



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

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



Insect protease inhibitors; promising inhibitory compounds against SARS-CoV-2 main protease

Seyed Ali Hemmati^{*}, Saeid Tabein

Department of Plant Protection, Faculty of Agriculture, Shahid Chamran University of Ahvaz, Ahvaz, Iran

ARTICLE INFO

Keywords:

SARS-CoV-2
Main protease (M^{pro})
Peptide inhibitor
Docking
Molecular dynamic simulation

ABSTRACT

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of coronavirus disease 2019 (COVID-19), has adversely affected global health since its emergence in 2019. The lack of effective treatments prompted worldwide efforts to immediately develop therapeutic strategies against COVID-19. The main protease (M^{pro}) of SARS-CoV-2 plays a crucial role in viral replication, and therefore it serves as an attractive target for COVID-19-specific drug development. Due to the richness and diversity of insect protease inhibitors, we docked SARS-CoV-2 M^{pro} onto 25 publicly accessible insect-derived protease inhibitors using the ClusPro server, and the regions with high inhibitory potentials against M^{pro} were used to design peptides. Interactions of these inhibitory peptides with M^{pro} were further assessed by two directed docking programs, AutoDock and Haddock. AutoDock analysis predicted the highest binding energy (−9.39 kcal/mol) and the lowest inhibition constant (130 nM) for the peptide 1KJ0-7 derived from SGCI (*Schistocerca gregaria* chymotrypsin inhibitor). On the other hand, Haddock analysis resulted in the discovery of a different peptide designated 2ERW-9 from *infestans*, a serine protease inhibitor of *Triatoma infestans*, with the best docking score (−131), binding energy (−11.7 kcal/mol), and dissociation constant (2.6E-09 M) for M^{pro}. Furthermore, using molecular dynamic simulations, 1KJ0-7 and 2ERW-9 were demonstrated to form stable complexes with M^{pro}. The peptides also showed suitable drug-likeness properties compared to commercially available drugs based on Lipinski's rule. Our findings present two peptides with possible protease inhibitor activities against M^{pro} and further demonstrate the potential of insect-derived peptides and computer-aided methods for drug discovery.

1. Introduction

Coronaviruses are large, enveloped, single-stranded RNA viruses with the ability to infect a wide range of animal species, including, but not limited to, bats and humans [1]. Human coronaviruses (HCoVs) were previously known to cause mild respiratory infections [2]. However, two new coronaviruses, designated as severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV), capable of causing fatal human respiratory infections, emerged in 2002 and 2012, respectively [3,4]. In late December 2019, a cluster of cases of pneumonia was reported from Wuhan, Hubei Province of China [5]. The etiologic agent responsible for pneumonia cases was identified as the 2019 novel coronavirus [6], which was later renamed as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [7]. The coronavirus disease 2019 (COVID-19) rapidly spread throughout the world, and it was officially declared as a global pandemic by the World Health Organization in March 2020.

Although patients with COVID-19 mostly manifest mild symptoms, the disease may progress to severe symptoms, including pneumonia and respiratory failure leading to the patient's death [8]. As of December 2021, more than 270 million coronavirus cases have been reported globally, and around 5.3 million have died [9]. Though several vaccines have been approved across the globe over the past months [10], there are currently no COVID-19-specific treatments available. The protective ability of the vaccines is also challenged [11] by new emerging viral clones, thus arguing for an urgent need to develop effective therapeutic strategies against the virus.

Like other coronaviruses, SARS-CoV-2 contains a positive-sense single-stranded RNA genome that resembles eukaryotic messenger RNA and is directly translated by host cell translation machinery to produce two overlapping polyproteins, pp1a and pp1ab [12]. These two polyproteins are processed into 16 non-structural proteins (nsp1-nsp16) to form the viral replicase-transcriptase complex, where new viral RNAs are synthesized [12]. The cleavage of polyproteins into individual

^{*} Corresponding author. Department of Plant Protection, Faculty of Agriculture, Shahid Chamran University of Ahvaz, P.O. Box 61357-43311, Ahvaz, Iran.
E-mail address: sa.hemmati@scu.ac.ir (S.A. Hemmati).

non-structural proteins, a critical step in viral replication, is primarily mediated by the virus main protease (Mpro, also called 3CLpro) [13]. Therefore, inhibiting the activity of Mpro would block viral replication, making M^{pro} an attractive target for drug development against SARS-CoV-2. Protease inhibitors have successfully been developed for the treatment of several viral infections such as human immunodeficiency virus (HIV) [14] and hepatitis C virus (HCV) [15] infections.

The discovery and development of novel drugs prove to be a costly and time-consuming process [16]. Using computer-aided drug discovery (CADD) approaches, which apply computational software and chemistry simulation techniques to identify novel hits or leads, helps accelerate the preliminary stage of drug discovery and minimize failures in the final stage [16]. One of the most frequently used CADD methods is molecular docking [17], which can predict the interaction of a target protein (e.g., Mpro) against a large library of chemical structures.

Natural compounds have always been rich sources for discovering new drugs [18]. Arthropods (insects, arachnids, myriapods, and crustaceans) make up the largest division of the animal kingdom, representing approximately 80% of all known animals [19]. Insects produce a wide range of protease inhibitors [20], which have been neglected for drug development in comparison with other natural sources, including plants, fungi, and microorganisms [21]. Due to being relatively unexplored, insect-derived protease inhibitors may provide opportunities to identify novel drug candidates against SARS-CoV-2 Mpro.

Considering the functional importance of Mpro for replication of SARS-CoV-2, we first employed bioinformatics tools to analyze the structure of the protein. Due to the richness and diversity of insect protease inhibitors, a blind docking approach was employed to screen for insect-derived compounds with potential inhibitory properties against Mpro. Based on peptide coverage of inhibitors with the catalytic domain of M^{pro}, multiple inhibitors were designed, and the binding affinities of the designed peptides towards Mpro were characterized using molecular docking and molecular dynamics simulations. The *in-silico* approach adopted in this study enabled the discovery of novel drug candidates with potential inhibitory effects against Mpro, mainly targeting the enzyme's active site. Our findings suggest that domain-specific M^{pro} inhibitory peptides may prove to be a new generation of drugs to be used against SARS-CoV-2.

2. Material and methods

2.1. Data retrieval and *in silico* analysis of M^{pro}

The amino acid sequence (Uniport code: P0DTD1) and the three-dimensional (3D) structure (PDB ID: 6LU7) [22] of SARS-CoV-2 M^{pro} were retrieved from the universal protein resource (Uniprot) database (www.uniprot.org/) and the protein data bank (PDB) archive (<https://www.rcsb.org/>), respectively. We estimated various physicochemical properties of M^{pro}, including protein length, molecular weight, the total number of negatively and positively charged residues, theoretical isoelectric point, instability index, aliphatic index, and grand average of hydropathicity index using ExPASy ProtParam tools (<http://web.expasy.org/protparam>) [23]. The secondary structures of M^{pro} were predicted using the self-optimized prediction method with alignment (SOPMA) (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html) [24]. Transmembrane topology prediction was performed using the transmembrane hidden Markov model (TMHMM) (<http://www.cbs.dtu.dk/services/TMHMM>, <https://services.healthtech.dtu.dk/service.php?TMHMM-2.0>) [25]. We identified the amino acid residues of ligand-binding and active sites of M^{pro} (as the potential binding site for inhibitors) by the computed atlas for surface topography of proteins (CASTp) (<http://sts.bioe.uic.edu/castp/>) to ensure that all the available residues were involved in our research [26].

2.2. Data retrieval and *in silico* analysis of insect-derived protease inhibitors

The pro-region of proteases is required for the proper folding of the protease domain and can also function as a potent inhibitor of the mature enzyme [27–29]. Amino acid sequences of multiple pro-regions and other naturally occurring protease inhibitors from different insect species were obtained from the national center for biotechnology information (NCBI) (<http://www.ncbi.nlm.nih.gov/>) and Uniprot databases. The 3D structures of these inhibitors obtained from the PDB archive. Additionally, the 3D structures of selected peptides were predicted by iterative threading assembly refinement (I-TASSER) (<https://zhanglab.cmb.med.umich.edu/I-TASSER/>, <https://zhanggroup.org/I-TASSER/>) [30] and PEP-FOLD3 (<https://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD3/>) [31].

2.3. Blind docking and peptide design

To investigate the interaction of insect-derived inhibitors with SARS-CoV-2 M^{pro}, we performed a blind docking between the inhibitors and the 3D structure of SARS-CoV-2 M^{pro} using the ClusPro server (<https://cluspro.org>) without changing the program default settings [32]. The ClusPro server combines conformational sampling, root mean square deviation (RMSD)-based clustering of the predicted protein-protein complexes, and energy refinement to generate a list of near-native structures [32]. These top-ranked near-native structures were run through the WHAT-IF server (<https://swift.cmbi.umcn.nl/servers/html/index.html>) [33] to refine the inhibitory peptide design. For this purpose, the change in accessible surface area (Δ ASA) of free and protein-bound ligands was calculated for the aforementioned top-ranked inhibitors in complex with residues in the active site of M^{pro}. The regions within inhibitors that showed higher Δ ASA values were considered to be actively involved in ligand-protein interaction and therefore exerted an inhibitory effect on M^{pro}.

2.4. Directed docking

This work used Merck molecular force field (MMFF94) in Ligand optimization, Avogadro version 1.2. PDBQT format file for the target molecule (protomer A of M^{pro}) was prepared using AutoDock 4.2 by removing water molecules and adding polar hydrogen atoms. A grid box of $126 \times 126 \times 126$ size with 0.375 Å spacing was centred to the active site of M^{pro}. The PDBQT files were generated for ligands (inhibitors) after detecting the torsion root. The optimal binding mode between inhibitors and the binding site of M^{pro} was determined by 100 runs of the genetic algorithm. Further docking analyses were carried out by both AutoDock 4.2 and Biovia Discovery Studio 4.5.

Additionally, peptide inhibitors were docked to the 3D structure of M^{pro} using an experimentally-based docking program called HADDOCK (high ambiguity driven protein-protein docking) (<https://milou.science.uu.nl/services/HADDOCK2.2/haddock.php>) [34] without changing the program default settings. To launch the docking process, two ambiguous interaction restraints (AIRs) were defined: (1) all residues of the ligands (or inhibitors) were classified as passive residues, (2) all residues within the active site of M^{pro} were classified active. The standard HADDOCK protocol generated 1000 complexes at the rigid body minimization stage. The best 200 structures out of these 1000 complexes were subjected to a semi-flexible simulated annealing and final energy minimization. The top-ranked complexes obtained from the docking process were clustered on the basis of RMSD and ranked by average HADDOCK scores. Furthermore, PRODIGY (protein binding energy prediction) webserver (<https://bianca.science.uu.nl/prodigy/>) [35] was used to calculate the binding energy (Δ G) and dissociation constant (K_d) at 25 °C, as indicators of binding affinity of designed peptides to M^{pro}.

Table 1

Summary of primary structure analysis and secondary structure prediction for SARS-CoV-2 M^{PRO} (6LU7).

| Tool | Parameter | Value | |
|-------------------|---|---------------------|-------|
| ProtParam | Number of amino acids (AA) | 306 | |
| | Molecular weight (Mw) | 33796.64 | |
| | Theoretical isoelectric point (pI) | 5.95 | |
| | Total number of negatively charged residues (Asp + Glu) | 26 | |
| | Total number of positively charged residues (Arg + Lys) | 22 | |
| | Instability index | 27.65 | |
| | Aliphatic index | 82.12 | |
| | GRAVY ^a | -0.019 | |
| | SOPMA | α -helix (%) | 29.08 |
| | | β -strand (%) | 27.12 |
| β -turn (%) | | 11.44 | |
| Random coil (%) | | 32.35 | |
| TMHMM | Number of predicted TMHs ^b | 0 | |
| | Expected number of AAs in TMHs ^c | 0.88093 | |
| | Expected number of helices, first 60 AAs ^d | 0.01239 | |
| | Total probability of N-in ^e | 0.01343 | |

^a Grand average of hydropathicity index.

^b The number of predicted transmembrane helices.

