Computational Approaches for Identifying Bioactive Compounds Inhibiting SARS-CoV-2 Main Protease

Miss Piyatida Pojtanadithee

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science in Bioinformatics and Computational Biology (Interdisciplinary Program)

Inter-Department of Bioinformatics and Computational Biology

GRADUATE SCHOOL

Chulalongkorn University

Academic Year 2022

วิธีการเชิงคอมพิวเตอร์สำหรับการระบุสารออกฤทธิ์ทางชีวภาพที่ยับยั้งเอนไซม์
โปรตีเอสหลักของไวรัสซาร์ส-โควี-2

น.ส.ปิยธิดา พจน์ธนาดิถี

Thesis Title   Computational Approaches for Identifying Bioactive

                                Compounds Inhibiting SARS-CoV-2 Main Protease

By   Miss Piyatida Pojtanadithee

Field of Study   Bioinformatics and Computational Biology

                                (Interdisciplinary Program)

Thesis Advisor   Associate Professor Dr. THANYADA RUNGROTMONGKOL

Thesis Co Advisor   Associate Professor Dr. PHORNPHIMON MAITARAD

        Accepted by the GRADUATE SCHOOL, Chulalongkorn University in Partial
Fulfillment of the Requirement for the Master of Science

                                                           Dean of the GRADUATE SCHOOL
        (Associate Professor Dr. YOOTTHANA CHUPPUNNARAT)

THESIS COMMITTEE

                                                           Chairman
        (Professor Dr. SUPACHITRA CHADCHAWAN)

                                                           Thesis Advisor
        (Associate Professor Dr. THANYADA RUNGROTMONGKOL)

                                                           Thesis Co-Advisor
        (Associate Professor Dr. PHORNPHIMON MAITARAD)

                                                           Examiner
        (Assistant Professor Dr. NATAPOL PORNPUTTAPONG)

                                                           External Examiner
        (Associate Professor Dr. NADTANET NUNTHABOOT)

ปิยธิดา พจน์ธนาดิถี : วิธีการเชิงคอมพิวเตอร์สำหรับการระบุสารออกฤทธิ์ทางชีวภาพที่ยับยั้งเอนไซม์โปรตีเอสหลักของไวรัสซาร์ส-โควี-2. (Computational Approaches for Identifying Bioactive Compounds Inhibiting SARS-CoV-2 Main Protease) อ.ที่ปรึกษาหลัก : รศ. ดร.ธัญญดา รุ่งโรจน์มงคล, อ.ที่ปรึกษาร่วม : รศ. ดร.พรพิมล ไมตรัตน์

การเกิดโรคติดเชื้อจากไวรัสโคโรนา 2019 มีผลกระทบอย่างรุนแรงต่อสังคมและเศรษฐกิจทั่วโลก การต่อสู้กับโรคนี้ยังคงเผชิญกับข้อจำกัดต่างๆ เช่น ต้องใช้ระยะเวลาและทรัพยากรจำนวนมากในการค้นหาตัวยาที่สามารถยับยั้งการแพร่กระจายของเชื้อไวรัส ผู้วิจัยตระหนักถึงความสำคัญนี้ จึงได้นำเสนอกระบวนการคัดกรองทางคอมพิวเตอร์ที่สามารถจัดการกับปัญหาดังกล่าวได้อย่างรวดเร็ว โดยกระบวนการที่ใช้ประกอบไปด้วยการคัดกรองด้วยสมบัติเสมือนยา การคัดกรองเสมือนจริงโดยใช้แบบจำลองฟามาโคฟอร์ การจำลองการจับกันระหว่างโมเลกุล การจำลองทางพลวัตเชิงโมเลกุล และการคำนวณออร์บิทัลแบบแยกส่วน สารที่ได้จากกระบวนการนี้ถูกคัดเลือกมาจากฐานข้อมูลที่ประกอบด้วยสารผลิตภัณฑ์จากธรรมชาติและสารอนุพันธ์ โดยโปรตีนที่เป็นเป้าหมายสำหรับงานวิจัยคือ เอนไซม์โปรตีเอสหลัก หรือเอนไซม์ทรี-ไคโมทริปซิน-ไลค์-โปรตีเอส ของไวรัสซาร์ส-โควี-2 ซึ่งมีความสำคัญต่อกระบวนการจำลองตัวเองของไวรัสในเซลล์โฮสต์ การคัดกรองนี้ได้มีการเปรียบเทียบผลลัพธ์กับสารยับยั้งอ้างอิงประเภทเปปไทโดมิเมติก (11a, 13b และ N3) และยาที่ได้รับการอนุมัติจากสำนักงานคณะกรรมการอาหารและยาสหรัฐ (เนอร์มาเทรลเวียร์) พบว่า สารอนุพันธ์ของซัลโฟนาไมด์ชาลโคน (SWC422, SWC423 และ SWC424) และสารอนุพันธ์เอสเทอร์ของกรดคาเฟอิก (4k และ 4l) เป็นตัวแทนของสารประกอบที่มีความเป็นไปได้มากที่สุด สามารถสังเกตได้จากค่าพลังงานในการจับและการเกิดอันตรกิริยาที่สำคัญกับโปรตีนที่เป็นเป้าหมาย อย่างไรก็ตาม แม้ว่าการค้นหาสารประกอบในงานวิจัยนี้แสดงให้เห็นว่าเทคนิคการตรวจคัดกรองด้วยวิธีการทางคอมพิวเตอร์จะมีศักยภาพในการระบุหาสารออกฤทธิ์ที่ยับยั้งการทำงานของเอนไซม์ทรี-ไคโมทริปซิน-ไลค์-โปรตีเอสของไวรัสโคโรนา แต่การทดลองอื่นๆ เพิ่มเติม เช่น การทดสอบฤทธิ์การยับยั้งเอนไซม์ และการทดสอบความเป็นพิษในเซลล์ก็มีความจำเป็นอย่างยิ่งในการยืนยันประสิทธิภาพสารที่ได้จากกระบวนการคัดกรองเหล่านี้

สาขาวิชา	ชีวสารสนเทศศาสตร์และชีววิทยาเชิงคอมพิวเตอร์ (สหสาขาวิชา)	ลายมือชื่อนิสิต ....
ปีการศึกษา	2565	ลายมือชื่อ อ.ที่ปรึกษาหลัก ....
		ลายมือชื่อ อ.ที่ปรึกษาร่วม ....

6382036320 : MAJOR BIOINFORMATICS AND COMPUTATIONAL BIOLOGY (INTERDISCIPLINARY PROGRAM)

KEYWORD: SARS-CoV-2 3CLpro, virtual screening, sulfonamide chalcones, caffeic acid derivatives, natural products, COVID-19

Piyatida Pojtanadithee : Computational Approaches for Identifying Bioactive Compounds Inhibiting SARS-CoV-2 Main Protease. Advisor: Assoc. Prof. Dr. THANYADA RUNGROTMONGKOL Co-advisor: Assoc. Prof. Dr. PHORNPHIMON MAITARAD

The infectious disease caused by the novel coronavirus 2019 has devastatingly affected the global economy and society. However, the drug discovery process from concept to approval requires a significant investment of time and resources. To address these challenges, we employed structure-based virtual screening techniques, including drug-likeness screening, pharmacophore-based virtual screening, molecular docking, molecular dynamics simulation, and fragment molecular orbital calculation. The protein target of this investigation was the main protease or 3-chymotrypsin-like-protease (3CLpro) of the coronavirus, given its pivotal role in the viral replication process. Using our in-house database of natural products and their derivatives, we aimed to identify potent compounds with the potential for further development as anti-SARS-CoV-2 medications. Notably, the derivatives of sulfonamide chalcone (SWC422, SWC423, and SWC424) and ester derivatives of caffeic acid (4k and 4l) exhibited exceptional binding energy and substantial interactions with the 3CLpro binding pocket compared to peptidomimetic inhibitors (11a, 13b, and N3) and an FDA-approved drug (nirmatrelvir). While our findings show that *in-silico* strategies have the potential to identify new potent compounds that inhibit the 3CLpro activity of coronavirus, further studies such as enzyme inhibition assay and cell-based assay are necessary to ensure their effectiveness from these virtual screenings.

Field of Study:	Bioinformatics and Computational Biology (Interdisciplinary Program)	Student's Signature ....
Academic Year:	2022	Advisor's Signature ....
		Co-advisor's Signature ....

ACKNOWLEDGEMENTS

Piyatida Pojtanadithee

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

CHAPTER I

INTRODUCTION

1.1 Coronavirus disease 2019

Coronavirus disease 2019 (COVID-19) emerged in December 2019 in Wuhan City, Hubei Province, China. The initial cases were associated with exposure to wildlife at the Huanan seafood wholesale market.(Decaro & Lorusso, 2020) Initially named 2019-nCoV, the virus was officially renamed Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) by the International Committee on Virus Taxonomy.(Lai, Shih, Ko, Tang, & Hsueh, 2020) Coronaviruses belong to the subfamily *Coronavirinae*, order *Nidovirales* and are common human pathogens. They are enveloped, positive-sense RNA viruses with a diameter of 60–140 nm and 29,903 base pair single stranded RNA genome. There are four structural proteins, namely the spike (S), membrane (M), envelope (E), and nucleocapsid (N).(Gupta & Gupta, 2020) SARS-CoV-2 shares approximately 79.5% genomic homology with SARS-CoV while only about 50% similarity with MERS-CoV, indicating that SARS-CoV is closer to SARS-CoV.

In the process of virus multiplication, the main protease or 3-chymotrypsin-like-protease, plays a vital role by cleaving polyproteins at multiple sites to produce functional proteins. This enzyme is highly conserved cysteine hydrolases from CoVs, are capable of cleaving polyproteins at multiple sites to yield multiple functional proteins.(Lu et al., 2020; Qiao et al., 2021) Considering that 3CLpro play a vital role in CoV replication, especially in the two of the most severe pandemics of the 21st century caused by SARS-CoV-2 and SARS-CoV, these key hydrolases have been validated as promising targets for developing broad-spectrum anti-CoV agents. Because no homolog of 3CLpro has been identified in humans, it is feasible to develop efficacious and specific 3CLpro inhibitors with extremely weak inhibitory effects on human proteases, thereby reducing the side effects caused by 3CLpro inhibitors.

The 3CLpro of SARS-CoV-2 has specific amino acid residues, H41 and C145, located in its active site. The structure of this protease is composed of two monomers, each consisting of three domains. Domain I (residues 8–101) and domain II (residues 102–184) are catalytic domains with a unique antiparallel β-barrel structure. Domain III (residues 201–303) facilitates the dimerization of the enzyme and is made up of five α-helices. One notable feature of 3CLpro is its distinct cleavage site, marked by the conserved L-Q$\downarrow$ (S/A/G) sequence. This cleavage site is absent in closely related human host proteases, which makes the side effects of 3CLpro inhibitors in human patients less likely. The first target-based drugs developed are 3CL[pro] peptidomimetic inhibitors (Figure 13), for example, aldehyde inhibitor (11a)(Dai et al., 2020), α-ketoamide inhibitors (13b)(Liang et al., 2020), and Michael acceptor inhibitor (N3)(H. Yang et al., 2005).

Moreover, authorized drug such nirmatrelvir showed a strong inhibitory activity against SARS-CoV-2 3CLpro and has been authorized for emergency use or granted regulatory approvals in several countries for treating COVID-19 in specific patient populations.(Joyce, Hu, & Wang, 2022) Although nirmatrelvir has curative properties, it must be co-administered with ritonavir to achieve adequate plasma concentrations. Beyond that, administering ritonavir-boosted nirmatrelvir may not be the best choice for all patients due to the possibility of significant drug-drug interactions with concomitant medications.("COVID-19 Treatment Guidelines Panel,") Additionally, the presence of the single mutations E166M and H172Y in the SARS-CoV-2 3CL[pro] significantly reduced its inhibitory activity, which affects drug resistance and treatment outcomes.(Clayton et al., 2023; Sasi et al., 2022)

1.2 Virtual screening in structure-based drug discovery

Structure-based drug design (SBDD) is increasingly essential for the efficient development of therapeutic agents and for studying metabolic processes. SBDD has been shown to be more efficient than traditional drug discovery since it tries to understand the molecular basis of disease and uses information on the biological target's three-dimensional (3D) structure in the process. We can now examine the underlying molecular interactions involved in ligand-protein binding and interpret experimental results in atomic detail by applying computational approaches and 3D structural information of the protein target. The use of computers in drug development has the added benefit of delivering new drug candidates more quickly and cost-efficiently.

Virtual screening (VS) is a cutting-edge structure-based drug design approach. In virtual screening, large libraries of commercially available drug-like compounds are computationally screened against targets of known structure, and those that are predicted to bind well are experimentally tested.(Lavecchia & Di Giovanni, 2013) However, database screening does not provide structurally "novel" molecules because these chemicals have already been synthesized by commercial suppliers. Existing molecules can only be copyrighted with a "method of use" patent that covers their use for a specific application rather than their chemical structure. The 3D structure of the receptor is used in the *de novo* drug design to develop structurally novel compounds that have never been synthesized before using ligand-growing programs and the medicinal chemist's intuition.

Recently, significant successes in computer-aided drug discovery have occurred: new biologically active compounds have been predicted along with their receptor-bound structures, and in several cases, hit rates (ligands discovered per molecules tested) have been significantly higher than in HTS.(Benod et al., 2013; T. Cheng, Li, Zhou, Wang, & Bryant, 2012; Lavecchia & Di Giovanni, 2013) Furthermore, while it is uncommon to supply lead candidates in the nM regime via VS, numerous recent publications detail the identification of nM leads directly from VS; these methodologies will be reviewed herein.(Heifetz et al., 2013; Kolb et al., 2009; Schröder et al.,

2013) As a result, computational methods play an important role in drug design and discovery within the context of pharmaceutical research.

1.2.1 Structure-based virtual screening (SBVS)

Structure-based virtual screening begins with processing the 3D target structural information of interest. Experiment data (X-ray, NMR, or neutron scattering spectroscopy), homology modeling, or molecular dynamics simulations can all be used to determine the target structure. When considering a biological target for SBVS, there are numerous fundamental issues to consider, such as receptor druggability, binding site selection, protein structure selection, incorporating receptor flexibility, appropriate assignment of protonation states, and consideration of water molecules in a binding site, to name a few. Indeed, determining ligand binding locations on biological targets is becoming increasingly important. The scientific community has recently focused on druggable allosteric binding sites in order to find novel modulators of protein/gene function. Another factor to consider for SBVS is the careful selection of the chemical library to be screened in the VS exercise based on the target in question, as well as the preprocessing of libraries to assign the right stereochemistry, tautomeric, and protonation states.(Lionta, Spyrou, Vassilatis, & Cournia, 2014)

1.3 Drug-likeness screening

"Drug-likeness" refers to a protein's ability to bind to a high-affinity ligand with drug-like properties, thereby regulating its activity. Initially applied to proteins, this concept has expanded to include other molecules, such as DNA and RNA, which can serve as potential drug targets. In the field of cheminformatics, the evaluation of drug-likeness holds significant importance as it aids medicinal chemists in handling hits and lead compounds and screening potential drug candidates that can effectively modulate targets.

In current drug development endeavors, the optimal strategy involves the continuous exploration of new chemical entities to identify candidate drugs that closely resemble existing drugs in terms of essential physicochemical and biological properties. Assessing the properties of drug-likeness facilitates the acquisition of more precise pharmacokinetic and pharmacodynamic data. However, the dynamic nature and adaptive range of molecular entities pose significant challenges to this work, as they greatly influence the prediction outcomes. Consequently, further research on the structural and physicochemical characterization of bioactive compounds is still necessary to enhance our understanding in this field.

1.3.1 Lipinski's Rules of Five

In order to advance the discovery and development of new drugs, great efforts are being made to evaluate the similar drug-like properties of molecules in the early stages of the discovery-research process. There are different approaches to solving this problem. Still, the

simplest and most used method is developed by Chris Lipinski and his colleagues at Pfizer(Lipinski, 2004), which is generally referred to either as the Lipinski Rules or the Rule of Five (ROF).

The Rule of Five is a practical guideline to evaluate drug-likeness and determine the potential oral activity of a chemical compound in humans. According to the ROF, for a biologically active molecule to be considered a potential orally administered drug, it should fulfill five specific conditions. Poor absorption or permeation is more likely if the molecule exhibits the following characteristics:

- A molecular mass greater than 500 daltons
- More than 10 hydrogen bond acceptors
- More than 5 hydrogen bond donors
- A calculated octanol-water partition coefficient (Clog P) greater than 5

Based on the ROF, the rating of an orally active drug ranges between 0 and 4, indicating that a potential drug should have no more than one violation of the specified criteria. However, Lipinski acknowledges that molecules failing to meet these criteria should not be entirely disregarded, as many successful drugs do not conform to the Rule of Five.

While the Rule of Five finds wide application, it also possesses certain limitations. Two significant weaknesses include the equal weight given to each rule and the rigid boundary that defines the violation of a specific rule. Another drawback of this rule is its omission of considerations for natural and biological compounds, as well as criteria relevant to metabolism.

1.3.2 Physicochemical properties

The term "physicochemical" combines the words "physico" and "chemical," referring to the physical and chemical aspects of a compound. In the context of drug development, physicochemical properties encompass all the physical and chemical attributes of a drug. These properties play a crucial role in eliciting the pharmacological response from the receptor, which can be a biological molecule or system that interacts with the prescription.

When drugs interact with receptors, they form a Drug-Receptor Complex, which is responsible for the pharmacological effects of the drug. The diverse range of physicochemical properties exhibited by drugs contributes to the pharmacologically varied impact they produce.

1.3.2.1 Lipophilicity

Lipophilicity, commonly known as LogP, represents the equilibrium ratio of a compound's concentration between oil and liquid phases. It is a crucial physicochemical parameter in drug development, as it significantly influences various pharmacokinetic properties such as absorption, distribution, permeability, and clearance routes. The demand for drugs with high lipophilicity has increased to meet the requirements of

selectivity and potency, mainly due to the lipid nature of biological targets such as neurotransmitter pathway targets, anatomical targets, and intracellular targets.(Chandrasekaran, Abed, Al-Attraqchi, Kuche, & Tekade, 2018)

Conversely, suitable drug formulations must exhibit both good aqueous solubility and an appropriate level of lipophilicity to ensure optimal oral absorption, deposition, and activity. Consequently, computational methods for measuring aqueous solubility, lipophilicity, and ionization degree have been integrated into the early stages of drug discovery.(Artursson & Bergström, 2003) Given the significant role of lipophilicity in understanding the pharmacokinetic properties of drug candidates, there is a persistent need for accurate and precise in silico models to predict lipophilicity. LogP prediction models have been developed and facilitated the drug design process, leading to other prediction approaches based on multiple fragments and atoms.(Wenlock & Barton, 2013)

1.3.2.2 Solubility

Aqueous solubility is crucial role in drug development, affecting drug uptake, transfer, and elimination from the body. Intrinsic solubility refers to a drug's thermodynamic solubility at a pH where it is completely in the unionized form.(Bergström, Charman, & Porter, 2016) The efficiency of drugs heavily relies on their aqueous solubility, as poor solubility or low dissolution rates result in inadequate pharmacological activity.(Soni et al., 2016) The prevalence of poorly soluble drugs has increased, leading to issues with absorbability, food effects, and pharmacokinetic.(Kuentz & Imanidis, 2013) Access to sufficient solubility data significantly aids drug development, but finding compounds with the desired solubility profile can be challenging. Computational approaches can be employed to predict solubility and enhance drug absorption.(Lüder, Lindfors, Westergren, Nordholm, & Kjellander, 2007) The solubility of chemical compounds is influenced by lipophilicity and crystalline structure tightness, with an inverse relationship between these parameters and solubility. While solubility is not always considered an ADMET property, it is a critical factor in determining oral absorption. Compounds with poor solubility in the gut experience low permeability and poor absorption. This has prompted researchers to focus on solubility prediction, an important aspect of drug development in recent years. Despite its significance, challenges exist in obtaining consistent and reliable solubility data for prediction purposes.(Wenlock & Barton, 2013)

1.3.2.3 Polar surface area

The polar surface area (PSA) is determined by subtracting the area covered by carbon atoms, halogens, and nonpolar hydrogen atoms (i.e., hydrogen atoms bonded to carbon atoms) from the molecular surface. In other words, the PSA represents the

surface associated with heteroatoms (such as oxygen, nitrogen, and phosphorus atoms) and polar hydrogen atoms.(McCracken & Lipkowitz, 1990) It has been noted that the polar area is sensitive to the three-dimensional conformation of a molecule, and therefore, a weighted dynamic average that considers all significant conformers provides a more accurate description than a single static PSA value.(Palm, Luthman, Unge, Strandlund, & Artursson, 1996) For rapid analysis of large datasets, Ertl and colleagues introduced an approach that calculates the PSA as the sum of fragment-based increments.(Ertl, 2000) Extensive studies on PSA values have shown that the polar area exhibits a stronger correlation with hydrogen bonding (both donor and acceptor groups) compared to lipophilicity (with R^2 values of 0.76 and 0.30, respectively).(Winiwarter et al., 2003)

1.3.3 Drug-likeness profiles

Drug likeness profiles were conducted using established rules that serve as high-throughput screening filters in leading pharmaceutical companies. The Abbott bioavailability score was also calculated to predict the likelihood of achieving 10% oral bioavailability or Caco-2 diffusion. These filters have been developed to assess drug-likeness, which involves predicting a chemical entity's potential pharmacokinetic properties based on parameters such as molecular weight, LogP, and the number of hydrogen bond acceptors and donors. Furthermore, the potential of the presented structures to serve as starting scaffolds or lead compounds in future synthetic drug discovery programs was evaluated using specific medicinal chemistry and lead-likeness filters.(Daina, Michielin, & Zoete, 2017)

1.4 Structure-based pharmacophore modeling

Pharmacophore models depict the molecular recognition of a biological target by a group of compounds rather than representing real molecules or functional group associations. They illustrate the spatial arrangement of essential interactions within a receptor-binding pocket.(Giordano, Biancaniello, Argenio, & Facchiano, 2022) Structure-based pharmacophores (SBPs) can be derived from either a free (apo) structure or a complex structure involving a macromolecule and ligand (holo). SBP methods based on protein-ligand complexes utilize observed interactions between the ligand and protein, while methods based on ligand-free proteins solely use information from the protein's active site. Therefore, SBPs overcome challenges related to ligand flexibility, molecular alignment, and proper selection of training set compounds encountered in ligand-based pharmacophore modeling.(Pirhadi, Shiri, & Ghasemi, 2013)

1.4.1 Pharmacophore model generation

Pharmacophore model generation is the initial step in the process. It involves identifying and defining essential chemical features and spatial arrangements for a compound to interact

with a biological target. These features can include hydrogen bond acceptors, hydrogen bond donors, hydrophobic regions, aromatic rings, and other key functional groups. Various methods, such as ligand-based or structure-based approaches, can generate pharmacophore models. Ligand-based methods utilize information from known active compounds, while structure-based methods utilize the 3D structure of the target or target-ligand complexes.(Vuorinen & Schuster, 2014)

1.4.2 Pharmacophore-based virtual screening

Pharmacophore-based virtual screening is used to identify potential compounds that match the defined pharmacophore model. Virtual screening involves searching large databases of chemical compounds and comparing their features and spatial arrangements with those of the pharmacophore model. This process helps prioritize and select compounds more likely to interact with the target. Virtual screening can significantly reduce the time and cost associated with experimental screening of a large number of compounds. (Thomas Seidel, Ibis, Bendix, & Wolber, 2010)

1.4.3 Pharmacophore model validation

Pharmacophore model validation is a step to assess the reliability and predictive power of the generated model. One commonly used method for pharmacophore model validation is the receiver operating characteristic (ROC) curve analysis, which measures the model's ability to discriminate between active and inactive compounds. The area under the ROC curve (AUC) value is a quantitative measure that indicates the overall performance of the model. (John et al., 2011)

The ROC curve is constructed by plotting the true positive rate (sensitivity) against the false positive rate (1-specificity) at various classification thresholds. The sensitivity represents the proportion of correctly identified active compounds, while specificity represents the proportion of correctly identified inactive compounds. The ROC curve visually represents the model's performance across different threshold values.

The AUC value ranges from 0 to 1, with a higher value indicating better performance. A model with an AUC value close to 1 demonstrates high discriminative power, meaning it can effectively distinguish between actives and inactives.(Molla et al., 2023) Conversely, an AUC value less than to 0.7 suggests a random or poor performance, where the model is no better than a random chance.

To interpret the results from the ROC curve analysis, the following guidelines can be used:

- AUC > 0.9: Excellent discrimination power
- AUC between 0.8 and 0.9: Good discrimination power
- AUC between 0.7 and 0.8: Fair discrimination power
- AUC between 0.6 and 0.7: Poor discrimination power

- AUC < 0.6: Random or low discrimination power

1.5 Molecular docking

Molecular docking is a powerful computational method used to predict the favored conformation and orientation (collectively referred to as the "pose") of molecules within the binding site of a target macromolecule (receptor).(Torres, Sodero, Jofily, & Silva-Jr, 2019) It plays an essential role in computer-aided drug design and facilitates the prediction of ligand binding modes. By exploring ligand conformations and ranking them based on binding affinities, docking programs such as AutoDock(Trott & Olson, 2010), GOLD(Jones, Willett, Glen, Leach, & Taylor, 1997), and FlexX(Rarey, Kramer, Lengauer, & Klebe, 1996) provide valuable insights into the molecular interactions between ligands and their target proteins. However, challenges arise in accurately modeling receptor flexibility, which is essential for capturing induced-fit effects during binding.

1.5.1 Methods and scoring

To simplify calculations, docking methods treat proteins and ligands as rigid bodies while employing advanced sampling techniques like genetic algorithms and Monte Carlo simulations to explore the vast conformational space.(Altuntaş, Bozkus, & Fraguela, 2016) The selection of the most favorable ligand conformations is guided by scoring functions, including empirical and force field-based approaches. These scoring functions evaluate binding affinity by considering known protein-ligand interactions or utilizing statistical observations from protein-ligand databases.

1.5.2 Performance and validation

The performance of docking programs and scoring functions can vary depending on the specific targets and ligands involved. Different systems may require tailored approaches to achieve accurate results.(C. Yang, Chen, & Zhang, 2022) The validation of docking outcomes often involves comparing them with experimental data or using complementary computational techniques. In addition to experimental validation, molecular dynamics simulations can refine and validate docking results.

1.6 Molecular dynamics (MD) simulation

Molecular dynamics (MD) simulation is a powerful computational technique used to study the time-dependent behavior of a system of interacting atoms. By integrating the equations of motion for the atoms, MD generates information about atomic positions and velocities, providing microscopic insights into the system's dynamics.(Hollingsworth & Dror, 2018) To accurately simulate the behavior of a system, an appropriate interaction potential or force field is required, which describes the inter-particle interactions. The choice of force field depends on the specific application and plays a crucial role in the quality of MD simulation results.

The force field used in MD simulations consists of two types of terms: bonded and nonbonded. Bonded terms account for intramolecular interactions, such as bond stretching, bending, and torsions. These terms are represented by harmonic potentials or cosine series, depending on the type of interaction. Nonbonded terms describe van der Waals and electrostatic interactions between atoms. A Lennard-Jones potential typically describes the van der Waals interactions, while the Coulomb potential represents the electrostatic interactions.(Kubincová, Riniker, & Hünenberger, 2020) Although MD simulations can accurately describe the behavior of small systems, analytical solutions for large systems are not feasible. Numerical methods integrate the equations of motion over discrete time intervals called time steps. Common integrators, such as the velocity-Verlet algorithm,(Verlet, 1967) propagate the positions and velocities of atoms in the system. To ensure reliable results, the integrator must satisfy specific requirements, such as preserving the total energy of the system.(Swope, Andersen, Berens, & Wilson, 1982)

MD simulations can be performed in different ensembles to mimic macroscopic behavior. The NVE ensemble represents an isolated system with a constant number of molecules/atoms in assembly, volume, and energy. The NVT (constant number of molecules/atoms in assembly, volume, and temperature) ensemble maintains a constant temperature through thermostats, allowing for kinetic energy fluctuations. The NPT (constant number of molecules/atoms in assembly, pressure, and temperature) ensemble controls temperature and pressure using barostats, which scale the system volume.(Oh & Klein, 2006) Periodic boundary conditions (PBC) are often employed to simulate bulk properties with finite-size systems. PBC replicates the system in all directions, forming an infinite lattice of image atoms. Efficient methods like the particle-mesh Ewald (PME) algorithm are used to compute the long-range electrostatic interactions in systems with PBC.(Darden, York, & Pedersen, 1993)

1.6.1 Simulation methodology

MD simulation begins with the knowledge of the system's potential energy to its position coordinates. The first derivative of the possible function concerning the position coordinates aids in calculating the force operating on each atom in the system. The following are the essential steps involved in MD simulations of proteins.

1.6.1.1 Simulation environment

Protein simulations aim to replicate experimental conditions, considering various parameters. Typically, simulations are performed in the canonical ensemble during the initial equilibration steps or the isothermal-isobaric ensemble. In these simulations, proteins are placed in a unit cell and solvated with solvent. Several explicit water models, including TIP3P, TIP4P(William L. Jorgensen, 1983), TIP5P(Mahoney & Jorgensen, 2000), SPC, and SPC/E(Berendsen, Grigera, & Straatsma, 1987), are commonly used to

mimic the hydration of molecules. These water models capture important aspects of solvent behavior, such as dipole orientation, electrostatic shielding, hydrogen bond rearrangements, and hydrophobic interactions. However, due to limited time resolution and quantum complexities, hydrogen bonds are often not explicitly treated. Instead, average energy contributions and shake algorithms handle solvent hydrogens.

