

In Silico Mutagenesis Based Remodeling of SARs-CoV-1 Peptide (ATLQAIAS) to Inhibit SARs-CoV-2: Structural-dynamics and Free Energy Calculations

Abbas Khan

Department of Bioinformatics and Biological Statistics, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, 200240, P.R. China.

Shaheena Umbreen

Department of Botany, University of Okara, Okara, Punjab, Pakistan.

Asma Hameed

Department of Botany, University of Azad Jammu & Kashmir, Muzaffarabad, Azad Jammu & Kashmir, Pakistan.

Rida Fatima

Department of Chemistry, Bahauddin Zakariya University, Multan, Pakistan.

Ujala Zahoor

Department of Botany, Women University, Bagh, Azad Jammu & Kashmir, Pakistan.

Zainib Babar

Center for Viticulture and Enology, School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai, 200240, China.

Muhammad Waseem

Faculty of Rehabilitation and Allied Health Science, Riphah International

Zahid Hussain

Center for Biotechnology and Microbiology, University of Swat, Swat, KP, Pakistan.

Muhammad Rizwan

Center for Biotechnology and Microbiology, University of Swat, Swat, KP, Pakistan.

Nasib Zaman

Center for Biotechnology and Microbiology, University of Swat, Swat, KP, Pakistan.

Shahid Ali

Center for Biotechnology and Microbiology, University of Swat, Swat, KP, Pakistan.

Muhammad Suleman

Center for Biotechnology and Microbiology, University of Swat, Swat, KP, Pakistan.

Abdullah Shah

Department of Biotechnology, Shaheed Benazir Bhutto University, Sheringal, Dir, Pakistan.

Liaqat Ali

Faculty of Biological Sciences, National University of Medical Sciences (NUMS), Rawalpindi, Pakistan.

Syed Shujait Ali

Center for Biotechnology and Microbiology, University of Swat, Swat, KP, Pakistan.

Dong-Qing Wei (✉ dqwei@sjtu.edu.cn)

Department of Bioinformatics and Biological Statistics, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, 200240, P.R. China.

Research

Keywords: Peptide, Docking, in silico mutagenesis, simulation, Free Energy, PCA

Posted Date: September 23rd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-76731/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background:

The prolific spread of COVID-19 caused by a novel coronavirus (SARS-CoV-2) from its epicenter in Wuhan, China, to every nook and cranny of the world after December 2019, jeopardize the prevailing health system in the world and has raised serious concerns about human safety. To date efforts are continuing to design small molecule inhibitor, vaccines and many other therapeutic options are practiced but their final therapeutic potential is still to be tested. Using the old drug or vaccine or peptides could aid this process to avoid such long experimental procedure.

Results:

Hence, here we have repurposed a small peptide (ATLQAIAS) from the previous study which reported the inhibitory effects of this peptide. We used in silico mutagenesis approach to design more peptides from the native wild peptide, which revealed that substitutions (T2W, T2Y, L3R and A5W) could increase the binding affinity of the peptide towards the 3CLpro. Furthermore, using MD simulation and free energy calculation confirmed its dynamics stability and stronger binding affinities. Per-residues energy decomposition analysis revealed that the specified substitution significantly increased the binding affinity at residue level.

Conclusion:

Our wide-ranging analyses of binding affinities disclosed that our designed peptide owns the potential to hinder the SARS-CoV-2 and will reduce the progression of SARs-CoV-2-borne pneumonia. Our analysis strongly suggests the experimental and clinical validation of these peptides to curtail the recent corona outbreak.

Background

The prolific spread of COVID-19 caused by a novel coronavirus (SARS-CoV-2) from its epicenter in Wuhan, China, to every nook and cranny of the world after December 2019, jeopardize the prevailing health system in the world and has raised serious concerns about human safety [1]. The World Health Organization (WHO), on March 20, 2020, declared COVID-19 as pandemic due to its worldwide penetration in a few months. The updates of August 07, 2020, have been reported 19.2 million infections and 717,754 death cases. The symptoms manifested by most patients with COVID-19 include fever, coughing, dyspnea, myalgia, shortness of breath, and radiological evidence of ground-glass lung opacities compatible with atypical pneumonia. However, some patients with asymptomatic or mildly symptomatic cases have also been reported [2–4].

The members of subfamily Orthocoronavirinae (family Coronaviridae) are distributed within four genera: alpha (α), beta (β), gamma (γ) & delta (δ) [5, 6]. The host of alpha and beta coronaviruses are mammals,

whereas the host of gamma and delta coronaviruses are birds [7, 8]. The major epidemics of 2003, 2012 and 2019 (caused by SARS-CoV-1, MERS and SARS-CoV-2, respectively) revealed the vulnerability of humanity before the growing threat and offensive of the members of beta-coronaviruses[9, 10]. The SARS-CoV-1, MERS-CoV, and SARS-CoV-2 transmitted from bats to palm civets, dromedary camels, and pangolin, respectively. These coronaviruses became lethal due to their ability of transmission from animals to humans as well as from humans to humans [11]. CFR (case fatality ratio: calculated by dividing the number of deaths from the disease by a number of diagnosed cases of disease and multiplied by 100) of SARS-CoV-1, MERS-CoV, and SARS-CoV-2 were reported 10%, 35%, and 5% respectively. The phenomenal spread of SARS-CoV-2 across the globe exposed everyone to this lethal pathogen, and investigations are underway to use various approaches such as novel vaccine development, drug repurposing and novel drugs development to mitigate risks and threats associated with SARS-CoV-2 [12].