^c The expected number of amino acids in transmembrane helices.

^d The expected number of amino acids in transmembrane helices in the first 60 amino acids of the protein.

^e The total probability that the N-terminal end is on the cytoplasmic side of the membrane.

2.5. Molecular dynamic simulations

The molecular dynamic (MD) simulation was carried out to analyze the dynamic interactions of screened peptides in complex with M^{PRO}. MD simulations were performed using the GROMACS simulation package version 5.1.4 within the gromos 54a7 force field [36]. Simulations were run using an Intel Core i7 Processor Extreme Edition on CentOS Linux 6.8 with graphics processing unit acceleration by NVIDIA GeForce GTX 970. We applied three similar MD simulations to refine the structure of M^{PRO} in apo and inhibitor-bound states. In each simulation, the initial structure was placed in the center of a cubic box and solvated by the random distribution of water molecules in an extended single-point charge (SPC/E) model followed by adding counter ions to reach a neutral system. The system was first subjected to energy minimization using the steepest descent energy minimization for 50 000 steps. The energy minimization step was followed by a pre-equilibration simulation for 500 ps in the NVT ensemble with a time constant of 0.1 ps. Next, the NVT equilibrium simulation was performed with the Berendsen thermostat for temperature control (300 K) [37]. Then, each system underwent a 500-ps run in an NPT ensemble, which used the Parrinello-Rahman barostat at 1 bar with the coupling constant set at 0.2 ps [38]. The linear constraint solver (LINCS) algorithm was utilized to constrain bonds during simulation [39]. The periodic boundary condition (PBC) was applied in x, y, and z directions to minimize the 'edge effects'. The Lennard-Jones (LJ) potential with a cut-off radius of 1.4 nm was used for the short-range van der Waals interactions. The particle-mesh Ewald (PME) algorithm was used to calculate long-range electrostatic interactions of Coulomb potential energies with the real space contribution to the Columbic interactions truncated at 0.9 nm applied to the system [40]. The initial velocity of particles was assigned according to Maxwell distributions. Finally, 10 ns MD simulation was produced to remove the structural clashes of the lone protein besides the 100 ns MD simulation for the peptides/M^{PRO} complexes to examine the binding phenomena. All MD simulations were carried out when RMSD values reached a plateau.

2.6. Pharmacokinetic, biochemical properties and Lipinski's rule of five

Lipinski's rule of five was used to evaluate drug-likeness of the designed peptides. Moreover, pharmacokinetic properties of peptides,

including absorption, distribution, metabolism, excretion and toxicity (ADMET) profiling of peptides, were determined using the admetSAR [41] and ProTox web tools [42]. Ritonavir and Lopinavir, as two FDA-approved protease inhibitors, were used as reference compounds.

3. Results

3.1. Primary and secondary structural analysis of M^{PRO}

M^{PRO} plays an essential role in SARS-CoV-2 replication [13] and is, therefore, an attractive target for drug development against COVID-19. In February 2020, the crystal structure of SARS-CoV-2 M^{PRO} in complex with an inhibitor (N3: N-[(5-Methylisoxazol-3-Yl)Carbonyl]Alanyl-L-Valyl-N1~((1r,2z)-4-(Benzyloxy)-4-Oxo-1-[[[3r]-2-Oxopyrrolidin-3-Yl]Methyl]But-2-Enyl)-L-Leucinamide) was made publicly available by Jin et al. [22], which was retrieved to perform primary and secondary structure analysis for M^{PRO} using ExPASy ProtParam, SOPMA, and TMHMM. The results of sequence analyses and secondary structure prediction are summarized in Table 1. The theoretical isoelectric point (pI) of M^{PRO} was calculated to be 5.95, which indicates to acidic nature of the protein (Table 1). Since proteins with instability indices smaller than 40 are predicted to be stable [43], ExPASy ProtParam classified M^{PRO} (with the instability index of 27.65) as a stable protein. The relatively high aliphatic index of M^{PRO} (82.12) points to the considerable thermostability of the protein, and the negative GRAVY value (-0.019) indicates to hydrophilic nature of the protein and its better interaction with water (Table 1). The predicted secondary structure of M^{PRO} was composed of 32.35% random coils, 29.08% α -helices, 27.12% β -strands, and 11.44% β -turns. The absence of any predicted transmembrane domain verifies that the enzyme is a cytoplasmic (not a membrane-bound) protein. Furthermore, amino acids in the binding and active sites of SARS-CoV-2 M^{PRO} (6LU7) determined using the CASTp server were as follows: T24, T25, T26, L27, H41, C44, T45, S46, M49, P52, Y54, F140, L141, N142, G143, S144, C145, H163, H164, M165, E166, L167, P168, H172, D187, R188, Q189, T190, Q192 (Supplementary Fig. 1).

3.2. ClusPro analysis and peptide design

Results of primary docking of 25 insect-derived protease inhibitors to SARS-CoV-2 M^{PRO} using ClusPro web server are presented in Table 2. ClusPro docking of protease inhibitors with M^{PRO} resulted in various clusters, and most of the protease inhibitors were demonstrated to interact with M^{PRO} through multiple regions. Inhibitor/M^{PRO} complexes were ranked based on the lowest binding energy and the cluster size (the number of members in clusters). The top-ranked cluster (cluster "0") was selected for further analysis. Within cluster 0, seven insect protease inhibitors, including 1CCV, 1KMA, 2OZQ, 2XXT, 3SSB, 2M5X, and 2ERW were predicted to have the lowest binding energy, which suggests the great potential of these peptides to inhibit SARS-CoV-2 M^{PRO}. Furthermore, selected regions within these 25 insect-derived protease inhibitors were predicted to be involved in binding to the active site of M^{PRO} based on changes in accessible surface areas (Δ ASA) of residues, which resulted in the generation of 60 peptide inhibitors with potential inhibitory effect against SARS-CoV-2 M^{PRO}. Finally, structural models of these 60 designed peptides were constructed using PEP-FOLD3 for peptide-protein docking.

3.3. AutoDock analysis

The molecular docking results of the 60 designed peptides with inhibitory properties against M^{PRO} are listed in Table 3. Peptide molecules were ranked based on the binding energy and inhibition constant. Most of the designed peptides showed the ability to bind to the enzyme's active site (i.e., M^{PRO}). Two peptides, 1KJ0-7 (RKGCPH) from the desert locust (*Schistocerca gregaria* Forsskål) and 3OZQ-6 (ATYIPV) from the

Table 2
Blind docking of insect-derived protease inhibitors to M^{Pro} (6LU7) using ClusPro server.

| No | Inhibitor PDB ID | Inhibitor length (AA) | Inhibitor characteristics | Insect species | ClusPro energy docking | | | | Selected region as peptide inhibitor | Designed peptide sequence | Δ ASA of peptide on M ^{Pro} |
|----|------------------|-----------------------|--|--------------------------------|------------------------|---------|----------------|---------------|--------------------------------------|---------------------------|---|
| | | | | | Cluster | Members | Central energy | Lowest energy | | | |
| 1 | 1CCV | 56 | Chymotrypsin inhibitor (AMCI) | <i>Apis mellifera</i> | 0 | 284 | -823.9 | -981.0 | 25-34 | TRICTMQCRI | 185 |
| | | | | | | | | | 26-34 | RICTMQCRI | 182 |
| | | | | | | | | | 28-34 | CTMQCRI | 150 |
| | | | | | | | | | 30-34 | MQCRI | 121 |
| 2 | 1GL1 | 36 | Protease inhibitor LCMI II | <i>Locusta migratoria</i> | 0 | 211 | -552.4 | -733.0 | 9-13 | FKDKC | 69 |
| | | | | | | | | | 27-34 | CTLKACPN | 151 |
| 3 | 1KGM | 35 | Serine protease inhibitor (SGCI) | <i>Schistocerca gregaria</i> | 0 | 182 | -624.4 | -771.6 | 1-6 | EVTCEP | 85 |
| | | | | | | | | | 10-13 | FKDK | 53 |
| | | | | | | | | | 30-35 | LKACPQ | 94 |
| 4 | 1KIO | 35 | Serine protease inhibitor (SGCI [L30R, K31 M]) | <i>Schistocerca gregaria</i> | 0 | 164 | -563.9 | -750.2 | 1-2 | EV | 29 |
| | | | | | | | | | 11-14 | KDKC | 33 |
| 5 | 1KJ0 | 35 | Serine protease inhibitor (SGTI) | <i>Schistocerca gregaria</i> | 0 | 123 | -573.3 | -646.7 | 30-35 | RMACPQ | 177 |
| | | | | | | | | | 11-15 | KQDCN | 75 |
| 6 | 1KMA | 55 | Thrombin inhibitor (Dipetalin) | <i>Dipetalogaster maximus</i> | 0 | 74 | -827.6 | -827.6 | 29-35 | RKGCPPH | 152 |
| | | | | | | | | | 7-14 | ECPPRALHR | 182 |
| | | | | | | | | | 10-14 | RALHR | 141 |
| 7 | 1WO9 | 35 | Trypsin inhibitor (HI) | <i>Locusta migratoria</i> | 0 | 99 | -583.8 | -647.5 | 51-55 | HDHDF | 93 |
| | | | | | | | | | 10-16 | KKQDCNT | 88 |
| | | | | | | | | | 29-35 | RKACRTT | 147 |
| 8 | 2ERW | 53 | Serine protease inhibitor infestin | <i>Triatoma infestans</i> | 0 | 158 | -647.5 | -787.2 | 1-9 | NPCACFRNY | 245 |
| | | | | | | | | | 6-10 | FRNYV | 178 |
| 9 | 2KSW | 66 | Oryctin | <i>Oryctes rhinoceros</i> | 0 | 73 | -699.9 | -767.5 | 8-18 | EPKLCITMDLVP | 200 |
| | | | | | | | | | 11-16 | LCTMDL | 164 |
| | | | | | | | | | 32-38 | HGGALS | 42 |
| 10 | 2M5X | 40 | Silk protease inhibitor 2 (GmSPI-2) | <i>Galleria mellonella</i> | 0 | 116 | -714.1 | -787.5 | 1-9 | EAAVCTTEW | 145 |
| | | | | | | | | | 16-22 | DGKTRSN | 52 |
| 11 | 2VU8 | 33 | Protease Inhibitor 3 | <i>Locusta migratoria</i> | 0 | 117 | -547.5 | -630.0 | 37-40 | GECL | 37.69 |
| | | | | | | | | | 10-13 | QDCN | 37.59 |
| 12 | 2XTT | 36 | Protease inhibitor SGPI-1 | <i>Schistocerca gregaria</i> | 0 | 107 | -678.4 | -795.3 | 27-33 | RKACRTT | 181.34 |
| | | | | | | | | | 1-4 | QECE | 57.21 |
| | | | | | | | | | 9-14 | KKQDCN | 62.90 |
| 13 | 3BT4 | 85 | Fungal protease inhibitor-1 | <i>Antheraea mylitta</i> | 0 | 121 | -700.5 | -755.2 | 28-35 | RMGCPPHA | 133.77 |
| | | | | | | | | | 20-32 | RASCRSPATYRAN | 220.70 |
| | | | | | | | | | 24-30 | RSPATYYR | 149.72 |
| | | | | | | | | | 24-32 | RSPATYRAN | 160.43 |
| 14 | 3OZQ | 364 | Serpins48 | <i>Tenebrio molitor</i> | 0 | 174 | -665.7 | -823.1 | 43-49 | CVTLRE | 93.46 |
| | | | | | | | | | 178-182 | PFHTR | 44.86 |
| | | | | | | | | | 328-333 | ATYIPV | 162.97 |
| 15 | 3SSB | 40 | Metalloproteinase inhibitor protein | <i>Galleria mellonella</i> | 0 | 91 | -679.4 | -789.0 | 1-4 | LICN | 119.05 |
| | | | | | | | | | 8-15 | EYECGGA | 115.79 |
| 16 | 4POF | 393 | Serine protease inhibitor 4 | <i>Drosophila melanogaster</i> | 0 | 73 | -627.7 | -782.0 | 1-4 | AAHQ | 71.06 |
| | | | | | | | | | 79-85 | AAHQILR | 100.44 |
| 17 | 4R9I | 378 | cysteine proteinase inhibitor (Serpins18) | <i>Bombyx mori</i> | 0 | 103 | -566.2 | -646.6 | 1-6 | HHHHHM | 114.80 |
| | | | | | | | | | | | |
| 18 | 5C98 | 382 | Serine protease inhibitor (SRPN18) | <i>Anopheles gambiae</i> | 0 | 71 | -741.2 | -741.2 | 198-205 | TAFVRRCL | 79.65 |
| | | | | | | | | | 254-260 | ERLQSCW | 97.41 |
| | | | | | | | | | 321-324 | SSEF | 44.11 |
| 19 | 5DAE | 65 | Kazal-type serine protease inhibitor | <i>Aedes aegypti</i> | 0 | 214 | -757.1 | -757.1 | 2-11 | VCACPRIYMP | 223.55 |
| | | | | | | | | | 21-25 | NNDCL | 33.63 |
| | | | | | | | | | 21-29 | NNDCLRCE | 40.85 |
| 20 | 6CJ7 | 390 | Serine protease inhibitor (Serpins)-12 | <i>Manduca sexta</i> | 0 | 99 | -609.8 | -760.7 | 88-94 | LSENFNL | 87.77 |
| | | | | | | | | | 111-117 | TPTYFGK | 84.59 |
| 21 | 1GL0 | 35 | Protease inhibitor LCMI I | <i>Locusta migratoria</i> | 0 | 194 | -579.0 | -731.9 | 9-14 | QQDCNT | 38.56 |
| | | | | | | | | | 26-32 | LMGCQPT | 158.15 |
| | | | | | | | | | 25-32 | CTLMGCCQP | 167.97 |
| 22 | PRO42 | 42 | Chymotrypsinogen | <i>Helicoverpa armigera</i> | 0 | 109 | -762.7 | -762.7 | 21-25 | TKFGI | 146.75 |
| | | | | | | | | | | | |
| 23 | PRO-S7 | 7 | Trypsinogen | <i>Spodoptera frugiperda</i> | 0 | 1000 | -545.6 | -545.6 | 1-7 | VPSNPQR | 192.54 |
| | | | | | | | | | | | |
| 24 | PRO13 | 13 | Trypsinogen | <i>Plodia interpunctella</i> | 0 | 158 | -411.9 | -522.8 | 1-6 | AEVPSD | 142.23 |
| | | | | | | | | | 8-13 | YPNAQR | 65.60 |
| | | | | | | | | | 1-13 | AEVPSDPYPNAAQR | 207.94 |
| 25 | PRO-P7 | 7 | Trypsinogen | <i>Plutella xylostella</i> | 0 | 364 | -414.7 | -530.4 | 1-7 | VPKNPQR | 240.98 |
| | | | | | | | | | | | |