Implicit solvent models approximate the solute potential of the mean force, representing the solute conformations' statistical weight. This is achieved by averaging over solvent degrees of freedom. To maintain charge neutrality, solvent molecules may be replaced with ions. Boundary models such as PBC can be employed to avoid interaction problems. PBC involves replicating the system in adjacent unit cells, allowing for the conservation of mass and particle numbers. Long-range non-bonded interactions are computed using molecules within image systems, and the Ewald summation method is commonly employed to calculate electrostatic interactions in solvated periodic boundary simulations of biomolecular systems.(Cheatham, Miller, Fox, Darden, & Kollman, 1995)

1.6.1.2 Energy minimization

During the energy minimization step of MD simulations, the objective is to find the global minimum energy by optimizing the positions of the side chain atoms. This optimization represents the geometric arrangement where the net attractive force on each atom is maximized. Several methods exist for computing the minimum energy, with the steepest descent and conjugate gradient methods being widely used. The steepest descent method is a first-order iterative descent method that utilizes the gradient of the potential energy surface. It is based on the forces in the molecular mechanical description of the system and guides the search path toward the nearest energy minimum. This method effectively minimizes the energy by iteratively updating the atom positions along the direction of the steepest descent. An important aspect of the energy minimization step is correcting the protonation state of titratable residues. This can be accomplished through free energy of perturbation (FEP) MD simulations or by employing continuum electrostatics models such as finite differences Poisson-Boltzmann (FDPB) or protein dipole-Langevin dipole (PDLD). These approaches help account for the effects of protonation changes and the electrostatic environment on the system's energy and stability.

1.6.1.3 Heating the system and equilibration

During the heating phase of MD simulations, the system is prepared by assigning initial velocities to each atom during energy minimization at 0 K. Newton's equations of motion are then numerically integrated to simulate the time evolution of the system. At

predefined intervals, the velocities are updated to correspond to slightly higher temperatures, allowing the simulation to progress until the desired temperature is reached. As the system is heated, force constraints on different subdomains are gradually released to dissipate structural tensions. Thermalization typically occurs at constant volume using Langevin dynamics.(Uberuaga, Anghel, & Voter, 2004) The equilibration stage aims to achieve a balance between kinetic and potential energies by redistributing the kinetic energy throughout all degrees of freedom. The protein positions are fixed in explicit solvent simulations while the water molecules move accordingly. Once the solvent has equilibrated, the constraints on the protein can be removed, enabling the entire system (protein+solvent) to evolve in time. This equilibration step ensures that the system reaches a stable state where the properties and energies of the system are well-distributed and representative of the desired conditions.

1.6.1.4 Production phase

The production phase serves as the final step in the simulation methodology, where constraints on the protein are removed. This phase is carried out over a desired time scale to generate a trajectory of the protein molecule that conforms to specific equilibrium conditions. The duration of the production phase can range from several hundred picoseconds to microseconds or longer, depending on the objectives of the study. To prevent significant trajectory artifacts during long simulation runs, recent versions of CHARMM protein parameter files incorporate a 2D grid correction map known as the CMAP correction. This correction map, obtained from the surfaces of alanine, proline, and glycine dipeptides, helps improve the accuracy of backbone φ and ψ parameters. Including the CMAP correction can enhance the quality and reliability of the protein trajectory generated during the production phase.(Best et al., 2012)

1.7 Fragment molecular orbital (FMO) calculation

The fragment molecular orbital (FMO) method is widely used in quantum mechanical (QM) calculations for biomolecule simulations. It involves dividing proteins or nucleic acids into substructures or "fragments" and reconstructing the overall structure based on these fragments. The FMO method represents the total energy of the system using the energies of fragment monomers and fragment dimers, taking into account the environmental electrostatic potentials from surrounding fragments. FMO calculations are performed using programs such as GAMESS(D. Fedorov, 2017), ABINIT-MP(Tanaka, Mochizuki, Komeiji, Okiyama, & Fukuzawa, 2014), and PAICS(Mochizuki, Tanaka, & Fukuzawa, 2021). These programs allow for automatic calculations starting from a Protein Data Bank (PDB)(Berman et al., 2000) structure and additional fragmentation information. Depending on the size of the target system, FMO calculations can be

conducted at various levels of electron correlation, such as MP4(Akisawa et al., 2021) and density-functional tight-binding combined with FMO (FMO-DFTB) methods.(Nishimoto, 2021) The MP2(D. G. a. K. Fedorov, K, 2004) and DFTB(Mochizuki et al., 2021) methods are commonly used due to their cost-performance ratio.

The FMO method is widely used in high-precision *in silico* drug screening and QM-based biomolecular function elucidation. Large-scale user groups, such as the FMO drug design consortium, have developed the FMO database (FMODB),(Takaya et al., 2021) a quantum chemical calculation database for biomolecules. In FMO calculations, the interfragment interaction energy (IFIE)(Tokutomi, Shimamura, Fukuzawa, & Tanaka, 2020) obtained from the FMO method can be analyzed using energy decomposition analysis known as PIEDA. PIEDA allows for calculating binding energies, analyzing intermolecular interactions, and subsystem analyses. Different components of the PIEDA analysis provide insights into specific types of interactions, such as hydrogen bonding, CH/π interactions, and π-π interactions.(Maghami & Abdelrasoul, 2020)

In the FMO method, the total energy is represented by the energies of fragment monomers (EI) and fragment dimers (EIJ) subject to the electrostatic potentials of the surrounding fragments,

$$E_{total} = \sum_I E_I + \sum_{I>J}(E_{IJ} - E_I - E_J) \tag{1}$$

Transforming Eq. (1) to interaction energy expresses the sum of the monomer energies, excluding the contribution of the electrostatic potential of the environment, and the IFIE between monomers.

$$E_{total} = \sum_I E'_I + \sum_{I>J}\Delta\tilde{E}_{IJ} \tag{2}$$

1.5.1 Fragment molecular orbital-based analysis

The IFIE (PIE) obtained from Eq. (2) and its PIEDAare widely used in FMO calculations of biomolecules. The PIEDA component of IFIE between fragments I and J is defined as follows:

$$\Delta\tilde{E}_{IJ} = \Delta\tilde{E}_{IJ}^{ES} + \Delta\tilde{E}_{IJ}^{EX} + \Delta\tilde{E}_{IJ}^{CT+mix} + \Delta\tilde{E}_{IJ}^{DI} + \Delta\tilde{E}_{IJ}^{sol} \tag{3}$$

where ES, EX, CT+mix, DI, and sol stand for the electrostatic, exchange repulsion, charge transfer with higher-order mixed terms, dispersion, and solvation contributions, respectively. The final term, $\Delta\tilde{E}_{IJ}^{sol}$, is only added when the solvent model is considered. PIEDA enables binding energy calculations, intramolecular and intermolecular interaction energy,(Okiyama et al., 2019)

and analyses of subsystems.(D. G. Fedorov & Kitaura, 2016) Regarding intermolecular interactions, hydrogen bonding consists primarily of ES and CT+mix components, whereas CH/π and π-π interactions are dominated by DI components.

1.8 Research rationality

In December 2019, the emergence of COVID-19 was reported by Chinese health authorities, which was later declared a pandemic by the World Health Organization. This disease is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and has had significant global social and economic repercussions. The duration of the epidemic has been uncertain due to various factors, including different modes of transmission, a high number of asymptomatic carriers, limited accessibility to affordable testing, ineffective therapies, and the emergence of vaccine-resistant variants. Multiple transmission modes present challenges in implementing effective prevention and control measures. Consequently, there is a critical need for effective treatments for individuals who are unvaccinated or may have insufficient vaccine responses, significantly immunocompromised. Furthermore, these treatments should address viral mutations that could potentially undermine the effectiveness of vaccines.

The 3CLpro has emerged as a promising target for addressing these challenges, given its crucial role in viral replication and the absence of human proteases with similar cleavage specificity. The development of target-based drugs has led to the identification of 3CLpro peptidomimetic inhibitors, such as the aldehyde inhibitor (11a), α-ketoamide inhibitors (13b), Michael acceptor inhibitor (N3), and nirmatrelvir, identified by their respective PDB codes 6Y2F, 6LU7, 6LZE, and 7VH8. Nirmatrelvir, which inhibits 3CLpro activity by tightly binding to its active site, has received emergency use authorization or regulatory approvals in several countries to treat COVID-19 in specific patient populations. However, nirmatrelvir requires co-administration with ritonavir to achieve adequate plasma concentrations. Furthermore, using ritonavir-boosted nirmatrelvir may not be suitable for all patients due to potential significant drug-drug interactions with concurrent medications. Additionally, the presence of single mutations, E166M and H172Y, in SARS-CoV-2 3CLpro has been shown to significantly reduce its inhibitory activity, affecting drug resistance and treatment outcomes.

In this study, we employed several computational strategies to identify effective inhibitors for SARS-CoV-2 3CLpro from a series of natural products and derivatives. Initially, a screening of compounds from our in-house database was conducted based on drug-likeness properties. Structure-based pharmacophore modeling and molecular docking techniques were utilized, using reported peptidomimetic inhibitors (11a, 13b, and N3) and FDA-approved drug (nirmatrelvir) in complex with 3CLpro as model templates to narrow the selection of the

promising compounds against the activity of SARS-CoV-2 3CLpro. It should be noted that the chemical structures of four inhibitors that have been identified are depicted in Figure 13.

1.9 Research objective

        To identify promising compounds collected from two in-house databases that strongly inhibit the activity of SARS-CoV-2 3CLpro using structure-based virtual screening

1.10 Expected beneficial outcome(s) from the thesis

        The expected outcomes of this research are to identify and characterize potential inhibitors of SARS-CoV-2 3CLpro. The findings will provide valuable insights into developing effective antiviral treatments for COVID-19, specifically by targeting the 3CLpro. These insights have the potential to offer various therapeutic options and serve as a scaffold for designing inhibitors to combat the ongoing global pandemic.

CHAPTER II

METHODOLOGY

2.1 Research 1

2.1.1 Data collection and preprocessing from database

2.1.1.1 Known 3CLpro inhibitors

The co-crystal structures of three peptidomimetic inhibitors (11a, 13b, and N3) and an approved drug (nirmatrelvir) bound to SARS-CoV-2 3CLpro were obtained from the Research Collaboratory for Structural Bioinformatics Protein Data Bank (https://www.rcsb.org). The corresponding PDB codes for these structures are 6Y2F, 6LU7, 6LZE, and 7VH8, respectively. UCSF Chimera(Pettersen et al., 2004) removed water molecules and other small molecules from the complexes. To eliminate the covalent bond at C145 of SARS-CoV-2 3CLpro, Accelrys Discovery Studio Client 4.0("Discovery Studio Modeling Environment, Release 2.5.1," 2009) was utilized. Subsequently, the 3D structure of each inhibitor was extracted from its associated protein (3CLpro).

2.1.1.2 Natural and synthetic compounds from in-house library

A total of 1,052 compounds from an in-house database, exclusively designed and developed by the Department of Chemistry at Chulalongkorn University (CHEM-CU), were subjected to *in silico* studies.

2.1.2 Virtual screening by drug-likeness analysis

This screening was performed using freely available software and tools such as DataWarrior 5.5.0(Sander, Freyss, von Korff, & Rufener, 2015) and SwissADME(Daina et al., 2017) to identify compounds with desirable drug-like characteristics. Filtering criteria included assessments based on Lipinski's rule of five, physicochemical properties, and drug-likeness profiles, as presented in Table 1. Only compounds meeting the established criteria were selected for pharmacophore-based virtual screening.

Table 1 Properties and criteria of drug-likeness analysis

Properties	Profile	Values
Lipinski's rules of five	Molecular weight	≤ 500
	H-bond acceptor	≤ 10
	H-bond donor	≤ 5
	Octanol water coefficient (cLogP)	≤ 5
Physicochemical properties	No. Rotatable bond	≤ 10
	No. Heavy atom	12< X < 40
	Polar surface area	≤ 140

Properties	Profile	Values
	Lipophilic ligand efficiency (LLE)	> 5
	Ligand efficiency lipophilic price (LELP)	< 16.5
	Water solubility	Soluble
Drug-likeness profiles	Drug-likeness	Positive
	Bioavailability score	0.55

2.1.3 Pharmacophore study

2.1.3.1 Pharmacophore model generation

The representative structures of the pharmacophore models (RPMs) for 11a, 13b, and N3, along with their associated predicted chemical features, were obtained from previous studies(K. Sanachai et al., 2022). On the other hand, the representative structures for nirmatrelvir were generated based on a total of 2,000 frames obtained from three independent runs conducted over the last 50 nanoseconds (250-300 ns). Noted that the specific details regarding the MD simulation study carried out on the nirmatrelvir/3CLpro complex are described in the later sections. The chemical features supported by the LigandScout models encompass hydrophobic interactions, hydrogen bond donors/acceptors, positive/negative ionizable areas, aromatic rings, and halogen bond donor features.

2.1.3.2 Pharmacophore-based virtual screening

Using KNIME 4.4.0(Berthold et al., 2007), a screening process was conducted on the drug-like compounds obtained from our in-house database to identify active molecules. The pharmacophore models for 11a, 13b, N3, and nirmatrelvir bound to SARS-CoV-2 3CLpro were utilized as templates, employing the program's default parameters. The resulting file containing the pharmacophore fit scores was sorted in descending order using Accelrys Discovery Studio Client 4.0("Discovery Studio Modeling Environment, Release 2.5.1," 2009). Only compounds that demonstrated optimal fitting with the 3CLpro of the coronavirus were selected for further analysis.

2.1.3.3 Pharmacophore model validation

The active compounds identified during the screening were subjected to DecoyFinder(Cereto-Massagué et al., 2012) to identify their corresponding decoy molecules. The decoys were obtained from a dataset from the ZINC15(Sterling & Irwin, 2015) and DrugBank(Wishart et al., 2006) databases. Subsequently, the screened active compounds (test set) and decoys underwent further screening using KNIME 4.4.0 for the model validation. Upon completing the screening procedure, a hit list of molecules was generated and displayed in the Library View, represented by a ROC curve. Validation

parameters such as AUC and EF (Enrichment Factor) were computed to assess the accuracy of the hit list. This process enabled the determination of the accuracy level achieved by the hit list of molecules, as indicated by previous studies.(Fei, Zhou, Liu, & Tang, 2013; Réau, Langenfeld, Zagury, Lagarde, & Montes, 2018)

2.1.4 Molecular docking

To conduct a docking study, the 3D structures of the SARS-CoV-2 3CLpro, mentioned in section **2.1.1.1**, were used as reference models. Two docking programs, namely AutoDock VinaXB(Trott & Olson, 2010) and GOLD 5.6.3(Jones et al., 1997), were utilized for this purpose, with 200 independent docking runs. The protonation states of the 3CLpro structures at a neutral pH of 7.4 were verified using the PDB2PQR web service available at https://server.poissonboltzmann.org/pdb2pqr. The known inhibitors, as well as the screened compounds, were identified using MarvinSketch from ChemAxon("Marvin 17.21.0, Chemaxon ").

Rigid docking was performed using AutoDock VinaXB, where both the ligands and receptors were converted to the pdbqt format using AutoDockTools(El-Hachem, Haibe-Kains, Khalil, Kobeissy, & Nemer, 2017). The docking process involved using a grid box with dimensions of 20×20×20 centered on specific XYZ coordinates for each model. The coordinates used for the grid box were as follows: (i) 10.87, -0.25, 20.75 for 6Y2F, (ii) -10.72, 12.41, 68.81 for 6LU7, (iii) -10.94, 12.69, 68.91 for 6LZE, and (iv) -18.76, 17.14, -25.14 for 7VH8. The docked conformations with the lowest ΔG (Gibbs free energy) were selected for subsequent analysis.

For flexible docking, a docking sphere was created using the GoldScore scoring function within GOLD 5.6.3, with a 6 Å radius around the known inhibitors. Compounds with a higher GOLD fitness score than their known inhibitors were chosen to evaluate their binding interactions and inhibitory activity. The interactions between 3CLpro and the compounds were visualized using UCSF ChimeraX 1.4(Pettersen et al., 2004), Accelrys Discovery Studio Client 4.0("Discovery Studio Modeling Environment, Release 2.5.1," 2009), and LigandScout 4.4.9(Wolber & Langer, 2005).

2.1.5 Molecular dynamics (MD) simulations

For the nirmatrelvir/3CLpro and SWC423/3CLpro complexes, all-atom molecular dynamics (MD) simulations were conducted using AMBER20(D.A. Case, 2020) under the periodic boundary condition. The force field parameters from ff19SB(Tian et al., 2020) were applied for the bonded and nonbonded interactions of the target protein. The parameters for nirmatrelvir and SWC423 were generated using the *tleap* module with the general AMBER force field 2 (GAFF2)(Wang, Wolf, Caldwell, Kollman, & Case, 2004), and

their RESP charges(Cornell, Cieplak, Bayly, & Kollman, 1993) were obtained following the standard procedure.(Kamonpan Sanachai, Mahalapbutr, Sanghiran Lee, Rungrotmongkol, & Hannongbua, 2021; Sripattaraphan et al., 2022) To neutralize the simulated systems, sodium ions (Na+) were randomly added.

Before the MD simulation, the added hydrogen atoms and water molecules (TIP3P model)(William L. Jorgensen, 1983) were subjected to energy minimization using 1,500 steps of the steepest descent (SD) method followed by 3,000 steps of the conjugate gradient (CG) method. In contrast, the remaining molecules were held fixed. Each system was then further minimized with 1,000 iterations of SD and 2,500 iterations of CG, with solvent molecules restrained. Finally, the entire complex was fully minimized using the same procedure.

During the MD simulation, a time step of 2 femtoseconds (fs) was employed. The temperature and pressure were controlled using a Langevin thermostat(Uberuaga et al., 2004) with a collision frequency of 2 picoseconds (ps) and a Berendsen barostat(Berendsen, Postma, van Gunsteren, DiNola, & Haak, 1984) with a pressure relaxation time of 1 ps. Electrostatic interactions were treated using the PME method(Darden et al., 1993) with a 10 Å cutoff for nonbonded interactions. The SHAKE algorithm(Hünenberger, 2005) was applied to constrain all hydrogen-containing covalent bonds. The system was heated from 10 to 310 K under 1 atm of pressure, followed by a 300 ns MD simulation for nirmatrelvir/3CLpro and a 400 ns simulation for SWC423/3CLpro. MD trajectories were saved every 1,000 ps for later analysis.

Each simulation was repeated in triplicate with different random seeds, starting from identical minimized structures. The root-mean-square deviation (RMSD), number of hydrogen bonds (# H-bonds), and number of atom contacts (# atom contacts) were calculated based on the MD trajectories for further analysis.

2.1.6 Fragment molecular orbital (FMO) calculation

To analyze the binding mechanism at 3CLpro, the representative structure was subjected to FMO calculations using the resolution of the identity MP2 method combined with a polarizable continuum model. The FMO-RIMP2/PCM approach was implemented in the GAMESS software(D. Fedorov, 2017), specifically using the 2022 R2 version. The protocol for FMO-RIMP2/PCM in this study followed similar procedures described in literature references(Hengphasatporn, Harada, et al., 2022; Hengphasatporn, Wilasluck, et al., 2022; Wansri et al., 2022). These references provide details on the specific steps and methodologies employed for the FMO-RIMP2/PCM calculations in this study.

2.2 Research 2

The in-house database utilized in this study comprised a total of 553 natural and synthetic compounds. These compounds were classified into 12 primary groups, including 4 α-mangostins(Pyae et al., 2023), 18 anticancer agents(Pithi, Supakarn, Chuanpit, & Preeyaporn Plaimee, 2016), 9 avicequinones(Karnsomwan, Netcharoensirisuk, Rungrotmongkol, De-Eknamkul, & Chamni, 2017), 42 caffeic acid derivatives(Lin et al., 2022), 159 acid homodimers, 92 xanthones, 78 avenalumic acid derivatives, 9 quinonoids, 30 steroids(Ke, Shi, Zhang, & Yang, 2017), 6 piperic acid derivatives(Wansri et al., 2022), 53 renieramycins(Yokoya et al., 2023), and 53 ureas(Thongsom et al., 2023). These compounds were developed and provided by the Department of Pharmacognosy and Pharmaceutical Botany at the Faculty of Pharmaceutical Sciences, Chulalongkorn University.

2.2.1 Drug-likeness analysis

The pharmacokinetics, drug-like properties, and medicinal chemistry suitability of the 553 compounds in the in-house database were assessed using the SwissADME web-based interface(Daina et al., 2017). The evaluation was conducted based on criteria including Lipinski's rule of five, physicochemical properties, and drug-likeness profiles. A radar chart was generated to illustrate the number of compounds meeting each criterion, as listed in Table 2. Only compounds that fulfilled all the predicted properties were selected for pharmacophore-based virtual screening.

Table 2 Evaluation criteria for drug-like compounds

Property	Characteristic	Criteria
Lipinski's rules of five	Molecular weight	≤ 500
	#H-bond acceptors	≤ 10
	#H-bond donors	≤ 5
	Octanol water coefficient (cLogP)	≤ 5
Physicochemical properties	#Heavy atoms	12 < x < 40
	#Rotatable bonds	≤ 10
	Topological polar surface area (TPSA)	20 < x < 130
	Water solubility (ESOL Log S)	≤ 6
	Solubility class	< 16.5
	Water solubility	No poorly soluble
Drug-likeness profiles	Gastrointestinal absorption	High
	Bioavailability score	0.55, 0.56

2.2.2 Structure-based pharmacophore modeling

In previous studies, RPMs were developed for inhibitors 11a, 13b, and N3 in complex with 3CLpro.(K. Sanachai et al., 2022) These models included predicted chemical features and spatial arrangements necessary for effective interaction with the target protein. These RPMs, along with the drug-like compounds, were defined as input datasets for pharmacophore-based virtual screening. The *iScreen* node of the KNIME program(Berthold et al., 2007) was utilized to perform the pharmacophore-based virtual screening using default parameters. The output compounds from the screening were then subjected to model validation to ensure the reliability and accuracy of the model and to aid in the interpretation and prioritization of the virtual screening results.(Fei et al., 2013)

The active compounds obtained from the screening process were analyzed using the DecoyFinder(Cereto-Massagué et al., 2012). The same datasets as previous research (ZINC(Sterling & Irwin, 2015) and DrugBank(Wishart et al., 2006)) were used for finding inactive molecules. Additionally, KNIME extensions(Berthold et al., 2007) were employed for model validation as well. The input datasets for model validation consisted of RPMs, known active compounds, and a set of decoys. The validation result of ROC curves and metrics, such as AUC and EF, were used for data analysis and visualization.

2.2.3 Molecular docking

The crystal structures of SARS-CoV-2 3CLpro complexed with 11a, 13b, and N3 were downloaded from the Protein Data Bank with PDB codes 6Y2F, 6LU7, and 6LZE, respectively. Water molecules and other ligands in the complexes were eliminated. The protein and ligand structures were then extracted and saved as separate files. The protonation state of ionizable amino acids and ligands at a neutral pH 7.4 was confirmed using PDB2PQR for the protein and MarvinSketch("Marvin 17.21.0, Chemaxon ") for the ligands. The protein and ligand files were converted to PDBQT format to prepare for redocking simulations using AutoDock Tools(El-Hachem et al., 2017). These PDBQT files were used as input for AutoDock VinaXB(Trott & Olson, 2010), with specific docking parameters configured. A grid box with dimensions of 20×20×20 Å was centered for the docking process.

The same procedure was applied to selected active compounds from the pharmacophore-based virtual screening. Only hit compounds with ΔG values lower than the known inhibitors were chosen for ligand-protein binding analysis using Accelrys Discovery Studio Client 4.0. To assess the binding strength of the compounds, the distance from the center of mass of each classified group to the center of mass of the H41 and C145 catalytic dyad (dCOM) was also considered using ChimeraX (Pettersen et al., 2021).

CHAPTER III

RESULTS AND DISCUSSION

3.1 Research 1

3.1.1 Identification of drug-like properties of natural product-based compounds

In the early stages of drug discovery, it is essential to consider a series of drug-like properties to identify compounds that are more likely to succeed in clinical trials. This approach is crucial for improving the overall success rate and reducing manufacturing costs18. In the present study, out of the 1,052 compounds evaluated, it was observed that 75 compounds fulfilled the bioavailability criterion, as listed in Table 3, indicating that they exhibited desirable drug-like properties. These 75 compounds were considered orally active compounds since they did not violate more than one of Lipinski's parameters, which include molecular weight (MW) ≤ 500 g/mol, H-bond acceptors (HBA) ≤ 10, H-bond donors (HBD) ≤ 5, and cLogP value ≤ 5. Furthermore, these compounds underwent physicochemical inspections to determine their suitability as oral drug candidates. The results indicated that these compounds displayed high membrane permeability, and their minimum permissible bioavailability score suggested good systemic absorption, with the potential for over 10% bioavailability in rats. This characteristic is crucial for the drug delivery system and is linked to the enzymatic and physical environment of the gastrointestinal tract.

Table 3 Molecular properties of drug-likeness analysis of 75 selected compounds

| Compound | Lipinski's rules of five | | | | Physicochemical | | | | | | Drug-likeness profiles | |
	MW	HBA	HBD	cLogP	#Rotatable bonds	#Heavy atoms	PSA	LLE	LELP	Water solubility	Drug-likeness	Bioavailability score
STK001	275.1	4	0	0.94	1	15	44.76	5.62	1.57	Soluble	16.37	0.55
STK002	214.22	5	1	-0.72	2	15	64.99	7.39	-1.18	Very soluble	11.9	0.55
STK004	277.07	5	1	0.11	4	15	72.83	6.44	0.19	Very soluble	7.04	0.55
STK005	260.24	5	1	1.01	5	19	72.83	5.58	2.12	Soluble	3.57	0.55
STK008	260.24	5	1	1.01	5	19	72.83	5.58	2.12	Soluble	3.57	0.55
NST005	478.53	7	3	7.21	4	35	105.5	-0.9	29.1	Soluble	2.57	0.55
STK013	275.1	4	0	0.94	1	15	44.76	5.62	1.57	Soluble	16.37	0.55
NPT018	302.24	7	5	1.49	1	22	131.4	5.03	3.67	Soluble	0.08	0.55
NPT019	290.27	6	5	1.51	1	21	110.4	5.03	3.53	Soluble	0.32	0.55

Compound	Lipinski's rules of five				Physicochemical						Drug-likeness profiles	
	MW	HBA	HBD	cLogP	#Rotatable bonds	#Heavy atoms	PSA	LLE	LELP	Water solubility	Drug-likeness	Bioavailability score
NPP002	250.25	5	1	1.25	1	18	57.15	5.36	2.48	Soluble	0.71	0.55
NWC014	182.17	4	2	1.57	2	13	66.76	5.14	2.39	Soluble	3.84	0.55
NWC015	210.18	5	2	1.16	3	15	83.83	5.52	1.9	Soluble	5.99	0.55
NWC016	196.2	4	2	1.57	2	14	66.76	5.14	2.39	Soluble	3.84	0.55
NWC017	194.18	4	1	0.64	1	14	55.76	6.07	0.98	Soluble	1.25	0.55
SWC019	176.17	3	0	1.43	1	13	39.44	5.33	2	Soluble	3.13	0.55
SWC025	206.19	4	1	0.91	3	15	59.67	5.78	1.48	Soluble	3.48	0.55
NPT042	204.18	4	0	1.36	2	15	56.51	5.33	2.22	Soluble	5.28	0.55
NPT044	470.51	8	0	1.03	1	34	104.6	5.3	4.03	Soluble	3	0.55
NPT052	352.38	6	1	0.87	2	26	71.36	5.58	2.57	Soluble	1.71	0.55
SPT054	217.22	4	1	0.97	4	16	64.11	5.69	1.7	Soluble	3.8	0.55
SPT058	215.25	3	0	1.43	1	16	38.25	5.23	2.51	Soluble	1.15	0.55
SPT061	218.21	4	0	0.37	3	16	61.19	6.29	0.65	Very soluble	0.56	0.55
SPT063	191.14	5	1	0.5	1	14	91.83	6.22	0.76	Very soluble	7.18	0.55
SPT066	206.2	5	1	1.28	2	15	64.47	5.41	2.09	Soluble	1.69	0.55
SPT076	260.25	5	0	0.51	2	19	84.07	6.07	1.08	Soluble	4.03	0.55
SPT078	233.24	4	0	1.53	1	17	38.25	5.1	2.87	Soluble	0.19	0.55
SPT079	275.3	5	0	1.29	3	20	56.71	5.27	2.88	Soluble	1.15	0.55
NWC054	286.28	5	0	1.42	1	21	69.67	5.13	3.32	Soluble	1.77	0.55
NTK015	245.27	3	2	1.2	2	18	54.62	5.41	2.39	Very soluble	2.95	0.55
NTK016	228.24	5	1	-0.25	2	16	72.83	6.89	-0.44	Very soluble	0.19	0.55
SWC054	204.18	4	0	1.48	2	15	56.51	5.21	2.43	Soluble	3.17	0.55
NPP004	176.17	3	0	1.72	2	13	35.53	5.03	2.42	Soluble	5.35	0.55
NTK028	408.49	5	2	5.36	4	30	75.99	1.03	18.3	Soluble	0.07	0.55
PPP032	205.21	5	5	-2.2	2	14	110	8.89	-3.35	Highly soluble	0.09	0.55

