^1\text{H}$  NMR Chemical Shifts (Calculated Values Are Given in Blue) of Limonin (18) and  $\alpha$ -Copaene (19) Based on the Performed Calculations

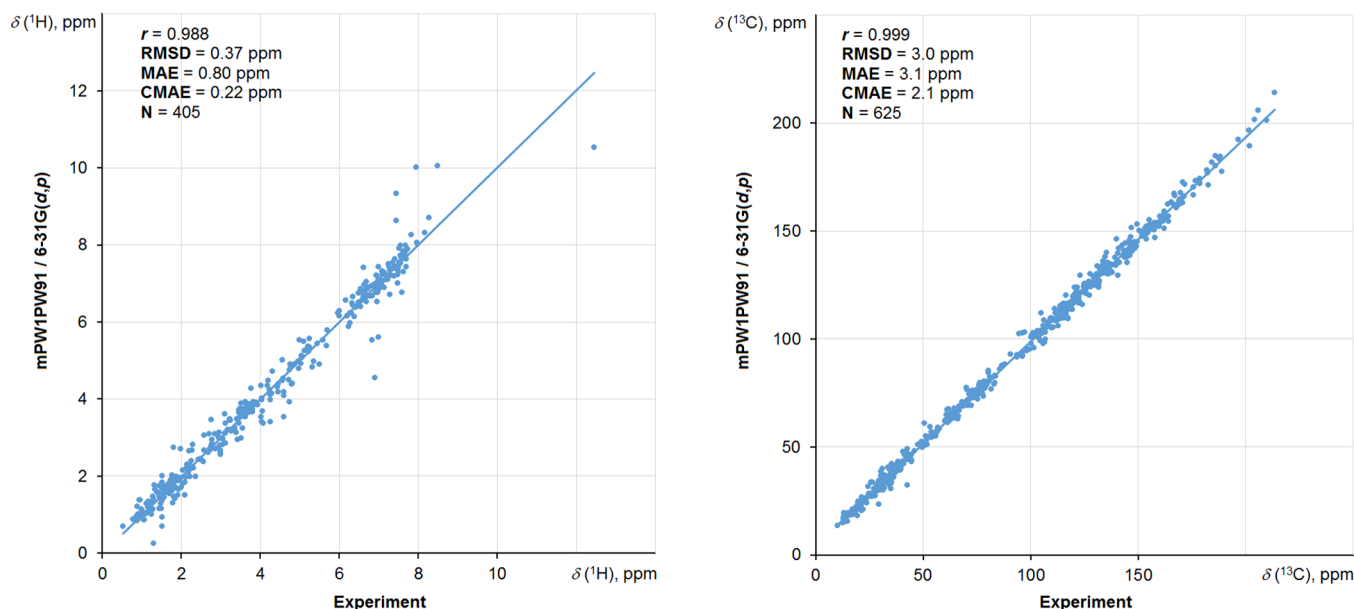
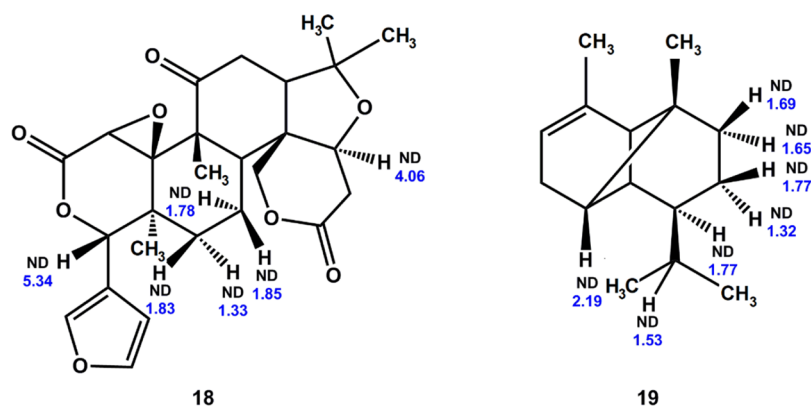


Figure 2. Correlation plots of calculated vs experimental  $^1\text{H}$  (left) and  $^{13}\text{C}$  (right) NMR chemical shifts of the natural products 1–24 of diverse chemical origins.

Itoaic acid (23), which belongs to seco-friedelolactones, was isolated from the bark and twigs of *Itoa orientalis*.<sup>81</sup> The structure of this alkaloid was elucidated by means of MS together with different 1D and 2D NMR techniques. Its anti-inflammatory activity was evaluated for several compounds extracted from the plant *Xylosma controversum*.

Anabsinthin (24) and accompanying natural products were isolated from *Artemisia absinthium* L.,<sup>82</sup> the latter commonly known as wormwood, a yellow-flowering perennial plant distributed throughout Europe and Siberia and used for the antiparasitic effects and for the treatment of anorexia and indigestion. The structure of anabsinthin was elucidated by means of various physicochemical methods, including optical rotation, Fourier transform infrared spectroscopy, NMR, and high-resolution mass spectrometry.

**Structural Validation.** Within the framework of the approach based on ANN,<sup>50</sup> the correctness of the initially established stereochemical structure of compounds 1–24 was verified. For this purpose, calculated shielding constants together with experimental  $^1\text{H}$  and  $^{13}\text{C}$  NMR chemical shifts were subjected to correlation analysis. Further, the resulting set

of 18 statistical descriptors for each of the 24 compounds was used as input vectors for the ANN input layer. Based on the pattern recognition results, it was revealed that stereochemical structures of all studied compounds were correctly established with a confidence of 59.4–99.9% with a mean value of 95.4%.