mealworm beetle (*Tenebrio molitor* L.), respectively, showed the best predicted binding energy of -9.39 and -9.26 kcal/mol, inhibition constant of 130.51 and 163.08 nM, and intermolecular energy of -16.55 and -14.93 kcal/mol suggesting the potential inhibitory effect of 1KJ0-7 and 3OZQ-6 against M^{Pro}.

3.4. HADDOCK analysis

To further elucidate the inhibitory effect of the designed peptides against M^{Pro}, docking studies were also carried out with the HADDOCK web server. The 60 designed peptides were docked into the active site of

Table 3Molecular docking (AutoDock) analysis of designed peptide originated from insects with inhibitory effect against M^{PRO}.

| No | Inhibitor (PDB ID) | Peptide sequence | Binding energy (ΔG) (Kcal/mol) | Inhibition constant | Intermolecular energy (Kcal/mol) | VDW-H bond desolvation energy (Kcal/mol) |
|----|--------------------|------------------|--|----------------------|----------------------------------|--|
| 1 | 1CCV | RICTMQCRI | -4.18 | 858.66 μM | -16.12 | -15.69 |
| | | CTMQCRI | -0.83 | 247.18 mM | -9.78 | -8.93 |
| | | MQCRI | -7.52 | 3.10 μM | -14.38 | -13.13 |
| 2 | 1GL1 | FKDKC | -5.48 | 95.81 μM | -12.94 | -10.51 |
| | | CTLKACPQ | -4.62 | 407.36 μM | -12.98 | -11.49 |
| 3 | 1KGM | EVTCEP | -4.05 | 1.07 mM | -10.32 | -9.38 |
| | | FKDK | -6.37 | 21.55 μM | -12.93 | -10.95 |
| | | LKACPQ | -5.48 | 95.77 μM | -12.05 | -11.97 |
| 4 | 1KIO | EV | -4.91 | 252.32 μM | -7.30 | -7.57 |
| | | KDKC | -4.32 | 684.28 μM | -10.58 | -9.16 |
| | | RMACPQ | -6.83 | 9.91 μM | -13.39 | -11.93 |
| 5 | 1KJ0 | KQDCN | -3.98 | 1.20 mM | -10.84 | -9.52 |
| | | RKGCPPH | -9.39 | 130.51 nM | -16.55 | -13.91 |
| 6 | 1KMA | ECPPRALHR | -3.75 | 1.79 mM | -13.00 | -12.07 |
| | | RALHR | -7.00 | 7.40 μM | -13.56 | -12.39 |
| | | HDHDF | -5.29 | 131.50 μM | -11.26 | -11.88 |
| 7 | 1WO9 | KKQDCNT | -2.26 | 22.07 mM | -12.40 | -11.21 |
| | | RKACRTT | -4.69 | 363.48 μM | -14.24 | -11.97 |
| 8 | 2ERW | NPCACFRNY | -3.99 | 1.19 mM | -13.54 | -12.79 |
| | | FRNYV | -7.27 | 4.67 μM | -13.83 | -12.64 |
| 9 | 2KSW | EPKLCMDLVP | -2.34 | 19.37 mM | -14.57 | -14.71 |
| | | LCTMDL | -4.02 | 1.13 mM | -11.18 | -11.12 |
| | | HGGCALS | -5.93 | 45.22 μM | -12.19 | -12.16 |
| 10 | 2M5X | EAAVCTTEW | -2.02 | 32.58 mM | -11.57 | -11.96 |
| | | DGKTRSN | -2.14 | 26.88 mM | -11.09 | -10.93 |
| | | GECL | -5.31 | 128.51 μM | -9.48 | -9.44 |
| 11 | 2VU8 | QDCN | -4.67 | 378.71 μM | -9.44 | -9.62 |
| | | RKACRTT | -2.24 | 22.81 mM | -11.79 | -9.63 |
| 12 | 2XTT | QECE | -3.18 | 4.64 mM | -8.55 | -9.25 |
| | | KKQDCN | -1.49 | 80.75 mM | -10.22 | -9.87 |
| | | RMGCPPHA | -7.87 | 1.70 μM | -15.03 | -14.58 |
| 13 | 3BT4 | RASCRSPATYRAN | -0.36 | 548.83 mM | -14.97 | -13.79 |
| | | RSPATYYR | -4.61 | 419.32 μM | -12.96 | -12.53 |
| | | RSPATYRAN | -1.92 | 39.45 mM | -12.06 | -10.07 |
| | | CVTLLE | -4.81 | 297.83 μM | -13.46 | -12.45 |
| 14 | 3OZQ | PFHTR | -8.96 | 269.07 nM | -14.93 | -14.18 |
| | | ATYIPV | -9.26 | 163.08 nM | -14.93 | -14.56 |
| 15 | 3SSB | LICN | -7.13 | 5.90 μM | -11.61 | -11.63 |
| | | EYIECGGA | -3.82 | 1.58 mM | -12.47 | -12.73 |
| 16 | 4POF | AAHQ | -6.95 | 8.00 μM | -10.83 | -10.69 |
| | | AAQILR | -4.21 | 818.32 μM | -12.56 | -12.31 |
| 17 | 4R9I | HHHHHM | -6.54 | 15.99 μM | -14.00 | -13.94 |
| 18 | 5C98 | ERLQSCW | -2.21 | 24.10 mM | -11.45 | -10.59 |
| 19 | 5DAE | VCACPRIYMP | -4.60 | 421.22 μM | -14.45 | -13.42 |
| | | NNDCL | -4.03 | 1.11 mM | -9.70 | -9.81 |
| | | NNDLLRCE | -1.05 | 170.99 mM | -12.08 | -12.49 |
| | | LSENFNL | -4.83 | 289.02 μM | -13.48 | -13.38 |
| 20 | 6CJ7 | TPTYFGK | -7.08 | 6.42 μM | -15.14 | -13.69 |
| | | QQDCNT | -1.67 | 60.03 mM | -9.12 | -9.30 |
| | | LMGCQPT | -6.77 | 10.94 μM | -12.73 | -12.55 |
| 21 | 1GL0 | CTLMGCQP | -5.24 | 143.08 μM | -13.30 | -13.24 |
| | | TKFGI | -7.41 | 3.72 μM | -13.67 | -11.67 |
| | | VPSNPQR | -7.81 | 1.88 μM | -14.97 | -14.63 |
| 22 | PRO42 | AEVPSD | -3.95 | 1.26 mM | -9.62 | -9.87 |
| 23 | PRO-S7 | YPNAQR | -7.15 | 5.75 μM | -14.01 | -13.16 |
| 24 | PRO13 | VPKNPQR | -7.60 | 2.71 μM | -15.05 | -13.37 |
| 25 | PRO-P7 | | | | | |

M^{PRO} (Table 4), and resulting poses were ranked based on HADDOCK score, cluster size, calculated RMSD, binding energy, and dissociation constant values. Similar to AutoDock results, it predicted most of the designed peptides to bind to the active site of the enzyme (M^{PRO}) by HADDOCK analysis. The peptide 2ERW-9 (NPCACFRNY) from the assassin bug (*Triatoma infestans* Klug) showed the best HADDOCK score of -131.2 with the binding energy of -11.7 kcal/mol and the dissociation constant of 2.6 nM. The buried surface area (BSA) of 2ERW-9 bound to active site residues of the enzyme was calculated to be about 1676 Å².

3.5. Binding of screened peptides to M^{PRO}

SARS-CoV-2 M^{PRO} is a cysteine protease whose active site has an

unusual catalytic dyad formed by C145 and H41 [41]. A catalytic water molecule forms three hydrogen bond interactions with H41, H164, and D187 in the active site of M^{PRO}. A salt bridge interaction between D187 and R40 is important to maintain the architecture of the catalytic cavity. It has been reported that L141, N142, S46, Q189, E166, P168, A191, and T190 in the solvent-exposed region of the M^{PRO} substrate-binding site are involved in trapping of the substrate [44]. The docking procedure was validated using the coordination information of the 6LU7 PDBID of M^{PRO} by manually removing and redocking the peptide-like N3 inhibitor following the same docking procedure used to run HADDOCK and AutoDock. The re-docked complex was then superimposed onto the reference co-crystallized complex using AutoDock tools 1.5.7, and the RMSD value was calculated. ΔRMSD values (the differences between the predicted dock structure to the reference ligand N3 position in the PDB

Table 4Statistics of HADDOCK results for top-ranked cluster of different designed peptides originated from insects with inhibitory effect against M^{Pro}.