Compound	Lipinski's rules of five				Physicochemical						Drug-likeness profiles	
	MW	HBA	HBD	cLogP	#Rotatable bonds	#Heavy atoms	PSA	LLE	LELP	Water solubility	Drug-likeness	Bioavailability score
PPP033	233.26	5	5	-1.29	4	16	110	7.92	-2.27	Highly soluble	0.34	0.55
PPP034	261.31	5	5	-0.38	6	18	110	6.96	-0.76	Very soluble	9.41	0.55
PPP035	261.31	5	5	-0.38	6	18	110	6.96	-0.76	Very soluble	9.41	0.55
PPP036	317.42	5	5	1.44	10	22	110	5.06	3.55	Very soluble	19.11	0.55
PPP037	317.42	5	5	1.44	10	22	110	5.06	3.55	Very soluble	19.11	0.55
PPP040	191.22	5	5	-1.85	2	13	92.95	8.57	-2.61	Highly soluble	1.59	0.55
PPP041	219.28	5	5	-0.94	4	15	92.95	7.6	-1.54	Very soluble	0.22	0.55
PPP042	219.28	5	5	-0.94	4	15	92.95	7.6	-1.54	Very soluble	0.22	0.55
PPP043	247.33	5	5	-0.03	6	17	92.95	6.64	-0.06	Very soluble	7.66	0.55
PPP044	247.33	5	5	-0.03	6	17	92.95	6.64	-0.06	Very soluble	7.66	0.55
PPP048	219.28	5	4	-1.18	3	15	84.16	7.83	-1.93	Very soluble	2.99	0.55
PPP049	219.28	5	4	-1.18	3	15	84.16	7.83	-1.93	Very soluble	2.99	0.55
PPP050	275.38	5	4	0.64	7	19	84.16	5.92	1.36	Very soluble	1.92	0.55
PPP051	275.38	5	4	0.64	7	19	84.16	5.92	1.36	Very soluble	1.92	0.55
SWC102	178.14	4	2	0.77	0	13	70.67	5.98	1.08	Soluble	2.83	0.55
SWC110	234.2	5	0	0.99	4	17	65.74	5.65	1.84	Soluble	5.18	0.55
SWC111	250.25	5	0	1.06	5	18	57.9	5.55	2.1	Soluble	10.88	0.55
SWC113	173.17	2	1	0.61	0	13	60.16	6.15	0.86	Very soluble	3.07	0.55
SWC122	231.27	2	2	1.33	2	16	99.57	5.3	2.34	Soluble	2.55	0.55
SWC123	260.2	5	0	-0.53	1	17	99.72	7.11	-0.94	Very soluble	5.35	0.55
NPP027	459.49	8	5	0.3	10	33	137.7	6.04	1.13	Soluble	2.74	0.55
NPP030	207.18	4	2	0.67	1	15	75.63	6.01	1.09	Very soluble	1.49	0.55
SWC128	220.22	4	1	1.36	4	16	59.67	5.3	2.39	Soluble	5.3	0.55
NWC084	302.24	7	5	1.49	1	22	131.4	5.03	3.67	Soluble	0.08	0.55
NWC088	290.27	6	5	1.51	1	21	110.4	5.03	3.53	Soluble	0.32	0.55

Compound	Lipinski's rules of five				Physicochemical						Drug-likeness profiles	
	MW	HBA	HBD	cLogP	#Rotatable bonds	#Heavy atoms	PSA	LLE	LELP	Water solubility	Drug-likeness	Bioavailability score
SWC195	224.21	5	3	0.71	2	16	94.83	5.94	1.24	Soluble	1.85	0.55
NWC097	182.17	4	2	1.23	2	13	66.76	5.51	1.72	Soluble	3.84	0.55
NWC098	210.18	5	2	1.16	3	15	83.83	5.52	1.9	Soluble	5.99	0.55
NWC099	196.2	4	2	1.57	2	14	66.76	5.14	2.39	Soluble	3.84	0.55
SWC219	322.33	4	1	0.25	1	24	51.8	6.25	0.66	Soluble	2.14	0.55
NWC100	302.24	7	5	1.49	1	22	131.4	5.03	3.67	Soluble	0.08	0.55
NTK049	273.24	5	3	0.15	0	20	103.7	6.42	0.32	Soluble	5.31	0.55
NTK064	192.17	4	1	1.08	1	14	59.67	5.63	1.64	Soluble	3.06	0.55
NTK066	280.27	6	1	0.43	3	20	74.22	6.12	0.96	Soluble	0.08	0.55
NTK073	318.32	6	1	1.42	7	23	89.9	5.08	3.67	Soluble	1.92	0.55
SWC272	254.28	5	0	1.27	9	18	53.99	5.33	2.52	Very soluble	15.75	0.55
SWC307	281.26	6	0	0.88	5	20	90.58	5.67	1.95	Soluble	14.15	0.55
SWC339	196.2	4	0	1.38	5	14	44.76	5.33	2.1	Very soluble	3.94	0.55
SWC422	447.93	3	1	5.4	6	31	71.62	0.95	19.2	Soluble	3.76	0.55
SWC423	458.49	5	1	3.88	7	33	117.4	2.46	14.7	Soluble	7.22	0.55
SWC424	463.55	3	1	5.99	6	34	71.62	0.34	23.5	Soluble	2.9	0.55

3.1.2 Screened Compounds from an Investigation of pharmacophore models

The pharmacophore models of inhibitors 11a, 13b, N3, and the nirmatrelvir/3CLpro complex, generated from MD trajectories, were analyzed using LigandScout. This analysis provided insights into the spatial arrangement and composition of the ligands at a molecular level. Figure 1 illustrates that all pharmacophore models have three chemical features: H-bond donor, H-bond acceptor, and hydrophobic interaction. Specifically, the interactions between SARS-CoV-2 3CLpro and inhibitors 11a, 13b, N3, and nirmatrelvir resulted in two, three, five, and three hydrophobic interactions, respectively. Moreover, H-bond donor and H-bond acceptor features were observed to interact with key binding residues, including T25, M49, C145, M165, E166, and A191, as previously described by Zhu et al.(Zhu et al., 2022). These interactions play a vital role in the binding mechanism and stability of the ligand-protein complex.

| 11a | 13b | N3 | Nirmatrelvir |

◄——— **H-bond donor** ◄—— **H-bond acceptor** ● **Hydrophobic interaction**
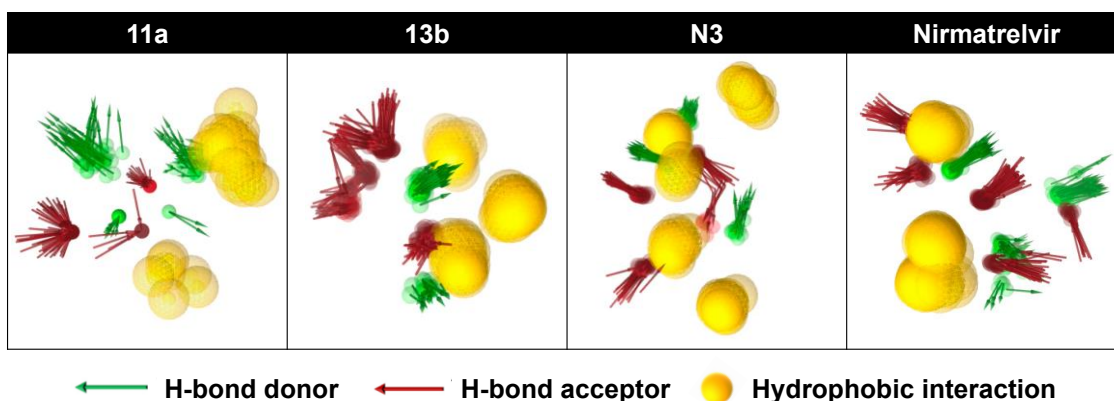
Figure 1 Representative pharmacophore models (RPMs) of the 3CLpro complex of three peptidomimetic inhibitors originated from a previous study and the nirmatrelvir/3CLpro complex from the current study. The yellow spheres represent hydrophobic interactions, while the red and green arrows indicate the H-bond acceptor and H-bond donor, respectively.

In pharmacophore-based virtual screening, the geometric fit of a molecule's features to each 3D structure-based pharmacophore model was evaluated using the pharmacophore fit score. This study determined the pharmacophore fit score of 75 compounds, selected based on drug-likeness screening, using KNIME. Virtual screening used the pharmacophore models of inhibitors 11a, 13b, N3, and the nirmatrelvir/3CLpro complex. As a result, 41 compounds were identified that matched the features of the 11a pharmacophore model, 49 compounds matched the features of the 13b pharmacophore model, 2 compounds matched the features of the N3 pharmacophore model, and 5 compounds matched the features of the nirmatrelvir/3CLpro pharmacophore model. Overall, this screening process demonstrated that 60 compounds from the focused library matched the features of the four pharmacophore models (Table 4).

Table 4 Pharmacophore fit score of 60 compounds derived from the pharmacophore-based virtual screening

Compound	Model			
	11a/3CLpro	13b/3CLpro	N3/3CLpro	nirmatrelvir/3CLpro
STK001		35.71		
STK002	35.86	36.78		
STK004		36.99		
STK005	38.05	36.21		
STK008	38.03	36.22		
NST005	37.84	46.2		46.42
NPT018	37.43			
NPT019	38.38			

Compound	Model			
	11a/3CLpro	13b/3CLpro	N3/3CLpro	nirmatrelvir/3CLpro
NPP002		37.17		
NWC014	36.74	36.02		
NWC015	36.71	37.18		
NWC016	36.74	36.02		
NWC017	36.95			
NPT044		37.81		
NPT052		36.34		
SPT054		43.91		
SPT058		36.02		
SPT061		36.26		
SPT063		35.71		
SPT076		35.71		
SPT078		37.6		
NTK015	36.64			
NTK016	35.61			
NTK028	38.44	37.86		45.42
PPP032	36.42	45.83		
PPP033	36.23			
PPP034	45.41	44.39		
PPP035	38.81	47.05		
PPP036	39.63	53.91		
PPP037	43.85	37.87		
PPP040		45.78		
PPP041	36.24	56.49		
PPP042		56.46		
PPP043	55.07	37.69		
PPP044	45.31	57.34		
PPP048		38.46		
PPP049	54.84	38.54		
PPP050	36.25	37.78		
PPP051		56.32		
SWC111		37.42		
SWC113		35.65		
SWC122		35.48		
SWC123		36.08		
NPP027	36.36	56.08		
NPP030	35.9	37.82		
NWC084	37.43			
NWC088	38.36			
SWC195	36.68	37.96		
NWC097	36.82	35.68		
NWC098	36.71	37.18		

Compound	Model			
	11a/3CLpro	13b/3CLpro	N3/3CLpro	nirmatrelvir/3CLpro
NWC099	36.74	36.02		
NWC100	37.43			
NTK049	35.54	44.52		
NTK066	36.37			
NTK073	35.9	36.91		
SWC272	36.02			
SWC307	38.54	36.58		46.99
SWC422	38.1	35.78	45.61	53.32
SWC423	34.31	35.94	43.22	55.43
SWC424	43.65	56.21		

Four datasets consisting of actives and decoys were used to validate the pharmacophore model. By comparing the features of the active compounds to those of the decoys, the model could learn the specific patterns and characteristics associated with active compounds. This enabled the model to differentiate between compounds with the desired biological activity and those without it.(Kaserer, Beck, Akram, Odermatt, & Schuster, 2015; T. Seidel, Wieder, Garon, & Langer, 2020) The model's performance distinguishing true active compounds from decoys was assessed using AUC and EF values. A high AUC value, typically above 0.7, indicates a strong capability of the four models to accurately classify true active compounds from decoys (Figure 2).(Hamza, Wei, & Zhan, 2012)

The pharmacophore model yielded significant hits, with 550, 694, 3, and 44 compounds identified for inhibitors 11a, 13b, N3, and nirmatrelvir, respectively. Among these models, the 13b pharmacophore model exhibited the highest performance, with an impressive EF of 22.6 and AUC values of 1.00, 1.00, 1.00, and 0.85 at 1%, 5%, 10%, and 100% of the screened database, respectively. These results indicate that the 13b model is highly sensitive and specific, successfully identifying 694 active compounds and 49 decoys. However, the other models (11a, N3, and nirmatrelvir) also demonstrated successful discrimination between true actives and decoys. Therefore, all the models possess the capability to accurately identify inhibitors with high accuracy and favorable quality for effective pharmacophore-based virtual screening. Notably, this test set validation demonstrates that the pharmacophore model can identify active molecules. Nonetheless, it does not guarantee the exclusion of compounds lacking binding affinity.
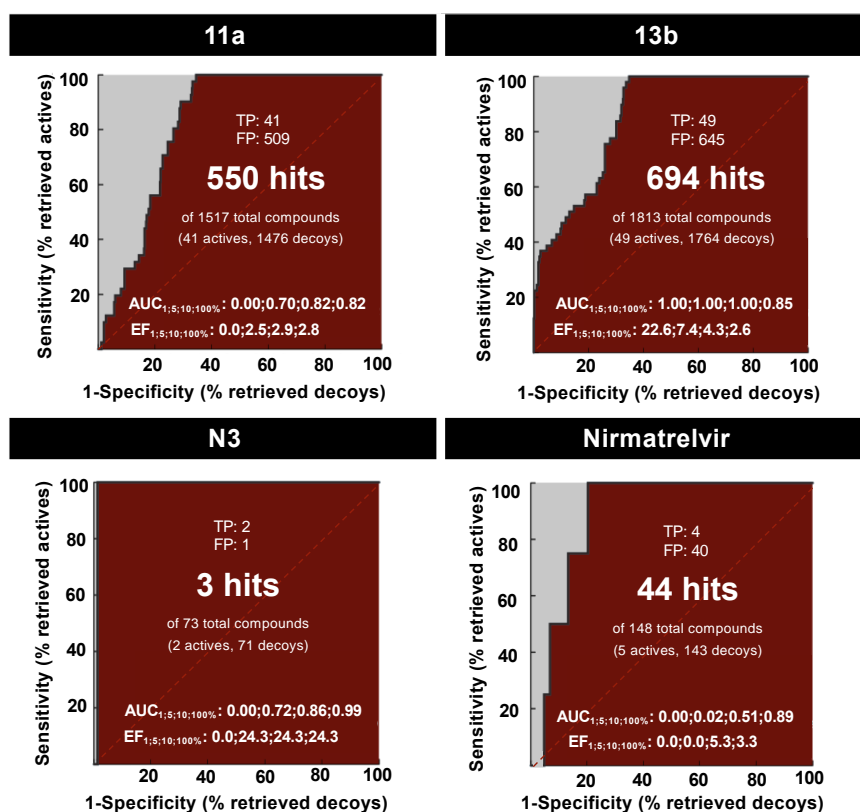
Figure 2 Curve validation of receiver operating characteristic (ROC). It was generated by KNIME 4.4.0 based on the recognized ability of actives to decoys of four structure-based pharmacophore models. Actives were selected hit compounds from pharmacophore-based virtual screening, and decoys were based on the dataset downloaded from ZINC15 and DrugBank databases.

3.1.3 Common hits identification

Molecular docking studies were conducted on candidate compounds to refine the hits obtained through pharmacophore screening. All 60 compounds, along with known inhibitors 11a, 13b, N3, and nirmatrelvir, were individually docked into the active site of SARS-CoV-2 3CLpro using AutoDock VinaXB(Trott & Olson, 2010). The binding free energy (ΔG) and RMSD below 2 Å were used to assess the agreement between experimental and predicted binding poses. The structural arrangement of the known inhibitors guided the evaluation of the orientation of the docked poses. The predicted ΔG values of all 60 compounds were compared with the reported inhibitors from crystal structures 6Y2F (11a, -7.2 kcal/mol), 6LU7 (13b, -7.4 kcal/mol), 6LZE (N3, -7.6 kcal/mol), and 7VH8 (nirmatrelvir, -7.3 kcal/mol). In Figure 3, compounds with a lower ΔG than the known inhibitors were highlighted in yellow, indicating a favorable binding affinity. On the other hand, compounds with a higher ΔG was shown in gray, suggesting weaker binding or unfavorable interactions with the target protein, as described by Pantsar and Poso(Pantsar & Poso, 2018).

	6Y2F	6LU7	6LZE	7VH8		6Y2F	6LU7	6LZE	7VH8
Inhibitor	-7.2	-7.4	-7.6	-7.3	PPP040	-5.0	-5.0	-4.7	-4.7
STK001	-5.2	-5.5	-5.5	-5.4	PPP041	-5.0	-4.8	-5.2	-5.0
STK002	-5.3	-5.7	-5.5	-5.3	PPP042	-5.3	-5.1	-5.3	-5.0
STK004	-5.2	-5.6	-5.7	-5.4	PPP043	-5.3	-5.7	-5.8	-5.5
STK005	-5.3	-5.9	-5.8	-6.4	PPP044	-5.6	-6.1	-5.6	-5.5
STK008	-5.6	-6.1	-5.8	-6.4	PPP048	-4.8	-5.0	-4.7	-4.8
NST005	-7.6	-9.0	-7.8	-7.7	PPP049	-4.9	-5.2	-5.1	-4.5
NPT018	-7.2	-7.4	-7.3	-7.6	PPP050	-4.9	-5.4	-4.4	-5.0
NPT019	-6.8	-7.3	-6.7	-7.5	PPP051	-4.9	-4.9	-5.2	-5.3
NPP002	-5.6	-6.2	-6.7	-6.6	SWC111	-5.7	-6.0	-6.1	-6.1
NWC014	-5.4	-5.0	-5.2	-5.3	SWC113	-5.0	-5.0	-5.4	-5.3
NWC015	-5.2	-5.3	-5.4	-5.5	SWC122	-5.2	-5.5	-5.6	-5.9
NWC016	-5.2	-5.0	-5.2	-5.3	SWC123	-5.4	-6.4	-5.6	-5.5
NWC017	-4.9	-5.2	-5.4	-5.1	NPP027	-7.7	-7.5	-7.5	-7.3
NPT044	-8.0	-7.9	-7.8	-7.8	NPP030	-5.4	-5.7	-5.7	-5.6
NPT052	-6.1	-6.0	-6.5	-6.9	NWC084	-7.2	-7.4	-7.2	-7.5
SPT054	-4.9	-5.5	-5.6	-5.5	NWC088	-7.2	-7.2	-6.9	-7.8
SPT058	-5.2	-5.6	-5.7	-5.4	SWC195	-5.8	-5.9	-5.5	-6.2
SPT061	-5.4	-5.6	-5.5	-5.7	NWC097	-4.7	-5.4	-5.3	-5.0
SPT063	-5.3	-5.4	-5.5	-5.9	NWC098	-5.5	-5.3	-5.4	-5.5
SPT076	-6.0	-6.5	-5.7	-5.9	NWC099	-5.3	-5.0	-5.1	-5.3
SPT078	-5.4	-5.7	-5.9	-5.8	NWC100	-7.2	-7.3	-7.2	-7.5
NTK015	-6.0	-6.4	-5.9	-6.7	NTK049	-6.2	-6.6	-6.4	-6.9
NTK016	-5.3	-5.1	-5.4	-5.8	NTK066	-6.1	-6.5	-6.3	-6.6
NTK028	-7.5	-7.5	-6.9	-7.5	NTK073	-6.6	-6.2	-6.5	-6.5
PPP032	-5.8	-5.3	-5.1	-5.6	SWC272	-4.8	-5.0	-5.3	-5.2
PPP033	-4.9	-5.5	-4.9	-5.2	SWC307	-6.4	-6.4	-6.4	-6.2
PPP034	-5.5	-5.9	-5.7	-6.0	SWC422	-7.7	-7.9	-6.9	-8.0
PPP035	-5.4	-5.5	-5.5	-5.5	SWC423	-7.3	-8.3	-8.3	-8.4
PPP036	-5.5	-5.9	-5.5	-5.5	SWC424	-7.6	-7.5	-8.8	-8.4
PPP037	-5.9	-5.8	-5.8	-6.0					

ΔG (kcal/mol)

Min ▮▮▮▮▮ Max

Reference inhibitor
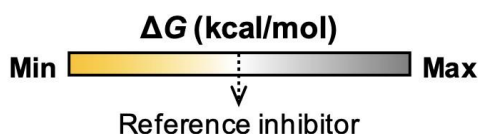
Figure 3 A heatmap generated using AutoDock VinaXB to display the ΔG (kcal/mol) values for the complex formed by 60 in-house compounds and SARS-CoV-2 3CLpro, with the first row (white box) showing ΔG values of re-docked inhibitors (11a, 13b, N3, and nirmatrelvir) in the crystal structures 6Y2F, 6LU7, 6LZE, and 7VH8, respectively. The compounds with a lower ΔG than the known inhibitor in each column were indicated by a yellow box, while those with a higher ΔG was represented by a gray box.

Among the compounds, NST005, NPT018, NPT019, NPT044, NTK028, NPP027, NWC084, NWC088, NWC100, SWC422, SWC423, and SWC424 exhibited more negative ΔG values than their

reference inhibitors. This observation suggests these compounds have a favorable binding affinity and a likelihood of tight binding to the 3CLpro target. Conversely, the remaining compounds displayed higher positive ΔG values, indicating weaker binding or unfavorable interactions with the target protein. To further evaluate the 12 compounds with lower ΔG, flexible docking using the GOLD 5.6.3 suite(Jones et al., 1997) was performed. The results of the GOLD fitness score, presented in Figure 4, revealed that a series of three sulfonamide chalcones (SWC422, SWC423, and SWC424) exhibited higher GOLD fitness scores, indicating better binding affinity to the active site of SARS-CoV-2 3CLpro compared to nirmatrelvir.

	6Y2F	6LU7	6LZE	7VH8
Inhibitor	72.7	78.0	70.2	63.9
NST005	63.7	69.4	60.3	62.5
NPT018	47.7	50.6	-	49.8
NPT019	-	-	-	46.7
NPT044	48.4	46.9	50.5	50.9
NTK028	57.6	62.2	-	60.9
NPP027	58.2	60.4	-	61.7
NWC084	46.8	50.9	-	48.6
NWC088	43.3	-	-	47.4
NWC100	47.2	-	-	49.9
SWC422	66.8	68.3	-	72.3
SWC423	63.8	65.5	63.0	68.6
SWC424	67.4	69.5	66.2	71.8

GOLD fitness score

Min ▮▮▮▮▮ Max

Reference inhibitor
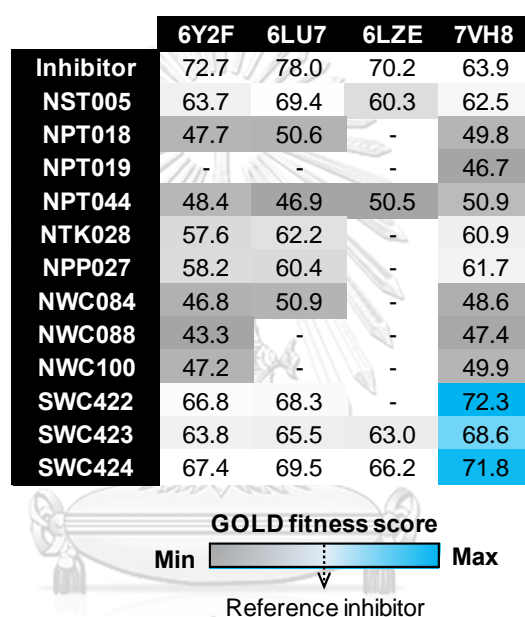
Figure 4 GOLD fitness score for 12 candidate compounds with their respective SARS-CoV-2 models relative to the known inhibitors 11a (6Y2F), 13b (6LU7), N3 (6LZE), and nirmatrelvir (7VH8).

Previous studies have highlighted the inhibitory potential of natural-based chalcones against the 3CLpro enzyme of SARS and MERS coronaviruses,[Elkhalifa, Al-Hashimi, Al Moustafa, & Khalil, 2021; Park et al., 2016; Valipour, 2022] as well as their antiviral activity against other viruses such as influenza(Dao et al., 2011), herpes simplex(Phrutivorapongkul et al., 2003), hepatitis C(Mateeva et al., 2017), HIV(Cole, Hossain, Cole, & Phanstiel, 2016), and Zika(Mottin et al., 2022). Given the wide range of natural-based and synthetic chalcone derivatives available, further investigation is warranted to discover more potent derivatives specifically targeting SARS-CoV-2.

The three sulfonamide chalcones were superimposed with nirmatrelvir in the active site of SARS-CoV-2 3CLpro, as shown in Figure 5A. The docked poses of the sulfonamide chalcones aligned in the same orientation as nirmatrelvir, indicating their potential as anti-SARS-CoV-2

3CLpro agents. Figure 5B illustrates the intermolecular interactions of the three docked complexes. All compounds shared binding residues (T25, L27, M49, N142, G143, C145, and E166) at the active site of SARS-CoV-2 3CLpro. Hydrophobic interactions were observed with residues T25, L27, M49, M165, and E166, similar to N3 and 13b inhibitors.(K. Sanachai et al., 2022) Hydrogen bond acceptors were found at N142 and G143, while a hydrogen bond donor interacted with C145, resembling the interaction observed with other compounds such as remdesivir, paritaprevir, glecaprevir, and lopinavir.(Hasan et al., 2021) It should be noted that there may be differences in the predicted binding residues compared to other studies, which could be attributed to variations in docking methods and experimental techniques for structural analysis.
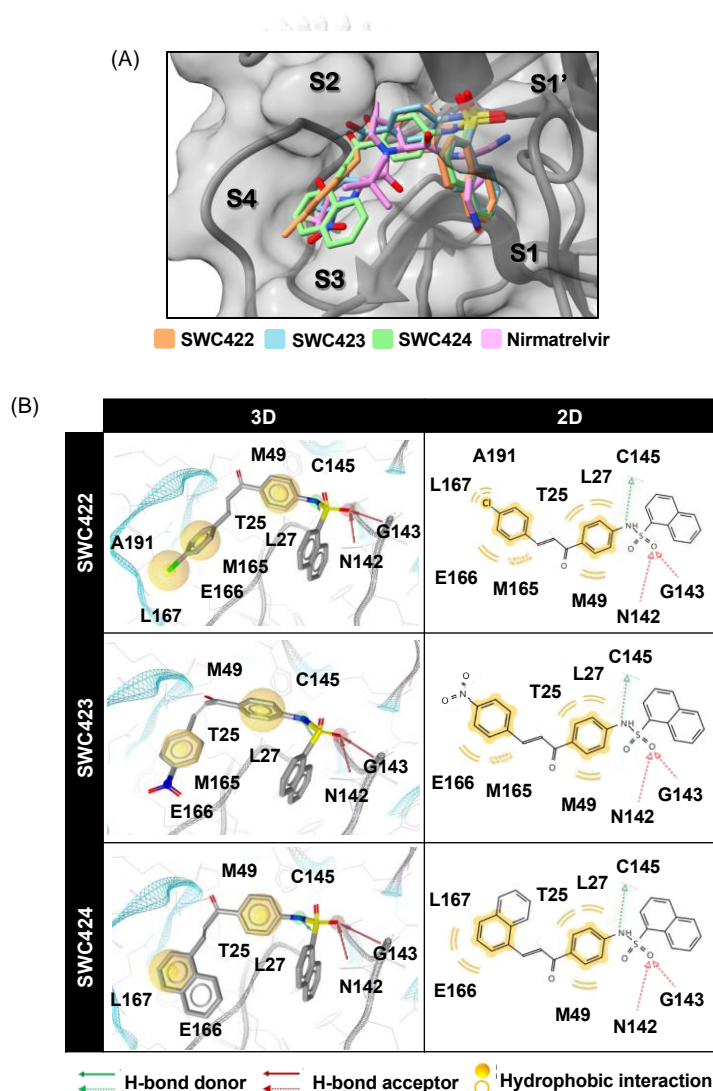
Figure 5 (A) Structural overlay of the docked structures (licorice three-dimensional model) of sulfonamide chalcones (SWC422-424) and nirmatrelvir in the substrate binding cleft of SARS-CoV-

2 3CLpro (PDB code: 7VH8). (B) Intermolecular interactions of three screened sulfonamide chalcones interacting with SARS-CoV-2 3CLpro obtained from LigandScout.

According to our findings, the sulfonamide moiety of the compounds formed three hydrogen bonds, one with the catalytic residue C145 and the other with residues N142 and G143. The phenyl ring of the α,β-unsaturated ketone interacted hydrophobically with residues T25, L27, and M49, while the aromatic substituents induced hydrophobic interactions with M165, E166, and L167. Additionally, the chlorine (Cl) attached to the aromatic ring of SWC422 interacted with L167 and A191 in the SARS-CoV-2 3CLpro binding pocket. The three sulfonamide chalcones exhibited similar intermolecular interactions with important residues (M49, G143, C145, and E166) as nirmatrelvir in complex with 3CLpro.(Zhao et al., 2022) Noting that binding free energies do not fully define binding equilibria because they do not consider entropic contributions and the influence of water polarity. Then the collaborators in research conducted experimental investigations (results shown in Figure 14 of the appendix), and we conducted MD simulations to overcome these limitations. Studies on experiments shed light on the behavior of hit compounds in practical circumstances. At the same time, MD simulations enable a more comprehensive analysis of the binding characteristics of our hits in a solution environment.(Gapsys et al., 2021)

3.1.4 Dynamics and stability of SWC423 binding to SARS-CoV-2 3CLpro

The binding dynamics of the three SWC inhibitors were investigated through a 500-ns MD simulation. The stability of the SWC423/SARS-CoV-2 3CLpro complex was assessed based on the all-atom RMSD, # H-bonds, and # Atom contacts over the simulation time, as depicted in Figure 6A. Initially, the RMSD values of the complex increased during the first 100 ns but subsequently reached a stable state, oscillating around an average value of 1.85 Å. This observation indicates a high level of overall stability throughout the simulation. The equilibrium phase was achieved at 300 ns and maintained until the end of the 500-ns simulation, with an average of 3 ± 1 # H-bonds and 16 ± 5 # atom contacts.

To further explore the intermolecular interactions, the last 200 ns of the MD trajectories were analyzed, focusing on the hydrogen bonding between SWC423 and SARS-CoV-2 3CLpro. Hydrogen bonds play a critical role in stabilizing ligands within the open conformational environment of protein structures. Our analysis revealed the formation of eight hydrogen bonds in the SWC423 inhibitor. Notably, these hydrogen bonds predominantly occurred at specific sites, including the sulfonamide moiety (N1...OG1(T25) at 65%, O1...N(T26) at 28%, O1...OG1(T25) at 15%, O2...OG1(T45) at 18%, and O2...OG1(T24) at 12.54%), the chalcone carbonyl group (O3...N(G143) at 81%), and the nitro benzyl group (O4...NE2(H163) at 32% and O5...NE2(H163) at 13%). For clarity, only hydrogen bonds with occupancies greater than 30% are depicted in Figure

6B. Interestingly, our results revealed that the binding orientation of SWC423 to the active site of 3CLpro differed from the initial prediction obtained through GOLD docking (Figure 5B). Specifically, the nitro group of SWC423 showed an affinity for residue H163, while the sulfonamide nitrogen moved closer to T26. Moreover, the oxygen atom of the α,β-unsaturated ketone formed strong hydrogen bonding with G143, which was not anticipated in the docking structure. These findings suggest that protein flexibility, solvent effects, differences in force fields between docking and MD simulation, and the ligand's adaptability to the protein environment during the MD simulation may contribute to these observed differences.