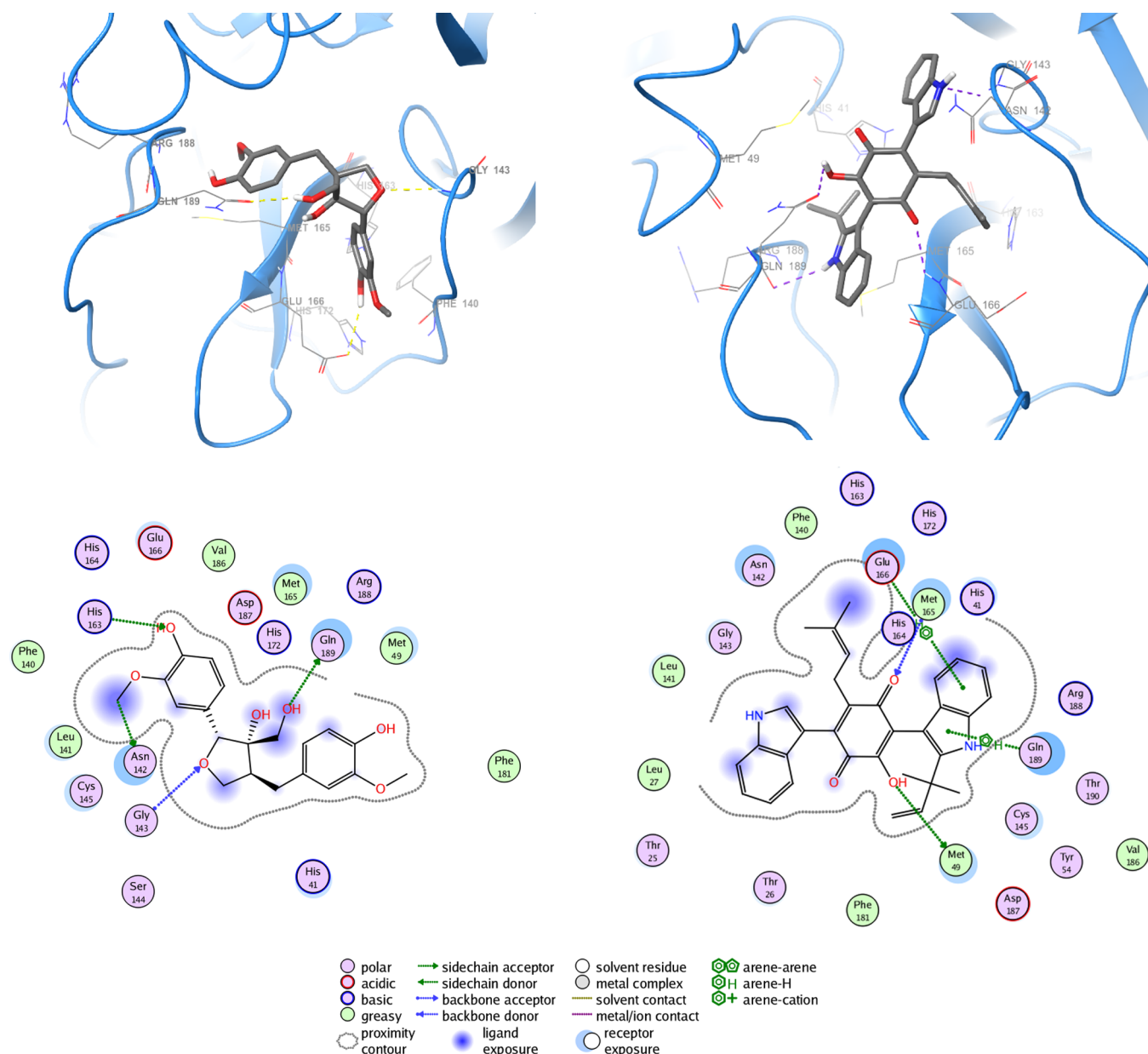
**Spectral Analyses and Reassignments.** A comparison of calculated and experimental NMR chemical shifts allowed us to perform a number of spectral assignments and possible reassignments in the series of studied natural products, as presented in Schemes 2–4. Among those are the reassignments of C-20 and C-33 in asperlicin (10), C-4'' and C-5''; H-2'', H-4'', H-5'', H-11'', H-13'', and H-17'' in mulberrofuran G (22); and C-13 and C-14 together with H-15 and H-21 in itoaic acid (23); see Scheme 2. The reassignments of C-14 and C-20 in iguesterin (20) together with H-18 and H-19 in betulinic acid (17), as well as a number of predicted  $^1\text{H}$  chemical shifts of those two compounds, are presented in Scheme 3. Finally, shown in Scheme 4 are the predicted additional assignments of  $^1\text{H}$  NMR chemical shifts of limonin (18) and  $\alpha$ -copaene (19), which were not reported before this study.

Table 1. Molecular Docking Results for the SARS-CoV-2 M<sup>pro</sup> Inhibitors 1–24

cmpd. <sup>a</sup>	ligand	binding energy, kcal/mol	inhibition constant, $\mu$ M	docking binding energy range, kcal/mol	residue	type of interactions
16	berchemol	-11.33	0.01	-11.33 to -7.90	Asn142, Gly143, His163, Glu166, Gln189	H-bond
					Gly143, Gln189, Gln189	vdW
					Met49, Met165	hydrophobic
					His41	$\pi$ -alkyl
4	terrequinone A	-10.39	0.02	-10.39 to -7.56	Gly143, Glu166, Arg188, Gln189	H-bond
					Thr26, Asn142, His163	vdW
					Met49, Met165	hydrophobic
					His41, Cys145, Glu166, Gln189	$\pi$ -alkyl
17	betulinic acid	-9.73	0.07	-9.73 to -9.35	Thr26, His164, Glu166	H-bond
					Leu141	hydrophobic
					His41	$\pi$ -alkyl
8	2,2-di(3-indolyl)-3-indolone	-9.59	0.09	-9.59 to -8.67	His164	H-bond
					Cys145, Glu166	$\pi$ -alkyl
					His41	$\pi$ -stacking
	native ligand	-9.03	0.24	-9.10 to -4.01	His164, His164, Glu166	H-bond
					His41	$\pi$ -alkyl
9	oriciacridone F	-8.72	0.40	-8.72 to -8.23	Met165, Met165	hydrophobic
24	anabsinthin	-8.65	0.46	-8.65 to -8.62	Asn142	vdW
					Leu141	hydrophobic
12	diosgenin glucoside	-8.56	0.53	-8.56 to -3.80	Glu166	H-bond
					Thr26, Asn119, Gly143	vdW
18	limonin	-8.51	0.58	-8.51 to -8.45	His164	H-bond
					Met165	hydrophobic
22	mulberrofuran G	-8.47	0.62	-8.47 to -7.33	Thr45, His189	vdW
2	baicalein-7-O-glucoside	-8.40	0.69	-8.40 to -5.60	His163, His163	H-bond
					Asn142, Asn142	vdW
5	zeylanone	-8.17	1.02	-8.17 to -7.93	His163, His164, Glu166	H-bond
					Gly143, Cys145, Gln189	vdW
7	cinchonain-1b	-8.13	1.09	-8.13 to -5.23	His164	H-bond
					Ser144	vdW
					Leu141, Met165	hydrophobic
20	iguesterin	-8.08	1.19	-8.08 to -7.98	Glu166	H-bond
23	itoaic acid	-7.76	2.06	-7.76 to -7.51	His163, His172	H-bond
					Asn142	vdW
11	physalin D	-7.62	2.58	-7.62 to -7.37	Asn142, Asn142, Gly143, Cys145, Gln189	vdW
3	procyanidin B2	-7.28	4.63	-7.28 to -4.25	His164	H-bond
					Thr26, Asn142, Gly143, Cys145, Gln189, Gln189	vdW
21	<i>N</i> -feruloyltyramine	-7.26	4.74	-7.26 to -5.23	Leu141	hydrophobic
19	$\alpha$ -copaene	-7.12	6.04	-7.12 to -6.98		
13	strictinin	-6.77	10.90	-6.77 to -3.29	Thr26, Thr26, Ser46, Ser46, Ser144, Cys145	vdW
					Leu141	hydrophobic
1	kaempferol	-6.76	11.13	-6.76 to -6.23	Arg188	H-bond
10	asperlicin	-6.44	18.90	-6.44 to -4.70	His163	H-bond
					Asn142, Gly143, Gln189	vdW
6	carminic acid	-6.22	27.42	-6.22 to -3.86	Glu166, Glu166	H-bond
					Thr24, Thr24, Thr26, Asn142, Gly143	vdW
					Met49	hydrophobic
15	phyllaemblicin B	-6.06	36.22	-6.06 to -0.77	His41	H-bond
					Asn142, Cys145, Gln189	vdW
					Met49	hydrophobic
					His41	$\pi$ -alkyl
14	pedunculagin	-4.66	382.24	-4.66 to -2.54	Glu166	H-bond
					Thr26, Ser46, Cys145	vdW

<sup>a</sup>Compounds arranged in the order of their decreasing binding energy.