| No | Inhibitor (PDB ID) | Peptide sequence | Cluster rank | HADDOCK score | Cluster size | RMSD | Energy (kcal/mol) | | | | Buried surface area (Å ²) | ΔG _{binding} (kcal/mol) | K _d (M) |
|----|--------------------|------------------|--------------|---------------|--------------|------|-------------------|---------------|-------------|---------|---------------------------------------|----------------------------------|--------------------|
| | | | | | | | Van der Waals | Electrostatic | Desolvation | Total | | | |
| 1 | 1CCV | RICTMQCRI | 1 | -92.4 | 51 | 0.4 | -58.00 | -96.16 | -17.6 | -154.17 | 1466.62 | -10.4 | 2.2E-08 |
| | | CTMQCRI | 1 | -76.0 | 75 | 1.2 | -71.18 | -53.22 | -22.9 | -124.41 | 1683.94 | -11.6 | 2.9E-09 |
| | | MQCRI | 4 | -74.3 | 22 | 0.4 | -53.42 | -104.14 | -8.5 | -157.57 | 1229.64 | -8.2 | 1.0E-06 |
| 2 | 1GL1 | FKDKC | 1 | -98.6 | 81 | 0.4 | -41.74 | -245.11 | -17.4 | -286.86 | 1139.57 | -9.1 | 2.0E-07 |
| | | CTLKACPN | 1 | -87.8 | 77 | 0.4 | -54.44 | -79.75 | -18.2 | -134.20 | 1292.04 | -9.8 | 6.4E-08 |
| 3 | 1KGM | EVTCEP | 1 | -78.2 | 101 | 1.2 | -45.94 | -158.26 | -13.3 | -204.21 | 1162.56 | -10.3 | 2.9E-08 |
| | | FKDK | 3 | -90.5 | 19 | 0.5 | -38.92 | -225.86 | -22.1 | -264.79 | 1014.6 | -8.7 | 4.2E-07 |
| | | LKACPQ | 3 | -82.8 | 40 | 0.3 | -46.28 | -142.49 | -10.4 | -188.77 | 1162.81 | -10.3 | 2.9E-08 |
| 4 | 1KIO | EV | 1 | -44.7 | 139 | 0.4 | -25.83 | -135.56 | -8.00 | -161.40 | 625.07 | -6.6 | 1.5E-05 |
| | | KDKC | 6 | -84.9 | 15 | 0.9 | -7.6 | -214.84 | -37.65 | -252.49 | 956.91 | -8.4 | 7.3E-07 |
| | | RMACPQ | 2 | -82.3 | 57 | 0.4 | -59.12 | -117.91 | -6.0 | -177.04 | 1256.34 | -8.8 | 3.3E-07 |
| 5 | 1KJO | KQDCN | 1 | -89.7 | 76 | 0.4 | -45.34 | -231.81 | -6.1 | -277.16 | 104.92 | -7.8 | 1.8E-06 |
| | | RKGCPPH | 1 | -89.7 | 114 | 0.5 | -53.83 | -170.27 | -13.3 | -224.10 | 1224.04 | -9.9 | 5.9E-08 |
| 6 | 1KMA | ECPPRALHR | 1 | -83.3 | 100 | 2.3 | -21.6 | -124.40 | -59.89 | -184.29 | 1386.74 | -11.0 | 8.2E-09 |
| | | RALHR | 2 | -80.2 | 78 | 1.2 | -51.57 | -136.26 | -13.5 | -187.84 | 1174.4 | -8.6 | 5.0E-07 |
| | | HDHDF | 3 | -106.0 | 29 | 0.4 | -55.43 | -98.30 | -35.5 | -153.74 | 1126.78 | -9.0 | 2.5E-07 |
| 7 | 1WO9 | KKQDCNT | 2 | -92.0 | 39 | 0.4 | -60.24 | -195.04 | -1.4 | -255.29 | 1328.25 | -7.9 | 1.6E-06 |
| | | RKACRTT | 2 | -91.0 | 25 | 1.3 | -45.63 | -239.62 | -5.1 | -285.25 | 1009.26 | -7.3 | 4.2E-06 |
| 8 | 2ERW | NPCACFRNY | 2 | -131.2 | 30 | 0.3 | -72.14 | -119.45 | -35.1 | -191.60 | 1676.11 | -11.7 | 2.6E-09 |
| | | FRNYV | 1 | -116.7 | 134 | 0.4 | -64.47 | -115.88 | -33.6 | -180.35 | 1204.51 | -6.2 | 2.7E-05 |
| 9 | 2KSW | EPKLCTMDLVP | 1 | -77.9 | 102 | 0.8 | -35.10 | -186.55 | -9.9 | -221.66 | 1096.04 | -8.8 | 3.3E-07 |
| | | LCTMDL | 6 | -64.7 | 13 | 1.5 | -39.02 | -122.41 | -10.5 | -161.43 | 1188.8 | -10.1 | 4.0E-08 |
| | | HGGCALS | 1 | -80.3 | 118 | 0.5 | -38.40 | -72.26 | -29.9 | -110.67 | 1056.27 | -9.7 | 7.8E-08 |
| 10 | 2M5X | EAAVCTTEW | 2 | -101.8 | 40 | 3.3 | -53.98 | -28.16 | -24.7 | -82.14 | 1138.79 | -9.2 | 1.7E-07 |
| | | DGKTRSN | 1 | -108 | 57 | 0.4 | -44.33 | -255.62 | -14.2 | -299.95 | 1211.56 | -9.3 | 1.4E-07 |
| | | GECL | 1 | -64.4 | 84 | 0.5 | -33.77 | -114.10 | -13.4 | -147.87 | 842.9 | -8.4 | 7.1E-07 |
| 11 | 2VU8 | QDCN | 3 | -64 | 22 | 0.5 | -43.93 | -132.03 | -7.4 | -175.96 | 825.74 | -7.9 | 1.7E-06 |
| | | RKACRTT | 1 | -94.2 | 53 | 0.4 | -69.77 | -126.92 | -12.6 | -196.69 | 1266.87 | -9.6 | 8.8E-08 |
| 12 | 2XIT | QECE | 1 | -75.8 | 86 | 0.5 | -54.33 | -98.88 | -12.4 | -153.22 | 996.64 | -8.8 | 3.7E-07 |
| | | KKQDCN | 3 | -88.7 | 25 | 0.4 | -51.82 | -179.07 | -7.7 | -230.89 | 1051.05 | -8.1 | 1.1E-06 |
| | | RMGCPPHA | 5 | -88 | 11 | 0.5 | -42.40 | -29.25 | -31.1 | -71.65 | 1012.8 | -10.4 | 2.4E-08 |
| 13 | 3BT4 | RASCRSPATYRAN | 3 | -91 | 17 | 1.8 | -70.00 | -167.44 | -5.2 | -237.44 | 1511.99 | -10.1 | 3.7E-08 |
| | | RSPATYR | 1 | -92.5 | 67 | 0.8 | -58.48 | -107.90 | -23.9 | -166.38 | 1308.31 | -9.5 | 1.0E-07 |
| | | RSPATYRAN | 1 | -88.1 | 60 | 0.5 | -63.75 | -95.52 | -14.1 | -159.27 | 1395.55 | -9.0 | 2.7E-07 |

(continued on next page)

Table 4 (continued)

| No | Inhibitor (PDB ID) | Peptide sequence | Cluster rank | HADDOCK score | Cluster size | RMSD | Energy (kcal/mol) | | | | Buried surface area (Å ²) | ΔG _{binding} (kcal/mol) | K _d (M) |
|----|--------------------|------------------|--------------|---------------|--------------|------|-------------------|---------------|-------------|---------|---------------------------------------|----------------------------------|--------------------|
| | | | | | | | Van der Waals | Electrostatic | Desolvation | Total | | | |
| | | CVTLRE | 4 | -70.2 | 17 | 0.4 | -61.90 | -105.33 | -1.9 | -167.24 | 1388.16 | -8.7 | 4.3E-07 |
| 14 | 3OZQ | PFHTR | 1 | -93.5 | 128 | 0.4 | -53.80 | -38.64 | -41 | -92.44 | 1133.42 | -7.9 | 1.7E-06 |
| | | ATYIPV | 3 | -90.7 | 12 | 0.4 | -69.89 | -52.43 | -24.6 | -122.33 | 1124.62 | -10.5 | 1.9E-08 |
| 15 | 3SSB | LICN | 1 | -66.7 | 106 | 0.4 | -47.18 | -39.03 | -21.8 | -81.21 | 985.57 | -8.3 | 7.9E-07 |
| | | EYYECGGA | 1 | -95.1 | 84 | 0.4 | -71.53 | -98.20 | -19.4 | -169.75 | 1295.48 | 11.1 | 7.6E-09 |
| 16 | 4POF | AAHQ | 1 | -71.8 | 71 | 0.8 | -38.15 | -82.55 | -21.7 | -120.70 | 821.02 | -7.1 | 6.1E-06 |
| | | AAQILR | 1 | -87.9 | 104 | 0.9 | -61.71 | -102.76 | -17.4 | -104.47 | 1418.28 | -10.2 | 3.5E-08 |
| 17 | 4R9I | HHHHHM | 1 | -114.2 | 96 | 0.4 | -68.28 | -74.45 | -35 | -142.73 | 1352.69 | -9.7 | 7.4E-08 |
| 18 | 5C98 | ERLQSCW | 5 | -84.5 | 10 | 2 | -52.13 | -53.58 | -34.5 | -105.72 | 1197.93 | -9.1 | 2.1E-07 |
| 19 | 5DAE | VCACPRIYMP | 1 | -98.6 | 98 | 1.9 | -61.93 | -136.51 | -25.5 | -198.44 | 1521.5 | -12.1 | 1.2E-09 |
| | | NNDCL | 1 | -72.7 | 52 | 0.5 | -50.16 | -92.33 | -9.5 | -142.49 | 1043.93 | -8.9 | 3.2E-07 |
| | | NNDCLRCE | 3 | -81.5 | 18 | 0.5 | -50.44 | -115.26 | -7.5 | -165.70 | 1237.95 | -9.1 | 2.2E-07 |
| 20 | 6CJ7 | LSENFNL | 2 | -86.2 | 57 | 0.4 | -60.33 | -58.95 | -21.4 | -119.29 | 1161.43 | -9.0 | 2.5E-07 |
| | | TPTYFGK | 1 | -90.9 | 89 | 0.4 | -43.72 | -158.54 | -32.6 | -202.26 | 1220.92 | -8.4 | 6.5E-07 |
| 21 | 1GL0 | QQDCNT | 1 | -78.9 | 92 | 0.3 | -50.85 | -147.67 | -10.8 | -198.53 | 1127.68 | -8.1 | 1.2E-06 |
| | | LMGCQPT | 1 | -64.3 | 164 | 0.5 | -45.01 | -32.87 | -20.0 | -77.88 | 1091.75 | -9.9 | 5.5E-08 |
| | | CTLMGCQP | 1 | -74.2 | 96 | 0.5 | -43.72 | -65.31 | -23.6 | -109.04 | 1247.31 | -11.0 | 8.9E-09 |
| 22 | PRO42 | TKFGI | 2 | -92.9 | 35 | 0.4 | -59.84 | -97.85 | -27.1 | -157.70 | 1117.14 | -10.8 | 1.3E-08 |
| 23 | PRO-S7 | VPSNPQR | 4 | -93.0 | 22 | 0.8 | -59.84 | -97.85 | -32.0 | -157.70 | 1117.14 | -10.8 | 1.3E-08 |
| 24 | PRO13 | AEVPSD | 1 | -66.2 | 89 | 0.4 | -52.41 | -103.88 | -10.7 | -156.29 | 1090.02 | -7.2 | 5.5E-06 |
| | | YPNAQR | 1 | -78.9 | 92 | 0.3 | -51.70 | -76.73 | -10.8 | -128.43 | 1080.78 | -8.4 | 7.1E-07 |
| 25 | PRO-P7 | VPKNPQR | 2 | -95.3 | 42 | 0.4 | -60.34 | -174.54 | -5.1 | -234.89 | 1361.72 | -10.4 | 2.2E-08 |

ID 6LU7 coordinate) were calculated to be 0.05 nm by AutoDock, and 0.08 nm by HADDOCK (Supplementary Fig. 2), indicating the accuracy of the docking methods.