To gain insights into the energy contributions of key binding residues interacting with SWC423, fragment molecular orbital calculation using PIE and decomposition analysis (PIEDA) was conducted. The analysis revealed that electrostatic interactions involving G23, R40, S139, and S1' played a significant role, as indicated by the ΔE_{ij}^{ES} shown in Figure 6B. These findings highlight the importance of electrostatic interactions in binding SWC423 to the target protein. In addition to electrostatic interactions, hydrophobic interactions involving T25, T45, M49, and N142 were observed, as supported by ΔE_{ij}^{DI}. These interactions are crucial in maintaining the compound within the binding pocket. Notably, the S-π interaction between M49 and the aromatic ring of SWC423 contributed to the binding complex's overall stability, which aligns with previous reports.(Hengphasatporn, Harada, et al., 2022). However, repulsion effects (ΔE_{ij}^{ES}) were observed in residues L27, E47, G146, E166, A173, and D187, as indicated by PIEDA values greater than 10 kcal/mol. These repulsion effects could be attributed to the specific arrangement of the ligand within the active site.
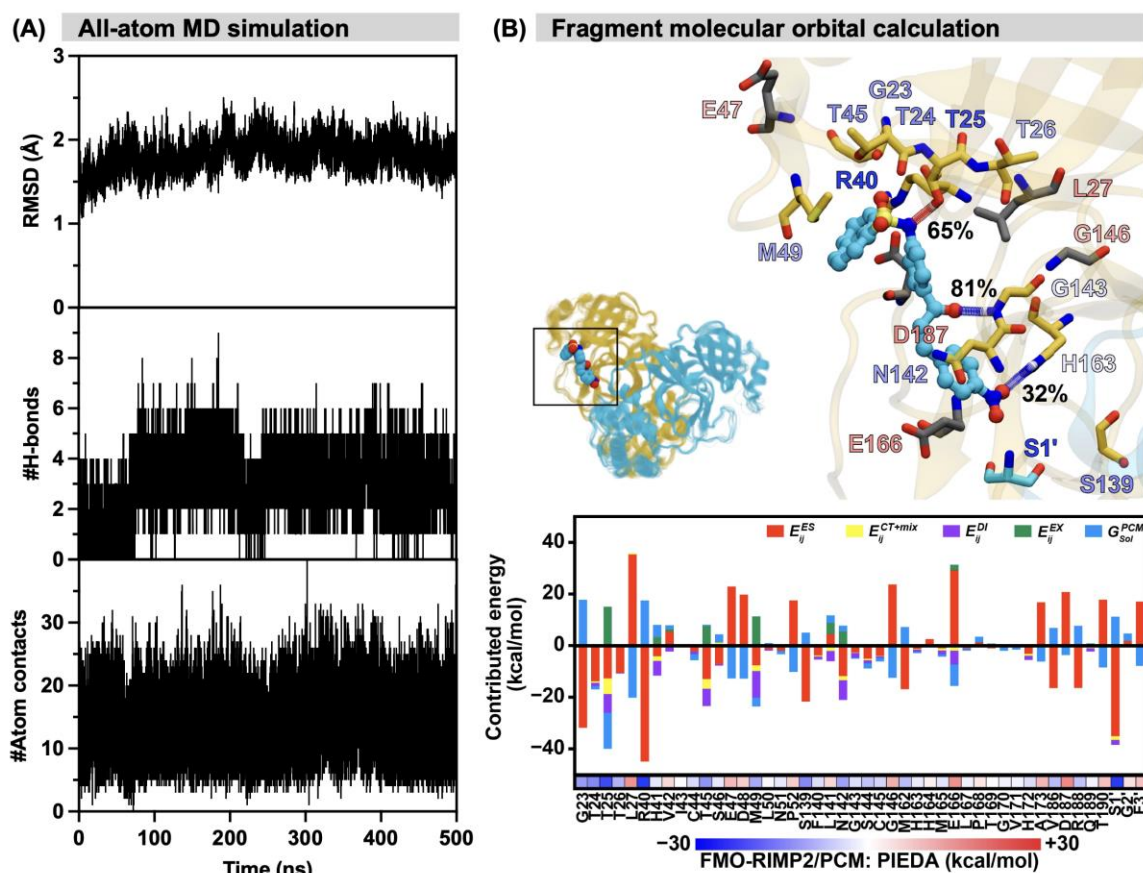
Figure 6 (A) Analysis of the structure and dynamics of SWC423 binding to SARS-CoV-2 3CLpro over 500 ns-MD trajectories, represented by all-atom RMSD, # H-bonds, and # atom contacts. (B) Binding pattern of SWC423/3CLpro complex, determined from fragment molecular orbital calculation with RIMP2/PCM on the last snapshot of the MD simulation. Strong H-bond interactions and residues with >10 kcal/mol and <-10 kcal/mol binding energy are labeled.

3.2 Research 2

3.2.1 Evaluation of drug-like properties

Evaluating compounds with drug-like properties is crucial in drug discovery as it helps identify potential and safe compounds for further development. According to Lipinski's rule, most compounds from our database met these criteria, with 490, 488, 549, and 515 compounds fulfilling the requirements for molecular weight, number of H-bond donors, number of H-bond acceptors, and cLogP, respectively (Figure 7). These results indicate that these compounds are likely to exhibit favorable drug-like properties, including oral bioavailability and permeability probability. However, some compounds did not meet the criteria for physicochemical properties and drug-likeness profiles. These provide insights into a compound's behavior in biological systems, such as solubility, stability, and safety.(Council, 2014; Daina et al., 2017; Martin, 2005) To

increase the chances of identifying compounds with improved pharmacokinetics, bioavailability, and therapeutic potential for further development, selecting compounds that fulfill all the criteria is crucial.
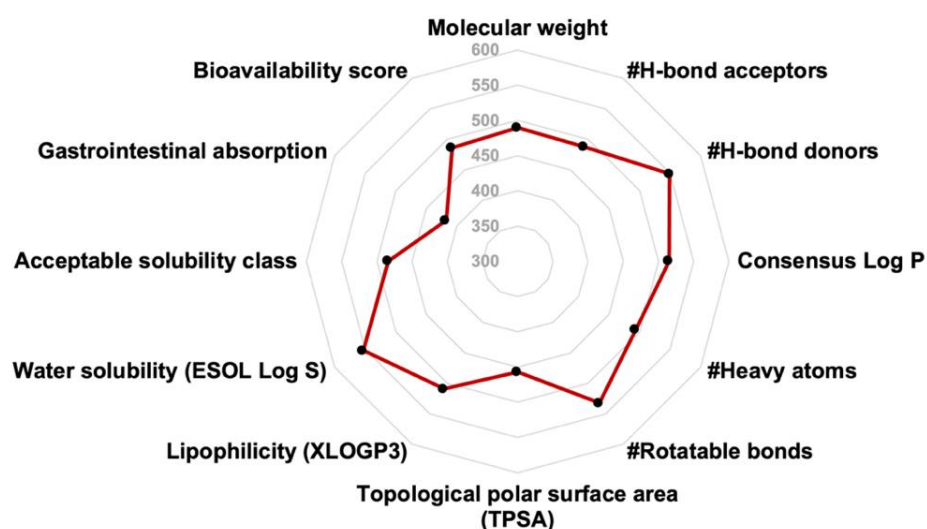
Figure 7 Radar chart with the compounds eligible for drug-like characteristic

In our subsequent studies, we focused on 358 compounds derived from 11 different groups, including α-mangostin, anticancer agents, avicequinones, caffeic acid derivatives, acid homodimers, xanthones, avenalumic acid derivatives, quinonoids, steroids, piperic acid derivatives, and ureas. These compounds were selected based on their properties predicted by SwissADME, listed in Table 5.

Table 5 Drug-like properties of 358 selected compounds predicted by SwissADME

Compound	Lipinski's rules of five				Physicochemical properties						Drug-likeness profiles	
	MW	#HBA	#HBD	cLogP	#Heavy atoms	#Rotatable bonds	TPSA	XLOGP3	ESOL Log S	ESOL Class	GI Absorption	Bioavai score
α-Mangostins												
1a	408.5	4	3	4.27	30	5	111.71	5.62	-5.93	Moderately	High	0.55
Anticancer agents												
2a	368.4	6	2	3.03	27	8	93.06	3.2	-3.94	Soluble	High	0.55
2b	270.3	4	0	3.31	20	3	52.58	3.5	-4	Moderately	High	0.55
2c	284.3	5	2	2.56	21	2	79.9	3.49	-4.23	Moderately	High	0.55
2d	290.3	5	3	2.61	21	5	79.15	3.02	-3.64	Soluble	High	0.55
2e	304.3	5	2	2.86	22	6	68.15	3	-3.62	Soluble	High	0.55
2f	274.3	4	2	2.86	20	5	58.92	3.03	-3.56	Soluble	High	0.55
2g	332.4	5	0	3.55	24	8	46.15	3.58	-4	Soluble	High	0.55
2h	318.4	5	1	3.19	23	7	57.15	3.25	-3.79	Soluble	High	0.55
2i	318.4	5	1	3.22	23	7	57.15	3.33	-3.84	Soluble	High	0.55
2j	198.2	4	1	1.65	14	3	47.92	1.38	-2.06	Soluble	High	0.55

Compound	Lipinski's rules of five				Physicochemical properties						Drug-likeness profiles	
	MW	#HBA	#HBD	cLogP	#Heavy atoms	#Rotatable bonds	TPSA	XLOGP3	ESOL Log S	ESOL Class	GI Absorption	Bioavai score
2k	344.4	6	0	2.25	25	3	78.9	2.4	-3.29	Soluble	High	0.55
2l	386.5	4	2	3.68	28	1	70.67	3.2	-4.35	Moderately	High	0.55
2m	270.2	5	3	2.24	20	1	90.9	3.16	-4.03	Moderately	High	0.55
2n	284.3	5	2	2.54	21	2	79.9	3.49	-4.23	Moderately	High	0.55
Avicequinones												
3a	256.3	4	1	1.99	19	1	67.51	2.01	-3.06	Soluble	High	0.55
3b	240.2	4	0	1.95	18	1	64.35	2.28	-3.15	Soluble	High	0.55
3c	254.2	4	0	2.3	19	2	64.35	2.75	-3.45	Soluble	High	0.55
3d	198.2	3	0	1.9	15	0	47.28	2.27	-3.04	Soluble	High	0.55
Caffeic acid derivatives												
4a	193.2	3	3	0.93	14	3	69.56	0.9	-1.72	Soluble	High	0.55
4b	207.2	3	3	1.29	15	4	69.56	1.4	-2.04	Soluble	High	0.55
4c	221.3	3	3	1.55	16	4	69.56	1.7	-2.3	Soluble	High	0.55
4d	235.3	3	3	1.93	17	5	69.56	2.37	-2.72	Soluble	High	0.55
4e	221.3	3	3	1.65	16	5	69.56	1.93	-2.38	Soluble	High	0.55
4f	249.3	3	3	2.23	18	5	69.56	2.76	-3.04	Soluble	High	0.55
4g	235.3	3	3	1.99	17	6	69.56	2.29	-2.61	Soluble	High	0.55
4h	323.3	6	3	3.29	23	5	69.56	3.36	-4.02	Moderately	High	0.55
4i	323.3	6	3	3.23	23	5	69.56	3.24	-3.94	Soluble	High	0.55
4j	323.3	6	3	3.24	23	5	69.56	3.34	-4	Moderately	High	0.55
4k	194.2	4	2	1.35	14	3	66.76	1.48	-2.1	Soluble	High	0.55
4l	208.2	4	2	1.82	15	4	66.76	2.56	-2.78	Soluble	High	0.55
4m	222.2	4	2	2	16	4	66.76	2.28	-2.67	Soluble	High	0.55
4n	236.3	4	2	2.42	17	5	66.76	3.52	-3.45	Soluble	High	0.55
4o	222.2	4	2	2.21	16	5	66.76	3.17	-3.16	Soluble	High	0.55
4p	250.3	4	2	2.68	18	5	66.76	3.92	-3.78	Soluble	High	0.55
4q	236.3	4	2	2.46	17	6	66.76	3.53	-3.39	Soluble	High	0.55
4r	324.3	7	2	3.61	23	5	66.76	3.93	-4.38	Moderately	High	0.55
4s	324.3	7	2	3.61	23	5	66.76	3.93	-4.38	Moderately	High	0.55
4t	324.3	7	2	3.62	23	5	66.76	3.92	-4.38	Moderately	High	0.55
4u	392.3	10	2	4.61	4.61	6	66.76	4.82	-5.24	Moderately	High	0.55
4v	392.3	10	2	4.61	4.61	6	66.76	4.82	-5.24	Moderately	High	0.55
Acid homodimers												
5a	318.3	7	2	2.15	23	6	102.29	2.45	-3.35	Soluble	High	0.56
5b	304.3	7	3	1.84	22	5	113.29	2.49	-3.37	Soluble	High	0.56
5c	304.3	7	3	1.78	22	5	113.29	2.12	-3.14	Soluble	High	0.56
5d	290.2	7	4	1.38	21	4	124.29	1.79	-2.93	Soluble	High	0.56
5e	290.2	7	4	1.43	21	4	124.29	1.78	-2.92	Soluble	High	0.56
5f	304.3	7	3	1.82	22	5	113.29	2.49	-3.37	Soluble	High	0.56
5g	288.3	6	2	2.07	21	5	93.06	2.54	-3.32	Soluble	High	0.56
5h	304.3	7	3	1.82	22	5	113.29	2.67	-3.48	Soluble	High	0.56
5i	304.3	7	3	1.75	22	5	113.29	2.18	-3.17	Soluble	High	0.56
5j	290.2	7	4	1.38	21	4	124.29	1.79	-2.93	Soluble	High	0.56
5k	274.2	6	3	1.81	20	4	104.06	2.35	-3.2	Soluble	High	0.56
5l	290.2	7	4	1.43	21	4	124.29	2.26	-3.22	Soluble	High	0.56
5m	290.2	7	4	1.37	21	4	124.29	1.96	-3.03	Soluble	High	0.56
5n	290.2	7	4	1.51	21	4	124.29	2.34	-3.27	Soluble	High	0.56
5o	290.2	7	4	1.43	21	4	124.29	2.34	-3.27	Soluble	High	0.56
5p	288.3	6	2	2.24	21	5	93.06	3.09	-3.67	Soluble	High	0.56

Compound	Lipinski's rules of five				Physicochemical properties						Drug-likeness profiles	
	MW	#HBA	#HBD	cLogP	#Heavy atoms	#Rotatable bonds	TPSA	XLOGP3	ESOL Log S	ESOL Class	GI Absorption	Bioavai score
5q	274.2	6	3	1.94	20	4	104.06	2.9	-3.55	Soluble	High	0.56
5r	304.3	7	3	1.87	22	5	113.29	2.73	-3.52	Soluble	High	0.56
5s	290.2	7	4	1.55	21	4	124.29	2.51	-3.38	Soluble	High	0.56
5t	304.3	7	3	1.99	22	5	113.29	2.73	-3.52	Soluble	High	0.56
5u	290.2	7	4	1.56	21	4	124.29	2.34	-3.27	Soluble	High	0.56
5v	304.3	7	3	1.8	22	5	113.29	2.34	-3.27	Soluble	High	0.56
5w	290.2	7	4	1.44	21	4	124.29	2.16	-3.16	Soluble	High	0.56
5x	274.2	6	3	1.87	20	4	104.06	2.9	-3.55	Soluble	High	0.56
5y	258.2	5	2	2.26	19	4	83.83	3.04	-3.56	Soluble	High	0.56
5z	274.2	6	3	1.96	20	4	104.06	3.16	-3.71	Soluble	High	0.56
5aa	274.2	6	3	1.84	20	4	104.06	2.9	-3.55	Soluble	High	0.56
5ab	290.2	7	4	1.51	21	4	124.29	2.51	-3.38	Soluble	High	0.56
5ac	274.2	6	3	1.89	20	4	104.06	2.9	-3.55	Soluble	High	0.56
5ad	290.2	7	4	1.66	21	4	124.29	2.81	-3.57	Soluble	High	0.56
5ae	290.2	7	4	1.5	21	4	124.29	2.51	-3.38	Soluble	High	0.56
5af	290.2	7	4	1.53	21	4	124.29	2.55	-3.4	Soluble	High	0.56
5ag	274.2	6	3	1.94	20	4	104.06	2.69	-3.41	Soluble	High	0.56
5ah	290.2	7	4	1.72	21	4	124.29	2.81	-3.57	Soluble	High	0.56
5ai	290.2	7	4	1.45	21	4	124.29	2.55	-3.4	Soluble	High	0.56
5aj	290.2	7	4	1.34	21	4	124.29	1.79	-2.93	Soluble	High	0.56
5ak	274.2	6	3	1.72	20	4	104.06	2.14	-3.07	Soluble	High	0.56
5al	290.2	7	4	1.6	21	4	124.29	2.8	-3.56	Soluble	High	0.56
5am	290.2	7	4	1.32	21	4	124.29	1.79	-2.93	Soluble	High	0.56
5an	304.3	7	3	1.88	22	5	113.29	2.73	-3.52	Soluble	High	0.56
5ao	290.2	7	4	1.54	21	4	124.29	2.51	-3.38	Soluble	High	0.56
5ap	304.3	7	3	1.89	22	5	113.29	2.73	-3.52	Soluble	High	0.56
5aq	290.2	7	4	1.43	21	4	124.29	2.34	-3.27	Soluble	High	0.56
5ar	330.3	7	3	2.13	24	6	113.29	2.43	-3.39	Soluble	High	0.56
5as	316.3	7	4	1.74	23	5	124.29	2.22	-3.26	Soluble	High	0.56
5at	290.2	7	4	1.52	21	4	124.29	2.51	-3.38	Soluble	High	0.56
5au	274.2	6	3	1.89	20	4	104.06	2.9	-3.55	Soluble	High	0.56
5av	290.2	7	4	1.53	21	4	124.29	2.81	-3.57	Soluble	High	0.56
5aw	290.2	7	4	1.41	21	4	124.29	2.51	-3.38	Soluble	High	0.56
5ax	290.2	7	4	1.54	21	4	124.29	2.55	-3.4	Soluble	High	0.56
5ay	274.2	6	3	1.9	20	4	104.06	2.69	-3.41	Soluble	High	0.56
5az	290.2	7	4	1.66	21	4	124.29	2.81	-3.57	Soluble	High	0.56
5ba	290.2	7	4	1.5	21	4	124.29	2.34	-3.27	Soluble	High	0.56
5bb	316.3	7	4	1.72	23	5	124.29	2.22	-3.26	Soluble	High	0.56
5bc	300.3	6	3	2.13	22	5	104.06	2.57	-3.39	Soluble	High	0.56
5bd	316.3	7	4	1.87	23	5	124.29	2.77	-3.6	Soluble	High	0.56
5be	316.3	7	4	1.72	23	5	124.29	2.22	-3.26	Soluble	High	0.56
5bf	304.3	7	3	1.76	22	5	113.29	2.67	-3.48	Soluble	High	0.56
5bg	304.3	7	3	1.73	22	5	113.29	2.12	-3.14	Soluble	High	0.56
5bh	304.3	7	3	1.82	22	5	113.29	2.49	-3.37	Soluble	High	0.56
5bi	304.3	7	3	1.91	22	5	113.29	2.67	-3.48	Soluble	High	0.56
5bj	290.2	7	4	1.45	21	4	124.29	2.34	-3.27	Soluble	High	0.56
5bk	290.2	7	4	1.32	21	4	124.29	1.79	-2.93	Soluble	High	0.56
5bl	290.2	7	4	1.46	21	4	124.29	2.21	-3.19	Soluble	High	0.56
5bm	290.2	7	4	1.55	21	4	124.29	2.26	-3.22	Soluble	High	0.56

Compound	Lipinski's rules of five				Physicochemical properties						Drug-likeness profiles	
	MW	#HBA	#HBD	cLogP	#Heavy atoms	#Rotatable bonds	TPSA	XLOGP3	ESOL Log S	ESOL Class	GI Absorption	Bioavai score
5bn	304.3	7	3	1.72	22	5	113.29	2.12	-3.14	Soluble	High	0.56
5bo	304.3	7	3	1.76	22	5	113.29	2.73	-3.52	Soluble	High	0.56
5bp	330.3	7	3	2.18	24	6	113.29	2.57	-3.48	Soluble	High	0.56
5bq	330.3	7	3	2.1	24	6	113.29	2.43	-3.39	Soluble	High	0.56
5br	290.2	7	4	1.36	21	4	124.29	1.79	-2.93	Soluble	High	0.56
5bs	290.2	7	4	1.45	21	4	124.29	2.34	-3.27	Soluble	High	0.56
5bt	316.3	7	4	1.75	23	5	124.29	2.22	-3.26	Soluble	High	0.56
5bu	316.3	7	4	1.67	23	5	124.29	2.22	-3.26	Soluble	High	0.56
5bv	274.2	6	3	1.98	20	4	104.06	3.25	-3.77	Soluble	High	0.56
5bw	274.2	6	3	1.89	20	4	104.06	2.7	-3.42	Soluble	High	0.56
5bx	274.2	6	3	1.79	20	4	104.06	2.61	-3.36	Soluble	High	0.56
5by	274.2	6	3	1.95	20	4	104.06	3.16	-3.71	Soluble	High	0.56
5bz	290.2	7	4	1.74	21	4	124.29	3.1	-3.75	Soluble	High	0.56
5ca	290.2	7	4	1.53	21	4	124.29	2.34	-3.27	Soluble	High	0.56
5cb	290.2	7	4	1.4	21	4	124.29	2.26	-3.22	Soluble	High	0.56
5cc	290.2	7	4	1.61	21	4	124.29	2.81	-3.57	Soluble	High	0.56
5cd	290.2	7	4	1.67	21	4	124.29	2.9	-3.63	Soluble	High	0.56
5ce	290.2	7	4	1.46	21	4	124.29	2.34	-3.27	Soluble	High	0.56
5cf	290.2	7	4	1.46	21	4	124.29	2.26	-3.22	Soluble	High	0.56
5cg	290.2	7	4	1.61	21	4	124.29	2.81	-3.57	Soluble	High	0.56
5ch	290.2	7	4	1.58	21	4	124.29	2.67	-3.48	Soluble	High	0.56
5ci	290.2	7	4	1.33	21	4	124.29	1.79	-2.93	Soluble	High	0.56
5cj	290.2	7	4	1.35	21	4	124.29	1.79	-2.93	Soluble	High	0.56
5ck	290.2	7	4	1.45	21	4	124.29	2.26	-3.22	Soluble	High	0.56
5cl	274.2	6	3	1.84	20	4	104.06	2.9	-3.55	Soluble	High	0.56
5cm	274.2	6	3	1.91	20	4	104.06	3.25	-3.77	Soluble	High	0.56
5cn	300.3	6	3	2.24	22	5	104.06	3.13	-3.75	Soluble	High	0.56
5co	300.3	6	3	2.23	22	5	104.06	3.13	-3.75	Soluble	High	0.56
5cp	290.2	7	4	1.54	21	4	124.29	2.51	-3.38	Soluble	High	0.56
5cq	290.2	7	4	1.34	21	4	124.29	3.1	-3.75	Soluble	High	0.56
5cr	316.3	7	4	1.85	23	5	124.29	2.77	-3.6	Soluble	High	0.56
5cs	316.3	7	4	1.78	23	5	124.29	2.77	-3.6	Soluble	High	0.56
5ct	290.2	7	4	1.57	21	4	124.29	2.55	-3.4	Soluble	High	0.56
5cu	290.2	7	4	1.62	21	4	124.29	2.9	-3.63	Soluble	High	0.56
5cv	316.3	7	4	1.89	23	5	124.29	2.77	-3.6	Soluble	High	0.56
5cw	316.3	7	4	1.91	23	5	124.29	2.77	-3.6	Soluble	High	0.56
5cx	290.2	7	4	1.32	21	4	124.29	1.79	-2.93	Soluble	High	0.56
5cy	290.2	7	4	1.22	21	4	124.29	2.34	-3.27	Soluble	High	0.56
5cz	316.3	7	4	1.72	23	5	124.29	2.24	-3.27	Soluble	High	0.56
5da	316.3	7	4	1.69	23	5	124.29	2.22	-3.26	Soluble	High	0.56
5db	290.2	7	4	1.64	21	4	124.29	3.1	-3.75	Soluble	High	0.56
5dc	290.2	7	4	1.53	21	4	124.29	2.34	-3.27	Soluble	High	0.56
5dd	290.2	7	4	1.46	21	4	124.29	2.26	-3.22	Soluble	High	0.56
5de	290.2	7	4	1.36	21	4	124.29	2.81	-3.57	Soluble	High	0.56
5df	290.2	7	4	1.67	21	4	124.29	2.9	-3.63	Soluble	High	0.56
5dg	290.2	7	4	1.55	21	4	124.29	2.34	-3.27	Soluble	High	0.56
5dh	290.2	7	4	1.49	21	4	124.29	2.34	-3.27	Soluble	High	0.56
5di	290.2	7	4	1.62	21	4	124.29	2.81	-3.57	Soluble	High	0.56
5dj	316.3	7	4	1.74	23	5	124.29	2.77	-3.6	Soluble	High	0.56

Compound	Lipinski's rules of five				Physicochemical properties						Drug-likeness profiles	
	MW	#HBA	#HBD	cLogP	#Heavy atoms	#Rotatable bonds	TPSA	XLOGP3	ESOL Log S	ESOL Class	GI Absorption	Bioavai score
5dk	316.3	7	4	1.76	23	5	124.29	2.22	-3.26	Soluble	High	0.56
5dl	316.3	7	4	1.72	23	5	124.29	2.21	-3.25	Soluble	High	0.56
5dm	316.3	7	4	1.86	23	5	124.29	2.77	-3.6	Soluble	High	0.56
5dn	290.2	7	4	1.52	21	4	124.29	2.51	-3.38	Soluble	High	0.56
5do	290.2	7	4	1.54	21	4	124.29	3.1	-3.75	Soluble	High	0.56
5dp	316.3	7	4	1.78	23	5	124.29	2.77	-3.6	Soluble	High	0.56
5dq	316.3	7	4	1.81	23	5	124.29	2.77	-3.6	Soluble	High	0.56
5dr	290.2	7	4	1.56	21	4	124.29	2.34	-3.27	Soluble	High	0.56
5ds	290.2	7	4	1.68	21	4	124.29	2.9	-3.63	Soluble	High	0.56
5dt	316.3	7	4	1.82	23	5	124.29	2.79	-3.61	Soluble	High	0.56
5du	316.3	7	4	1.91	23	5	124.29	2.77	-3.6	Soluble	High	0.56
5dv	316.3	7	4	1.72	23	5	124.29	2.22	-3.26	Soluble	High	0.56
5dw	316.3	7	4	1.57	23	5	124.29	2.77	-3.6	Soluble	High	0.56
5dx	342.3	7	4	2.1	25	6	124.29	2.64	-3.58	Soluble	High	0.56
5dy	342.3	7	4	2.11	25	6	124.29	2.64	-3.58	Soluble	High	0.56
Xanthones												
6a	426.5	7	3	3.77	31	3	109.36	4.71	-5.59	Moderately	High	0.55
6b	394.4	6	3	4.02	29	2	100.13	5.3	-5.85	Moderately	High	0.55
6c	274.2	6	3	1.53	20	1	100.13	1.87	-3.17	Soluble	High	0.55
6d	274.2	6	3	1.71	20	1	100.13	2.42	-3.52	Soluble	High	0.55
6e	302.3	6	1	2.3	22	3	78.13	2.52	-3.57	Soluble	High	0.55
6f	260.2	6	4	1.3	19	0	111.13	2.09	-3.32	Soluble	High	0.55
6g	326.3	5	2	3.54	24	3	79.9	4.7	-5.06	Moderately	High	0.55
6h	356.4	6	2	3.58	26	4	89.13	4.67	-5.13	Moderately	High	0.55
6i	326.3	5	2	3.54	24	3	79.9	4.7	-5.06	Moderately	High	0.55
6j	326.3	5	2	3.54	24	3	79.9	4.7	-5.06	Moderately	High	0.55
6k	342.3	6	3	3.21	25	3	100.13	4.35	-4.92	Moderately	High	0.55
6l	442.5	8	4	3.13	32	4	129.59	4.15	-5.26	Moderately	High	0.55
6m	440.5	7	3	4.1	32	6	109.36	5.02	-5.66	Moderately	High	0.55
6n	410.5	6	2	4.33	30	3	89.13	5.13	-5.76	Moderately	High	0.55
6o	396.4	6	3	3.78	29	3	100.13	4.74	-5.44	Moderately	High	0.55
6p	426.5	7	4	3.81	31	5	120.36	4.69	-5.44	Moderately	High	0.55
6q	426.5	7	3	3.51	31	3	109.36	4.16	-5.24	Moderately	High	0.55
6r	440.5	7	2	3.89	32	4	98.36	4.48	-5.45	Moderately	High	0.55
6s	398.5	6	4	3.89	29	5	111.13	4.93	-5.44	Moderately	High	0.55
6t	412.5	6	3	4.29	30	6	100.13	5.25	-5.65	Moderately	High	0.55
6u	396.4	6	3	4.2	29	2	100.13	5.36	-5.9	Moderately	High	0.55
6v	412.4	7	4	3.41	30	2	120.36	4.38	-5.37	Moderately	High	0.55
6w	426.5	7	4	3.62	31	6	120.36	5.22	-5.71	Moderately	High	0.55
6x	428.5	7	3	3.52	31	4	109.36	3.75	-4.93	Moderately	High	0.55
6y	428.5	7	3	3.76	31	4	109.36	4.31	-5.28	Moderately	High	0.55
6z	378.4	5	2	4.39	28	2	79.9	5.65	-5.98	Moderately	High	0.55
6aa	396.4	6	2	4.11	29	0	89.13	4.76	-5.65	Moderately	High	0.55
6ab	426.5	7	3	3.73	31	4	109.36	4.71	-5.52	Moderately	High	0.55
6ac	394.4	6	3	4.06	29	2	100.13	5.3	-5.85	Moderately	High	0.55
6ad	408.4	6	1	4.33	30	1	78.13	5.03	-5.82	Moderately	High	0.55
6ae	396.4	6	3	3.7	29	3	100.13	4.74	-5.44	Moderately	High	0.55
6af	424.4	7	3	3.87	31	6	117.2	4.93	-5.52	Moderately	High	0.55
6ag	324.3	5	3	2.77	24	0	90.9	3.65	-4.58	Moderately	High	0.55