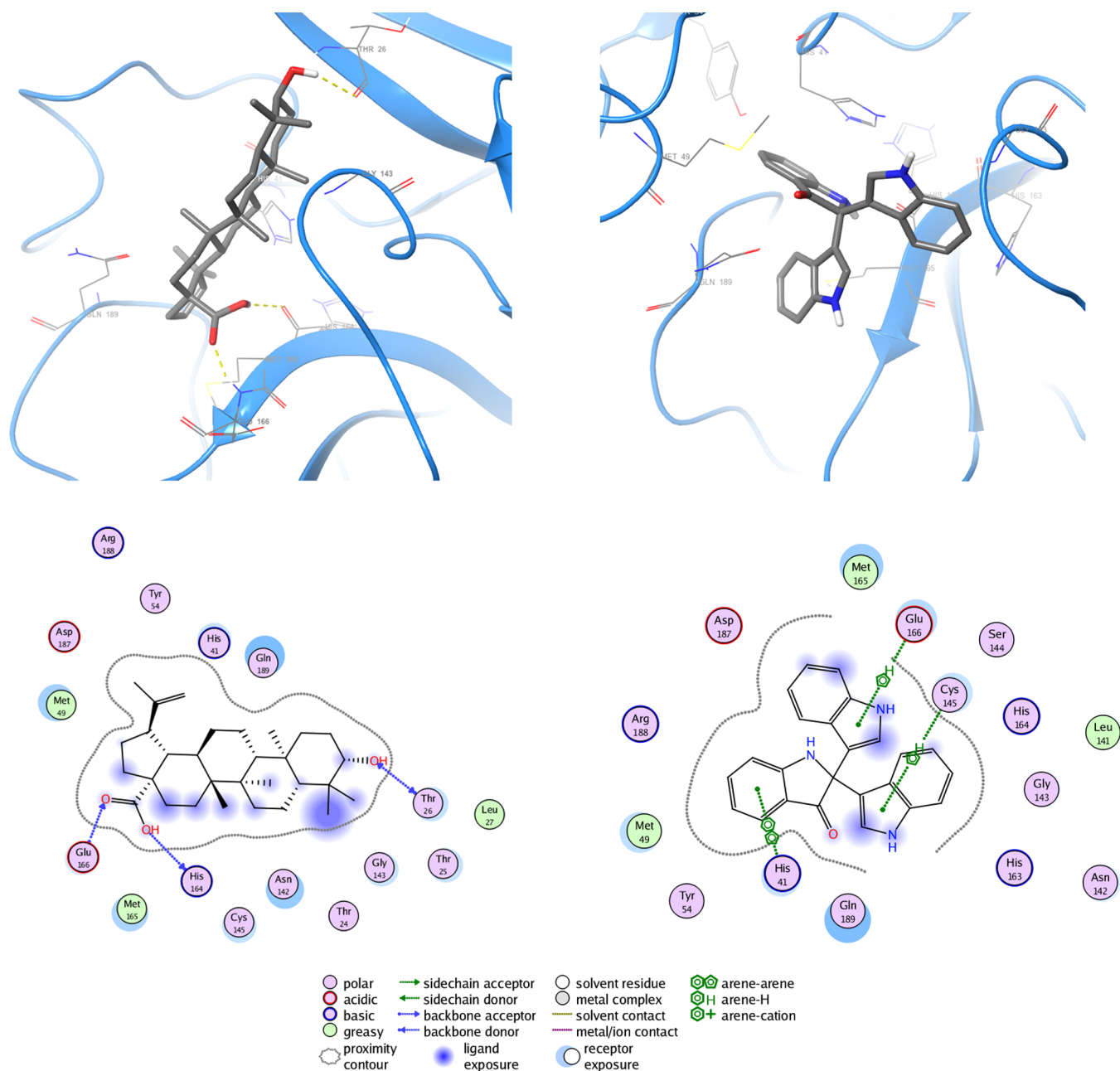


**Figure 3.** Top: docking poses and binding interactions of berchemol (**16**) (left) and terrequinone A (**4**) (right) inside the  $M^{\text{Pro}}$  active site; bottom: 2D interaction diagrams of the same compounds with  $M^{\text{Pro}}$ .

Shown in Figure 2 are the final correlation plots of calculated and experimental  $^1\text{H}$  and  $^{13}\text{C}$  NMR chemical shifts of the studied natural products of diverse chemical origin **1–24**, providing anti-SARS-CoV-2  $M^{\text{Pro}}$  activity. For protons, the RMSD was 0.37 ppm for the range of about 10 ppm, while MAE and CMAE were found to be 0.80 and 0.22 ppm (405 points), respectively. For carbons, RMSD was 3.0 ppm, while MAE and CMAE were 3.1 and 2.1 ppm, respectively, for the range of about 200 ppm and 625 data points. Presented correlations indicate the adequate level of theory [mPW1PW91/6-31G(d,p)] applied for the calculation of  $^1\text{H}$  and  $^{13}\text{C}$  NMR chemical shifts in this series.

**Molecular Docking Simulations.** To analyze structural features, which are critical for binding of the SARS-CoV-2 main protease, we selected 24 compounds (Scheme 1), which had previously demonstrated high  $M^{\text{Pro}}$  inhibition potential. Compounds **1–24** were all docked with the selected three-

dimensional (3D) structure of the  $M^{\text{Pro}}$  protease (PDB ID: 6LU7). The results of these simulations were analyzed based on the ability of the docked ligands to reproduce the N3– $M^{\text{Pro}}$  interaction pattern. Table 1 shows the results of performed calculations in the most ranked binding position arranged in the order of decreasing predicted free binding energy,  $\Delta G$ . It is seen that docking energies in this series of ligands are in the range from  $-11.33$  to  $-4.66$  kcal/mol. In addition, for comparison with the studied ligands, redocking into the SARS-CoV-2  $M^{\text{Pro}}$  was also performed to validate the used docking protocol and to estimate the binding affinity of the native ligand N3 to the catalytic dyad of Cys145–His41. The performed analysis showed that compounds of the selected series are predicted to fit into the  $M^{\text{Pro}}$  active site, repeating the docked pose, which significantly matches the interaction pattern of the native ligand, as illustrated in Figures 3 and 4.



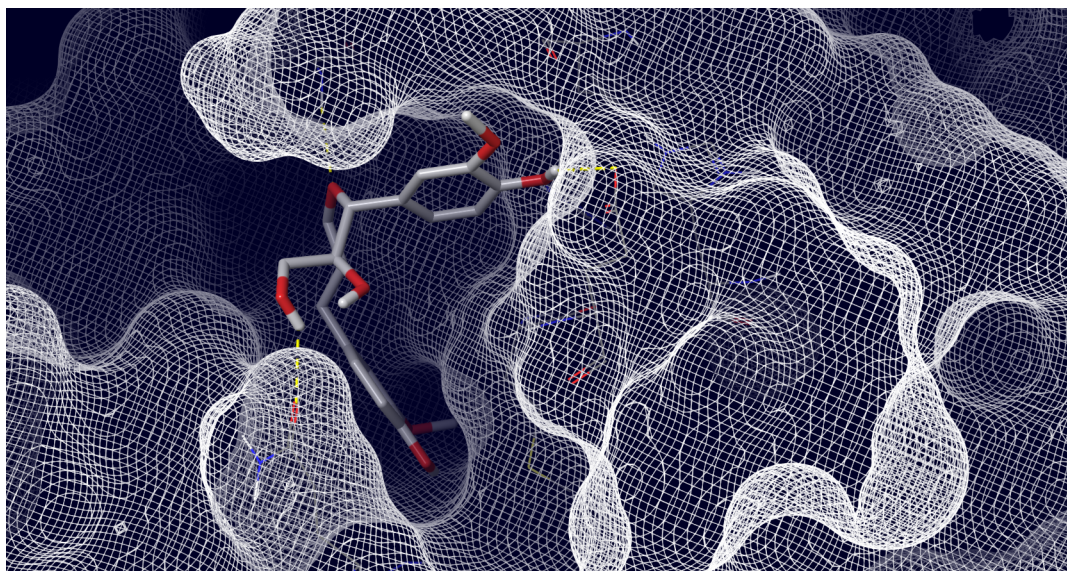
**Figure 4.** Top: docking poses and binding interactions of betulinic acid (**17**) (left) and 2,2-di(3-indolyl)-3-indolone (**8**) (right) inside the M<sup>Pro</sup> active site; bottom: 2D interaction diagrams of the same compounds with M<sup>Pro</sup>.

It followed that four compounds, namely, berchemol (**16**), terrequinone A (**4**), betulinic acid (**17**), and 2,2-di(3-indolyl)-3-indolone (**8**), have a binding energy lower than that of the native ligand N3. As a result, they showed good coordination with the binding affinity of  $-11.33$ ,  $10.39$ ,  $-9.73$ , and  $9.59$  kcal/mol, respectively. Analysis of their interactions with M<sup>Pro</sup> residues showed that they are located deeply within the binding pocket of the M<sup>Pro</sup> active site, being in the cleft between the two protease domains and constituting the catalytic dyad Cys145–His41. This finding indicates the fact that they have the potential to covalently bind to amino acid residues in this region of 6LU7. This ability to interact with a main protease provides additional benefits in suppressing viral activity.