The SARS-CoV-2 has a genome of ~ 30 kb and is responsible for encoding the whole proteome of the pathogen. Both the structural and non-structural protein-coding region and accessory protein-coding region are the major parts of the intact coding RNA[13]. The four structural proteins constituting the main envelope of the virus, and eight accessory proteins are encoded by the genes located on the 3'-terminus. The structural protein encodes for the spike surface glycoprotein (S), small envelope protein (E), membrane protein (M) and nucleocapsid protein (N) whereas the eight accessory proteins codes for 3a, 3b, p6, 7a, 7b, 8b, 9b and ORF14[14]. The non-structural proteins constitute the replication/transcription complex (RTC) that are encoded by two overlapping genes; ORF1a and ORF1ab, located on the 5'-terminus of the viral genome. The OFR1a and OFR1ab encode two long polypeptides, pp1a and pp1ab, respectively. These polypeptides, pp1a and pp1ab, are directly translated and then proteolytically processed by two main viral proteases, papain-like protease (PLpro) and 3-chymotrypsin-like protease (3CLpro) also known as main proteases (Mpro)[15]. The former one is responsible for the cleavage of non-structural proteins (nsp) 1, 2, and 3 while the later (3CLpro) cleaves the polyprotein at 11 discrete sites downstream of nsp4 to produce different non-structural proteins that play a pivotal role in the viral life cycle. On account of these multi-faceted roles, the 3CLpro is considered an attractive target for anti-coronaviruses drug development[16].

Earlier studies suggested an important role of 3CLpro (Mpro) in cell replication and maturation; therefore, inhibition of this target will greatly contribute in controlling the COVID-19. 3CLpro composed of three domains I (8-101), II (102–184), and III (201–303). Domain II and III three are connected through a loop (185–200)[17]. In this study in Silico mutagenesis-based remodeling of SARs-CoV-1 peptide (ATLQAIAS) was performed to inhibit the activities of 3CLpro of SARs-CoV-2 for the curtailment of COVID-19. The activity was confirmed by molecular dynamics simulation and free energy calculations. This study provides a basis for designing peptide inhibitors against the SARs-CoV-2 to confront this pandemic.

Materials And Methods

Structure retrieval and Modeling of Peptide (ATLQAIAS)

The crystal structure of the 3CLpro complex was retrieved from the protein databank (<http://www.rcsb.org/>) and assessed for Structural deformities[18]. The peptide sequence (ATLQAIAS) was obtained from the previous literature and modeled using PEPFOLD3(<https://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD3/>)[19], which is an online web server for peptides modeling. Previous computational and experimental studies reported that this peptide, along with others, remained active against the SARS virus and potentially inhibited the virus replication[20, 21].

Protein-peptide docking

To test the activity of the (ATLQAIAS) peptide against the SARs-CoV-2, molecular docking of the peptide and 3CLpro from SARs-CoV-2 was carried out to understand the interaction pattern. For this purpose, ZDOCK (<http://zdock.umassmed.edu/>)[22] was used with a blind docking option to increase and test the reliability of the server. The top ten docked conformations were retrieved and analyzed for the interaction with the key residues such as His41 and Cys145, which form a catalytic dyad and do the proteolysis.

Interface analysis and Peptide library construction

Understanding the binding interface of two interacting protein or peptide is vital for designing small peptide inhibitors. The use of small peptides derived from the native peptide has been widely used in the designing of short peptide derivatives. A similar strategy to manipulate the interface residues of the reference-native-peptide-3CLpro and to design decoy peptides against SARs-CoV-2 using the machine learning protocol implemented in MOE[23]. The initial docking of the peptide (ATLQAIAS) with the 3CLpro confirmed that it also blocks the catalytic dyad and form supplementary interaction with the key active site residues. After this confirmation, information regarding the key residues in the interface is important for mutants modeling. For this purpose, the Alanine scanning strategy was exercised to calculate the impact of each residue in the interaction with the 3CLpro. To grasp this strategy, the dAffinity and dStability parameters were monitored in the ASM (alanine scanning module) of MOE. The dAffinity and dStability values show the relative change in binding energy when a particular amino acid is changed into alanine. These scores reveal important information regarding the importance of a particular residue, after the confirmation of key residues selection of the hotspot residues for residue scan to be replaced by the 19 amino acids. For this purpose, the residue scan module of MOE was used using the Unary Quadratic Optimization (UQO) parameter under the LowModeMD ensemble. This residue scan generated a database of mutant peptides with their respective scores. The detailed mechanism of this alanine scanning mutagenesis and residue scan approaches has been discussed previously[24].