Compound	Lipinski's rules of five				Physicochemical properties						Drug-likeness profiles	
	MW	#HBA	#HBD	cLogP	#Heavy atoms	#Rotatable bonds	TPSA	XLOGP3	ESOL Log S	ESOL Class	GI Absorption	Bioavai score
6ah	394.4	6	3	4.02	29	2	100.13	5.3	-5.85	Moderately	High	0.55
6ai	376.4	5	1	4.24	28	0	68.9	5.01	-5.7	Moderately	High	0.55
6aj	326.3	5	2	3.3	24	3	79.9	4.15	-4.71	Moderately	High	0.55
6ak	342.3	6	3	3.16	25	3	100.13	4.35	-4.92	Moderately	High	0.55
6al	458.5	8	4	3.56	33	7	129.59	4.25	-5.21	Moderately	High	0.55
6am	384.4	7	3	3.09	28	3	109.36	4.19	-5.03	Moderately	High	0.55
6an	460.5	8	4	3.66	33	8	129.59	4.45	-5.28	Moderately	High	0.55
6ao	386.4	7	3	3.23	28	4	109.36	4.39	-5.11	Moderately	High	0.55
Avenalumic acid derivatives												
7a	285.3	3	0	3.03	21	4	38.77	3.46	-3.74	Soluble	High	0.55
7b	293.3	3	1	3.43	22	5	47.56	4.54	-4.59	Moderately	High	0.55
7c	353.4	5	1	3.3	26	7	66.02	3.58	-4.17	Moderately	High	0.55
7d	353.4	5	1	3.28	26	7	66.02	3.58	-4.17	Moderately	High	0.55
7e	353.4	5	1	3.23	26	7	66.02	3.58	-4.17	Moderately	High	0.55
7f	280.3	3	2	3.26	21	5	57.53	3.56	-3.91	Soluble	High	0.55
7g	286.4	3	2	3.59	21	5	57.53	4.64	-4.42	Moderately	High	0.55
7h	288.3	4	2	3.21	21	5	66.76	3.94	-3.99	Soluble	High	0.55
7i	315.8	3	3	3.29	22	5	69.56	3.74	-4.23	Moderately	High	0.55
7j	295.3	3	3	3.06	22	5	69.56	3.48	-3.94	Soluble	High	0.55
7k	295.3	3	3	3.06	22	5	69.56	3.48	-3.94	Soluble	High	0.55
7l	295.3	3	3	2.98	22	5	69.56	3.48	-3.94	Soluble	High	0.55
7m	315.8	3	3	3.23	22	5	69.56	3.63	-4.16	Moderately	High	0.55
7n	311.3	4	3	2.75	23	6	78.79	2.97	-3.63	Soluble	High	0.55
7o	311.3	4	3	2.69	23	6	78.79	2.97	-3.63	Soluble	High	0.55
7p	311.3	4	3	2.69	23	6	78.79	2.97	-3.63	Soluble	High	0.55
7q	315.8	3	3	3.23	22	5	69.56	3.63	-4.16	Moderately	High	0.55
7r	297.3	4	4	2.35	22	5	89.79	2.76	-3.5	Soluble	High	0.55
7s	297.3	4	4	2.3	22	5	89.79	2.65	-3.43	Soluble	High	0.55
7t	350.2	3	3	3.81	23	5	69.56	4.37	-4.82	Moderately	High	0.55
7u	299.3	4	3	3.06	22	5	69.56	3.22	-3.8	Soluble	High	0.55
7v	299.3	4	3	3.01	22	5	69.56	3.1	-3.72	Soluble	High	0.55
7w	299.3	4	3	3	22	5	69.56	3.1	-3.72	Soluble	High	0.55
7x	350.2	3	3	3.72	23	5	69.56	4.26	-4.75	Moderately	High	0.55
7y	349.3	6	3	3.86	25	6	69.56	4	-4.48	Moderately	High	0.55
7z	349.3	6	3	3.78	25	6	69.56	4	-4.48	Moderately	High	0.55
7aa	349.3	6	3	3.78	25	6	69.56	4	-4.48	Moderately	High	0.55
7ab	350.2	3	3	3.72	23	5	69.56	4.26	-4.75	Moderately	High	0.55
7ac	310.4	4	4	2.25	23	6	95.58	2.52	-3.34	Soluble	High	0.55
7ad	310.4	4	4	2.16	23	6	95.58	1.97	-3	Soluble	High	0.55
7ae	310.4	4	4	2.15	23	6	95.58	1.97	-3	Soluble	High	0.55
7af	350.2	3	3	3.8	23	5	69.56	4.26	-4.75	Moderately	High	0.55
7ag	309.4	3	3	3.38	23	5	69.56	3.84	-4.23	Moderately	High	0.55
7ah	309.4	3	3	3.36	23	5	69.56	3.84	-4.23	Moderately	High	0.55
7ai	309.4	3	3	3.36	23	5	69.56	3.84	-4.23	Moderately	High	0.55
7aj	350.2	3	3	3.78	23	5	69.56	4.26	-4.75	Moderately	High	0.55
7ak	341.4	5	3	2.62	25	7	88.02	2.94	-3.7	Soluble	High	0.55
7al	341.4	5	3	2.67	25	7	88.02	2.94	-3.7	Soluble	High	0.55
7am	341.4	5	3	2.67	25	7	88.02	2.94	-3.7	Soluble	High	0.55
7an	350.2	3	3	3.75	23	5	69.56	4.37	-4.82	Moderately	High	0.55

Compound	Lipinski's rules of five				Physicochemical properties						Drug-likeness profiles	
	MW	#HBA	#HBD	cLogP	#Heavy atoms	#Rotatable bonds	TPSA	XLOGP3	ESOL Log S	ESOL Class	GI Absorption	Bioavai score
7ao	313.3	5	5	1.99	23	5	110.02	2.41	-3.36	Soluble	High	0.55
7ap	350.2	3	3	3.76	23	5	69.56	4.37	-4.82	Moderately	High	0.55
7aq	317.3	5	3	3.37	23	5	69.56	3.32	-3.95	Soluble	High	0.55
7ar	317.3	5	3	3.3	23	5	69.56	3.2	-3.88	Soluble	High	0.55
7as	317.3	5	3	3.3	23	5	69.56	3.2	-3.88	Soluble	High	0.55
7at	339.4	5	5	1.73	25	7	121.6	1.38	-2.71	Soluble	High	0.55
7au	339.4	5	5	1.58	25	7	121.6	0.83	-2.36	Soluble	High	0.55
7av	339.4	5	5	1.58	25	7	121.6	0.83	-2.36	Soluble	High	0.55
7aw	309.4	3	3	3.4	23	5	69.56	3.84	-4.23	Moderately	High	0.55
7ax	309.4	3	3	3.41	23	5	69.56	3.84	-4.23	Moderately	High	0.55
7ay	309.4	3	3	3.39	23	5	69.56	3.84	-4.23	Moderately	High	0.55
7az	339.4	5	5	1.58	25	7	121.6	0.83	-2.36	Soluble	High	0.55
7ba	309.4	3	3	3.4	23	5	69.56	3.84	-4.23	Moderately	High	0.55
7bb	309.4	3	3	3.41	23	5	69.56	3.84	-4.23	Moderately	High	0.55
7bc	309.4	3	3	3.39	23	5	69.56	3.84	-4.23	Moderately	High	0.55
7bd	309.4	3	3	3.39	23	5	69.56	3.84	-4.23	Moderately	High	0.55
7be	341.4	5	3	2.7	25	7	88.02	2.94	-3.7	Soluble	High	0.55
7bf	341.4	5	3	2.75	25	7	88.02	2.94	-3.7	Soluble	High	0.55
7bg	341.4	5	3	2.71	25	7	88.02	2.94	-3.7	Soluble	High	0.55
7bh	341.4	5	3	2.73	25	7	88.02	3.06	-3.78	Soluble	High	0.55
7bi	313.3	5	5	1.93	23	5	110.02	2.41	-3.36	Soluble	High	0.55
7bj	313.3	5	5	2.02	23	5	110.02	2.41	-3.36	Soluble	High	0.55
7bk	317.3	5	3	3.33	23	5	69.56	3.2	-3.88	Soluble	High	0.55
7bl	317.3	5	3	3.34	23	5	69.56	3.2	-3.88	Soluble	High	0.55
7bm	317.3	5	3	3.33	23	5	69.56	3.32	-3.95	Soluble	High	0.55
7bn	317.3	5	3	3.31	23	5	69.56	3.32	-3.95	Soluble	High	0.55
7bo	339.4	5	5	1.65	25	7	121.6	1.38	-2.71	Soluble	High	0.55
7bp	339.4	5	5	1.67	25	7	121.6	1.38	-2.71	Soluble	High	0.55
7bq	339.4	5	5	1.61	25	7	121.6	0.83	-2.36	Soluble	High	0.55
7br	339.4	5	5	1.68	25	7	121.6	1.38	-2.71	Soluble	High	0.55
Quinonoids												
8a	240.2	4	0	1.95	18	1	64.35	2.28	-3.15	Soluble	High	0.55
8b	254.2	4	0	2.3	19	2	64.35	2.75	-3.45	Soluble	High	0.55
8c	242.2	5	1	1.48	18	1	84.58	2.13	-3.07	Soluble	High	0.56
8d	228.2	3	3	2.48	17	2	60.69	3.13	-3.62	Soluble	High	0.55
8e	368.4	6	2	3.03	27	8	93.06	3.2	-3.94	Soluble	High	0.55
Steroids												
9a	372.5	2	2	3.31	27	3	58.2	3.03	-3.86	Soluble	High	0.55
9b	392.5	2	2	3.62	29	3	58.2	3.6	-4.5	Moderately	High	0.55
9c	406.6	2	1	3.77	30	3	49.41	3.79	-4.7	Moderately	High	0.55
9d	440.5	6	2	3.46	31	5	75.27	3.32	-4.33	Moderately	High	0.55
9e	440.5	6	2	3.46	31	5	75.27	3.32	-4.33	Moderately	High	0.55
9f	364.5	1	2	4.12	27	2	41.13	5.33	-5.49	Moderately	High	0.55
9g	488.6	5	2	4.79	35	5	58.2	4.88	-5.74	Moderately	High	0.55
9h	488.6	5	2	4.79	35	5	58.2	4.88	-5.74	Moderately	High	0.55
9i	454.5	6	2	3.64	32	5	75.27	3.51	-4.54	Moderately	High	0.55
9j	383.5	3	2	2.71	28	3	81.99	2.42	-3.54	Soluble	High	0.55
9k	464.6	3	2	4.13	34	5	67.43	4.1	-5.1	Moderately	High	0.55
9l	371.6	2	1	4.16	27	3	46.17	4.11	-4.53	Moderately	High	0.55

Compound	Lipinski's rules of five				Physicochemical properties						Drug-likeness profiles	
	MW	#HBA	#HBD	cLogP	#Heavy atoms	#Rotatable bonds	TPSA	XLOGP3	ESOL Log S	ESOL Class	GI Absorption	Bioavai score
9m	396.6	3	1	4.2	29	3	69.96	4.51	-4.94	Moderately	High	0.55
9n	331.5	3	1	3.03	24	2	55.4	3.52	-3.98	Soluble	High	0.55
9o	446.6	2	2	4.45	33	4	58.2	5.14	-5.72	Moderately	High	0.55
9p	350.9	2	2	2.73	24	1	72.19	3.38	-4.08	Moderately	High	0.55
9q	385.6	2	0	4.58	28	3	37.38	5.48	-5.49	Moderately	High	0.55
9r	315.5	3	1	3.83	23	2	49.66	4.05	-4.22	Moderately	High	0.55
9s	403.6	2	1	4.59	28	3	84.97	4.94	-5.26	Moderately	High	0.55
9t	406	2	1	4.9	28	3	46.17	5.42	-5.57	Moderately	High	0.55
9u	450.5	2	1	4.94	28	3	46.17	5.48	-5.89	Moderately	High	0.55
Piperic acid derivatives												
10a	285.3	3	0	3.03	21	4	38.77	3.46	-3.74	Soluble	High	0.55
10b	293.3	3	1	3.43	22	5	47.56	4.54	-4.59	Moderately	High	0.55
10c	353.4	5	1	3.3	26	7	66.02	3.58	-4.17	Moderately	High	0.55
10d	353.4	5	1	3.28	26	7	66.02	3.58	-4.17	Moderately	High	0.55
10e	353.4	5	1	3.23	26	7	66.02	3.58	-4.17	Moderately	High	0.55
Ureas												
11a	220.3	3	2	1.45	15	4	95.15	1.77	-2.6	Soluble	High	0.55
11b	282.4	1	2	3.71	21	6	41.13	3.87	-4.06	Moderately	High	0.55
11c	426.9	5	3	2.81	30	8	115.57	2.75	-4.09	Moderately	High	0.55
11d	375.4	3	4	3.68	28	5	95.83	3.86	-4.82	Moderately	High	0.55
11e	454.9	7	2	3.79	32	8	107.74	4.03	-5.16	Moderately	High	0.55
11f	242.3	2	2	2.4	18	5	50.36	2.06	-2.8	Soluble	High	0.55
11g	226.3	1	2	2.8	17	4	41.13	2.85	-3.3	Soluble	High	0.55
11h	242.3	2	2	2.52	18	5	50.36	2.69	-3.2	Soluble	High	0.55
11i	226.3	1	2	2.67	17	4	41.13	2.24	-2.91	Soluble	High	0.55
11j	246.7	1	2	3.03	17	4	41.13	3.11	-3.59	Soluble	High	0.55
11k	226.3	1	2	2.85	17	4	41.13	3.07	-3.44	Soluble	High	0.55
11l	230.2	2	2	2.77	17	4	41.13	2.42	-3.05	Soluble	High	0.55
11m	230.2	2	2	2.8	17	4	41.13	2.59	-3.16	Soluble	High	0.55
11n	230.2	2	2	2.89	17	4	41.13	3.04	-3.44	Soluble	High	0.55
11o	246.7	1	2	2.93	17	4	41.13	2.67	-3.31	Soluble	High	0.55
11p	257.2	3	2	2.04	19	5	86.95	3.05	-3.49	Soluble	High	0.55
11q	246.7	1	2	3.22	17	4	41.13	4.07	-4.19	Moderately	High	0.55
11r	242.3	2	2	2.49	18	5	50.36	2.42	-3.03	Soluble	High	0.55
11s	237.3	2	2	1.75	18	5	53.49	2.47	-3.03	Soluble	High	0.55
11t	288.3	1	2	3.83	22	5	41.13	4.11	-4.49	Moderately	High	0.55
11u	262.3	1	2	3.37	20	4	41.13	3.4	-3.94	Soluble	High	0.55
11v	256.3	3	2	2.21	19	4	59.59	2.15	-2.99	Soluble	High	0.55
11w	314.7	4	2	4.2	21	5	41.13	4.72	-4.86	Moderately	High	0.55
11x	178.2	1	2	1.79	13	4	41.13	1.92	-2.23	Soluble	High	0.55
11y	218.3	1	2	2.54	16	4	41.13	2.77	-2.95	Soluble	High	0.55
11z	226.3	1	2	2.51	17	5	41.13	2.42	-2.96	Soluble	High	0.55
11aa	240.3	1	2	2.88	18	6	41.13	3.23	-3.46	Soluble	High	0.55
11ab	238.3	1	2	2.75	18	5	41.13	2.91	-3.31	Soluble	High	0.55
11ac	202.2	2	2	1.78	15	4	54.27	1.88	-2.56	Soluble	High	0.55
11ad	218.3	1	2	2.43	15	4	69.37	2.49	-3.04	Soluble	High	0.55
11ae	251.3	1	3	2.55	19	4	56.92	2.61	-3.36	Soluble	High	0.55
11af	228.3	3	2	1.28	17	4	66.91	1.08	-2.19	Soluble	High	0.55
11ag	228.3	2	3	2.07	17	4	61.36	2.11	-2.84	Soluble	High	0.55

Compound	Lipinski's rules of five				Physicochemical properties						Drug-likeness profiles	
	MW	#HBA	#HBD	cLogP	#Heavy atoms	#Rotatable bonds	TPSA	XLOGP3	ESOL Log S	ESOL Class	GI Absorption	Bioavai score
11ah	241.3	2	3	1.91	18	5	67.15	1.34	-2.34	Soluble	High	0.55
11ai	244.3	1	2	2.7	17	4	79.93	2.64	-3.28	Soluble	High	0.55
11aj	278.3	2	3	3.01	21	4	61.36	3.38	-3.99	Soluble	High	0.55
11ak	256.3	3	2	2.21	19	4	59.59	2.15	-2.99	Soluble	High	0.55
11al	305.2	1	2	3.4	18	4	41.13	3.42	-4.12	Moderately	High	0.55
11am	264.7	2	2	3.52	18	4	41.13	4.17	-4.34	Moderately	High	0.55
11an	296.4	1	2	4.14	22	6	41.13	3.8	-4.08	Moderately	High	0.55
11ao	262.7	2	3	2.8	18	4	61.36	3.72	-4.04	Moderately	High	0.55
11ap	206.3	1	2	2.62	15	7	41.13	3.07	-2.89	Soluble	High	0.55
11aq	216.2	2	2	1.85	16	5	54.27	1.52	-2.32	Soluble	High	0.55
11ar	227.3	2	2	1.67	17	5	54.02	1.39	-2.32	Soluble	High	0.55
11as	279.3	1	3	2.69	21	6	56.92	2.34	-3.18	Soluble	High	0.55
11at	213.2	2	2	1.73	16	4	54.02	1.8	-2.59	Soluble	High	0.55
11au	263.3	2	2	2.83	20	4	54.02	3.3	-3.88	Soluble	High	0.55
11av	251.3	1	3	2.45	19	4	56.92	1.94	-2.94	Soluble	High	0.55
11aw	329.4	1	2	3.87	25	5	46.06	4.2	-4.76	Moderately	High	0.55
11ax	192.3	1	1	2.03	14	5	32.34	1.87	-2.2	Soluble	High	0.55
11ay	192.3	1	1	2.04	14	5	32.34	1.84	-2.18	Soluble	High	0.55

3.2.2 Pharmacophore-based virtual screening and model validation

The pharmacophore models of previous work were used to identify compounds that share similar chemical features and spatial arrangements. The virtual screening results, depicted in Figure 8, revealed the identification of 258, 230, and 157 active compounds using models 11a, 13b, and N3, respectively. These active compounds were further categorized into 11 groups, including α-mangostin, anticancer agents, avicequinones, caffeic acid derivatives, acid homodimers, xanthones, avenalumic acid derivatives, quinonoids, steroids, piperic acid derivatives, and ureas. The identified compounds exhibited good alignment with the key chemical features, such as hydrogen bond donor, hydrogen bond acceptor, and hydrophobic interaction, as defined by their corresponding reference model.(K. Sanachai et al., 2022) They were considered promising candidates for further exploration due to their higher likelihood of exhibiting the desired biological activity against 3CLpro. So, these screened compounds were also utilized in the model validation step to identify decoys.

Table 1

	11a	13b	N3
1a	36	36	46
2a	38	37	
2b		36	46
2c		46	
2d	37	38	
2e	37	38	
2f	37	36	
2g		37	
2h	36	38	
2i	37	38	
2j		37	
2k		36	46
2n	46		46
3a	36		
3b	36		
4d	38		44
4e	38		44
4f	38		44
4g	38		45
4h	38	36	45
4i	37	44	46
4j	38	45	46
4k	35	42	
4l	38	41	
4m	37		
4n	38		44
4o	38		44
4p	38		46
4q	38		
4r	38	38	
4s	38	38	46
4t	38		45
4u	38	37	47
4v	38	46	46
5a	37	36	
5b	36	35	
5c	38	36	
5d	37	36	
5e	35	36	
5f	38	36	
5g	36	38	
5h	36	36	
5i	36	36	
5j	37	36	
5k	35	46	
5l	38	46	
5m	38	46	
5n	36	45	
5o	36	36	
5p	38		46

Table 2

	11a	13b	N3
5q	36	36	46
5r	38		46
5s	36	36	
5t	37	45	46
5u	36	36	46
5y	37	37	
5aa	36	37	46
5ab	38	36	46
5ac	38	46	
5ad	38		
5ae	37	36	47
5af	38	36	
5ag	35	47	
5ah	37	46	
5ai	37	47	
5ak	35	56	
5am	36	55	
5an	37		46
5ao	38	36	
5ap	38		46
5aq	37	36	
5ar	39		
5as	36	37	
5at	38	36	46
5au	37	36	
5av	37		
5aw	37	36	46
5ax	37		46
5ay	37	37	47
5az	38	36	46
5ba	37	37	47
5bb	39	36	
5bc	37	35	
5bd	35	36	
5be	36	38	
5bf	36	36	
5bg	37		
5bh	36	46	
5bi	37	36	
5bj	38	36	47
5bl	38	36	48
5bm	38	36	
5bn	37		
5bo	36	38	
5bp	36	37	
5bq	37	37	
5br	38		
5bs	37	46	
5bt	36		
5bu	38	37	

Table 3

	11a	13b	N3
5bv	39	36	46
5bw	36	36	
5bx	37	36	
5by	38		
5bz	38	36	47
5ca	36	38	45
5cb	38	36	
5cc	38	36	46
5cd	39	46	46
5ce	39	46	
5cf	38	36	
5cg	38	36	46
5ch	37		
5ci	36	56	
5cl	37	36	
5cm	38	37	47
5cn	36	36	
5co	36	36	
5cp	37	36	
5cq	38	36	45
5cr	38	36	
5cs	37		
5ct	38	46	
5cu	39	46	46
5cv	37	46	44
5cw	37	36	44
5cx	38		
5cy	37	56	
5cz	36	37	
5da	37	37	
5db	39	36	45
5dc	38	46	45
5dd	37	36	
5df	37	36	
5dg	37	37	45
5dh	38	36	
5di	37	36	46
5dj	39	37	
5dk	38	38	
5dl	38	37	
5dm	39		
5dn	37	46	
5do	36	46	46
5dp	38	35	
5dq	38	36	
5dr	36	36	46
5ds	39	37	45
5dt	37		
5dv	36	38	
5dw	38	38	

Table 4

	11a	13b	N3
5dx	38	37	44
5dy	37	36	44
6a	38	46	46
6b	37	39	46
6g	36		45
6h	38	37	55
6i	38		46
6j	38		46
6k	38	36	46
6l	37	36	47
6m	36	36	47
6n	38		45
6o	38	56	
6p	37	37	46
6q	39	37	45
6r	47	37	46
6s	48	56	46
6t	36	35	46
6u	37	36	47
6v	38		47
6w	46	56	46
6x	36		46
6y	37	36	
6z		38	47
6aa		45	46
6ab	38	36	
6ac	38	36	56
6ad	37	36	45
6ae	37	38	
6af	37	37	47
6ag	46		
6ah	37	39	46
6ai			66
6aj	37	38	
6ak	38	36	
6al	38	37	56
6am	37	36	
6an	46	37	46
6ao	38		45
7f	35	38	46
7i		38	46
7j		38	46
7k		38	
7l	35	38	46
7m	36	38	46
7n	36	38	46
7o	36	38	46
7p	38	38	46
7q	36	38	47
7r	35	38	46

Table 5

	11a	13b	N3
7s	35	38	46
7t	35	38	47
7u	35	38	47
7v	35	38	46
7w	38	38	47
7x	38	38	47
7y		44	46
7z		44	46
7aa	56	38	46
7ab		38	47
7ac	36	38	45
7ad	36	38	45
7ae	35	38	46
7af		38	47
7ag	55	38	48
7ah	36	38	45
7ai		38	46
7aj		38	46
7ak		38	46
7al	38	38	
7am	35	38	46
7an		38	46
7ao		38	46
7ap	55	38	
7aq	37	38	48
7ar	38	38	47
7as	38	38	47
7at		38	
7au	35	38	
7av	35	38	
7aw		38	46
7ax		38	46
7bd		38	46
7be	36	44	46
7bf	38	38	
7bg	37	38	
7bh		38	46
7bi	36	38	
7bj	37	38	
7bk	37	38	45
7bl	38	38	47
7bm	36	38	
7bn		38	46
7bo	36	38	
7bp		38	
7bq	35	38	
7br		38	
8a	36		
8b	36		
8c	36	36	

Table 6

	11a	13b	N3
8e	38	37	
9c	37		
9d	56	38	
9e	56	39	
9g	34	36	
9h	37	36	
9i	56	38	46
9j	36	36	
9m	46		
9n	36		
9s	37		
10b			46
10c	38		46
10d	36		45
10e	35		
11a	36	37	
11b	33		47
11c	35	38	
11d	38	54	45
11e		37	46
11f		45	44
11g			46
11h	36	36	44
11l	45	36	46
11m	38	37	46
11n	36	37	55
11o			46
11r	38		44
11t	37		44
11u	37		55
11v	39		44
11w	36	44	46
11aa			44
11ac			44
11ad			44
11af	35	74	
11ag	44	36	
11ai	44	45	44
11aj	38		46
11al	37		46
11am	38		54
11ar	36	36	
11as	44		
11at	36	37	44
11av			46
11aw			55

Legend: ☐ α-Mangostins ☐ Anticancer agents ☐ Avicequinones ☐ Caffeic acid derivatives ☐ Acid homodimers ☐ Xanthones ☐ Avenalumic acid derivatives ☐ Quinonoids ☐ Steroids ☐ Piperic acid derivatives ☐ Ureas
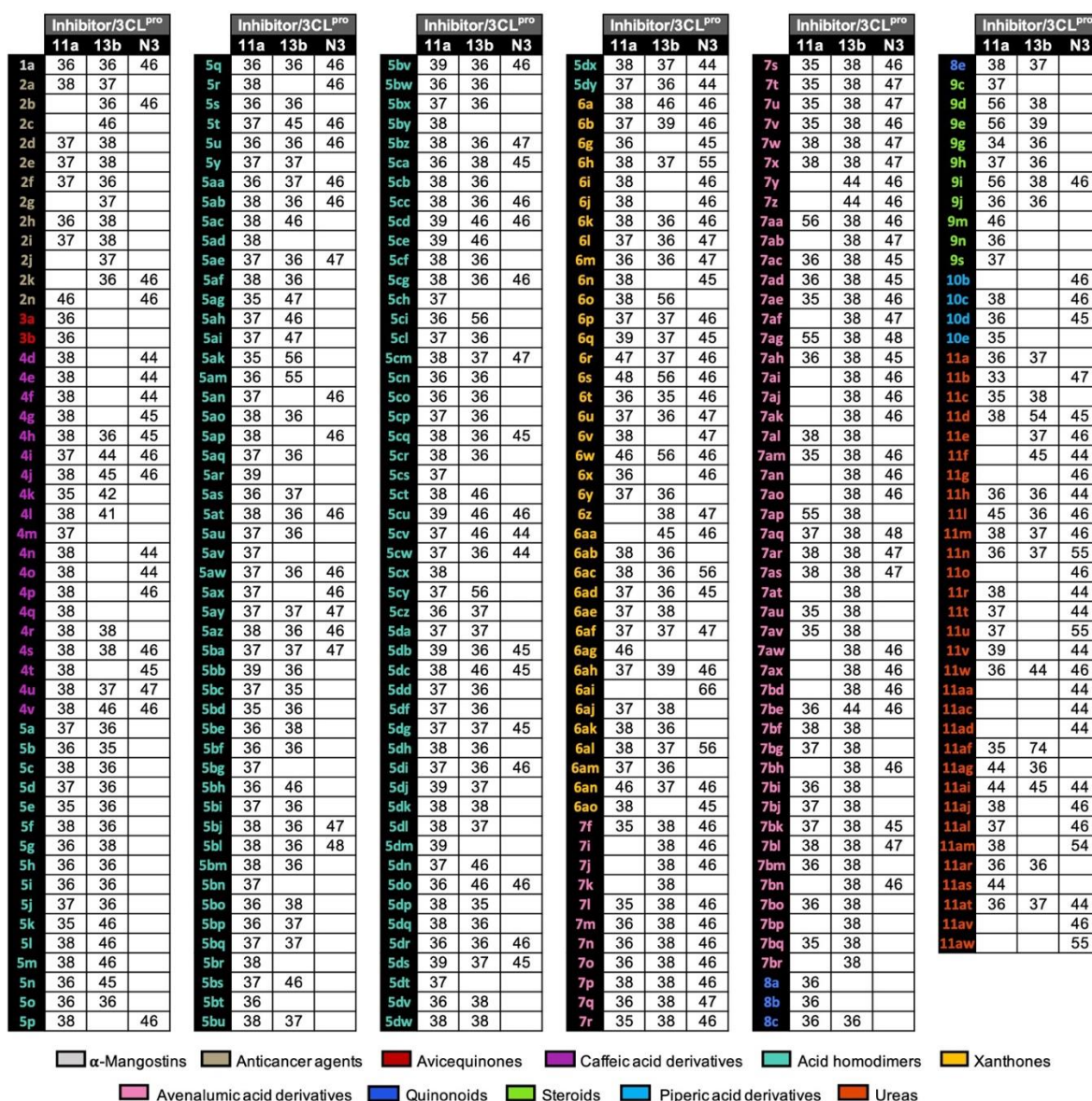
Figure 8 Pharmacophore fit scores of 296 active compounds derived from pharmacophore-based virtual screening

In order to comprehensively evaluate the performance of the pharmacophore models used in this study, we conducted a thorough validation process. This involved assessing the models using ROC curves, which offer their ability to accurately classify compounds as either active or inactive. Figure 9 represents the obtained ROC curves. The AUC values were obtained by considering the top 5%, 10%, and 100% of the ranked compounds from each model, which ranged between 0.5 and 1.0. These values indicate that all the models have a moderate to strong ability to classify compounds as active or inactive correctly. The EF value measures the degree of enrichment the screening method achieves compared to random sampling.(Gan et al., 2023; Li et al., 2009) When the EF value is greater than 1, it indicates a higher probability of

detecting active compounds than random selection. Based on the AUC and EF values evaluation, the N3 model demonstrated superior discrimination and enrichment compared to the other two models. The 11a model showed moderate discrimination and enrichment, while the 13b model exhibited lower performance. However, further analysis and validation should be conducted to ensure the reliability and robustness of the findings. Validation methods such as cross-validation, external testing on independent datasets, and experimental assays are necessary to confirm the performance of the models and validate their predictive capabilities.
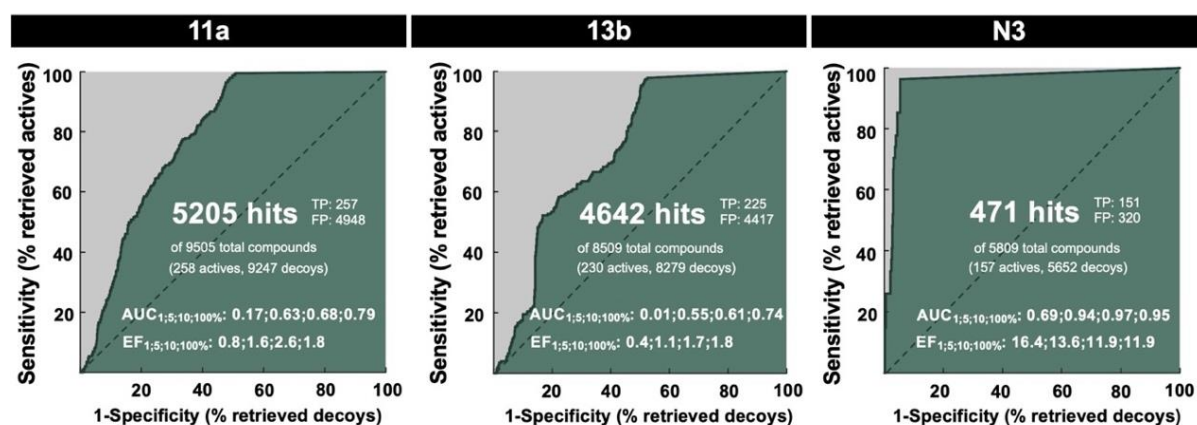

Figure 9 ROC curves of pharmacophore model validation generated by KNIME

3.2.3 Hits identification from molecular docking

Molecular docking was performed on the 296 active compounds obtained from the pharmacophore-based virtual screening to predict their binding modes and affinities with the SARS-CoV-2 3CLpro protein target. The predicted binding energies (ΔG) of the best pose compounds were analyzed based on three criteria: RMSD < 2 Å, low ΔG, and similar binding orientation to their reference compounds. Figure 10A shows the predicted ΔG values in kcal/mol for the 43 compounds that exhibited stronger binding affinity with the SARS-CoV-2 3CLpro than the known inhibitors. The variations in ΔG can be attributed to the specific interactions between the protein and the ligands, such as hydrogen bonding, electrostatic interactions, hydrophobic interactions, and van der Waals forces. These interactions play a critical role in determining the strength and stability of the ligand-protein complex, ultimately influencing the overall binding energy. The 43 selected hit compounds belong to six groups: 8 caffeic acid derivatives, 10 acid homodimers, 14 xanthones, 6 avenalumic acid derivatives, 2 steroids, and 3 ureas. Figure 10B depicts the intermolecular interactions between these hit compounds and the 3CLpro protein. It should be noted that Table 6 provides detailed data on the identification of the selected hit compounds, corresponding to the compound numbers in Figure 10A.