Figures 3 and 4 also demonstrate the extensive interactions between these four compounds and amino acid residues, which

form the binding cavity in 2D diagrams. Binding diagrams for the rest of the studied compounds are given in the Supporting Information. These contacts include hydrogen bonds (HB) and van der Waals,  $\pi$ -alkyl, and  $\pi$ - $\pi$  stacking interactions. In a more extended manner, the main interactions of compounds **1–24** with the M<sup>Pro</sup> active site are presented in Table 1. It should be noted that among the studied series of potential SARS-CoV-2 M<sup>Pro</sup> inhibitors **1–24**, berchemol (**16**) showed the best coordination with the main protease, with the lowest binding energy and, accordingly, the value of inhibition constant.

The molecule of berchemol (**16**) has five HB, including that between the pyrrole hydrogen atom of His163 and oxygen of the hydroxy group of one of the methoxyphenyl moieties. The second HB is formed between the amino acid Glu166 and the hydrogen of the same hydroxy group. The third HB is between



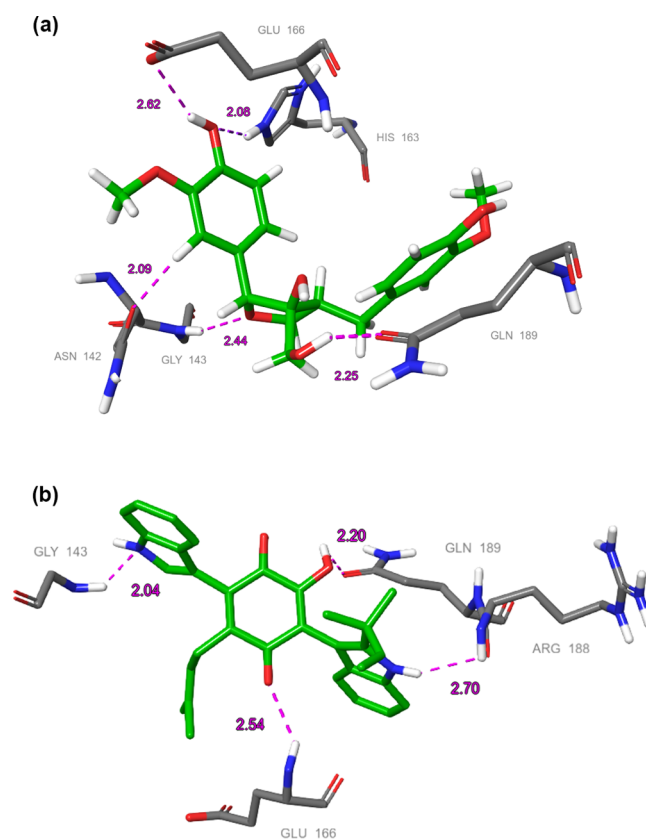
**Figure 5.** Interaction of SARS-CoV-2 M<sup>Pro</sup> with berchemol (**16**), the latter aligned with the active site environment.

the carbonyl oxygen Asn142 and one of the aromatic protons of the methoxyphenyl fragment, while the fourth HB is formed between the oxygen atom of the methoxytetrahydrofuranol ring and the proton of the amino group of Gly143. The fifth HB is located between the hydrogen atom of the hydroxy group of the methoxytetrahydrofuranol cycle and one of the carbonyl oxygen atoms of the Gln189 amino acid. In addition, this compound has two weak  $\pi$ -sulfur interactions between Met49 and Met165 and one of the methoxyphenyl moieties.

Likewise, the amino acid His41 has a  $\pi$ -alkyl contact with the aromatic moiety of berchemol. Other weaker van der Waals interactions are formed by amino acids Cys145, Gly143, Gln189, Asp187, His172, Arg188, and others of the binding pocket. The ligand–receptor coordination of berchemol is illustrated on the left in the 2D diagram in Figure 3 and is shown in general in Figure 5. The detailed H-binding interactions with the amino acid residues Asn142, Gly143, His163, Glu166, and Gln189 of the M<sup>Pro</sup> active site are presented in Figure 6a.

Terrequinone A (**4**) forms the interaction network of four HBs, with the strongest one located between the proton of the amino group of Gly143 and the nitrogen atom of one of the indole fragments, as well as between the hydroxyl hydrogen of the quinone core and the carbonyl oxygen atom of Gln189. The other two weaker HBs belong to the oxygen atom of the quinone and the amine hydrogen of Glu166, as well as the carbonyl oxygen atom of Arg188 and one of the pyrrole protons of terrequinone A. The hydrogen bonding is presented in more detail in Figure 6b. It should be noted that terrequinone A forms two  $\pi$ -alkyl interactions with the catalytic dyad Cys145–His41. There are also  $\pi$ -alkyl interactions with the amino acid residues Glu166 and Gln189. Terrequinone A forms other non-covalent contacts of the hydrophobic type with Met49 and Met165. The rest van der Waals-type interactions are observed with Thr26, Asn142, His163, and other amino acids of M<sup>Pro</sup> binding site.

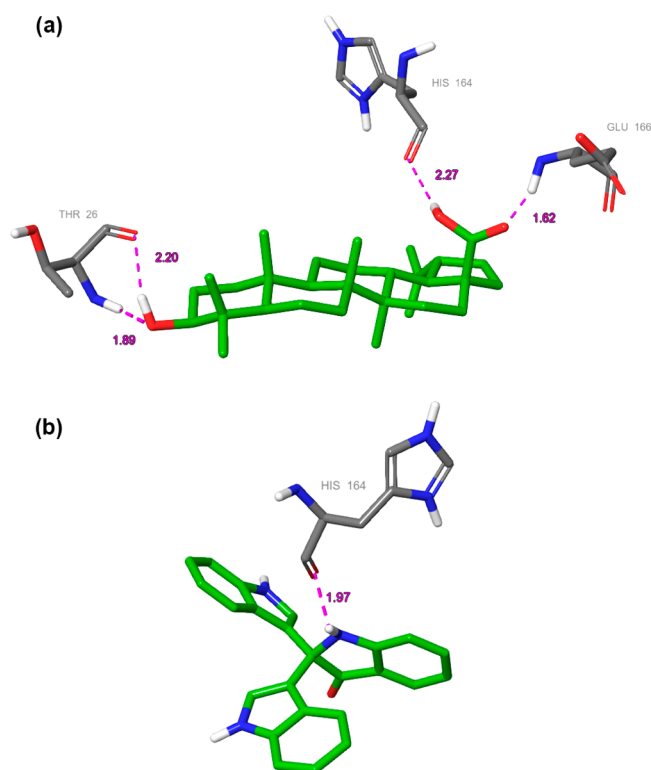
As for betulinic acid (**17**), as in the case of the previous compound, there are four HBs, two of which are formed by the amino group and a carbonyl oxygen atom of Thr26 with the hydroxy group of the dimethylcyclohexyl fragment; see Figure 7a. The third strongest HB is formed between the hydrogen atom of the amino group of Glu166 and the carboxyl oxygen atom of betulinic acid. The last HB is located between the



**Figure 6.** Hydrogen bonding of (a) berchemol (**16**) and (b) terrequinone A (**4**) with the SARS CoV-2 M<sup>Pro</sup> (PDB ID: 6LU7). HB lengths with key amino acids are given in Angstroms and shown as dashed purple lines.

proton of the same carboxyl group and the carbonyl oxygen atom of His164. It should be noted that  $\pi$ -alkyl interactions are formed in this molecule only by His41. Among the nonpolar interactions, the influence of leucine 141 can be noted.