To further investigate the stability of the peptides bound to the predicted binding site of M^{PrO} and identify the residues involved in forming the intermolecular interactions, we subjected peptide/M^{PrO} complexes to 100 ns MD simulations. In addition, a 10-ns MD simulation was carried out for the apo M^{PrO} to refine the reference structure and remove steric clashes. RMSD values were used to examine whether the simulation systems were thermodynamically converged (Fig. 1). Further analyses were performed between 10 and 100 ns for the generated peptides/M^{PrO} simulation trajectories to acquire accurate and reproducible data where the systems were at equilibrium. Low fluctuations of RMSD are associated with greater stability of the protein structure [45, 46]. The fluctuation of RMSD along the trajectory was found to be 0.15 to 0.25, 0.15–0.3, and ~0.25–0.35 nm for apo M^{PrO}, 1KJ0-7/M^{PrO} and 2ERW-9/M^{PrO} complexes, respectively (Fig. 1A). We also calculated the radius of gyration (Rg) (Fig. 1B) and solvent accessible surface area (SASA) (Fig. 1C) of the protein to determine the compactness of M^{PrO} in complex with selected peptides and the contributions of solvation to the

electrostatic energy of the system, respectively. The Rg and SASA mean values were estimated to be 2.17 nm and 149.42 nm²/S2N for M^{PrO} in 2ERW-9/M^{PrO} complex and 2.21 nm and 150.92 nm²/S2N for M^{PrO} in 1KJ0-7/M^{PrO} complex, respectively. Conversely, the mean values of the Rg and SASA were 2.22 nm and 151.41 nm²/S2N for the apo M^{PrO} as the reference for comparison. Overall, RMSD, Rg, and SASA values indicated that both peptides, especially 2ERW-9, increased the compactness of M^{PrO} and decreased the solvent-exposed area and structural fluctuations of the protein. It should be noted that the secondary structure analysis of the peptides from the extracted structures along the trajectory revealed that both peptides consistently remained in a random coil conformation (Data not shown).

Biovia Discovery Studio software was used to predict molecular interactions between the screened peptides and M^{PrO} ligand-binding residues [47]. We examined the binding mode of peptides in the resulting peptide/protein complex structures along trajectories by visual inspection. Both peptides did not exhibit considerable conformational changes during trajectory with ΔRMSD values (the differences in RMSD between the initial and final refined models in MD simulation) being 0.26 nm for IKJ0-7 and 0.38 nm for 2ERW-9 (Fig. 2A and B and 3A, B). Therefore,

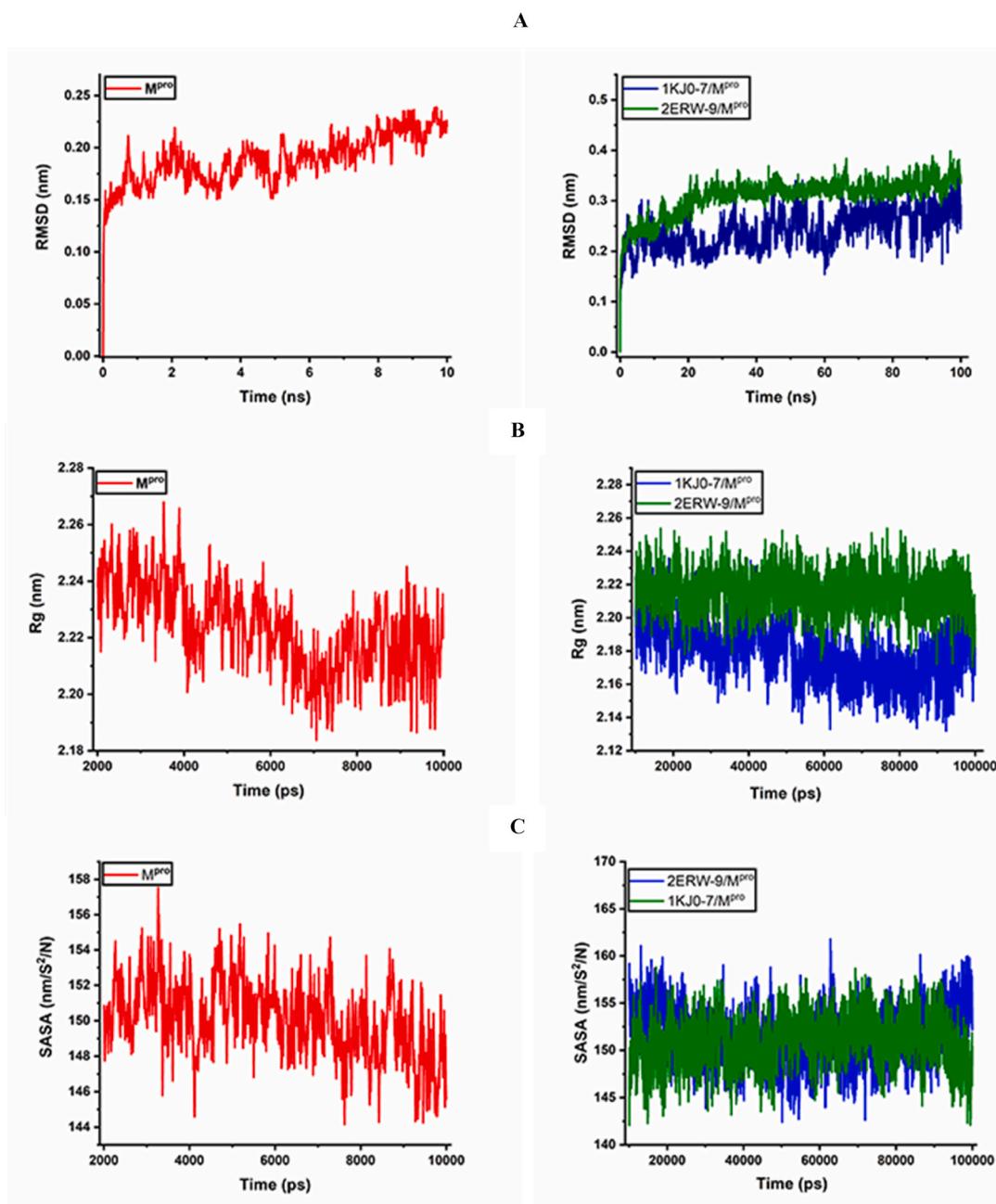


Fig. 1. Structural functions analysis of the apo M^{pro} (Red), 1KJ0-7/M^{pro} complex (Green), and 2ERW-9/M^{pro} complex (Blue) in the MD simulation predicting values of RMSD (A), Rg (B), and SASA (C).

two peptides maintained the conformation of their binding site throughout the 100-ns MD simulation, and Biovia Discovery Studio demonstrated that, upon binding of peptide candidates, the active site was adequately inaccessible by the natural substrates. 1KJ0-7 was shown to form interactions with H41, T45, E47, D48, G143, C145, M165, E166, Q189, and A191 in the M^{pro} catalytic site (Fig. 2D). Accordingly, the catalytic residue H41 of M^{pro} interacts with the C5 residue of 1KJ0-7 through a Pi-Sulfur interaction, and catalytic residue C145 of M^{pro} forms two hydrogen bonds with the H7 residue in 1KJ0-7. The binding of 1KJ0-7 to M^{pro} catalytic residues suggests that the designed inhibitor could largely inhibit the enzymatic reaction. In addition, E166, Q189, and A191, which are responsible for substrate trapping in M^{pro}, were shown to be blocked with (P5, H7, C4), (G3), and (K2) residues of 1KJ0-7, respectively (Fig. 2D). 2ERW-9 was shown to interact with T25, H41, T45, D48, N51, N142, C145, M165, E166, P168,

and A191 residues from the M^{pro} catalytic cavity (Fig. 3D). Catalytic dyad H41 and C145 from M^{pro} are simultaneously blocked with the A4 residue in 2ERW-9. Since the M^{pro} D187 was in contact with the R7 residue of 2ERW-9, we can conclude that the compactness of the active site is destabilized due to the loss or weakening of the M^{pro} internal salt bridge (between D187 and R40). Furthermore, the substrate trapping residues of M^{pro} including N142, E166, P168, and A191 were in close interactions with (N8 and V10), (N8, Y9), (V10), and (Y9) residues of 2ERW-9, respectively (Fig. 3D). Notwithstanding, to investigate the stability and dynamics of the interactions, the binding site were compared in terms of protein root mean square fluctuation (RMSF) measurements concerning carbon α (C α) only. As depicted in Figs. 2C and 3C, for both peptides, the RMSF values of the bonded residues of M^{pro} decreased along the trajectory, while the other residues remained constant and were similar to the values predicted for the apo M^{pro} as the

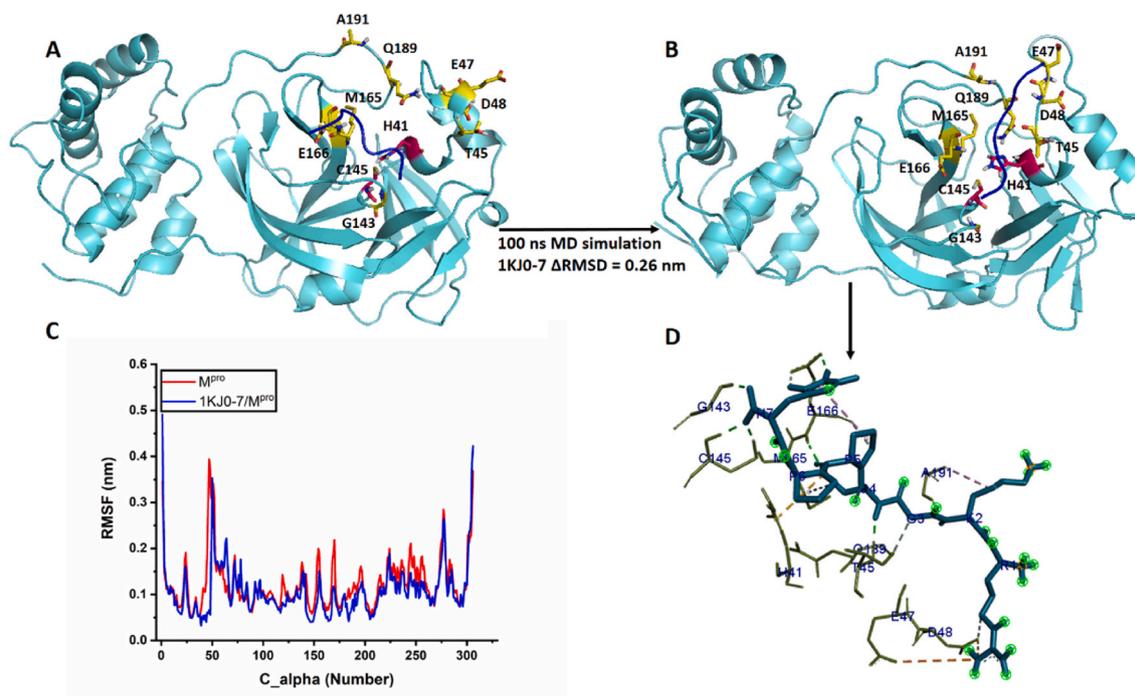


Fig. 2. (A) The initial structure 1KJ0-7/M^{pro} complex in MD simulation, (B) the final state of the 1KJ0-7/M^{pro} complex in MD simulation, (C) RMSF analysis for free M^{pro} (Red) and the 1KJ0-7/M^{pro} complex (Blue), (D) 3D representation of the residues involved in the binding of the 1KJ0-7 peptide (blue licorice stick) and M^{pro} (olive lines) in DS visualizer. Hydrogen interactions (green dashed lines); Pi-Alkyl interactions (light purple dashed lines); Pi-Pi and Pi-Sigma interactions (dark purple dashed lines); Pi-Sulfur, Pi-Anion, and Pi-Cation interactions (orange dashed lines) are displayed.