Table 6 Correspondence between compound number in molecular docking (Figure 10A) and compound name from visual inspection (Figure 10B)

Compound No.	Compound name
Caffeic acid derivatives	
21	4i
22	4j
23	4k
24	4l
26	4n
28	4p
33	4u
34	4v
Acid homodimers	
42	5h
44	5j
73	5as
84	5bd
85	5be
91	5bl
95	5bp
99	5bt
121	5cr
129	5cz
Xanthones	
154	6b
162	6n
165	6q
166	6r
167	6s
169	6u
172	6x
174	6z
175	6aa
177	6ac
178	6ad
179	6ae
182	6ah
183	6ai
Avenalumic acid derivatives	
203	7u

Compound No.	Compound name
208	7z
209	7aa
225	7aq
242	7bm
243	7bn
Steroids	
253	9d
257	9i
Ureas	
269	11d
270	11e
296	11aw

Previous studies have reported the potential inhibitory activities of compounds from the discovered groups against the 3CLpro. For example, caffeic acid phenethyl esters have shown comparable binding efficacy and binding energies to the known N3 inhibitor.(Kumar, Dhanjal, Kaul, Wadhwa, & Sundar, 2021) Xanthone derivatives, particularly rubraxanthone, have been found to bind to the allosteric site stably and potentially the active site of 3CLpro. Steroids and ureas, such as glycyrrhizin, withanolides, curcumin, and quercetin, have also demonstrated potential inhibitory activities against the 3CLpro enzyme.(Alves et al., 2022; Bahun et al., 2022; Dhanjal et al., 2021; van de Sand et al., 2021) These findings from previous studies align with the computational results of this study, further supporting the potential of the identified compounds as promising hit candidates for inhibiting the activity of the 3CLpro.

The qualitative and quantitative interactions between compounds and the protein target play a significant role in their binding ability and the stability of the resulting complexes.(Suriya, Mahalapbutr, & Rungrotmongkol, 2022) Figure 10B provides information on the proportion of intermolecular interactions between each compound and the SARS-CoV-2 3CLpro protein target, including hydrogen bonding, π-alkyl interaction, halogen interaction, and alkyl-alkyl interaction. Hydrogen bonds are found to be highly involved in the binding of derivatives of caffeic acid, acid homodimer, avenalumic acid derivatives, and steroids to the SARS-CoV-2 3CLpro. These hydrogen bonds contribute to the stabilization of the ligand-protein complex and facilitate specific interactions at the binding site. Furthermore, π-alkyl interactions play a crucial role in ensuring the complementary shape and fit of all screened compounds, particularly xanthones and ureas, within the active site of the enzyme. These interactions involve the stacking of aromatic moieties and alkyl groups, contributing to the overall binding affinity and specificity. Halogen interactions are observed in four caffeic acid derivatives, three avenalumic acid derivatives, a steroid, and

urea. These interactions involve the interaction of halogen atoms (e.g., chlorine or bromine) with specific residues in the protein, contributing to the stabilization of the ligand-protein complex. Lastly, alkyl-alkyl interactions are likely present in some caffeic acid derivatives, xanthones, and avenalumic acid derivatives. These interactions involve the interaction of alkyl groups, such as methyl or ethyl groups, contributing to the overall binding stability and hydrophobic interactions within the binding site. These various types of intermolecular interactions play a crucial role in the binding of the screened compounds to the SARS-CoV-2 3CLpro, ensuring favorable binding affinity and stability of the ligand-protein complex.
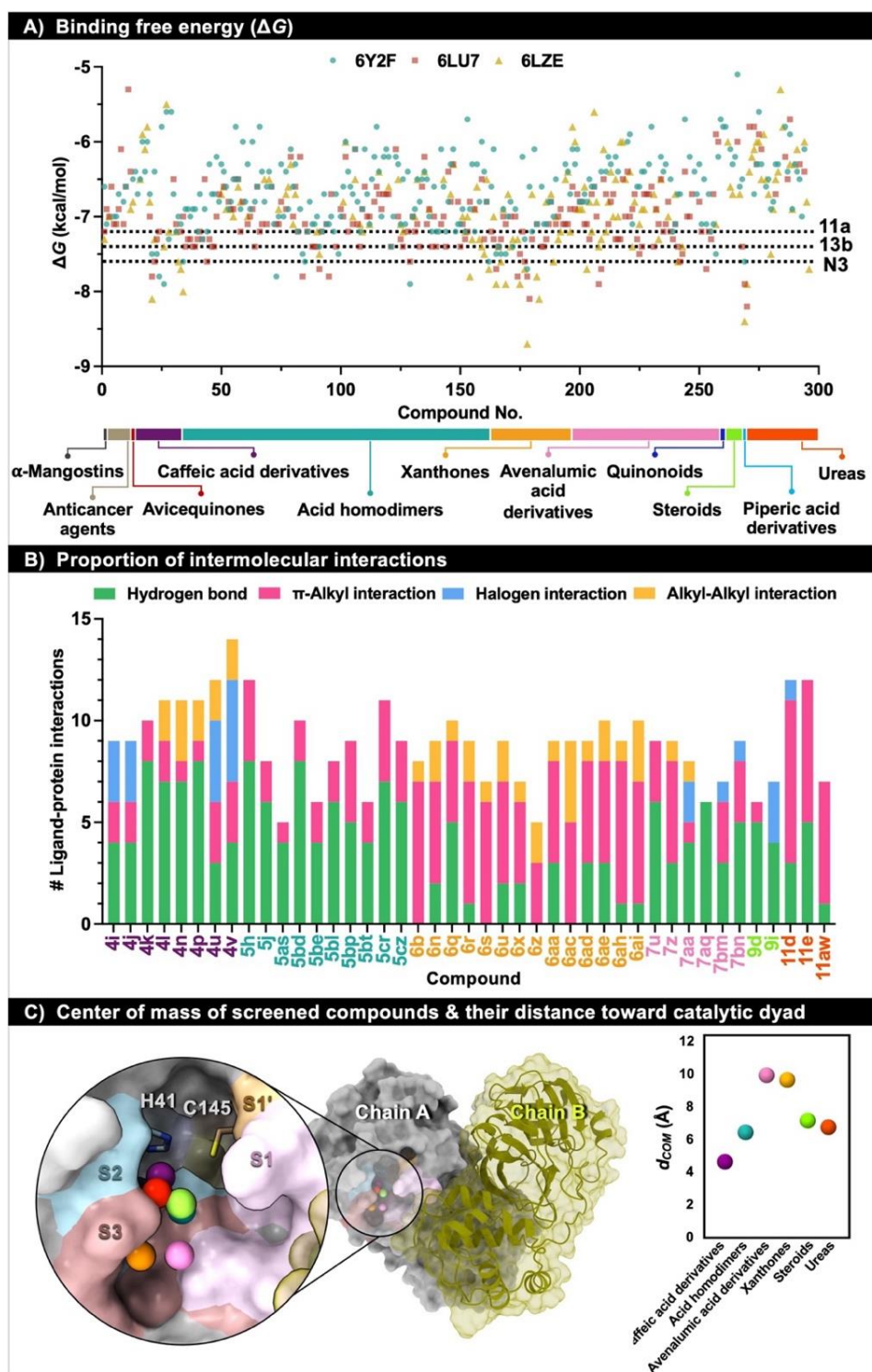
Figure 10 (A) The calculated ΔG (kcal/mol) from molecular docking study via AutoDock VinaXB. Each dot indicates the ΔG of a particular compound/3CLpro complex, whereas the dashed line represents the ΔG of known inhibitors derived from the redocking of the crystal structures. (B) A stacked bar chart displaying the proportion of intermolecular interactions observed in the selected docking complexes resulted from Accelrys Discovery Studio Client 4.0. (C) The center of

mass calculated from the docked poses for each group (left) and its distance to the catalytic dyad H41 and C145 (right)

Figure 11 illustrates the aligned structure of the selected hit compounds, revealing variations and similarities among compound groups in their binding to the target protein. Avenalumic acid derivatives and steroids primarily occupied the S1 and S3 subsites, while acid homodimers and ureas preferred binding to the S1, S2, and S3 subsites. Caffeic acid derivatives showed potential binding to all four subsites: S1, S2, S3, and S1'. Xanthones exhibited diverse binding interactions with 3CLpro. To assess the positioning of hit compounds within the active site, we calculated the center of mass for each compound group and measured its distance from the catalytic dyad (H41 and C145), referred to as dCOM. In Figure 10C, caffeic acid derivatives displayed a dCOM value of 4.7 Å, indicating closer proximity to the catalytic site compared to others (≥ 6.5 Å). These findings suggest that caffeic acid derivatives are well-positioned for strong interactions with 3CLpro, prioritizing their synthesis and subsequent experimental testing, particularly those demonstrating at least ten ligand-protein interactions.
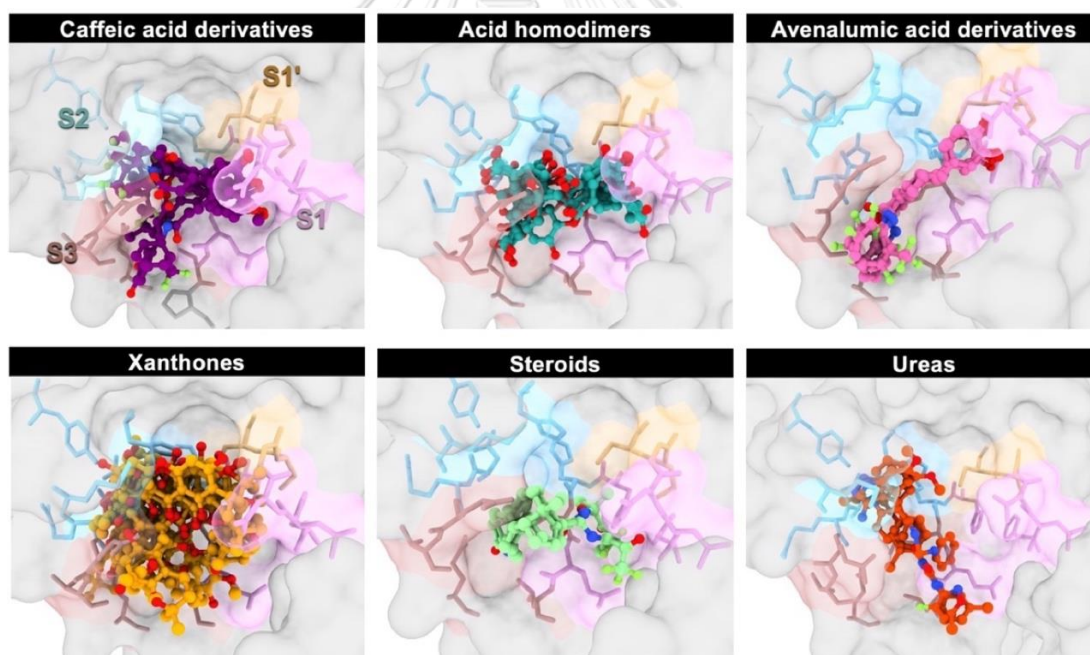

Figure 11 Alignment of selected hit compounds in the active site of 3CLpro according to their optimal binding orientation from molecular docking

The laboratories collaborated on the synthesis of six caffeic acid derivatives and the *in vitro* investigation of an enzyme inhibition assay, as depicted in Figures 15 and 16. Compounds 4k and 4l showed better inhibitory activity than rutin, the standard compound. At 100 μM concentration, compounds 4k and 4l decreased enzymatic activity to 68.8% and 58.0%,

respectively. These results indicate the potential of these two ester derivatives of caffeic acid as promising 3CLpro inhibitors, highlighting their suitability for additional testing and development as SARS-CoV-2 antiviral agents.

Figure 12 displays the intermolecular interactions between two ester caffeic acid derivatives (4k and 4l) and SARS-CoV-2 3CLpro. Both compounds interact with key residues H41, F140, L141, N142, S144, H164, and E166, which are similar to the critical contact residues of the orally active inhibitors nirmatrelvir and ensitrelvir.(Lee et al., 2022; Unoh et al., 2022) Hydrogen bonds are formed between two hydroxyl groups of the phenyl moiety and residues F140, L141, N142, S144, and H163. The ester carbonyl oxygen forms a hydrogen bond with H164, and in the case of 4k, it also interacts with C145. Additionally, two π-alkyl interactions with H41 and E166 contribute to stabilizing the binding of the caffeic acid ester derivatives at the methyl and cinnamyl catechol groups, respectively. The ethyl ester of compound 4l favorably contacts residues F181 and D187 *via* alkyl-alkyl interactions. These interactions provide information on the particular molecular interactions of these two potent compounds with the active site of SARS-CoV-2 3CLpro. Although their potency may not be as strong as other compounds identified in virtual screening, the findings suggest that ester derivatives of caffeic acid could be a starting point for developing more potent inhibitors.(Ang, Kendall, & Atamian, 2023; Onyango, Odhiambo, Angwenyi, & Okoth, 2022; K. Sanachai et al., 2022)
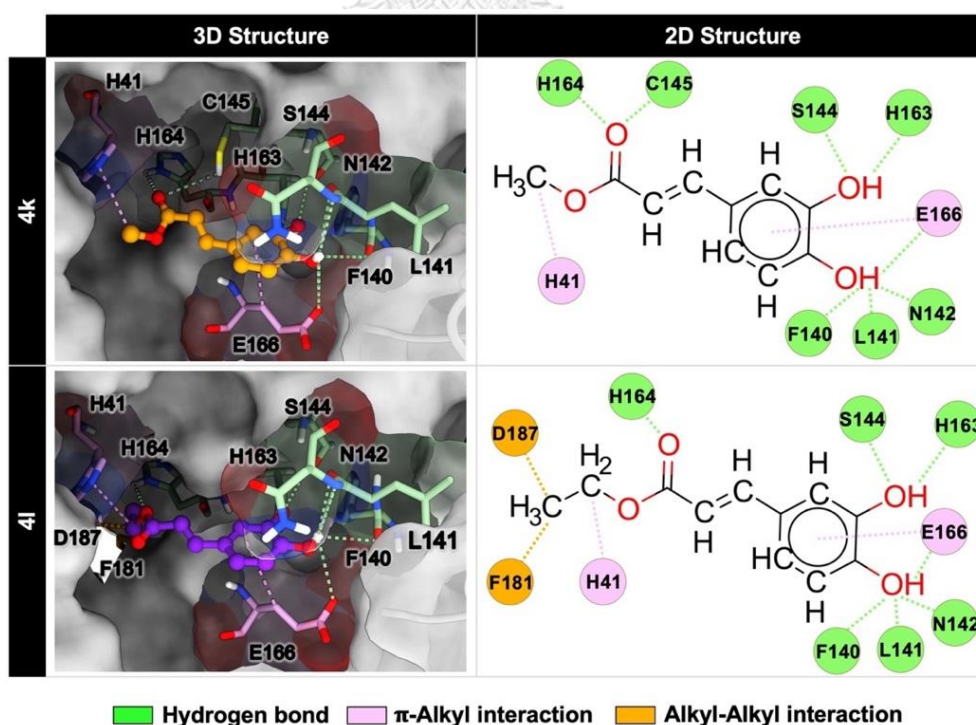

Figure 12 Intermolecular interactions of the two ester derivatives of caffeic acid, 4k and 4l, interacting with SARS-CoV-2 3CLpro active site

CHAPTER IV
CONCLUSIONS

Conclusions

This study focused on identifying potential inhibitors for the SARS-CoV-2 3CLpro using computational approaches. Through screening of in-house compounds, three sulfonamide chalcones (SWC422, SWC423, and SWC424) were identified as promising inhibitors with strong binding affinity and comparable inhibitory efficiency to the authorized drug nirmatrelvir. Further MD trajectory analysis of SWC423 revealed its stability in the active site of 3CLpro and supported by various interactions that contribute to its binding efficacy. Moreover, a comprehensive investigation involving 553 compounds, including natural products and their derivatives, identified 43 hits from six distinct groups (caffeic acid derivatives, acid homodimers, xanthones, avenalumic acid derivatives, steroids, and ureas). These compounds displayed superior binding affinity to the target protein, with caffeic acid derivatives exhibiting the most favorable binding free energy and positioning closest to the catalytic dyad of 3CLpro (H41 and C145). These findings underscore the potential of targeting the SARS-CoV-2 3CLpro enzyme using inhibitors such as sulfonamide chalcones and caffeic acid derivatives. These findings highlight the potential of targeting the SARS-CoV-2 3CLpro with inhibitors such as sulfonamide chalcones and caffeic acid derivatives, providing valuable insights and exciting scaffolds for the development of effective treatments against COVID-19.

Limitations of research

The *in vitro* cytotoxicity testing and MD simulation of caffeic acid derivatives in Vero E6 cells were not performed.

Suggestions for future research

i) The MD simulations of SWC422, SWC424, 4k, and 4l should be run.

ii) Although two ester derivatives of caffeic acid (4k and 4l) exhibited better inhibitory activity than the standard rutin, they still have a lower inhibition rate than many previously published inhibitors.

REFERENCES

(1) Decaro, N.; Lorusso, A. Novel human coronavirus (SARS-CoV-2): A lesson from animal coronaviruses. *Veterinary microbiology* **2020**, *244*, 108693.

(2) Lai, C.-C.; Shih, T.-P.; Ko, W.-C.; Tang, H.-J.; Hsueh, P.-R. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *International journal of antimicrobial agents* **2020**, *55* (3), 105924.

(3) Gupta, T.; Gupta, S. K. Potential adjuvants for the development of a SARS-CoV-2 vaccine based on experimental results from similar coronaviruses. *International immunopharmacology* **2020**, *86*, 106717.

(4) Lu, R.; Zhao, X.; Li, J.; Niu, P.; Yang, B.; Wu, H.; Wang, W.; Song, H.; Huang, B.; Zhu, N.; et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* **2020**, *395* (10224), 565-574. DOI: 10.1016/s0140-6736(20)30251-8 From NLM.

(5) Qiao, J.; Li, Y. S.; Zeng, R.; Liu, F. L.; Luo, R. H.; Huang, C.; Wang, Y. F.; Zhang, J.; Quan, B.; Shen, C.; et al. SARS-CoV-2 M(pro) inhibitors with antiviral activity in a transgenic mouse model. *Science* **2021**, *371* (6536), 1374-1378. DOI: 10.1126/science.abf1611 From NLM.

(6) Dai, W.; Zhang, B.; Jiang, X. M.; Su, H.; Li, J.; Zhao, Y.; Xie, X.; Jin, Z.; Peng, J.; Liu, F.; et al. Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. *Science* **2020**, *368* (6497), 1331-1335. DOI: 10.1126/science.abb4489.

(7) Liang, J.; Pitsillou, E.; Karagiannis, C.; Darmawan, K. K.; Ng, K.; Hung, A.; Karagiannis, T. C. Interaction of the prototypical α-ketoamide inhibitor with the SARS-CoV-2 main protease active site in silico: Molecular dynamic simulations highlight the stability of the ligand-protein complex. *Comput Biol Chem* **2020**, *87*, 107292. DOI: 10.1016/j.compbiolchem.2020.107292 From NLM.

(8) 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* (10), e324. DOI: 10.1371/journal.pbio.0030324 From NLM.

(9) Joyce, R. P.; Hu, V. W.; Wang, J. The history, mechanism, and perspectives of nirmatrelvir (PF-07321332): an orally bioavailable main protease inhibitor used in combination with ritonavir to reduce COVID-19-related hospitalizations. *Med Chem Res* **2022**, *31* (10), 1637-1646. DOI: 10.1007/s00044-022-02951-6 From NLM.

(10) *COVID-19 Treatment Guidelines Panel*. National Institutes of Health, https://www.covid19treatmentguidelines.nih.gov/ (accessed May 30, 2023).

(11) Sasi, V. M.; Ullrich, S.; Ton, J.; Fry, S. E.; Johansen-Leete, J.; Payne, R. J.; Nitsche, C.; Jackson, C. J. Predicting Antiviral Resistance Mutations in SARS-CoV-2 Main Protease with Computational and Experimental Screening. *Biochemistry* **2022**, *61* (22), 2495-2505. DOI: 10.1021/acs.biochem.2c00489 From NLM.

(12) Clayton, J.; de Oliveira, V. M.; Ibrahim, M. F.; Sun, X.; Mahinthichaichan, P.; Shen, M.; Hilgenfeld, R.; Shen, J. Integrative Approach to Dissect the Drug Resistance Mechanism of the H172Y Mutation of SARS-CoV-2 Main Protease. *Journal of Chemical Information and Modeling* **2023**. DOI: 10.1021/acs.jcim.3c00344.

(13) Lavecchia, A.; Di Giovanni, C. Virtual screening strategies in drug discovery: a critical review. *Curr Med Chem* **2013**, *20* (23), 2839-2860. DOI: 10.2174/09298673113209990001 From NLM.

(14) Benod, C.; Carlsson, J.; Uthayaruban, R.; Hwang, P.; Irwin, J. J.; Doak, A. K.; Shoichet, B. K.; Sablin, E. P.; Fletterick, R. J. Structure-based discovery of antagonists of nuclear receptor LRH-1. *J Biol Chem* **2013**, *288* (27), 19830-19844. DOI: 10.1074/jbc.M112.411686 From NLM.

(15) Cheng, T.; Li, Q.; Zhou, Z.; Wang, Y.; Bryant, S. H. Structure-based virtual screening for drug discovery: a problem-centric review. *Aaps j* **2012**, *14* (1), 133-141. DOI: 10.1208/s12248-012-9322-0 From NLM.

(16) Heifetz, A.; Barker, O.; Verquin, G.; Wimmer, N.; Meutermans, W.; Pal, S.; Law, R. J.; Whittaker, M. Fighting obesity with a sugar-based library: discovery of novel MCH-1R antagonists by a new computational-VAST approach for exploration of GPCR binding sites. *J Chem Inf Model* **2013**, *53* (5), 1084-1099. DOI: 10.1021/ci4000882 From NLM.

(17) Schröder, J.; Klinger, A.; Oellien, F.; Marhöfer, R. J.; Duszenko, M.; Selzer, P. M. Docking-based virtual screening of covalently binding ligands: an orthogonal lead

discovery approach. *J Med Chem* **2013**, *56* (4), 1478-1490. DOI: 10.1021/jm3013932 From NLM.

(18) Kolb, P.; Rosenbaum, D. M.; Irwin, J. J.; Fung, J. J.; Kobilka, B. K.; Shoichet, B. K. Structure-based discovery of beta2-adrenergic receptor ligands. *Proc Natl Acad Sci U S A* **2009**, *106* (16), 6843-6848. DOI: 10.1073/pnas.0812657106 From NLM.

(19) Lionta, E.; Spyrou, G.; Vassilatis, D. K.; Cournia, Z. Structure-based virtual screening for drug discovery: principles, applications and recent advances. *Curr Top Med Chem* **2014**, *14* (16), 1923-1938. DOI: 10.2174/1568026614666140929124445 From NLM.

(20) Lipinski, C. A. Lead- and Drug-like Compounds: The Rule-of-Five Revolution. *Drug Discovery Today: Technologies* **2004**, *1*, 337-341.

(21) Chandrasekaran, B.; Abed, S. N.; Al-Attraqchi, O.; Kuche, K.; Tekade, R. K. Chapter 21 - Computer-Aided Prediction of Pharmacokinetic (ADMET) Properties. In *Dosage Form Design Parameters*, Tekade, R. K. Ed.; Academic Press, 2018; pp 731-755.

(22) Artursson, P.; Bergström, C. Computational approaches to drug absorption and bioavailability. 17. Intestinal absorption; the role of polar surface area. **2003**.

(23) Wenlock, M. C.; Barton, P. In Silico Physicochemical Parameter Predictions. *Molecular Pharmaceutics* **2013**, *10* (4), 1224-1235. DOI: 10.1021/mp300537k.

(24) Bergström, C. A. S.; Charman, W. N.; Porter, C. J. H. Computational prediction of formulation strategies for beyond-rule-of-5 compounds. *Advanced Drug Delivery Reviews* **2016**, *101*, 6-21. DOI: https://doi.org/10.1016/j.addr.2016.02.005.

(25) Soni, N.; Soni, N.; Pandey, H.; Maheshwari, R.; Kesharwani, P.; Tekade, R. K. Augmented delivery of gemcitabine in lung cancer cells exploring mannose anchored solid lipid nanoparticles. *Journal of Colloid and Interface Science* **2016**, *481*, 107-116. DOI: https://doi.org/10.1016/j.jcis.2016.07.020.

(26) Kuentz, M.; Imanidis, G. In silico prediction of the solubility advantage for amorphous drugs – Are there property-based rules for drug discovery and early pharmaceutical development? *European Journal of Pharmaceutical Sciences* **2013**, *48* (3), 554-562. DOI: https://doi.org/10.1016/j.ejps.2012.11.015.

(27) Lüder, K.; Lindfors, L.; Westergren, J.; Nordholm, S.; Kjellander, R. In Silico Prediction of Drug Solubility. 3. Free Energy of Solvation in Pure Amorphous Matter. *The Journal of Physical Chemistry B* **2007**, *111* (25), 7303-7311. DOI: 10.1021/jp071687d.

(28) McCracken, R. O.; Lipkowitz, K. B. Experimental and Theoretical Studies of Albendazole, Oxibendazole, and Tioxidazole. *The Journal of Parasitology* **1990**, *76* (2), 180-185. DOI: 10.2307/3283011 (acccessed 2023/07/17/).JSTOR.

(29) Palm, K.; Luthman, K.; Unge, A.-L.; Strandlund, G.; Artursson, P. Correlation of Drug Absorption with Molecular Surface Properties. *Journal of Pharmaceutical Sciences* **1996**, *85* (1), 32-39. DOI: https://doi.org/10.1021/js950285r.

(30) Ertl, P. a. R., B. and Selzer, P. Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. *Journal of medicinal chemistry* **2000**, *43*, 3714-3717.

(31) Winiwarter, S.; Ax, F.; Lennernäs, H.; Hallberg, A.; Pettersson, C.; Karlén, A. Hydrogen bonding descriptors in the prediction of human in vivo intestinal permeability. *Journal of Molecular Graphics and Modelling* **2003**, *21* (4), 273-287. DOI: https://doi.org/10.1016/S1093-3263(02)00163-8.

(32) Daina, A.; Michielin, O.; Zoete, V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific Reports* **2017**, *7* (1), 42717. DOI: 10.1038/srep42717.

(33) Giordano, D.; Biancaniello, C.; Argenio, M. A.; Facchiano, A. Drug Design by Pharmacophore and Virtual Screening Approach. *Pharmaceuticals (Basel)* **2022**, *15* (5). DOI: 10.3390/ph15050646 From NLM.

(34) Pirhadi, S.; Shiri, F.; Ghasemi, J. Methods and Applications of Structure Based Pharmacophores in Drug Discovery. *Current topics in medicinal chemistry* **2013**, *13*. DOI: 10.2174/1568026611313090006.

(35) Vuorinen, A.; Schuster, D. Methods for generating and applying pharmacophore models as virtual screening filters and for bioactivity profiling. *Methods* **2014**, *71*. DOI: 10.1016/j.ymeth.2014.10.013.