In the molecule of 2,2-di(3-indolyl)-3-indolone (**8**), only one HB is observed, namely, that between the carbonyl oxygen of His164 and the pyrrolidine hydrogen atom of the indolone core



**Figure 7.** Hydrogen bonding of (a) betulinic acid (17) and (b) 2,2-di(3-indolyl)-3-indolone (8) with the SARS-CoV-2  $M^{\text{Pro}}$  (PDB ID: 6LU7). HB lengths with key amino acids are given in Angstroms and shown as dashed purple lines.

of 8, as can be seen in Figure 7b. Here, it should be noted that the stabilization of the considered complex originates mainly in the formation of two  $\pi$ -alkyl contacts between indolyl residues and Cys145 and Glu166 simultaneously. On the other hand, it is due to the formation of the attractive interaction of the  $\pi$ -stacking between the aromatic systems of the imidazole moiety of the His41 and the benzyl ring of indolone (8).

It is interesting to note that in all cases considered herewith (as it has already been indicated by many authors), the base for the stabilization of potential inhibitors in the  $M^{\text{Pro}}$  binding pocket is a distributed network of HBs, predominantly those

involving residues Gly143, His163, Glu166, and Gln189. On the other hand, it is a non-valence coordination of the aromatic systems of ligands and the catalytic dyad Cys145–His41 within the formation of stable  $\pi$ -contacts.

Finally, in a more detailed presentation, the frequencies (by way of the corresponding fractions) of the main types of interactions of the active site residues with ligands of the studied series 1–24 (namely, HBs together with the van der Waals, hydrophobic, and  $\pi$ -alkyl interactions) are shown in the share diagram in Figure 8.

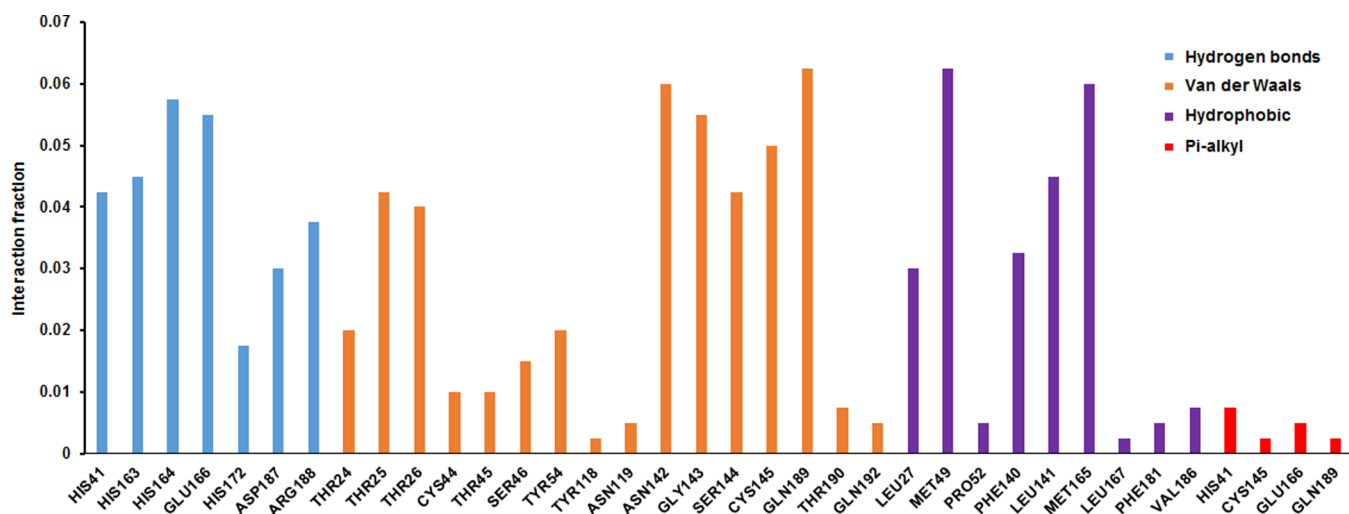
## CONCLUSIONS

In the background of the ongoing COVID-19 pandemic, quantum-chemical methods used for the elucidation of chemical structures can be effectively implemented to the target potential inhibitors of this virus, accelerating the development of appropriate drugs. In the present study, we proposed and used a combined approach, based on the modeling of NMR chemical shifts and molecular docking to identify the potential SARS-CoV-2  $M^{\text{Pro}}$  inhibitors among a series of natural products of diverse origins.

The well-known challenge related to the incorrect initial assignment of the stereochemical structure was solved in this study within the framework of the formalism of the ANN, which allows, based on the correlations of calculated and experimental chemical shifts, to determine the *true* stereochemistry of the molecule with high reliability. The modeling of spectral NMR parameters, performed in the present study, made it possible to predict potential NMR reassignments in some of the studied natural products.

The molecular docking of the established structures was used to study binding interactions in the active site of the main protease. For several compounds with high binding energies, the nonpolar and electrostatic interactions were found to stabilize these ligands in the binding pocket. Undoubtedly, one of the main roles was played by the distributed networks of the ligand–receptor HBs.

As a result of this study, it was found that berchemol, which is a natural extract from the seeds of *C. cyanus*, showed the best affinity with the SARS-CoV-2 main protease. Several other tested natural products, such as terrequinone A, betulinic acid, and 2,2-di(3-indolyl)-3-indolone, also showed better binding



**Figure 8.** Distribution share diagram of different types of interactions of the  $M^{\text{Pro}}$  residues with compounds 1–24.

affinity with the protease than the native ligand. Selected potential inhibitor candidates identified in this work showed improved interaction energies in relation to SARS-CoV-2 M<sup>PRO</sup> and have increased specificity due to additional HBs with the active site residues. Presented results are expected to stimulate further research aimed toward the development of specific drugs against COVID-19.

## DATA AND SOFTWARE AVAILABILITY

The PDB files were downloaded from the RCSB protein data bank (<https://rcsb.org>). All structures of the investigated compounds have been provided in the main text of the article and in the **Supporting Information**. The following free and commercial software packages were used in this article: Gaussian 09 (<http://gaussian.com>), Dalton 2016 (<https://daltonprogram.org>), Schrödinger Maestro 11.5 (<https://schrodinger.com/products/maestro>), MOE 2014 (<https://chemcomp.com>), AutoDockTools 1.5.7 (<http://autodock.scripps.edu/resources/adt>), and ANN-PRA workflow (<https://rsc.org/suppdata/ob/c3/c3ob40843d/c3ob40843d.xls>).

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpcc.1c10489>.

Cartesian coordinates (Angstroms) of **1–24** optimized at the M06-2X/cc-pVTZ//aug-cc-pVTZ level in the liquid phase, modulated within the IEF-PCM scheme. 2D interaction diagrams of compounds **1–24** (PDF)

## AUTHOR INFORMATION

### Corresponding Authors

Valentin A. Semenov – A. E. Favorsky Irkutsk Institute of Chemistry, Siberian Branch of the Russian Academy of Sciences, 664033 Irkutsk, Russia; [orcid.org/0000-0002-8076-2245](https://orcid.org/0000-0002-8076-2245); Email: [semenovval@inbox.ru](mailto:semenovval@inbox.ru)

Leonid B. Krivdin – A. E. Favorsky Irkutsk Institute of Chemistry, Siberian Branch of the Russian Academy of Sciences, 664033 Irkutsk, Russia; [orcid.org/0000-0003-2941-1084](https://orcid.org/0000-0003-2941-1084); Email: [krivdin55@gmail.com](mailto:krivdin55@gmail.com)

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.jpcc.1c10489>

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

All calculations were performed at the Irkutsk Supercomputer Center of the Siberian Branch of the Russian Academy of Sciences using the HPC cluster “Academician V.M. Matrosov” (<http://hpc.icc.ru>) and at the A. E. Favorsky Irkutsk Institute of Chemistry using the facilities of the Baikal Analytical Center (<http://ckp-ir.ru/ckp/3050>).