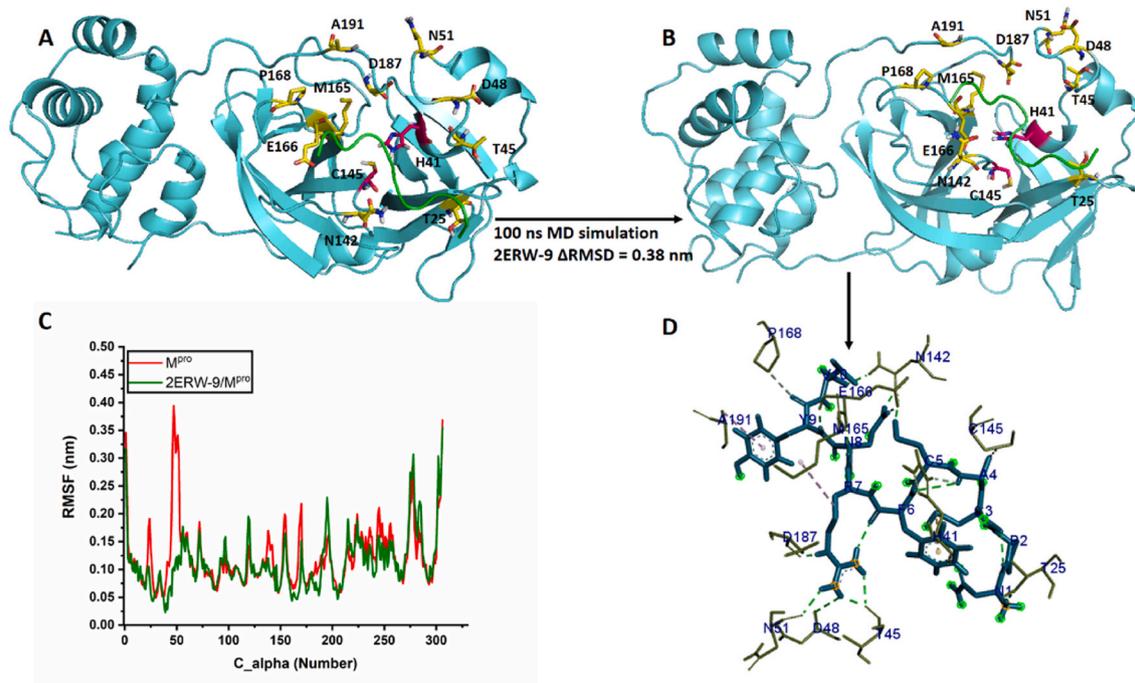


Fig. 3. (A) The initial structure 2ERW-9/M^{pro} complex in MD simulation, (B) the final state of the 2ERW-9/M^{pro} complex in MD simulation, (C) RMSF analysis for free M^{pro} (Red) and the 2ERW-9/M^{pro} complex (Green), (D) 3D representation of the residues involved in the binding of the 2ERW-9 7 peptide (blue licorice stick) and M^{pro} (olive lines) in DS visualizer. Hydrogen interactions (green dashed lines); Pi-Alkyl interactions (light purple dashed lines); Pi-Pi and Pi-Sigma interactions (dark purple dashed lines); Pi-Sulfur, Pi-Anion, and Pi-Cation interactions (orange dashed lines) are displayed.

reference structure.

3.6. Pharmacokinetic, toxicity, biochemical parameters (ligand efficiency), and Lipinski properties of peptides

The drug-likeness of a novel compound is investigated using

Table 5

Lipinski properties of the two screened peptides and two FDA-approved protease inhibitors including Ritonavir and Lopinavir as control compounds.

| Compound | Decoy peptides in SMILES format | Lipinski properties |
|------------------|--|--|
| 2ERW-9 | <chem>CC(C)C(NC(=O)C(C1=CC=C(O)C=C1)NC(=O)C(CCN) = O)NC(=O)C(CCC[NH+] = C(N)N)NC(=O)C(CC2=CC=CC=C2)NC(=O)C(CS)NC(=O)C(C)NC(=O)C(CS)NC(=O)C3CCCN3C(=O)C([NH3+])CC(N) = O C([O-]) = O</chem> | Molecular weight: 1187.39 ALog P: 3.31 H-bond donor: 17 9 H-bond acceptor: 16 11 Molar refractivity: 304.73 195.54 Topological Polar Surface Area: 570.9220.14 Number of rotatable bonds: 33 16 |
| 1KJ0-7 | <chem>NC(N) = [NH+]CCCC([NH3+])C(=O)NC(CCCC[NH3+])C(=O)NCC(=O)NC(CS)C(=O)N1CCCC1C(=O)N2CCCC2C(=O)NC(CC3=C[NH]C=N3)C([O-]) = O</chem> | Molecular weight: 795.97 ALog P: 5.82 H-bond donor: 11 8 H-bond acceptor: 10 6 Molar refractivity: 210.22 Topological Polar Surface Area: 385.92 169.15 Number of rotatable bonds: 22 11 |
| Lopinavir | <chem>CC(C)C(N1CCCNC1 = O)C(=O)NC(CC(O)C(CC2=CC=CC=C2)NC(=O)COC3=C(C)C=CC=C3C)CC4=CC=CC=C4</chem> | Molecular weight: 628.81 ALog P: 4.33 H-bond donor: 5 H-bond acceptor: 4 Molar refractivity: 187.92 Topological Polar Surface Area: 120 Number of rotatable bonds: 15 |
| Ritonavir | <chem>CC(C)C(NC(=O)N(C)CC1=CSC(=N1)C(C)C(=O)NC(C(C)O)C(CC2=CC=CC=C2)NC(=O)OCC3=CN=CS3)CC4=CC=CC=C4</chem> | Molecular weight: 720.94 ALog P: 5.91 H-bond donor: 9 H-bond acceptor: 4 Molar refractivity: 197.82 Topological Polar Surface Area: 202.26 Number of rotatable bonds: 17 |

Lipinski's rule of five in the drug discovery process [48]. The rule determines essential pharmacokinetic properties of drug molecules, including the absorption, distribution, metabolism, and excretion (ADMET) [41,48]. Designed drug candidates that comply with the Lipinski's are considered to have ideal pharmacokinetic properties. Lipinski's rule is based on four physicochemical characteristics, including (1) molecular weight (≤ 500 Da), (2) number of hydrogen bond donors (≤ 5) (sum of OH and NH groups), (3) number of hydrogen bond acceptors (≤ 10) (sum of N and O atoms), (4) lipophilicity ($\text{Log } P \leq 5$) [48, 49]. In addition to calculating these four characteristics, other related criteria that were predicted [50] included the polar surface area $< 140 \text{ \AA}^2$, the number of rotatable bonds (≤ 10), and the molar refractivity (< 130). The values predicted for the two screened peptides and two FDA-approved protease inhibitors (Ritonavir and Lopinavir) as control compounds are listed in Tables 5–7. Shivanika et al. (2020) used the AutoDock docking method and reported that Lopinavir and Ritonavir had the potency to inhibit the M^{pro} active site with binding affinities of -9.70 kcal/mol and Ritonavir with -11.15 kcal/mol, respectively [51]. In *in-silico* studies, Lopinavir was proved to exhibit the effective IC₅₀

Table 6

Toxicity properties of the two screened peptides and two FDA-approved protease inhibitors including Ritonavir and Lopinavir as control compounds.

| Compound | AMES toxicity | Acute oral toxicity (kg/mol) | Carcinogenicity | Hepatotoxicity | Tetrahymena Pyriformis Toxicity pIC ₅₀ , mg/L | Rat acute toxicity LD ₅₀ , mol/kg | Biodegradation |
|------------------|---------------|------------------------------|-----------------|----------------|--|--|----------------|
| 2ERW-9 | No | 3.496 | No | No | 0.45 | – | No |
| 1KJ0-7 | No | 2.684 | No | No | 0.364 | – | No |
| Lopinavir | No | 2.994 | No | Yes | 0.875 | 2.2503 | No |
| Ritonavir | No | 2.281 | No | Yes | 0.841 | 2.6154 | No |

value of $13.7 \mu\text{M}$ when used alone, and combining Lopinavir with Ritonavir resulted in improvement of IC₅₀ ($10.9 \mu\text{M}$) in HEK-293 T cell cultures [52]. We followed the same AutoDock steps that were previously applied for the two FDA-approved drugs. Results showed that Lopinavir with the binding affinity of -11.26 kcal/mol and Ritonavir with that of -11.12 kcal/mol blocked M^{pro} in the active site. The complex and interaction analysis are depicted in Supplementary Figs. 3 and 4. Lopinavir and Ritonavir are FDA-approved drugs characterized as M^{pro} inhibitors. In our study, AutoDock results showed that both ligands interacted with residues in the catalytic binding site of M^{pro} , including P168, E166, M165, A191, H41, T26, C145, and Q189 for Ritonavir, and P168, M165, L167, E166, H41, C145, and T25 for Lopinavir, which were similar to our findings for the binding site of the designed peptides.

4. Discussion

SARS-CoV-2, a human coronavirus, is the causative agent of coronavirus disease 2019 (COVID-19). The pandemic spread of SARS-CoV-2 and the continually increasing number of COVID-19-related deaths have made it necessary to develop effective therapeutic strategies against the virus. M^{pro} of SARS-CoV-2 plays a pivotal role in viral replication through mediating the cleavage of replicase polyproteins [13]. Therefore, SARS-CoV-2 M^{pro} is considered an attractive target for drug development against COVID-19.

Insects are known to produce a wide range of protease inhibitors [20]. However, in comparison with other natural sources, insects have been relatively neglected for drug development [21]. Therefore, this study aimed to adopt a bioinformatics approach to screen for insect-derived compounds with potential inhibitory properties against M^{pro} and to further predict interactions of these inhibitors with the enzyme *in silico*.

The crystal structure of SARS-CoV-2 M^{pro} in complex with the inhibitor N3 was determined by Jin and colleagues in 2020 [22] and was made publicly accessible in the protein data bank (PDB-ID: 6LU7). It is worth noting that amino acid sequences of M^{pro} encoded by SARS-CoV-2 and SARS-CoV (PDB-ID: 2GTB) have been previously shown to be 96% identical [53]. Further primary and secondary structure analysis, performed in our study, revealed a high similarity between the two proteins (6LU7 and 2GTB). For instance, random coils were predicted to be predominant secondary structures followed by α -helices in both proteins. 6LU7 and 2GTB were predicted to be stable, acidic, and hydrophilic proteins (Supplementary Table 1). Taken together, M^{pro} seems to be highly conserved among coronaviruses, as was also demonstrated through the superposition of 12 crystal structures of M^{pro} by Jin et al. [22]. Being highly conserved among coronaviruses, M^{pro} is believed to be a promising target for developing wide-spectrum inhibitors [54].

In the present study, the publicly available 3D structure of SARS-CoV-2 M^{pro} was run through molecular docking experiments. 60 inhibitor peptides were designed by blind docking of the protomer A of SARS-CoV-2 M^{pro} (6LU7) to various insect protease inhibitors using the ClusPro server followed by structural model prediction of inhibitor/ M^{pro} complexes by PEP-FOLD3. Upon blind docking, directed docking of these 60 peptides was performed by two independent docking programs: AutoDock and HADDOCK. The use of two different docking programs enabled us to evaluate our adopted methodology by two programs that

Table 7

Pharmacokinetic properties of the two screened peptides and two FDA-approved protease inhibitors including Ritonavir and Lopinavir as control compounds.

| Part B | 2ERW-9 | 1KJ0-7 | Lopinavir | Ritonavir |
|---------------------------------|--------------|--------------|--------------|-----------|
| Human | - | - | + | + |
| Intestinal Absorption | | | | |
| Blood Brain Barrier | + | + | + | - |
| Plasma protein binding (100%) | 0.42 | 0.127 | 1.157 | 1.113 |
| Water solubility (LogS) | -3.308 | -2.702 | -3.414 | -3.225 |
| Estrogen receptor binding | + | + | + | + |
| Androgen receptor binding | + | + | + | + |
| Thyroid receptor binding | + | + | + | + |
| Glucocorticoid receptor binding | + | - | + | + |
| Aromatase binding | + | + | - | + |
| PPAR gamma | + | + | + | + |
| Subcellular localization | Mitochondria | Mitochondria | Mitochondria | Lysosomes |
| Caco-2 Permeability | | | | |
| OATP2B1 inhibitor | - | - | + | + |
| OATP1B1 inhibitor | + | + | + | + |
| OATP1B3 inhibitor | + | + | + | + |
| MATE1 inhibitor | - | - | - | - |
| OCT2 inhibitor | - | - | - | - |
| BSEP inhibitor | + | + | + | + |
| P-glycoprotein inhibitor | + | + | + | + |
| P-glycoprotein substrate | + | + | + | + |
| CYP3A4 substrate | + | + | + | + |
| CYP2C9 substrate | - | - | - | + |
| CYP2D6 substrate | - | - | - | - |
| CYP3A4 inhibition | - | - | - | + |
| CYP2C9 inhibition | - | - | - | - |
| CYP2C19 inhibition | - | - | - | + |
| CYP2D6 inhibition | - | - | - | - |
| CYP1A2 inhibition | - | - | - | - |
| CYP inhibitory promiscuity | - | - | - | + |
| UGT catalyzed | - | - | - | - |

function on the basis of two different search algorithms, scoring functions, and pose selection schemes [55]. Furthermore, the binding affinity of the designed peptides to the active site of the enzyme, as calculated by the two docking programs, can be compared to one another. Predicting the binding affinity of inhibitors to the active site is particularly important, as enzyme inhibitors modify the catalytic properties of the target enzyme through binding to and blocking the active site [56].