(36) Seidel, T.; Ibis, G.; Bendix, F.; Wolber, G. Strategies for 3D pharmacophore-based virtual screening. *Drug Discovery Today: Technologies* **2010**, *7* (4), e221-e228. DOI: https://doi.org/10.1016/j.ddtec.2010.11.004.

(37) John, S.; Thangapandian, S.; Arooj, M.; Hong, J. C.; Kim, K. D.; Lee, K. W. Development, evaluation and application of 3D QSAR Pharmacophore model in the discovery of potential human renin inhibitors. *BMC Bioinformatics* **2011**, *12* (14), S4.

DOI: 10.1186/1471-2105-12-S14-S4.

(38) Molla, M. H. R.; Aljahdali, M. O.; Sumon, M. A. A.; Asseri, A. H.; Altayb, H. N.; Islam, M. S.; Alsaiari, A. A.; Opo, F.; Jahan, N.; Ahammad, F.; et al. Integrative Ligand-Based Pharmacophore Modeling, Virtual Screening, and Molecular Docking Simulation Approaches Identified Potential Lead Compounds against Pancreatic Cancer by Targeting FAK1. *Pharmaceuticals (Basel)* **2023**, *16* (1). DOI: 10.3390/ph16010120 From NLM.

(39) Torres, P. H. M.; Sodero, A. C. R.; Jofily, P.; Silva-Jr, F. P. Key Topics in Molecular Docking for Drug Design. *Int J Mol Sci* **2019**, *20* (18). DOI: 10.3390/ijms20184574 From NLM.

(40) 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* (2), 455-461. DOI: 10.1002/jcc.21334.

(41) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking11Edited by F. E. Cohen. *Journal of Molecular Biology* **1997**, *267* (3), 727-748. DOI: https://doi.org/10.1006/jmbi.1996.0897.

(42) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J Mol Biol* **1996**, *261* (3), 470-489. DOI: 10.1006/jmbi.1996.0477 From NLM.

(43) Altuntaş, S.; Bozkus, Z.; Fraguela, B. B. GPU Accelerated Molecular Docking Simulation with Genetic Algorithms. In *Applications of Evolutionary Computation*, Cham, 2016//, 2016; Squillero, G., Burelli, P., Eds.; Springer International Publishing: pp 134-146.

(44) Yang, C.; Chen, E. A.; Zhang, Y. Protein-Ligand Docking in the Machine-Learning Era. *Molecules* **2022**, *27* (14). DOI: 10.3390/molecules27144568 From NLM.

(45) Hollingsworth, S. A.; Dror, R. O. Molecular Dynamics Simulation for All. *Neuron* **2018**, *99* (6), 1129-1143. DOI: 10.1016/j.neuron.2018.08.011 From NLM.

(46) Kubincová, A.; Riniker, S.; Hünenberger, P. H. Reaction-field electrostatics in molecular dynamics simulations: development of a conservative scheme compatible with an atomic cutoff. *Physical Chemistry Chemical Physics* **2020**, *22* (45), 26419-26437, 10.1039/D0CP03835K. DOI: 10.1039/D0CP03835K.

(47) Verlet, L. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review* **1967**, *159* (1), 98-103. DOI: 10.1103/PhysRev.159.98.

(48) Swope, W.; Andersen, H.; Berens, P.; Wilson, K. A Computer-Simulation Method for the Calculation of Equilibrium-Constants for the Formation of Physical Clusters of Molecules—Application to Small Water Clusters. *The Journal of Chemical Physics* **1982**, *76*. DOI: 10.1063/1.442716.

(49) Oh, K. J.; Klein, M. L. A general purpose parallel molecular dynamics simulation program. *Computer Physics Communications* **2006**, *174* (7), 560-568. DOI: https://doi.org/10.1016/j.cpc.2005.12.002.

(50) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N•log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **1993**, *98*.

(51) William L. Jorgensen, J. C. a. J. D. M. e. a. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983**, *79(2)*.

(52) Mahoney, M.; Jorgensen, W. A five-site model for liquid water and the reproduction of the density anomaly by rigid. *The Journal of Chemical Physics* **2000**, *112*, 8910-8922. DOI: 10.1063/1.481505.

(53) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The missing term in effective pair potentials. *The Journal of Physical Chemistry* **1987**, *91* (24), 6269-6271. DOI: 10.1021/j100308a038.

(54) Cheatham, T. E., III; Miller, J. L.; Fox, T.; Darden, T. A.; Kollman, P. A. Molecular Dynamics Simulations on Solvated Biomolecular Systems: The Particle Mesh Ewald Method Leads to Stable Trajectories of DNA, RNA, and Proteins. *Journal of the American Chemical Society* **1995**, *117* (14), 4193-4194. DOI: 10.1021/ja00119a045.

(55) Uberuaga, B. P.; Anghel, M.; Voter, A. F. Synchronization of trajectories in canonical molecular-dynamics simulations: observation, explanation, and exploitation. *J Chem Phys* **2004**, *120* (14), 6363-6374. DOI: 10.1063/1.1667473.

(56) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M.; Mackerell, A. D., Jr. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain $\chi(1)$ and $\chi(2)$ dihedral angles. *J Chem*

Theory Comput **2012**, *8* (9), 3257-3273. DOI: 10.1021/ct300400x From NLM.

(57) Fedorov, D. The fragment molecular orbital method: theoretical development, implementation in GAMESS, and applications: Fragment molecular orbital method in GAMESS. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2017**, *7*, e1322. DOI: 10.1002/wcms.1322.

(58) Tanaka, S.; Mochizuki, Y.; Komeiji, Y.; Okiyama, Y.; Fukuzawa, K. Electron-correlated fragment-molecular-orbital calculations for biomolecular and nano systems. *Physical Chemistry Chemical Physics* **2014**, *16* (22), 10310-10344, 10.1039/C4CP00316K. DOI: 10.1039/C4CP00316K.

(59) Mochizuki, Y.; Tanaka, S.; Fukuzawa, K. Recent Advances of the Fragment Molecular Orbital Method. *(No Title)* **2021**, 616.

(60) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res* **2000**, *28* (1), 235-242. DOI: 10.1093/nar/28.1.235.

(61) Akisawa, K.; Hatada, R.; Okuwaki, K.; Mochizuki, Y.; Fukuzawa, K.; Komeiji, Y.; Tanaka, S. Interaction analyses of SARS-CoV-2 spike protein based on fragment molecular orbital calculations. *RSC Advances* **2021**, *11* (6), 3272-3279, 10.1039/D0RA09555A. DOI: 10.1039/D0RA09555A.

(62) Nishimoto, Y. a. I., Stephan. *The FMO-DFTB Method*. 2021. https://www.osti.gov/biblio/1771915 (accessed.

(63) Fedorov, D. G. a. K., K. Second order Møller-Plesset perturbation theory based upon the fragment molecular orbital method. *Journal of Chemical Physics* **2004**, *121*, 2483-2490.

(64) Takaya, D.; Watanabe, C.; Nagase, S.; Kamisaka, K.; Okiyama, Y.; Moriwaki, H.; Yuki, H.; Sato, T.; Kurita, N.; Yagi, Y.; et al. FMODB: The World's First Database of Quantum Mechanical Calculations for Biomacromolecules Based on the Fragment Molecular Orbital Method. *Journal of Chemical Information and Modeling* **2021**, *61* (2), 777-794. DOI: 10.1021/acs.jcim.0c01062.

(65) Tokutomi, S.; Shimamura, K.; Fukuzawa, K.; Tanaka, S. Machine learning prediction of inter-fragment interaction energies between ligand and amino-acid residues on the fragment molecular orbital calculations for Janus kinase – inhibitor complex. *Chemical*

Physics Letters **2020**, *757*, 137883. DOI: https://doi.org/10.1016/j.cplett.2020.137883.

(66) Maghami, M.; Abdelrasoul, A. Pair interaction energy decomposition analysis (PIEDA) and experimental approaches for investigating water interactions with hydrophilic and hydrophobic membranes. *Journal of Molecular Graphics and Modelling* **2020**, *96*, 107540. DOI: https://doi.org/10.1016/j.jmgm.2020.107540.

(67) Okiyama, Y.; Watanabe, C.; Fukuzawa, K.; Mochizuki, Y.; Nakano, T.; Tanaka, S. Fragment Molecular Orbital Calculations with Implicit Solvent Based on the Poisson–Boltzmann Equation: II. Protein and Its Ligand-Binding System Studies. *The Journal of Physical Chemistry B* **2019**, *123* (5), 957-973. DOI: 10.1021/acs.jpcb.8b09326.

(68) Fedorov, D. G.; Kitaura, K. Subsystem Analysis for the Fragment Molecular Orbital Method and Its Application to Protein–Ligand Binding in Solution. *The Journal of Physical Chemistry A* **2016**, *120* (14), 2218-2231. DOI: 10.1021/acs.jpca.6b00163.

(69) 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* (13), 1605-1612. DOI: 10.1002/jcc.20084.

(70) *Discovery Studio Modeling Environment, Release 2.5.1*; Accelrys Software Inc: San Diego, CA. , 2009. (accessed October 6, 2022).

(71) Sander, T.; Freyss, J.; von Korff, M.; Rufener, C. DataWarrior: an open-source program for chemistry aware data visualization and analysis. *J Chem Inf Model* **2015**, *55* (2), 460-473. DOI: 10.1021/ci500588j.

(72) Sanachai, K.; Somboon, T.; Wilasluck, P.; Deetanya, P.; Wolschann, P.; Langer, T.; Lee, V. S.; Wangkanont, K.; Rungrotmongkol, T.; Hannongbua, S. Identification of repurposing therapeutics toward SARS-CoV-2 main protease by virtual screening. *PLoS One* **2022**, *17* (6), e0269563. DOI: 10.1371/journal.pone.0269563.

(73) *KNIME: The Konstanz Information Miner. In Studies in Classification, Data Analysis, and Knowledge Organization (GfKL 2007)*; University of Konstanz, Nycomed Chair for Bioinformatics and Information Mining, Box 712, 78457 Konstanz, Germany, 2007. (accessed October 14, 2022).

(74) Cereto-Massagué, A.; Guasch, L.; Valls, C.; Mulero, M.; Pujadas, G.; Garcia-Vallvé, S. DecoyFinder: an easy-to-use python GUI application for building target-specific decoy sets. *Bioinformatics* **2012**, *28* (12), 1661-1662. DOI: 10.1093/bioinformatics/bts249.

(75) Sterling, T.; Irwin, J. J. ZINC 15 – Ligand Discovery for Everyone. *Journal of Chemical Information and Modeling* **2015**, *55* (11), 2324-2337. DOI: 10.1021/acs.jcim.5b00559.

(76) Wishart, D. S.; Knox, C.; Guo, A. C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z.; Woolsey, J. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res* **2006**, *34* (Database issue), D668-672. DOI: 10.1093/nar/gkj067.

(77) Fei, J.; Zhou, L.; Liu, T.; Tang, X. Y. Pharmacophore modeling, virtual screening, and molecular docking studies for discovery of novel Akt2 inhibitors. *Int J Med Sci* **2013**, *10* (3), 265-275. DOI: 10.7150/ijms.5344.

(78) Réau, M.; Langenfeld, F.; Zagury, J. F.; Lagarde, N.; Montes, M. Decoys Selection in Benchmarking Datasets: Overview and Perspectives. *Front Pharmacol* **2018**, *9*, 11. DOI: 10.3389/fphar.2018.00011.

(79) *Marvin 17.21.0, Chemaxon* https://www.chemaxon.com (accessed August 19, 2022).

(80) El-Hachem, N.; Haibe-Kains, B.; Khalil, A.; Kobeissy, F. H.; Nemer, G. AutoDock and AutoDockTools for Protein-Ligand Docking: Beta-Site Amyloid Precursor Protein Cleaving Enzyme 1(BACE1) as a Case Study. *Methods Mol Biol* **2017**, *1598*, 391-403. DOI: 10.1007/978-1-4939-6952-4_20.

(81) Wolber, G.; Langer, T. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J Chem Inf Model* **2005**, *45* (1), 160-169. DOI: 10.1021/ci049885e.

(82) *Amber 2020*; University of California, San Francisco: 2020. (accessed March 4, 2023).

(83) Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Migues, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; et al. ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *Journal of Chemical Theory and Computation* **2020**, *16* (1), 528-552. DOI: 10.1021/acs.jctc.9b00591.

(84) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J Comput Chem* **2004**, *25* (9), 1157-1174. DOI: 10.1002/jcc.20035.

(85) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of

solvation. *Journal of the American Chemical Society* **1993**, *115* (21), 9620-9631. DOI: 10.1021/ja00074a030.

(86) Sanachai, K.; Mahalapbutr, P.; Sanghiran Lee, V.; Rungrotmongkol, T.; Hannongbua, S. In Silico Elucidation of Potent Inhibitors and Rational Drug Design against SARS-CoV-2 Papain-like Protease. *The Journal of Physical Chemistry B* **2021**, *125* (50), 13644-13656. DOI: 10.1021/acs.jpcb.1c07060.

(87) Sripattaraphan, A.; Sanachai, K.; Chavasiri, W.; Boonyasuppayakorn, S.; Maitarad, P.; Rungrotmongkol, T. Computational Screening of Newly Designed Compounds against Coxsackievirus A16 and Enterovirus A71. In *Molecules*, 2022; Vol. 27.

(88) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **1984**, *81*, 3684-3690. DOI: 10.1063/1.448118.

(89) Hünenberger, P. H. Thermostat Algorithms for Molecular Dynamics Simulations. *Advances in Polymer Science* **2005**, *173*, 105-147.

(90) Wansri, R.; Lin, A. C.; Pengon, J.; Kamchonwongpaisan, S.; Srimongkolpithak, N.; Rattanajak, R.; Wilasluck, P.; Deetanya, P.; Wangkanont, K.; Hengphasatporn, K.; et al. Semi-Synthesis of N-Aryl Amide Analogs of Piperine from Piper nigrum and Evaluation of Their Antitrypanosomal, Antimalarial, and Anti-SARS-CoV-2 Main Protease Activities. *Molecules* **2022**, *27* (9). DOI: 10.3390/molecules27092841.

(91) Hengphasatporn, K.; Wilasluck, P.; Deetanya, P.; Wangkanont, K.; Chavasiri, W.; Visitchanakun, P.; Leelahavanichkul, A.; Paunrat, W.; Boonyasuppayakorn, S.; Rungrotmongkol, T.; et al. Halogenated Baicalein as a Promising Antiviral Agent toward SARS-CoV-2 Main Protease. *Journal of Chemical Information and Modeling* **2022**, *62* (6), 1498-1509. DOI: 10.1021/acs.jcim.1c01304.

(92) Hengphasatporn, K.; Harada, R.; Wilasluck, P.; Deetanya, P.; Sukandar, E. R.; Chavasiri, W.; Suroengrit, A.; Boonyasuppayakorn, S.; Rungrotmongkol, T.; Wangkanont, K.; et al. Promising SARS-CoV-2 main protease inhibitor ligand-binding modes evaluated using LB-PaCS-MD/FMO. *Scientific Reports* **2022**, *12* (1), 17984. DOI: 10.1038/s41598-022-22703-1.

(93) Pyae, N. Y. L.; Maiuthed, A.; Phongsopitanun, W.; Ouengwanarat, B.; Sukma, W.; Srimongkolpithak, N.; Pengon, J.; Rattanajak, R.; Kamchonwongpaisan, S.; Ei, Z. Z.; et al.

N-Containing α-Mangostin Analogs via Smiles Rearrangement as the Promising Cytotoxic, Antitrypanosomal, and SARS-CoV-2 Main Protease Inhibitory Agents. *Molecules* **2023**, *28* (3), 1104.

(94) Pithi, C.; Supakarn, C.; Chuanpit, N.; Preeyaporn Plaimee, P. Potential Anti-metastasis Natural Compounds for Lung Cancer. *Anticancer Research* **2016**, *36* (11), 5707.

(95) Karnsomwan, W.; Netcharoensirisuk, P.; Rungrotmongkol, T.; De-Eknamkul, W.; Chamni, S. Synthesis, Biological Evaluation and Molecular Docking of Avicequinone C Analogues as Potential Steroid 5α-Reductase Inhibitors. *Chemical and Pharmaceutical Bulletin* **2017**, *65* (3), 253-260. DOI: 10.1248/cpb.c16-00727.

(96) Lin, A. C. K.; Netcharoensirisuk, P.; Sanachai, K.; Sukma, W.; Chansriniyom, C.; Chaotham, C.; De-Eknamkul, W.; Rungrotmongkol, T.; Chamni, S. Caffeic acid N-[3,5-bis(trifluoromethyl)phenyl] amide as a non-steroidal inhibitor for steroid 5α-reductase type 1 using a human keratinocyte cell-based assay and molecular dynamics. *Scientific Reports* **2022**, *12* (1), 20858. DOI: 10.1038/s41598-022-25335-7.

(97) Ke, S.; Shi, L.; Zhang, Z.; Yang, Z. Steroidal[17,16-d]pyrimidines derived from dehydroepiandrosterone: A convenient synthesis, antiproliferation activity, structure-activity relationships, and role of heterocyclic moiety. *Scientific Reports* **2017**, *7* (1), 44439. DOI: 10.1038/srep44439.

(98) Wansri, R.; Lin, A. C. K.; Pengon, J.; Kamchonwongpaisan, S.; Srimongkolpithak, N.; Rattanajak, R.; Wilasluck, P.; Deetanya, P.; Wangkanont, K.; Hengphasatporn, K.; et al. Semi-Synthesis of N-Aryl Amide Analogs of Piperine from Piper nigrum and Evaluation of Their Antitrypanosomal, Antimalarial, and Anti-SARS-CoV-2 Main Protease Activities. *Molecules* **2022**, *27* (9), 2841.

(99) Yokoya, M.; Yamazaki-Nakai, M.; Nakai, K.; Sirimangkalakitti, N.; Chamni, S.; Suwanborirux, K.; Saito, N. Transformation of Renieramycin M into Renieramycins T and S by Intramolecular Photoredox Reaction of 7-Methoxy-6-methyl-1,2,3,4-tetrahydroisoquinoline-5,8-dione Derivatives. *Journal of Natural Products* **2023**, *86* (1), 222-231. DOI: 10.1021/acs.jnatprod.2c00974.

(100) Thongsom, S.; Racha, S.; Ei, Z. Z.; Petsri, K.; Aksorn, N.; Chamni, S.; Panpuang, V.;

Zou, H.; Chanvorachote, P. N,N'-Diarylurea Derivatives (CTPPU) Inhibited NSCLC Cell Growth and Induced Cell Cycle Arrest through Akt/GSK-3β/c-Myc Signaling Pathway. *Int J Mol Sci* **2023**, *24* (2). DOI: 10.3390/ijms24021357 From NLM.

(101) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Meng, E. C.; Couch, G. S.; Croll, T. I.; Morris, J. H.; Ferrin, T. E. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* **2021**, *30* (1), 70-82. DOI: 10.1002/pro.3943 From NLM.

(102) Zhu, J.; Zhang, H.; Lin, Q.; Lyu, J.; Lu, L.; Chen, H.; Zhang, X.; Zhang, Y.; Chen, K. Progress on SARS-CoV-2 3CLpro Inhibitors: Inspiration from SARS-CoV 3CLpro Peptidomimetics and Small-Molecule Anti-Inflammatory Compounds. *Drug Des Devel Ther* **2022**, *16*, 1067-1082. DOI: 10.2147/dddt.S359009.

(103) Kaserer, T.; Beck, K. R.; Akram, M.; Odermatt, A.; Schuster, D. Pharmacophore Models and Pharmacophore-Based Virtual Screening: Concepts and Applications Exemplified on Hydroxysteroid Dehydrogenases. *Molecules* **2015**, *20* (12), 22799-22832. DOI: 10.3390/molecules201219880 From NLM.

(104) Seidel, T.; Wieder, O.; Garon, A.; Langer, T. Applications of the Pharmacophore Concept in Natural Product inspired Drug Design. *Mol Inform* **2020**, *39* (11), e2000059. DOI: 10.1002/minf.202000059 From NLM.

(105) Hamza, A.; Wei, N. N.; Zhan, C. G. Ligand-based virtual screening approach using a new scoring function. *J Chem Inf Model* **2012**, *52* (4), 963-974. DOI: 10.1021/ci200617d From NLM.

(106) Pantsar, T.; Poso, A. Binding Affinity via Docking: Fact and Fiction. *Molecules* **2018**, *23* (8). DOI: 10.3390/molecules23081899 From NLM.

(107) Elkhalifa, D.; Al-Hashimi, I.; Al Moustafa, A.-E.; Khalil, A. A comprehensive review on the antiviral activities of chalcones. *Journal of Drug Targeting* **2021**, *29* (4), 403-419. DOI: 10.1080/1061186X.2020.1853759.

(108) Valipour, M. Recruitment of chalcone's potential in drug discovery of anti-SARS-CoV-2 agents. *Phytother Res* **2022**, *36* (12), 4477-4490. DOI: 10.1002/ptr.7651 From NLM.

(109) Park, J. Y.; Ko, J. A.; Kim, D. W.; Kim, Y. M.; Kwon, H. J.; Jeong, H. J.; Kim, C. Y.; Park,

K. H.; Lee, W. S.; Ryu, Y. B. Chalcones isolated from Angelica keiskei inhibit cysteine proteases of SARS-CoV. *J Enzyme Inhib Med Chem* **2016**, *31* (1), 23-30. DOI: 10.3109/14756366.2014.1003215 From NLM.

(110) Dao, T. T.; Nguyen, P. H.; Lee, H. S.; Kim, E.; Park, J.; Lim, S. I.; Oh, W. K. Chalcones as novel influenza A (H1N1) neuraminidase inhibitors from Glycyrrhiza inflata. *Bioorg Med Chem Lett* **2011**, *21* (1), 294-298. DOI: 10.1016/j.bmcl.2010.11.016 From NLM.

(111) Phrutivorapongkul, A.; Lipipun, V.; Ruangrungsi, N.; Kirtikara, K.; Nishikawa, K.; Maruyama, S.; Watanabe, T.; Ishikawa, T. Studies on the chemical constituents of stem bark of Millettia leucantha: isolation of new chalcones with cytotoxic, anti-herpes simplex virus and anti-inflammatory activities. *Chem Pharm Bull (Tokyo)* **2003**, *51* (2), 187-190. DOI: 10.1248/cpb.51.187 From NLM.

(112) Mateeva, N.; Eyunni, S. V. K.; Redda, K. K.; Ononuju, U.; Hansberry, T. D., 2nd; Aikens, C.; Nag, A. Functional evaluation of synthetic flavonoids and chalcones for potential antiviral and anticancer properties. *Bioorg Med Chem Lett* **2017**, *27* (11), 2350-2356. DOI: 10.1016/j.bmcl.2017.04.034 From NLM.

(113) Cole, A. L.; Hossain, S.; Cole, A. M.; Phanstiel, O. t. Synthesis and bioevaluation of substituted chalcones, coumaranones and other flavonoids as anti-HIV agents. *Bioorg Med Chem* **2016**, *24* (12), 2768-2776. DOI: 10.1016/j.bmc.2016.04.045 From NLM.

(114) Mottin, M.; Caesar, L. K.; Brodsky, D.; Mesquita, N. C. M. R.; de Oliveira, K. Z.; Noske, G. D.; Sousa, B. K. P.; Ramos, P. R. P. S.; Jarmer, H.; Loh, B.; et al. Chalcones from Angelica keiskei (ashitaba) inhibit key Zika virus replication proteins. *Bioorganic Chemistry* **2022**, *120*, 105649. DOI: https://doi.org/10.1016/j.bioorg.2022.105649.

(115) Hasan, M.; Parvez, M. S. A.; Azim, K. F.; Imran, M. A. S.; Raihan, T.; Gulshan, A.; Muhit, S.; Akhand, R. N.; Ahmed, S. S. U.; Uddin, M. B. Main protease inhibitors and drug surface hotspots for the treatment of COVID-19: A drug repurposing and molecular docking approach. *Biomedicine & Pharmacotherapy* **2021**, *140*, 111742. DOI: https://doi.org/10.1016/j.biopha.2021.111742.

(116) Zhao, Y.; Fang, C.; Zhang, Q.; Zhang, R.; Zhao, X.; Duan, Y.; Wang, H.; Zhu, Y.; Feng, L.; Zhao, J.; et al. Crystal structure of SARS-CoV-2 main protease in complex with protease inhibitor PF-07321332. *Protein Cell* **2022**, *13* (9), 689-693. DOI: 10.1007/s13238-021-00883-2.

(117) Gapsys, V.; Yildirim, A.; Aldeghi, M.; Khalak, Y.; van der Spoel, D.; de Groot, B. L. Accurate absolute free energies for ligand–protein binding based on non-equilibrium approaches. *Communications Chemistry* **2021**, *4* (1), 61. DOI: 10.1038/s42004-021-00498-y.

(118) Martin, Y. C. A bioavailability score. *J Med Chem* **2005**, *48* (9), 3164-3170. DOI: 10.1021/jm0492002.

(119) Council, N. R. *A Framework to Guide Selection of Chemical Alternatives*. 2014. https://www.ncbi.nlm.nih.gov/books/NBK253956/ (accessed June 20, 2023).

(120) Li, H.; Zhang, H.; Zheng, M.; Luo, J.; Kang, L.; Liu, X.; Wang, X.; Jiang, H. An effective docking strategy for virtual screening based on multi-objective optimization algorithm. *BMC Bioinformatics* **2009**, *10* (1), 58. DOI: 10.1186/1471-2105-10-58.

(121) Gan, J.-h.; Liu, J.-x.; Liu, Y.; Chen, S.-w.; Dai, W.-t.; Xiao, Z.-X.; Cao, Y. DrugRep: an automatic virtual screening server for drug repurposing. *Acta Pharmacologica Sinica* **2023**, *44* (4), 888-896. DOI: 10.1038/s41401-022-00996-2.

(122) Kumar, V.; Dhanjal, J. K.; Kaul, S. C.; Wadhwa, R.; Sundar, D. Withanone and caffeic acid phenethyl ester are predicted to interact with main protease (M(pro)) of SARS-CoV-2 and inhibit its activity. *J Biomol Struct Dyn* **2021**, *39* (11), 3842-3854. DOI: 10.1080/07391102.2020.1772108 From NLM.

(123) van de Sand, L.; Bormann, M.; Alt, M.; Schipper, L.; Heilingloh, C. S.; Steinmann, E.; Todt, D.; Dittmer, U.; Elsner, C.; Witzke, O.; et al. Glycyrrhizin Effectively Inhibits SARS-CoV-2 Replication by Inhibiting the Viral Main Protease. *Viruses* **2021**, *13* (4). DOI: 10.3390/v13040609 From NLM.

(124) Dhanjal, J. K.; Kumar, V.; Garg, S.; Subramani, C.; Agarwal, S.; Wang, J.; Zhang, H.; Kaul, A.; Kalra, R. S.; Kaul, S. C.; et al. Molecular mechanism of anti-SARS-CoV2 activity of Ashwagandha-derived withanolides. *Int J Biol Macromol* **2021**, *184*, 297-312. DOI: 10.1016/j.ijbiomac.2021.06.015 From NLM.

(125) Alves, D. R.; Rocha, M. N. d.; Passos, C. C. O.; Marinho, M. M.; Marinho, E. S.; Morais, S. M. d. Curcumins and its derivatives as potential inhibitors of New Coronavirus (COVID-19) main protease: an in silico strategy. *Research, Society and Development* **2022**, *11* (1), e6511124334. DOI: 10.33448/rsd-v11i1.24334 (acccessed 2023/06/18).

(126) Bahun, M.; Jukić, M.; Oblak, D.; Kranjc, L.; Bajc, G.; Butala, M.; Bozovičar, K.;

Bratkovič, T.; Podlipnik, Č.; Poklar Ulrih, N. Inhibition of the SARS-CoV-2 3CLpro main protease by plant polyphenols. *Food Chemistry* **2022**, *373*, 131594. DOI: https://doi.org/10.1016/j.foodchem.2021.131594.

(127) Suriya, U.; Mahalapbutr, P.; Rungrotmongkol, T. Integration of In Silico Strategies for Drug Repositioning towards P38α Mitogen-Activated Protein Kinase (MAPK) at the Allosteric Site. *Pharmaceutics* **2022**, *14* (7). DOI: 10.3390/pharmaceutics14071461 From NLM.

(128) Lee, J. T.; Yang, Q.; Gribenko, A.; Perrin, B. S., Jr.; Zhu, Y.; Cardin, R.; Liberator, P. A.; Anderson, A. S.; Hao, L. Genetic Surveillance of SARS-CoV-2 M(pro) Reveals High Sequence and Structural Conservation Prior to the Introduction of Protease Inhibitor Paxlovid. *mBio* **2022**, *13* (4), e0086922. DOI: 10.1128/mbio.00869-22 From NLM.

(129) Unoh, Y.; Uehara, S.; Nakahara, K.; Nobori, H.; Yamatsu, Y.; Yamamoto, S.; Maruyama, Y.; Taoda, Y.; Kasamatsu, K.; Suto, T.; et al. Discovery of S-217622, a Noncovalent Oral SARS-CoV-2 3CL Protease Inhibitor Clinical Candidate for Treating COVID-19. *Journal of Medicinal Chemistry* **2022**, *65* (9), 6499-6512. DOI: 10.1021/acs.jmedchem.2c00117.

(130) Ang, D.; Kendall, R.; Atamian, H. S. Virtual and In Vitro Screening of Natural Products Identifies Indole and Benzene Derivatives as Inhibitors of SARS-CoV-2 Main Protease (Mpro). *Biology* **2023**, *12* (4), 519.

(131) Onyango, H.; Odhiambo, P.; Angwenyi, D.; Okoth, P. In Silico Identification of New Anti-SARS-CoV-2 Main Protease (M^{pro}) Molecules with Pharmacokinetic Properties from Natural Sources Using Molecular Dynamics (MD) Simulations and Hierarchical Virtual Screening. *Journal of Tropical Medicine* **2022**, *2022*, 3697498. DOI: 10.1155/2022/3697498.