## REFERENCES

- (1) Zumla, A.; Chan, J. F. W.; Azhar, E. I.; Hui, D. S. C.; Yuen, K.-Y. Coronaviruses – drug discovery and therapeutic options. *Nat. Rev. Drug Discovery* **2016**, *15*, 327–347.
- (2) de Wit, E.; van Doremalen, N.; Falzarano, D.; Munster, V. J. SARS and MERS: recent insights into emerging coronaviruses. *Nat. Rev. Microbiol.* **2016**, *14*, 523–534.

- (3) Xu, J.; Zhao, S.; Teng, T.; Abdalla, A. E.; Zhu, W.; Xie, L.; Wang, Y.; Guo, X. Systematic Comparison of Two Animal-to-Human Transmitted Human Coronaviruses: SARS-CoV-2 and SARS-CoV. *Viruses* **2020**, *12*, 244.

- (4) Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C.; Duan, Y.; et al. Structure of M<sup>PRO</sup> from SARS-CoV-2 and discovery of its inhibitors. *Nature* **2020**, *582*, 289–293.

- (5) Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.; Sauerhering, L.; Becker, S.; Rox, K.; Hilgenfeld, R. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved  $\alpha$ -ketoamide inhibitors. *Science* **2020**, *368*, 409–412.

- (6) Jiménez-Avalos, G.; Vargas-Ruiz, A. P.; Delgado-Pease, N. E.; Olivos-Ramirez, G. E.; Sheen, P.; Fernández-Díaz, M.; Quiliano, M.; Zimic, M. COVID-19 Working Group in Perú. Comprehensive virtual screening of 4.8 k flavonoids reveals novel insights into allosteric inhibition of SARS-CoV-2 M<sup>PRO</sup>. *Sci. Rep.* **2021**, *11*, 15452.

- (7) Ngo, S. T.; Quynh Anh Pham, N.; Thi Le, L.; Pham, D.-H.; Vu, V. V. Computational Determination of Potential Inhibitors of SARS-CoV-2 Main Protease. *J. Chem. Inf. Model.* **2020**, *60*, 5771–5780.

- (8) Rutwick Surya, U.; Praveen, N. A molecular docking study of SARS-CoV-2 main protease against phytochemicals of Boerhavia diffusa Linn. for novel COVID-19 drug discovery. *VirusDis.* **2021**, *32*, 46–54.

- (9) Gupta, M. K.; Vemula, S.; Donde, R.; Gouda, G.; Behera, L.; Vadde, R. In-silico approaches to detect inhibitors of the human severe acute respiratory syndrome coronavirus envelope protein ion channel. *J. Biomol. Struct. Dyn.* **2020**, *39*, 2617–2627.

- (10) Sepay, N.; Sekar, A.; Halder, U. C.; Alarifi, A.; Afzal, M. Anti-COVID-19 terpenoid from marine sources: A docking, admet and molecular dynamics study. *J. Mol. Struct.* **2021**, *1228*, 129433.

- (11) Aanouz, I.; Belhassan, A.; El-Khatibi, K.; Lakhlifi, T.; El-Ldrissi, M.; Bouachrine, M. Moroccan Medicinal plants as inhibitors against SARS-CoV-2 main protease: Computational investigations. *J. Biomol. Struct. Dyn.* **2021**, *39*, 2971–2979.

- (12) Manandhar, A.; Blass, B. E.; Colussi, D. J.; Almi, I.; Abou-Gharbia, M.; Klein, M. L.; Elokely, K. M. Targeting SARS-CoV-2 M3CLpro by HCV NS3/4a Inhibitors: In Silico Modeling and In Vitro Screening. *J. Chem. Inf. Model.* **2021**, *61*, 1020–1032.

- (13) Yoshino, R.; Yasuo, N.; Sekijima, M. Identification of Key Interactions between SARS-CoV-2 Main Protease and Inhibitor Drug Candidates. *Sci. Rep.* **2020**, *10*, 12493.

- (14) Abdusalam, A. A. A.; Murugaiyah, V. Identification of Potential Inhibitors of 3CL Protease of SARS-CoV-2 From ZINC Database by Molecular Docking-Based Virtual Screening. *Front. Mol. Biosci.* **2020**, *7*, 603037.

- (15) Paul, A. S.; Islam, R.; Parves, M. R.; Al Mamun, A.; Shahriar, I.; Hossain, M. I.; Hossain, M. N.; Ali, M. A.; Halim, M. A. Cysteine Focused Covalent Inhibitors against the Main Protease of SARS-CoV-2. *J. Biomol. Struct. Dyn.* **2020**, *40*, 1639.

- (16) Frecer, V.; Miertus, S. Antiviral agents against COVID-19: structure-based design of specific peptidomimetic inhibitors of SARS-CoV-2 main protease. *RSC Adv.* **2020**, *10*, 40244–40263.

- (17) Amendola, G.; Ettari, R.; Previti, S.; Di Chio, C.; Messere, A.; Di Maro, S.; Hammerschmidt, S. J.; Zimmer, C.; Zimmermann, R. A.; Schirmeister, T.; et al. Lead Discovery of SARS-CoV-2 Main Protease Inhibitors through Covalent Docking-Based Virtual Screening. *J. Chem. Inf. Model.* **2021**, *61*, 2062–2073.

- (18) Deeks, H. M.; Walters, R. K.; Barnoud, J.; Glowacki, D. R.; Mulholland, A. J. Interactive Molecular Dynamics in Virtual Reality Is an Effective Tool for Flexible Substrate and Inhibitor Docking to the SARS-CoV-2 Main Protease. *J. Chem. Inf. Model.* **2020**, *60*, 5803–5814.

- (19) Sheik Amamuddy, O.; Verkhivker, G. M.; Tastan Bishop, Ö. Impact of Early Pandemic Stage Mutations on Molecular Dynamics of SARS-CoV-2 Mpro. *J. Chem. Inf. Model.* **2020**, *60*, 5080–5102.

- (20) Balaramnavar, V. M.; Ahmad, K.; Saeed, M.; Ahmad, I.; Kamal, M.; Jawed, T. Pharmacophore-based approaches in the rational repurposing technique for FDA approved drugs targeting SARS-CoV-2 Mpro. *RSC Adv.* **2020**, *10*, 40264–40275.