As expected, the two docking programs generated different

outcomes. AutoDock analysis introduced 1KJ0-7 and 3OZQ-6 with binding energies of -9.39 and -9.26 kcal/mol, whereas HADDOCK analysis resulted in the discovery of 2ERW-9 with the binding energy of -11.70 kcal/mol. 1KJ0-7 (RKGCPH) was predicted to have the highest binding affinity to the target enzyme with desirable binding energy, inhibition constant, and intermolecular energy using AutoDock (Table 3). However, 2ERW-9 (NPCACFRNY) was ranked as the best M^{PTO} inhibitor by HADDOCK calculations based on several criteria including cluster size, HADDOCK score, RMSD, binding energy and dissociation constant (Table 4). 2ERW-9 showed the highest buried surface area value in comparison with other peptide inhibitors (Table 4). 2ERW-9 was shown to form 12 intermolecular hydrogen bonding interactions and 127 non-bonded contacts to M^{PTO} . A high number of hydrogen bonding interactions between inhibitors and target molecules plays a vital role in selecting potent and specific inhibitor peptides. 1KJ0-7 was shown to form nine intermolecular hydrogen bonding interactions and 91 non-bonded contacts to M^{PTO} . Considering the number of predicted hydrogen bonds between 1KJ0 and 7 or 2ERW-9 with the M^{PTO} active site, these peptides have a great potential to inhibit SARS-CoV-2 M^{PTO} .

The peptides are bound in two orientations, where the N-terminal and the C-terminal of peptide sequences are directed toward the binding pocket. All docking results binding affinities reflected the negative binding energies in all models, indicating favorable binding in all complexes, ranging from -1 to -10 kcal/mol in AutoDock results and -410 to -980 kcal/mol for HADDOCK values. The optimum binding interaction belonged to the hepta-to deca-mer peptides. As expected, due to the presence of charged and neutral residues like E, M, H, T, D, and Q in the M^{PTO} catalytic site, both positively and negatively charged amino acids can interact with the catalytic cavity; however, our best screened docked structures showed positively charged and basic residues at N- (R1K2 for 1KJ0 and N1 for 2ERW) and C-terminal (H7 for 1KJ0 and F6R7N8Y9 for 2ERW).

In 2020, the WHO recommended the combination of Lopinavir and Ritonavir as an antiviral treatment option for COVID-19 [57]. Lopinavir and Ritonavir are protease inhibitors currently used for the treatment of human immunodeficiency virus (HIV) [58]. A previous study reported the binding energy of -9.41 kcal/mol for Lopinavir, when docked against SARS-CoV-2 M^{PTO} (6LU7) [59]. This value of binding energy is relatively close to those reported in this study for 1KJ0-7 and 3OZQ-6 (-9.39 and -9.26 kcal/mol). The other peptide designed in this study (2ERW-9) was calculated to have the binding energy of -11.70 kcal/mol, which was significantly higher than that of Lopinavir. It should also be noted that the WHO solidarity trial and, later, other research groups concluded that Lopinavir alone or in combination with Ritonavir has limited effect on COVID-19 progression [60–62].

In summary, our study introduced two novel peptides with potential inhibitory properties against COVID-19. The methodology used in this study would possibly contribute to the discovery of other novel anti-COVID-19 compounds from natural sources, particularly from insect protease inhibitors or pro-regions of insect proteases.

Ligand efficiency scores are used to escape the affinity-biased selection and optimization towards larger ligands. Recently the application of ligand efficiency has been widely increased in the selection and optimization of newly introduced candidates. In particular, optimization of lipophilic ligand efficiency reveals whether increased affinity is mediated by an increase in lipophilic characteristics or not. This method provides a way to compare the affinity of ligands corrected for their size, even with challenging ‘lipophile-preferring’ targets. We calculated LE values (Ligand Efficiency: binding affinity divided by the number of heavy atoms (HA), $LE = -\Delta G/HA$) of 0.15 and 0.18 for 2ERW and 1KJ0 peptides, respectively. In a recent study, the mean LE values of 480 target-assay pairs that included 329 human drug targets obtained from recent medicinal chemistry literature span a broad range of ~ 0.15 – 0.60 [63].

5. Conclusion

The COVID-19 pandemic outbreak and the absence of effective drugs against the disease introduced a new challenge to researchers to urgently develop lead compounds or precursors with antiviral potential. The present study aimed to discover insect-derived protease inhibitors as potent agents against SARS-CoV-2 M^{Pro} through an *in silico* procedure. Molecular docking was applied using the AutoDock and HADDOCK methods for sixty insect protease inhibitors along with supporting MD simulations and Biovia Discovery Studio to elucidate the interactions phenomena between the selected peptides and M^{Pro}. This study also estimated the pharmacokinetics, toxicity, and Lipinski properties of the screened peptides and compared these peptides with the currently FDA-approved drugs. The dynamic interactions between the peptides and SARS-CoV-2 M^{Pro} confirmed the promising potential of our selected peptides for inhibiting M^{Pro}. More specifically, molecular modelling and simulation results suggested that 1KJ0-7 and 2ERW-9 are promising candidates as antivirals against SARS-CoV-2, which might help reduce COVID-19 infections and death cases in the near future. Natural compounds like peptides, flavonoids, alkaloids, tannins, and others usually do not possess any mutagenic and carcinogenic properties with little or no side effects, and the selected peptides (1KJ0-7 and 2ERW-9) are derived from insect proteases and satisfy these criteria. This research can be used as a promising road map for the discovery of novel SARS-CoV-2 M^{Pro} inhibitors. However, *in vitro* and *in vivo* analyses are further required to evaluate the safety and efficiency of selected inhibitors against M^{Pro}. Bio-nanotechnology and target delivery tools can be used to improve the bioavailability and therapeutic efficiency of drugs and preferential accumulation at the target site. Future studies will also focus on natural compounds as capping and reducing agents onto metal nanoparticles, which will provide positive insights towards the cure of infection.

Acknowledgment

This study was supported by Shahid Chamran University of Ahvaz, Ahvaz, Iran [Grant No. SCU.AP99.39134].

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.compbiomed.2022.105228>.

References

- J.S.M. Peiris, Coronaviruses, in: D.D. Richman, R.J. Whitley, F.J. Hayden (Eds.), *Clinical Virology*, ASM Press, Washington, DC, USA, 2016, pp. 1243–1265, <https://doi.org/10.1128/9781555819439>.
- S. Payne, Family Coronaviridae. *Viruses* 149–58 (2017), <https://doi.org/10.1016/B978-0-12-803109-4.00017-9>.
- Wk Lam, N.S. Zhong, W.C. Tan, Overview on SARS in Asia and the world, *Respirology* 8 (2003), <https://doi.org/10.1046/j.1440-1843.2003.00516.x> S2–S5.
- A.M. Zaki, S. van Boheemen, T.M. Bestebroer, A.D.M.E. Osterhaus, R.A. M. Fouchier, Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia, *N. Engl. J. Med.* 367 (19) (2012) 1814–1820, <https://doi.org/10.1056/NEJMoa1211721>.
- A.L. Phelan, R. Katz, L.O. Gostin, The novel coronavirus originating in Wuhan, China: challenges for global health governance, *JAMA* 323 (8) (2020) 709–710, <https://doi.org/10.1001/jama.2020.1097>.
- N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, P. Niu, A novel coronavirus from patients with pneumonia in China, 2019, *N. Engl. J. Med.* 382 (8) (2020) 727–733, <https://doi.org/10.1056/NEJMoa2001017>.
- Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2, *Nat Microbiol* 5 (4) (2020) 536–544, <https://doi.org/10.1038/s41564-020-0695-z>.
- World Health Organization, *Novel Coronavirus (2019-nCoV): Situation Report-8, 2020*.
- COVID-19 Map, Johns Hopkins Coronavirus Resource Center, 2021. <https://coronavirus.jhu.edu/map.html>.
- J.Y. Chung, M.N. Thone, Y.J. Kwon, COVID-19 vaccines: the status and perspectives in delivery points of view, *Adv. Drug Deliv. Rev.* 170 (2021) 1–25, <https://doi.org/10.1016/j.addr.2020.12.011>.
- S. Cele, I. Gazy, L. Jackson, S.H. Hwa, H. Tegally, G. Lustig, J. Giandhari, S. Pillay, E. Wilkinson, Y. Naidoo, F. Karim, Escape of SARS-CoV-2 501Y.V2 from neutralization by convalescent plasma, *Nature* (2021) 1–9, <https://doi.org/10.1038/s41586-021-03471-w>.
- A.R. Fehr, S. Perlman, Coronaviruses: an overview of their replication and pathogenesis, in: H.J. Maier, E. Bickerton, P. Britton (Eds.), *Coronaviruses*, Springer, New York, 2015, pp. 1–23, https://doi.org/10.1007/978-1-4939-2438-7_1.
- J. Ziebuhr, E.J. Snijder, A.E. Gorbalenya, Virus-encoded proteinases and proteolytic processing in the nidovirales, *J. Gen. Virol.* 81 (4) (2000) 853–879, <https://doi.org/10.1099/0022-1317-81-4-853>.
- J.V. Fernández-Montero, P. Barreiro, V. Soriano, HIV protease inhibitors: recent clinical trials and recommendations on use, *Expert Opin. Pharmacother.* 10 (10) (2009) 1615–1629, <https://doi.org/10.1517/14655660902980202>.
- I.A. Rowe, D.J. Mutimer, Protease inhibitors for treatment of genotype 1 hepatitis C virus infection, *BMJ* 343 (7832) (2011) 1060–1063, <https://doi.org/10.1136/bmj.d6972>.
- S.J.Y. Macalino, V. Gosu, S. Hong, S. Choi, Role of computer-aided drug design in modern drug discovery, *Arch Pharm. Res. (Seoul)* 38 (9) (2015) 1686–1701, <https://doi.org/10.1007/s12272-015-0640-5>.
- X.Y. Meng, H.X. Zhang, M. Mezei, M. Cui, Molecular docking: a powerful approach for structure-based drug discovery, *Curr. Comput. Aided Drug Des.* 7 (2) (2011) 146–157, <https://doi.org/10.2174/157340911795677602>.
- D. Kumar, M. Shahid, *Natural Materials and Products from Insects: Chemistry and Applications*, Springer International Publishing, Cham, 2020, <https://doi.org/10.1007/978-3-030-36610-0>.
- J. Powers, R.H. McDowell, *Insect Bites*. Treasure Island (FL), StatPearls Publishing, 2020. <http://www.ncbi.nlm.nih.gov/books/NBK537235/>.
- D. Gubb, A. Sanz-Parra, L. Barcena, L. Troxler, A. Fullaondo, Protease inhibitors and proteolytic signalling cascades in insects, *Biochimie* 92 (12) (2010) 1749–1759, <https://doi.org/10.1016/j.biochi.2010.09.004>.
- N.A. Ratcliffe, C.B. Mello, E.S. Garcia, T.M. Butt, P. Azambuja, Insect natural products and processes: new treatments for human disease, *Insect Biochem. Mol. Biol.* 41 (10) (2011) 747–769, <https://doi.org/10.1016/j.ibmb.2011.05.007>.
- Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, Y. Duan, Structure of M^{Pro} from SARS-CoV-2 and discovery of its inhibitors, *Nature* 582 (7811) (2020) 289–293, <https://doi.org/10.1038/s41586-020-2223-y>.
- E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R. Appel, A. Bairoch, Protein identification and analysis tools on the ExPASy server, in: J. M. Walker (Ed.), *The Proteomics Protocols Handbook*, Humana Press, Totowa, NJ, 2005, pp. 571–607, <https://doi.org/10.1385/1-59259-890-0:571>.
- C. Geourjon, G. Deléage, SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments, *Bioinformatics* 11 (6) (1995) 681–684, <https://doi.org/10.1093/bioinformatics/11.6.681>.
- A. Krogh, B. Larsson, G. von Heijne, E.L.L. Sonnhammer, Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes, *J. Mol. Biol.* 305 (3) (2001) 567–580, <https://doi.org/10.1006/jmbi.2000.4315>.
- W. Tian, C. Chen, X. Lei, J. Zhao, J. Liang, CASTp 3.0: computed atlas of surface topography of proteins, *Nucleic Acids Res.* 46 (2018) 363–367, <https://doi.org/10.1093/nar/gky473>.
- J. Jitnonom, K. Lomthaisong, V.S. Lee, Computational design of peptide inhibitor based on modifications of proregion from *Plutella xylostella* midgut trypsin, *Chem. Biol. Drug Des.* 79 (4) (2012) 583–593, <https://doi.org/10.1111/j.1747-0285.2011.01312.x>.
- S.A. Hemmati, Z. Takaloo, M. Taghdir, M. Mehrabadi, S. Balalaei, S. Moharrampour, R.H. Sajedi, The trypsin inhibitor pro-peptide induces toxic effects in Indianmeal moth, *Plodia interpunctella*, *Pestic. Biochem. Physiol.* 171 (2021) 104730, <https://doi.org/10.1016/j.pestbp.2020.104730>.
- S.A. Hemmati, N. Karam Kiani, J.E. Serrão, J. Jitnonom, Inhibitory potential of a designed peptide inhibitor based on zymogen structure of trypsin from *Spodoptera frugiperda*: *in silico* insights, *Int. J. Pept. Res. Therapeut.* 27 (2021) 1677–1687, <https://doi.org/10.1007/s10989-021-10200-4>.
- J. Yang, Y. Zhang, I-TASSER Server: New development for protein structure and function predictions, *Nucleic Acids Res.* 43 (2015) 174–181, <https://doi.org/10.1093/nar/gkv342>.
- P. Thevenet, Y. Shen, J. Maupetit, F. Guyon, P. Derreumaux, P. Tuffery, PEP-FOLD: an updated de Novo structure prediction server for both linear and disulfide bonded cyclic peptides, *Nucleic Acids Res.* 40 (2012) 288–293, <https://doi.org/10.1093/nar/gks419>.
- D. Kozakov, D.R. Hall, B. Xia, K.A. Porter, D. Padhorny, C. Yueh, D. Beglov, S. Vajda, The ClusPro web server for protein–protein docking, *Nat. Protoc.* 12 (2) (2017) 255–278, <https://doi.org/10.1038/nprot.2016.169>.
- G. Vriend, What IF: a molecular modeling and drug design program, *J. Mol. Graph.* 8 (1) (1990) 52–56, [https://doi.org/10.1016/0263-7855\(90\)80070-V](https://doi.org/10.1016/0263-7855(90)80070-V).
- S.J. de Vries, M. van Dijk, A.M.J.J. Bonvin, The HADDOCK web server for data-driven biomolecular docking, *Nat. Protoc.* 5 (5) (2010) 883–897, <https://doi.org/10.1038/nprot.2010.32>.
- L.C. Xue, J.P. Rodrigues, P.L. Kastriitis, AMjj Bonvin, A. Vangone, PRODIGY: a web server for predicting the binding affinity of protein–protein complexes, *Bioinformatics* 32 (23) (2016) 3676–3678, <https://doi.org/10.1093/bioinformatics/btw514>.