(132) Xue, X.; Yang, H.; Shen, W.; Zhao, Q.; Li, J.; Yang, K.; Chen, C.; Jin, Y.; Bartlam, M.; Rao, Z. Production of Authentic SARS-CoV Mpro with Enhanced Activity: Application as a Novel Tag-cleavage Endopeptidase for Protein Overproduction. *Journal of Molecular Biology* **2007**, *366* (3), 965-975. DOI: https://doi.org/10.1016/j.jmb.2006.11.073.

(133) *GraphPad Prism*; San Diego CA, 2007. www.graphpad.com. (accessed September 11, 2022).

(134) Cheng, Y.; Prusoff, W. H. Relationship between the inhibition constant (Ki) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. *Biochem Pharmacol* **1973**, *22* (23), 3099-3108. DOI: 10.1016/0006-2952(73)90196-2.

(135) Deetanya, P.; Hengphasatporn, K.; Wilasluck, P.; Shigeta, Y.; Rungrotmongkol, T.; Wangkanont, K. Interaction of 8-anilinonaphthalene-1-sulfonate with SARS-CoV-2 main protease and its application as a fluorescent probe for inhibitor identification. *Comput Struct Biotechnol J* **2021**, *19*, 3364-3371. DOI: 10.1016/j.csbj.2021.05.053.

(136) Lei, C.; Yang, J.; Hu, J.; Sun, X. On the Calculation of TCID(50) for Quantitation of Virus Infectivity. *Virol Sin* **2021**, *36* (1), 141-144. DOI: 10.1007/s12250-020-00230-5.

(137) Bullen, C. K.; Davis, S. L.; Looney, M. M. Quantification of Infectious SARS-CoV-2 by the 50% Tissue Culture Infectious Dose Endpoint Dilution Assay. *Methods Mol Biol* **2022**, *2452*, 131-146. DOI: 10.1007/978-1-0716-2111-0_9.

(138) Shin, H.-C.; Wayas, I.; Choi, B.; Lee, B. Identification and Biological Activities of the Phenolic Compounds in Eisenia arborea. *American Journal of Plant Sciences* **2021**, *12*, 259-265. DOI: 10.4236/ajps.2021.122015.

APPENDIX

1. Enzyme inhibition and kinetic studies of SWC inhibitors

According to the previous descriptions for the SARS-CoV-1 3CLpro, SARS-CoV-2 3CLpro was expressed, purified and stored.[132] In each experiment, a concentration of 0.2 μM of 3CLpro was utilized. The initial rate of cleavage of the fluorogenic substrate E(EDANS)TSAVLQSGFRK(DABCYL) was used to determine enzymatic activity, with excitation and emission wavelengths recorded at 340 and 490 nm, respectively. For the preliminary screening of inhibitory activity, enzymatic activity was evaluated with and without 10 and 100 μM of SWC inhibitors. The initial rate since there was no inhibitor present was employed for the normalization. The IC50 value was determined by measuring the initial rate of substrate cleavage at varying concentrations of 25 μM SWC423, followed by fitting the data with GraphPad Prism 9.5.0.[133] The K_i value was subsequently obtained by applying the Cheng-Prusoff equation[134] to the previously reported K_m value (51 μM).[135]

2. Cell-based assay of SWC inhibitors

2.1 Cells and virus culture

Vero E6 cells (ATCC, CRL-1587) were maintained in minimal essential medium (MEM) (Gibco, Langley, OK, USA) supplemented with 10% fetal bovine serum (Gibco® , Langley, OK, USA), 100 I.U./ml penicillin (Bio Basic Canada, Ontario, CA), and 100 μg/ml streptomycin (Bio Basic Canada, Ontario, CA), 10 mM HEPES (4-(2- hydroxyethyl)-1-piperazineethanesulfonic acid) (Sigma Aldrich, St. Louis, MO, USA), Non-essential amino acid (NEAA) (Gibco, Langley, OK, USA), and sodium pyruvate (Gibco, Langley, OK, USA). Cells were incubated at 37 °C humidified chamber under 5% CO_2.

The SARS-CoV-2 (accession number: pending) was isolated from clinical specimens. The virus was propagated in Vero E6 cells with MEM supplemented with 1% fetal bovine serum, 100 I.U./ml penicillin, and 100 μg/ml streptomycin, 10 mM HEPES, NEAA, and sodium pyruvate at 37 °C humidified chamber under 5% CO_2. Virus titers were determined as TCID 50 /ml in confluent cells in 96-well cell culture plates. All experiment with live SARS-CoV-2 was performed in a certified biosafety level 3 facility of the research affair-Medical Research Center (MRC), Faculty of Medicine, Chulalongkorn University. The study was conducted according to the guidelines of the Declaration of Helsinki, and Chulalongkorn University Institutional Biosafety Committee (CU-IBC 003/2021). The Institutional Review Board of Faculty of Medicine, Chulalongkorn University certified the protocol exemption for the use of a leftover specimen (COE 017/2021, IRB No. 297/64).

2.2 Efficacy study

Three sulfonamide chalcones were tested against SARS-CoV-2 (accession number: pending). Briefly, Vero E6 cells at 5 x 104 cells per well were seeded into a 24-well plate and

incubated overnight at 37 °C under 5% CO_2 . Cells were infected with SARS-CoV-2 at the 100 TCID50. After infection, cells were washed with phosphate buffer saline (PBS) and incubated with 1 ml of maintenance medium. The compounds were prepared at the indicated concentrations in 0.1% DMSO in the maintenance medium during infection and after infection. Cells were incubated at 37 °C for 72 h under 5% CO_2 humidified chamber. Supernatants were collected for analysis of the viral infectivity by TCID 50 /ml (v2.1 - 20-01-2017_MB* by Marco Binder; adapted @ TWC.5.6, accessed on 16 May 2022). The compound was serially diluted to 6-8 different concentrations and was added to final concentrations into SARS-CoV2-infected cells. Dihethylsulfoxide at 0.1% was used as a vehicle, no inhibition control. Cells were incubated for 72 h and supernatants were collected for subsequent TCID 50 /ml analysis (Lei et al.[136] and Davis et al.[137]). Data were plotted and EC_{50} values were calculated from nonlinear regression analysis. Results were reported as means and standard error of the means (SEM) of three independent experiments.

2.3 Cytotoxicity in cell-based assay

The cytotoxicity of the active compounds was tested with Vero E6 cell lines (Wansri et al.).[98] Results were reported as means and SEM of three independent experiments.

3. Preparation of ester derivatives of caffeic acid

Caffeic acid and reagents were purchased from Sigma-Aldrich (Missouri, USA). Analytical grade solvents were obtained from Carlo Erba (Michigan, USA) and Honeywell (North Carolin, USA) and distilled before use. Anhydrous solvents were dried over 4 Å molecular sieves. Reactions were carried out in oven-dried glassware and magnetically stirred under a nitrogen atmosphere using a small balloon unless otherwise noted. Room temperature was 25 °C unless otherwise stated. Thin-layer chromatography (TLC) was used to monitor all reactions by utilizing aluminum silica gel 60 F254 from Merck (Darmstadt, Germany) and observed under ultraviolet (UV) light at 254 nm. Flash column chromatography was also used to purify all synthetic compounds using silica gel (60 Å, 230-400 mesh) as the stationary phase and the suitable mixtures of ethyl acetate, dichloromethane, and hexane as the mobile phases. Spectroscopic methods were used to elucidate the structures of all synthetic compounds. Nuclear magnetic resonance (NMR) spectra, both proton (^{1}H) and carbon (^{13}C), were measured on Bruker Avance NEO 400 MHz spectrometer. ^{1}H-NMR chemical shifts are reported as δ values in ppm relative to residual $CHCl_3$ (7.27 ppm). ^{1}H-NMR coupling constants (J) are reported in Hertz (Hz). Unless otherwise indicated, deuterochloroform ($CDCl_3$) served as an internal standard (77.0 ppm) for all ^{13}C spectra. Infrared (IR) spectra were recorded on Perkin Elmer Frontier Fourier-transform IR Spectrometer. High-resolution mass spectra were performed on Bruker Daltonics microTOF mass spectrometer.

Synthesis of caffeic acid methyl ester (4k) and 3-hydroxyl-4-methoxy cinnamic acid methyl ester (4l). Caffeic acid (40 mg, 0.22 mmol) was added to an oven-dried round-bottomed flask and dissolved in a 4 mL mixture of dry tetrahydrofuran: dry methanol (1:1) under the inert atmosphere. The reaction mixture was cooled at 0 °C using an ice bath. Then trimethylsilyl diazomethane (0.6 M in hexane, 1.1 mL, 0.66 mmol) was slowly added dropwise to the reaction mixture until the yellow color persisted. The reaction was stirred at room temperature for 30 minutes. The reaction was monitored by TLC. After that, the volatile solvent was removed under reduced pressure to obtain the crude product. The synthesized compounds were purified by column chromatography with silica gel as the stationary phase and the mixture of ethyl acetate and hexane as the mobile phase. The chemical structures of the purified compounds were elucidated by spectroscopic techniques. Caffeic acid methyl ester (4k) was obtained as the brown oil at 15 mg (35%). ^{1}H-NMR (CDCl$_3$, 400 MHz) δ 7.58 (1H, d, J = 16.0 Hz, 7-H), 7.08 (1H, d, J = 2.0 Hz, 3-H), 7.00 (1H, dd, J = 8.4, 2.0 Hz, 5-H), 6.87 (1H, d, J = 8.0 Hz, 6-H), 6.26 (1H, d, J = 16.0 Hz, 8-H), 3.79 (3H, s, 10-H) ppm; IR (ATR, V_{max}) 3361, 2927, 1698, 1634, 1601, 1511, 1263, 1159, 854, 807 cm^{-1} [138]. 3-Hydroxyl-4-methoxycinnamic acid methyl ester (4l) was obtained as the brown oil at 32 mg (70%) ^{1}H-NMR (CDCl$_3$, 400 MHz) δ 7.58 (1H, d, J = 16.0 Hz, 7-H), 7.00 (1H, dd, J = 8.4, 2.0 Hz, 5-H), 6.88 (1H, d, J = 8.4 Hz, 3-H), 6.81 (1H, d, J = 7.2 Hz, 6-H), 6.26 (1H, d, J = 16.0 Hz, 8-H), 3.87 (3H, s, 10-H), 3.76 (3H, s, 11-H) ppm; IR (ATR, V_{max}) 3354, 2949, 1692, 1631, 1599, 1509, 1258, 1157, 852, 807 cm^{-1}; HRMS–ESI m/z 231.0625 ([M+Na]$^+$, calcd for C$_{11}$H$_{12}$NaO$_4$ 231.0628) and 209.0807 ([M+H]$^+$, calcd for C$_{11}$H$_{13}$O$_4$ 209.0808).

Synthesis of 3,4-diacetyloxy cinnamic acid (4n). Caffeic acid (500 mg, 2.78 mmol), acetic anhydride (5 mL, 52.82 mmol), and pyridine (0.1 mL) were added to an oven-dried round-bottomed flask. Then, the reaction mixture was refluxed at 120 °C for 5 h and monitored by TLC. After that, the reaction was cooled in the ice bath and quenched by adding water. The product was precipitated and filtered to obtain 4n as a pale light brown solid at 589.3 mg (80%). ^{1}H-NMR (CDCl$_3$, 400 MHz) δ 7.74 (1H, d, J = 16.0 Hz, 11-H), 7.46 (1H, dd, J = 8.4, 2.0 Hz, 7-H), 7.41 (1H, d, J = 2.0 Hz, 5-H), 7.27 (1H, d, J = 8.4 Hz, 8-H), 6.42 (1H, d, J = 16.0 Hz, 12-H), 2.34 (3H, s, 1-H), 2.33 (3H, s, 10-H) ppm; ^{13}C-NMR (CDCl$_3$, 100 MHz) δ 170.8 (C-13), 168.1 (C-9), 168.0 (C-2), 145.0 (C-11), 143.9 (C-3), 142.5 (C-4), 132.9 (C-6), 126.7 (C-7), 124.0 (C-8), 123.0 (C-5), 118.3 (C-12), 20.7 (C-1), 20.6 (C-10); IR (ATR, V_{max}) 2822, 2522, 1755, 1674, 1628, 1328, 1290, 1251, 878. 824 cm^{-1}; HRMS–ESI m/z 265.0702 ([M+H]$^+$, calcd for C$_{13}$H$_{13}$O$_6$ 265.0707).

Synthesis of 3,4-diacetyloxy cinnamic acid-2,2-dimethyl-1-propyl ester (4p). Caffeic acid (30 mg, 0.11 mmol) and 2,2-dimethyl-1-propanol (9.69 mg, 0.11 mmol) were added to an oven-dried round-bottomed flask and dissolved in 20 mL dry dichloromethane under the inert atmosphere. The reaction mixture was cooled at 0 °C in an ice bath. Then, 1-ethyl-3-(3-dimethyl

aminopropyl) carbodiimide (EDCI HCl: 24.9 mg, 0.13 mmol), hydroxybenzotriazole (HOBt: 17.6 mg, 0.13 mmol), and *N,N*-diisopropylethylamine (DIPEA: 0.038 mL, 0.22 mmol) were added to the reaction mixture. The reaction was stirred at room temperature for 4 h. TLC was used to monitor the reaction. After that, the volatile solvent was removed under reduced pressure to obtain the crude product. The synthesized compounds were purified by column chromatography with silica gel as the stationary phase and dichloromethane as the mobile phase to obtain 4p as a white amorphous solid at 3.9 mg (11%). ^{1}H-NMR (CDCl$_3$, 400 MHz) δ 7.65 (1H, d, *J* = 16.0 Hz, 11-H), 7.44 (1H, dd, *J* = 8.4, 2.0 Hz, 7-H), 7.40 (1H, d, *J* = 2.0 Hz, 5-H), 7.25 (1H, d, *J* = 8.4 Hz, 8-H), 6.44 (1H, d, *J* = 16.0 Hz, 12-H), 3.93 (2H, s, 14-H), 2.33 (3H, s, 1-H), 2.33 (3H, s, 10-H), 1.02 (9H, s, 16-H) ppm; ^{13}C-NMR (CDCl$_3$, 100 MHz) δ 168.1 (C-9), 168.0 (C-2), 166.7 (C-13), 143.4 (C-3), 142.6 (C-11), 142.4 (C-4), 133.4 (C-6), 126.4 (C-7), 123.9 (C-8), 122.7 (C-5), 119.5 (C-12), 74.0 (C-14), 68.2 (C-15), 26.5 (C-16), 20.7 (C-1), 20.6 (C-10) ppm; IR (ATR, $\mathbf{v}_{max}$) 2921, 2851, 1763, 1658, 1632, 1368, 1252, 1209, 1172, 1146, 1106, 874, 836, 803 cm^{-1}; HRMS–ESI *m/z* 357.1319 ([M+Na]$^+$, calcd for C$_{18}$H$_{22}$NaO$_6$ 357.1309) and 335.1492 ([M+H]$^+$, calcd for C$_{18}$H$_{23}$O$_6$ 335.1489).

Synthesis of 3,4-diacetyloxy cinnamic acid-2,5-bis(trifluoromethyl) phenyl ester (4u). Caffeic acid (30 mg, 0.11 mmol) and 2,5-bis(trifluoromethyl)phenol (25.3 mg, 0.11 mmol) were added to an oven-dried round-bottomed flask and dissolved in 20 mL dry dichloromethane under the inert atmosphere. The reaction mixture was cooled down to 0 °C using an ice bath. Then, EDCI HCl (24.9 mg, 0.13 mmol), HOBt (17.6 mg, 0.13 mmol), and DIPEA (0.038 mL, 0.22 mmol) were added. The reaction was stirred at room temperature for 4 h and monitored by TLC. After that, the reaction was stopped by concentration under reduced pressure. The crude product was worked up by redissolving in ethyl acetate and washing with 0.5% HCl, 2.5% NaHCO$_3$, and brine, respectively, to remove the by-products. The organic layers were combined and dried over anhydrous sodium sulfate. Then, the organic filtrate was dried under reduced pressure to obtain the crude product. The synthesized compounds were purified by column chromatography with silica gel as the stationary phase and the mixture of ethyl acetate and hexane as the mobile phase. Ester 4u was obtained as a white amorphous solid at 7.0 mg (13%). ^{1}H-NMR (CDCl$_3$, 400 MHz) δ 7.88 (1H, d, *J* = 16.0 Hz, 11-H), 7.88 (1H, d, *J* = 8.5 Hz, 16-H), 7.68 (1H, s, 19-H), 7.66 (1H, d, *J* = 8.5 Hz, 17-H), 7.53 (1H, dd, *J* = 8.4, 2.0 Hz, 7-H), 7.48 (1H, d, *J* = 2.0 Hz, 5-H), 7.31 (1H, d, *J* = 8.4 Hz, 8-H), 6.60 (1H, d, *J* = 16.0 Hz, 12-H), 2.35 (3H, s, 1-H), 2.34 (3H, s, 10-H) ppm; ^{13}C-NMR (CDCl$_3$, 100 MHz) δ 168.1 (C-9), 168.0 (C-2), 163.8 (C-13), 148.6 (q, *J* = 1.0 Hz, C-14), 146.4 (C-11), 144.3 (C-3), 142.6 (C-4), 135.2 (q, *J* = 34.0 Hz, C-15), 132.5 (C-6), 127.8 (q, *J* = 5.0 Hz, C-16), 126.9 (C-7), 126.3 (q, *J* = 33.0 Hz, C-18), 124.2 (C-8), 123.2 (C-5), 122.8 (q, *J* = 5.0 Hz, C-17), 121.9 (q, *J* = 4.0 Hz, C-19), 116.8 (C-12), 124.2 & 123.6 & 121.4 & 120.9 (*J* = 330.0 Hz, C-15a &

C18a), 20.7 (C-1), 20.6 (C-10) ppm; IR (ATR, $\mathbf{V}_{max}$) 2921, 2851, 1753, 1212, 1176, 1125, 873, 835 cm^{-1}; HRMS–ESI m/z 499.0591 ([M+Na]$^+$, calcd for $C_{21}H_{14}F_6NaO_6$ 499.0587).

      Synthesis of 3,4-diacetyloxy cinnamic acid-3,4-bis(trifluoromethyl) phenyl ester (4v). Caffeic acid (30 mg, 0.11 mmol) and 3,4-bis(trifluoromethyl)phenol (25.3 mg, 0.11 mmol) were employed under the same esterification protocol as described above. The ester 4v was obtained as a white amorphous solid at 7.5 mg (14%). ^{1}H-NMR (CDCl$_3$, 400 MHz) δ 7.93 (1H, d, J = 8.7 Hz, 18-H), 7.88 (1H, d, J = 16.0 Hz, 11-H), 7.71 (1H, d, J = 2.1 Hz, 15-H), 7.57 (1H, dd, J = 8.7, 2.1 Hz, 19-H), 7.51 (1H, dd, J = 8.4, 2.0 Hz, 7-H), 7.47 (1H, d, J = 2.0 Hz, 5-H), 7.31 (1H, d, J = 8.4 Hz, 8-H), 6.59 (1H, d, J = 16.0 Hz, 12-H), 2.35 (3H, s, 1-H), 2.35 (3H, s, 10-H) ppm; ^{13}C-NMR (CDCl$_3$, 100 MHz) δ 168.1 (C-9), 168.0 (C-2), 163.9 (C-13), 153.1 (C-14), 146.2 (C-11), 144.3 (C-3), 142.6 (C-4), 132.5 (C-6), 130.1 (dd, J = 30.0, 1.0 Hz, C-17), 129.6 (q, J = 6.0 Hz, C-18), 126.9 (C-7), 125.5 (dd, J = 32.0, 1.5 Hz, C-16), 125.1 (C-19), 124.2 (C-8), 128.9 & 123.5 & 121.3 & 120.8 (J = 307.0 Hz, C-16a & C17a), 123.2 (C-5), 121.8 (q, J = 7.0 Hz, C-15), 117.1 (C-12), 20.7 (C-1), 20.6 (C-10) ppm; IR (KBr) 2918, 2849, 1763, 1275, 1181, 1141, 903, 841 cm^{-1}; HRMS–ESI m/z 499.0595 ([M+Na]$^+$, calcd for $C_{21}H_{14}F_6NaO_6$ 499.0587).

4. Evaluation of caffeic acid derivatives for SARS-CoV-2 3CLpro inhibitory activity

      In order to evaluate the enzymatic activity of SARS-CoV-2 3CLpro using an enzyme inhibition assay, the enzyme was expressed, purified, and stored following similar procedures described for SARS-CoV-1 3CLpro [132]. A concentration of 0.2 μM of 3CLpro was used in each experiment. The enzymatic activity was monitored by measuring the initial rate of cleavage of the fluorogenic substrate E(EDANS)TSAVLQSGFRK(DABCYL) (25 μM). The excitation and emission wavelengths were recorded at 340 and 490 nm, respectively. To assess inhibitory activity, the enzymatic activity was examined in the presence of 100 μM of caffeic acid compounds, with a group of control that did not include any inhibitor. The initial rate of the enzymatic reaction in the absence of inhibitors was used as a normalization factor, allowing for the relative assessment of inhibitory effects induced by the caffeic acid compounds.
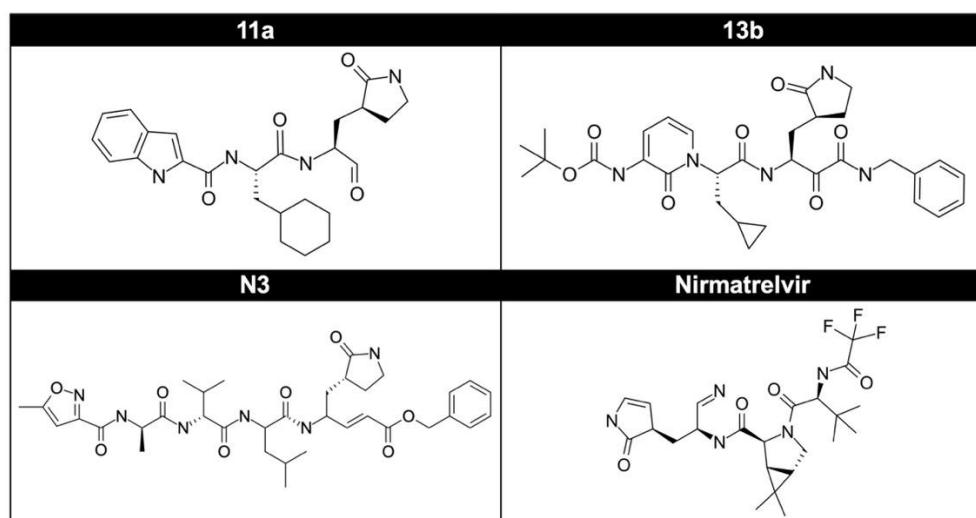
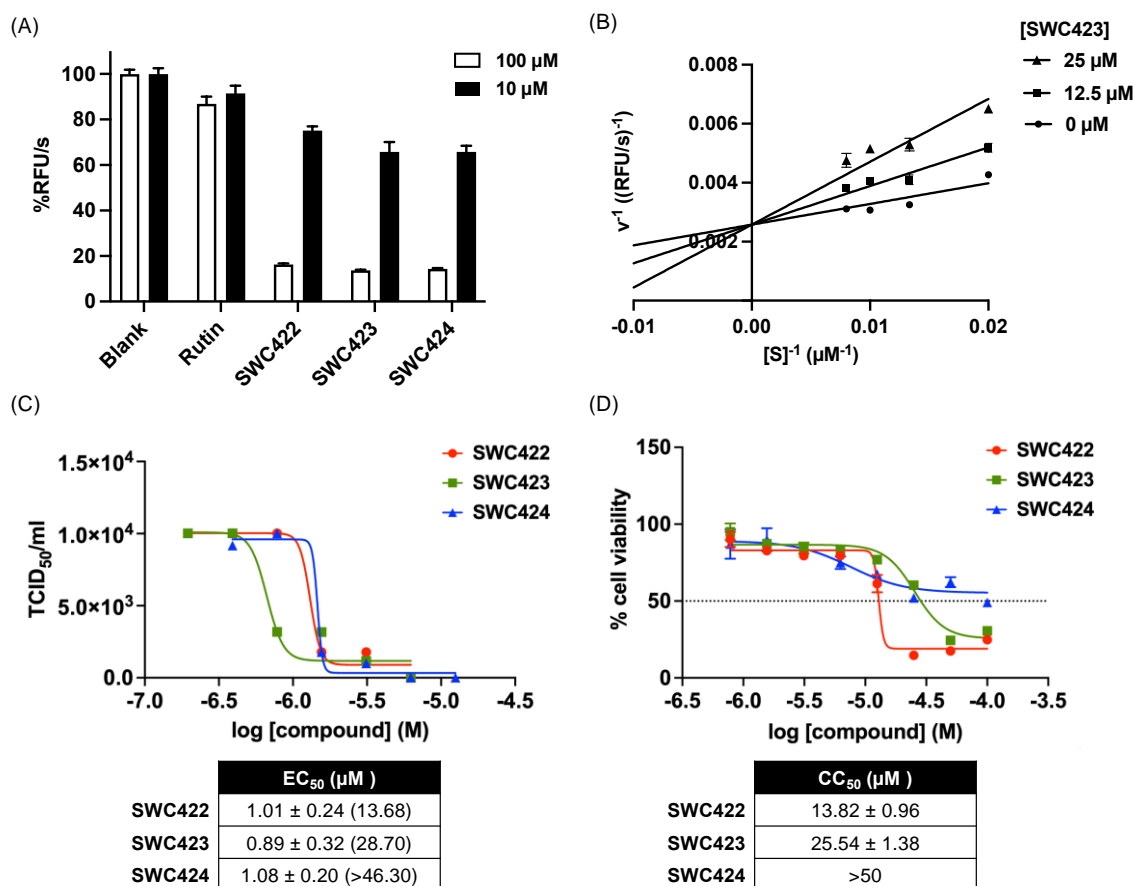Figure 13 Chemical structure of SARS-CoV-2 3CLpro peptidomimetic inhibitors

Figure 14 *In vitro* enzymatic and cell-based assays. (A) Relative activity of SARS-CoV-2 3CLpro in the presence of 100 μM and 10 μM of three sulfonamide chalcones, blank, and rutin. (B) Lineweaver-Burk plot demonstrating competitive inhibition of SWC423 inhibitor with the fluorogenic peptide substrate. The K_i value of SWC423 is 10.0 ± 1.8 μM. (C) Dose-response curves for reduction of SARS-CoV-2 titers. (D) Dose-response curves for cell viability of three sulfonamide chalcones in Vero E6 cells. Data in the table are presented as mean ± SEM (Selectivity Index).
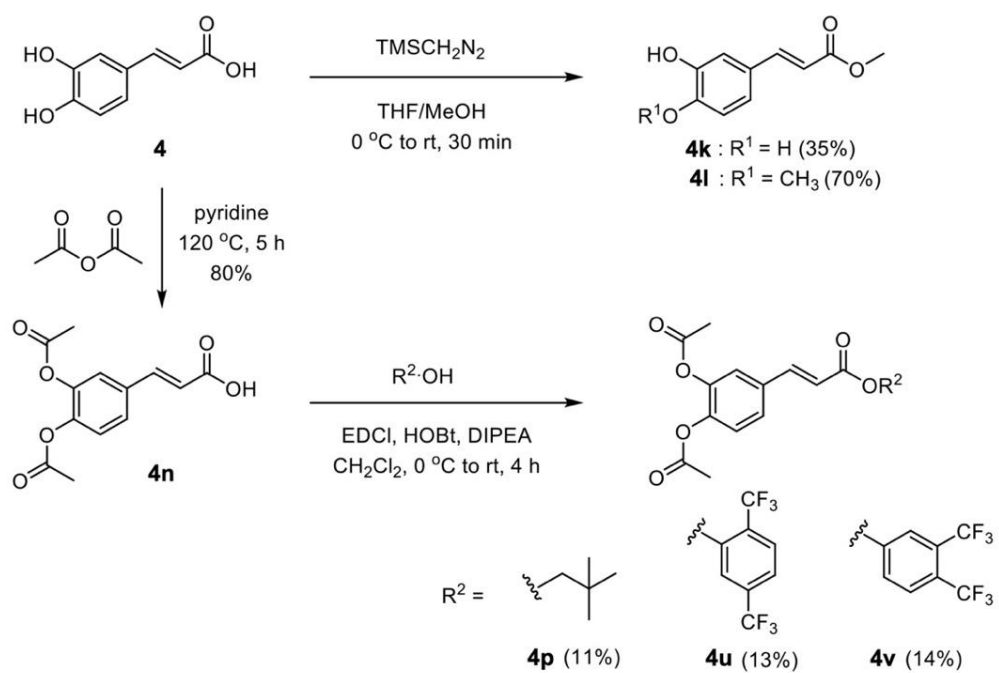
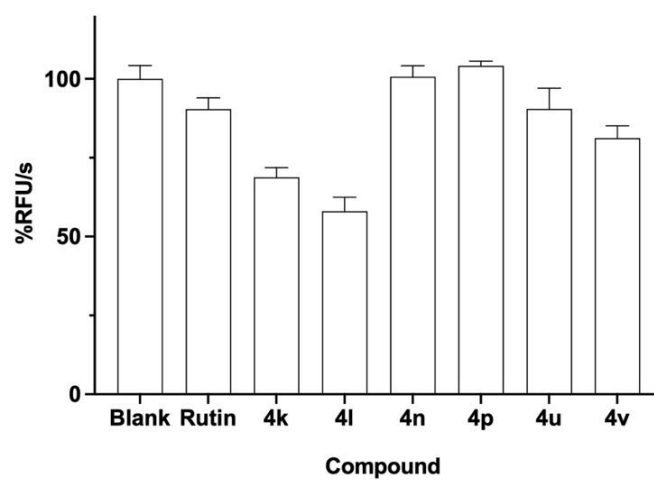Figure 15 Synthesis of six ester derivatives of caffeic acid

Figure 16 Relative activity of SARS-CoV-2 3CLpro with ester derivatives of caffeic acid and rutin at 100 μM concentration and a blank

VITA

NAME	Piyatida Pojtanadithee
DATE OF BIRTH	09 April 1998
PLACE OF BIRTH	Bangkok, Thailand
HOME ADDRESS	40/274 Village No.1, Nanthaya Village, Lumlukka-Wangnoi Rd, Khlong Chet, Khlong Luang, Pathum-Thani, 12120, Thailand