- (21) Abel, R.; Paredes Ramos, M.; Chen, Q.; Pérez-Sánchez, H.; Coluzzi, F.; Rocco, M.; Marchetti, P.; Mura, C.; Simmaco, M.; Bourne, P. E.; et al. Computational Prediction of Potential Inhibitors of the Main Protease of SARS-CoV-2. *Front. Chem.* **2020**, *8*, 590263.
- (22) Gahlawat, A.; Kumar, N.; Kumar, R.; Sandhu, H.; Singh, I. P.; Singh, S.; Sjöstedt, A.; Garg, P. Structure-Based Virtual Screening to Discover Potential Lead Molecules for the SARS-CoV-2 Main Protease. *J. Chem. Inf. Model.* **2020**, *60*, 5781–5793.
- (23) Gupta, A.; Zhou, H.-X. Profiling SARS-CoV-2 Main Protease (MPRO) Binding to Repurposed Drugs Using Molecular Dynamics Simulations in Classical and Neural Network-Trained Force Fields. *ACS Comb. Sci.* **2020**, *22*, 826–832.
- (24) Keretsu, S.; Bhujbal, S. P.; Cho, S. J. Rational approach toward COVID-19 main protease inhibitors via molecular docking, molecular dynamics simulation and free energy calculation. *Sci. Rep.* **2020**, *10*, 17716.
- (25) Sauer, S. P. A. *Molecular Electromagnetism: A Computational Chemistry Approach*; University Press: Oxford, 2012.
- (26) Krivdin, L. B. Computational  $^1\text{H}$  NMR: Part 2. Chemical applications. *Magn. Reson. Chem.* **2020**, *58*, 5–14.
- (27) Krivdin, L. B. Computational  $^1\text{H}$  NMR: Part 3. Biochemical studies. *Magn. Reson. Chem.* **2020**, *58*, 15–30.
- (28) Krivdin, L. B. Computational protocols for  $^{13}\text{C}$  NMR chemical shifts. *Prog. NMR Spectrosc.* **2019**, *112–113*, 103–156.
- (29) Krivdin, L. B. Computational NMR of carbohydrates: Theoretical background, applications, and perspectives. *Molecules* **2021**, *26*, 2450.
- (30) Semenov, V. A.; Krivdin, L. B. DFT computational schemes for  $^1\text{H}$  and  $^{13}\text{C}$  NMR chemical shifts of natural products, exemplified by strychnine. *Magn. Reson. Chem.* **2020**, *58*, 56–64.
- (31) Semenov, V. A.; Samultsev, D. O.; Krivdin, L. B.  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra of strychnos alkaloids: Selected NMR updates. *Int. J. Quant. Chem.* **2020**, *120*, No. e26348.
- (32) Semenov, V. A.; Samultsev, D. O.; Krivdin, L. B. The  $^1\text{H}$  and  $^{13}\text{C}$  NMR chemical shifts of Strychnos alkaloids revisited at the DFT level. *Magn. Reson. Chem.* **2020**, *58*, 532–539.
- (33) Semenov, V. A.; Krivdin, L. B. Computational  $^1\text{H}$  and  $^{13}\text{C}$  NMR of strychnobailonine: On the way to larger molecules calculated at lower computational costs. *Magn. Reson. Chem.* **2021**, *59*, 108–116.
- (34) Semenov, V. A.; Krivdin, L. B. Computational  $^1\text{H}$  and  $^{13}\text{C}$  NMR of the trimeric monoterpene indole alkaloid strychnohexamine: Selected spectral updates. *Magn. Reson. Chem.* **2021**, *59*, 691–700.
- (35) Semenov, V. A.; Krivdin, L. B. Benchmark density functional theory calculations of  $^{13}\text{C}$  NMR chemical shifts of the natural antimalarial compounds with a new basis set 3z-S. *Int. J. Quant. Chem.* **2021**, *121*, No. e26639.
- (36) Joshi, R. S.; Jagdale, S. S.; Bansode, S. B.; Shankar, S. S.; Tellis, M. B.; Pandya, V. K.; Chugh, A.; Giri, A. P.; Kulkarni, M. J. Discovery of potential multi-target-directed ligands by targeting host-specific SARS-CoV-2 structurally conserved main protease. *J. Biomol. Struct. Dyn.* **2021**, *39*, 3099–3114.
- (37) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al. *Gaussian 09*, Revision C.01; Gaussian, Inc.: Wallingford CT, USA, 2009; <http://www.gaussian.com>.
- (38) Tomasi, J.; Mennucci, B.; Cancès, E. The IEF version of the PCM solvation method: an overview of a new method addressed to study molecular solutes at the QM ab initio level. *J. Mol. Struct.: THEOCHEM* **1999**, *464*, 211–226.
- (39) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum mechanical continuum solvation models. *Chem. Rev.* **2005**, *105*, 2999–3094.
- (40) Aidas, K.; Angeli, C.; Bak, K. L.; Bakken, V.; Bast, R.; Boman, L.; Christiansen, O.; Cimiraglia, R.; Coriani, S.; Dahle, P.; et al. *DALTON, a Molecular Electronic Structure Program, Release Dalton 2016*, 2016; WIREs Comput. Mol. Sci. 2014, *4*, 269–284; <http://daltonprogram.org>.
- (41) Adamo, C.; Barone, V. Exchange functionals with improved long-range behavior and adiabatic connection methods without adjustable parameters: The mPW and mPW1PW models. *J. Chem. Phys.* **1998**, *108*, 664–675.
- (42) Frisch, M. J.; Pople, J. A.; Binkley, J. S. Self-consistent molecular orbital methods 25. Supplementary functions for Gaussian basis sets. *J. Chem. Phys.* **1984**, *80*, 3265–3269.
- (43) Harris, R. K.; Becker, E. D.; Cabral de Menezes, S. M.; Granger, P.; Hoffman, R. E.; Zilm, K. W. Further conventions for NMR shielding and chemical shifts. *Pure Appl. Chem.* **2008**, *80*, 59–84.
- (44) Protein Data Bank, <https://www.rcsb.org> (accessed December 21, 2021).
- (45) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. Autodock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *30*, 2785–2791.
- (46) Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity – a rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219–3228.
- (47) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639–1662.
- (48) *Schrödinger Release 2018-1*; Maestro, Schrödinger, LLC: New York, NY, 2018; <https://www.schrodinger.com/freemaestro>.
- (49) *Molecular Operating Environment (MOE) 2014.09*; Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2014.
- (50) Sarotti, A. M. Successful combination of computationally inexpensive GIAO  $^{13}\text{C}$  NMR calculations and artificial neural network pattern recognition: a new strategy for simple and rapid detection of structural misassignments. *Org. Biomol. Chem.* **2013**, *11*, 4847–4859.
- (51) Itoh, T.; Ninomiya, M.; Yasuda, M.; Koshikawa, K.; Deyashiki, Y.; Nozawa, Y.; Akao, Y.; Koketsu, M. Inhibitory effects of flavonoids isolated from *Fragaria ananassa* Duch on IgE-mediated degranulation in rat basophilic leukemia RBL-2H3. *Bioorg. Med. Chem.* **2009**, *17*, 5374–5379.
- (52) Chen, L.-J.; Games, D. E.; Jones, J. Isolation and identification of four flavonoid constituents from the seeds of *Oroxylum indicum* by high-speed counter-current chromatography. *J. Chromatogr. A* **2003**, *988*, 95–105.
- (53) Yuan, Y.; Hou, W.; Tang, M.; Luo, H.; Chen, L.-J.; Guan, Y. H.; Sutherland, I. A. Separation of flavonoids from the leaves of *oroxylum indicum* by HSCCC. *Chromatographia* **2008**, *68*, 885–892.
- (54) Lokman Khan, M.; Haslam, E.; Williamson, M. P. Structure and conformation of the procyanidin B-2 dimer. *Magn. Reson. Chem.* **1997**, *35*, 854–858.
- (55) He, J.; Wijeratne, E. M. K.; Bashyal, B. P.; Zhan, J.; Seliga, C. J.; Liu, M. X.; Pierson, E. E.; Pierson, L. S., III; VanEtten, H. D.; Gunatilaka, A. A. L. Cytotoxic and other metabolites of *aspergillus* inhabiting the rhizosphere of sonoran desert plants. *J. Nat. Prod.* **2004**, *67*, 1985–1991.
- (56) Gu, J.-Q.; Graf, T. N.; Lee, D.; Chai, H.-B.; Mi, Q.; Kardono, L. B. S.; Setyowati, F. M.; Ismail, R.; Riswan, S.; Farnsworth, N. R.; et al. Cytotoxic and antimicrobial constituents of the bark of *Diospyros maritima* collected in two geographical locations in Indonesia. *J. Nat. Prod.* **2004**, *67*, 1156.
- (57) Overeem, J. C.; van der Kerk, G. J. M. Mollisin, a naturally occurring chlorine-containing quinone: Part IV. Revised structures for cochenillic acid and for the insect pigments, carminic and kermesic acids. *Recl. Trav. Chim. Pays-Bas* **1964**, *83*, 1023–1035.
- (58) Schmitt, P.; Günther, H.; Hägele, G.; Stilke, R. A  $^1\text{H}$  and  $^{13}\text{C}$  NMR study of carminic acid. *Org. Magn. Reson.* **1984**, *22*, 446–449.
- (59) Nonaka, G.; Nishioka, I. Tannins and related compounds. VII. Phenylpropanoid-substituted epicatechins, cinchonins from *Cinchona succirubra*. (1). *Chem. Pharm. Bull.* **1982**, *30*, 4268.
- (60) Kobayashi, M.; Aoki, S.; Gato, K.; Matsunami, K.; Kurosu, M.; Kitagawa, I. Marine natural products. XXXIV. Trisindoline, a new antibiotic indole trimer, produced by a bacterium of *vibrio* sp. separated from the marine sponge *Hyrtios altum*. *Chem. Pharm. Bull.* **1994**, *42*, 2449–2451.