- [36] E. Lindahl, B. Hess, D. Van Der Spoel, Gromacs 3.0: a package for molecular simulation and trajectory analysis, *Mol Model Annu* 7 (8) (2001) 306–317, <https://doi.org/10.1007/s008940100045>.
- [37] H.J. Berendsen, J.V. Postma, W.F. van Gunsteren, A.R. DiNola, J.R. Haak, Molecular dynamics with coupling to an external bath, *J. Chem. Phys.* 81 (8) (1984) 3684–3690, <https://doi.org/10.1063/1.448118>.
- [38] M. Parrinello, A. Rahman, Crystal structure and pair potentials: a molecular-dynamics study, *Phys. Rev. Lett.* 45 (14) (1980) 1196, <https://doi.org/10.1103/PhysRevLett.45.1196>.
- [39] B. Hess, H. Bekker, H.J. Berendsen, J.G. Fraaije, LINCS: a linear constraint solver for molecular simulations, *J. Comput. Chem.* 18 (12) (1997) 1463–1472, [https://doi.org/10.1002/\(SICI\)1096-987X\(199709\)18:12<1463::AID-JCC4>3.0.CO;2-H](https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H).
- [40] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: an N-log (N) method for Ewald sums in large systems, *J. Chem. Phys.* 98 (12) (1993) 10089–10092, <https://doi.org/10.1063/1.464397>.
- [41] H. Yang, C. Lou, L. Sun, J. Li, Y. Cai, Z. Wang, W. Li, G. Liu, Y. Tang, admetSAR 2.0: web-service for prediction and optimization of chemical ADMET properties, *Bioinformatics* 35 (6) (2019) 1067–1069, <https://doi.org/10.1093/bioinformatics/bty707>.
- [42] M.N. Drwal, P. Banerjee, M. Dunkel, M.R. Wettig, R. Preissner, ProTox: a web server for the in silico prediction of rodent oral toxicity, *Nucleic Acids Res.* 42 (W1) (2014) W53–W58, <https://doi.org/10.1093/nar/gku401>.
- [43] K. Guruprasad, B.V.B. Reddy, M.W. Pandit, Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting *in vivo* stability of a protein from its primary sequence, *Protein Eng. Des. Sel.* 4 (2) (1990) 155–161, <https://doi.org/10.1093/protein/4.2.155>.
- [44] D.W. Kneller, G. Phillips, H.M. O'Neill, R. Jedrzejczak, L. Stols, P. Langan, A. Joachimiak, L. Coates, A. Kovalevsky, Structural plasticity of SARS-CoV-2 3CLM pro active site cavity revealed by room temperature X-ray crystallography, *Nat. Commun.* 11 (1) (2020) 1–6, <https://doi.org/10.1038/s41467-020-16954-7>.
- [45] I. Aier, P.K. Varadwaj, U. Raj, Structural insights into conformational stability of both wild-type and mutant EZH2 receptor, *Sci. Rep.* 6 (1) (2016) 1–10, <https://doi.org/10.1038/srep34984>.
- [46] K. Pouraghajan, H. Mahdiuni, S. Ghobadi, R. Khodarahmi, LRH-1 (liver receptor homolog-1) derived affinity peptide ligand to inhibit interactions between β -catenin and LRH-1 in pancreatic cancer cells: from computational design to experimental validation, *J. Biomol. Struct. Dyn.* 11 (2020) 1–16, <https://doi.org/10.1080/07391102.2020.1845241>.
- [47] D. Systèmes, BIOVIA Discovery Studio (Version 4.5). San Diego, 2020. <https://discover.3ds.com/discovery-studio-visualizer-download/>.
- [48] C.A. Lipinski, Lead-and drug-like compounds: the rule-of-five revolution, *Drug Discov. Today Technol.* 1 (4) (2004) 337–341, <https://doi.org/10.1016/j.ddtec.2004.11.007>.
- [49] C.A. Lipinski, Drug-like properties and the causes of poor solubility and poor permeability, *J. Pharmacol. Toxicol. Methods* 44 (1) (2000) 235–249, [https://doi.org/10.1016/S1056-8719\(00\)00107-6](https://doi.org/10.1016/S1056-8719(00)00107-6).
- [50] D.F. Veber, S.R. Johnson, H.Y. Cheng, B.R. Smith, K.W. Ward, K.D. Kopple, Molecular properties that influence the oral bioavailability of drug candidates, *J. Med. Chem.* 45 (12) (2002) 2615–2623, <https://doi.org/10.1021/jm020017n>.
- [51] C. Shivanika, D. Kumar, V. Ragunathan, P. Tiwari, A. Sumitha, Molecular docking, validation, dynamics simulations, and pharmacokinetic prediction of natural compounds against the SARS-CoV-2 main-protease, *J. Biomol. Struct. Dyn.* (2020) 1–27, <https://doi.org/10.1080/07391102.2020.1815584>.
- [52] M. Mahdi, J.A. Mótýán, Z.I. Szojka, M. Golda, M. Miczi, J. Tózsér, Analysis of the efficacy of HIV protease inhibitors against SARS-CoV-2's main protease, *Virol. J.* 17 (1) (2020) 1–8, <https://doi.org/10.1186/s12985-020-01457-0>.
- [53] Z. Xu, C. Peng, Y.S.Z. Zhu, K. Mu, X. Wang, W. Zhu, Nelfinavir was predicted to be a potential inhibitor of 2019-NCov main protease by an integrative approach combining homology modelling, molecular docking and binding free energy calculation, *Pharmacol. Toxicol.* (2020), <https://doi.org/10.1101/2020.01.27.921627>.
- [54] H. Yang, W. Xie, X. Xue, K. Yang, J. Ma, W. Liang, Q. Zhao, Z. Zhou, D. Pei, J. Ziebuhr, R. Hilgenfeld, Design of wide-spectrum inhibitors targeting coronavirus main proteases, *PLoS Biol.* 3 (10) (2005) 324, <https://doi.org/10.1371/journal.pbio.0030324>.
- [55] A. Basciu, G. Mallocci, F. Pietrucci, A.M.J.J. Bonvin, A.V. Vargiu, An enhanced-sampling MD-based protocol for molecular docking. *Biophysics.* <https://doi.org/10.1101/434092>, 2018.
- [56] M. Miró, P. Worsfold, C. Poole, A. Townshend, *Encyclopedia of analytical science.* <https://doi.org/10.1016/B978-0-12-409547-2.14331-8>, 2019, 1.
- [57] Solidarity clinical trial for COVID-19 treatments. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov/solidarity-clinical-trial-for-covid-19-treatments>, 2020.
- [58] Z. Lv, Y. Chu, Y. Wang, HIV protease inhibitors: a review of molecular selectivity and toxicity, *HIV/AIDS* 7 (2015) 95, <https://doi.org/10.2147/HIV.S79956>.
- [59] S. Khaerunnisa, H. Kurniawan, R. Awaluddin, S. Suhartati, S. Soetjipto, Potential inhibitor of COVID-19 main protease (M^{pro}) from several medicinal plant compounds by molecular docking study, *Med Pharmacol* (2020), <https://doi.org/10.20944/preprints202003.0226.v1>.
- [60] T.R.C. Group, Dexamethasone in hospitalized patients with Covid-19—preliminary report, *NEJM* (2020), <https://doi.org/10.1056/NEJMoa2023184>.
- [61] P.W. Horby, M. Mafham, J.L. Bell, L. Linsell, N. Staplin, J. Emberson, Recovery collaborative group. lopinavir-ritonavir in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial, *Lancet* 396 (10259) (2020) 1345–1352, [https://doi.org/10.1016/S0140-6736\(20\)32013-4](https://doi.org/10.1016/S0140-6736(20)32013-4).
- [62] W.J. Guan, Z.Y. Ni, Y. Hu, W.H. Liang, C.Q. Ou, J.X. He, L. Liu, H. Shan, C.L. Lei, D. S. Hui, B. Du, Clinical characteristics of coronavirus disease 2019 in China, *NEJM* 382 (18) (2020) 1708–1720, <https://doi.org/10.1056/NEJMoa2002032>.
- [63] A.L. Hopkins, G.M. Keserü, P.D. Leeson, D.C. Ress, C.H. Reynolds, The role of ligand efficiency metrics in drug discovery, *Nat. Rev. Drug Discov.* 13 (2014) 105–121, <https://doi.org/10.1038/nrd4163>.