- (61) Song, J.-G.; Su, J.-C.; Song, Q.-Y.; Huang, R.-L.; Tang, W.; Hu, L.-J.; Huang, X.-J.; Jiang, R.-W.; Li, Y.-L.; Ye, W.-C.; et al. Cleistocalones A and B, antiviral phloroglucinol-terpenoid adducts from *Cleistocalyx operculatus*. *Org. Lett.* **2019**, *21*, 9579–9583.
- (62) Wansi, J. D.; Wandji, J.; Mbase Meva'a, L.; Kamdem Waffo, A. F.; Ranjit, R.; Khan, S. N.; Asma, A.; Iqbal, C. M.; Lallemand, M.-C.; Tillequin, F.; et al.  $\alpha$ -Glucosidase inhibitory and antioxidant acridone alkaloids from the stem bark of *oriciopsis glaberrima* ENGL. (Rutaceae). *Chem. Pharm. Bull.* **2006**, *54*, 292–296.
- (63) Chang, R.; Lotti, V.; Monaghan, R.; Birnbaum, J.; Stapley, E.; Goetz, M.; Albers-Schonberg, G.; Patchett, A.; Liesch, J.; Hensens, O.; et al. A potent nonpeptide cholecystokinin antagonist selective for peripheral tissues isolated from *Aspergillus alliaceus*. *Science* **1985**, *230*, 177–179.
- (64) Sun, H. H.; Byard, S. J.; Cooper, R. Revised NMR assignments for the cholecystokinin antagonist asperlicin. *J. Antibiot.* **1994**, *47*, 599–601.
- (65) Januário, A. H.; Filho, E. R.; Pietro, R. C.; Kashima, S.; Sato, D. N.; França, S. C. Antimycobacterial physalins from *Physalis angulata* L. (Solanaceae). *Phytother. Res.* **2002**, *16*, 445.
- (66) Feng, B.; Kang, L.-P.; Ma, B.-P.; Quan, B.; Zhou, W.-B.; Wang, Y.-Z.; Zhao, Y.; Liu, Y.-X.; Wang, S.-q. The substrate specificity of a glucoamylase with steroidal saponin-rhamnosidase activity from *Curvularia lunata*. *Tetrahedron* **2007**, *63*, 6796–6812.
- (67) Yagi, K.; Goto, K.; Nanjo, F. Identification of a major polyphenol and polyphenolic composition in leaves of *Camellia irrawadiensis*. *Chem. Pharm. Bull.* **2009**, *57*, 1284–1288.
- (68) Satomi, H.; Umemura, K.; Ueno, A.; Hatano, T.; Okuda, T.; Noro, T. Carbonic anhydrase inhibitors from the pericarps of *Punica granatum* L. *Biol. Pharm. Bull.* **1993**, *16*, 787–790.
- (69) Tanaka, T.; Nonaka, G.-I.; Nishioka, I. Tannins and related compounds. XL. Revision of the structures of punicalin and punicalagin, and isolation and characterization of 2-O-galloylpunicalin from the bark of *Punica granatum* L. *Chem. Pharm. Bull.* **1986**, *34*, 650–655.
- (70) Zhang, Y.-J.; Tanaka, T.; Iwamoto, Y.; Yang, C.-R.; Kouno, I. Novel norsesquiterpenoids from the roots of *Phyllanthus emblica*. *J. Nat. Prod.* **2000**, *63*, 1507–1510.
- (71) Shoeb, M.; Jaspars, M.; MacManus, S. M.; Majinda, R. R. T.; Sarker, S. D. Epoxy lignans from the seeds of *Centaurea cyanus* (Asteraceae). *Biochem. Syst. Ecol.* **2004**, *32*, 1201–1204.
- (72) Siddiqui, S.; Hafeez, F.; Begum, S.; Siddiqui, B. S. Oleanderol, a new pentacyclic triterpene from the leaves of *Nerium oleander*. *J. Nat. Prod.* **1988**, *51*, 229–233.
- (73) Bisoli, E.; Garcez, W.; Hamerski, L.; Tieppo, C.; Garcez, F. Bioactive pentacyclic triterpenes from the stems of *Combretum laxum*. *Molecules* **2008**, *13*, 2717–2728.
- (74) Min, Y. D.; Kwon, H. C.; Yang, M. C.; Lee, K. H.; Choi, S. U.; Lee, K. R. Isolation of limonoids and alkaloids from *Phellodendron amurense* and their multidrug resistance (MDR) reversal activity. *Arch. Pharm. Res.* **2007**, *30*, 58.
- (75) Nishida, R.; Shelly, T. E.; Whittier, T. S.; Kaneshiro, K. Y.  $\alpha$ -Copaene, a potential rendezvous cue for the Mediterranean fruit fly, *Ceratitidis capitata*? *J. Chem. Ecol.* **2000**, *26*, 87.
- (76) Wenkert, E.; Bookser, B. C.; Arrhenius, T. S. Total syntheses of ( $\pm$ )- $\alpha$ - and ( $\pm$ )- $\beta$ -copaene and formal total syntheses of ( $\pm$ )-sativene, ( $\pm$ )-cis-sativenediol, and ( $\pm$ )-helminthosporal. *J. Am. Chem. Soc.* **1992**, *114*, 644.
- (77) Ryu, Y. B.; Park, S.-J.; Kim, Y. M.; Lee, J.-Y.; Seo, W. D.; Chang, J. S.; Park, K. H.; Rho, M.-C.; Lee, W. S. SARS-CoV 3CL<sup>pro</sup> inhibitory effects of quinone-methide triterpenes from *Tripterygium regelii*. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 1873–1876.
- (78) Tahir ul Qamar, M.; Alqahtani, S. M.; Alamri, M. A.; Chen, L.-L. Structural basis of SARS-CoV-2 3CL<sup>pro</sup> and anti-COVID-19 drug discovery from medicinal plants. *J. Pharm. Anal.* **2020**, *10*, 313.
- (79) Park, J. B. Isolation and characterization of N-feruloyltyramine as the P-selectin expression suppressor from garlic (*Allium sativum*). *J. Agric. Food Chem.* **2009**, *57*, 8868–8872.
- (80) Fukai, T.; Hano, Y.; Hirakura, K.; Nomura, T.; Uzawa, J.; Fukushima, K. Structures of two natural hypotensive Diels-Alder Type adducts, malberrofuranes F and G, from the cultivated mulberry tree (*Morus lhou Koidz*). *Chem. Pharm. Bull.* **1985**, *33*, 3195.
- (81) Chai, X.-Y.; Xu, Z.-R.; Bai, C.-C.; Zhou, F.-R.; Tu, P.-F. A new seco-friedelolactone acid from the bark and twigs of *Itoa orientalis*. *Fitoterapia* **2009**, *80*, 408–410.
- (82) Aberham, A.; Cicek, S. S.; Schneider, P.; Stuppner, H. Analysis of sesquiterpene lactones, lignans, and flavonoids in wormwood (*Artemisia absinthium* L.) using high-performance liquid chromatography (HPLC)-mass spectrometry, reversed phase HPLC, and HPLC-solid phase extraction-nuclear magnetic resonance. *J. Agric. Food Chem.* **2010**, *58*, 10817–10823.