Molecular Dynamics Simulation

A comprehensive protein dynamics technique was used to evaluate the dynamic features of the 3CLpro and the designed peptides via the residue scan method. Rationally developed peptides with fierce dAffinity and dStability scores were obtained using the Amber18 package and subjected to molecular dynamics simulation (MDS) with the ff14SB force field [25]. Using the TIP3P water model with box dimension 10.0Å, each structure was correctly solved and neutralized by incorporating Na⁺ counter ions.

Two steps of energy minimization were performed. For the first energy minimization, 6000 steps, while for the second 3000 steps of conjugate gradient minimization, was completed. Following the minimization, each system was heated using default parameters (300 K for 200 ps). Weak restraint was used for density equilibration for 2 ns, while the whole system at a constant pressure for 2 ns. A 100 ns MD under constant pressure was performed. For the temperature control, Langevin thermostat (1 atm, 300 K) was used [26]. Particle Mesh Ewald (PME) algorithm was used to compute long-range interactions [27, 28]. The cutoff distances were set to 10 Å. For the covalent bonds involving hydrogen, the SHAKE algorithm was used [29]. GPU accelerated simulation using (PMEMD.CUDA) was used for all the processes.

Post-simulation Analyses

Post simulation analyses such as RMSD, RMSF, and radius of gyration (Rg) were calculated by using CPPTRAJ and PTRAJ modules of Amber [30].

Binding affinity calculations

To estimate the binding free energy of the designed peptides towards the 3CLpro MMPBSA.PY script was used (Miller et al., 2012). Using 5000 frames from the 100 ns trajectory, the free energy of binding was estimated by using the following equation:

$$\Delta G_{bind} = \Delta G_{complex} - [\Delta G_{receptor} + \Delta G_{ligand}]$$

In the above equation, the ΔG_{bind} represent the total BFE while the other components in the equation represent BFE of the complex, the protein, and ligand molecules. Each term in the binding free energy was estimated using the following equation:

$$G = G_{bond} + G_{ele} + G_{vdW} + G_{pol} + G_{npol} - TS$$

The above equation represents the polar, nonpolar, electrostatic, solvent-accessible surface area SASA, and van der Waals interactions, respectively.

Per-residues Energy Decomposition analysis

Per-residue energy decomposition is the best approach to understand the energetic contribution at the residues level significantly. Herein, to understand the impact of each substitution on the binding of the designed peptides, we used per-residue energy decomposition analysis.

Clustering of MD trajectories using PCA

To comprehend the motion of MD trajectories, an unsupervised learning method known as Principal Component Analysis (PCA) [31, 32] was performed to acquire knowledge regarding the internal motion of the system. For this purpose, an Amber module known as CPPTRAJ was used. The spatial covariance matrix was determined for eigenvector and their atomic co-ordinates. Using the orthogonal coordinate transformation, a diagonal matrix of eigenvalues was generated. Based on the eigenvectors and

eigenvalues, the principal components were extracted. Using these PCs, the dominant motions during the simulation were plotted[33, 34].

Results

Structure Retrieval and Preparation

The crystal structure of the SARs-COV-2 3CLpro was retrieved from RCSB using accession number 6LU7. The structure was prepared and minimized using Galaxyweb server[35]. The peptide sequence (ATLQAIAS) was retrieved from previous literature, which reported that ATLQAIAS inhibits the SARs-CoV-1 by targeting the main protease 3CLpro. The peptide was modeled and prepared for docking. 3CLpro has three domains with a connecting loop, which connects domains II and III. The structure of 3CLpro and the peptide (ATLQAIAS) is given in Fig. 1.

The interaction analysis of the wild (ATLQAIAS) and 3CLpro structure

The ZDOCK server predicted different conformations for the peptide docked at different positions. The conformation surrounds the important catalytic dyad (His and Cys45) was selected. Using ZDOCK the initial interaction pattern was obtained. It can be seen that the native peptide ATLQAIAS formed eight hydrogen bonds with the active site residues. It can be seen that Thr26, Thr45, Ser46, Asn142, Gly143, Cys145, and Gln189 are involved in hydrogen bonding with the octapeptide. Among these residues, Cys145 acts as a catalytic dyad and is vital for the function of the 3CLpro. Also, Gln189 from the active site formed two hydrogen bonds. Hence, this confirms that the octapeptide acts as an inhibitor for the 3CLpro and ultimately blocks the SARs-CoV-2 protease function. The interaction pattern of the native peptide (ATLQAIAS) is given in Fig. 1.

Residue scan to design a peptide library

The interface residues of the peptide and the 3CLpro were identified by using the residue scan approach, an integrated module in the Molecular Operating Environment (MOE). In the native peptide, which was taken as reference peptide, the key hotspot residues include Ala1, Thr2, Leu3, and Ala5, which forms significant interactions with the 3CLpro however the other residues were involved as supplementary and were found out to be less significant at the interface to make interactions. In the first step, we used alanine scanning approach to select the most appropriate residue for fixed amino acid substitution to increase the binding affinity of the native peptide. To scan for the most important residues, alanine scanning reported the dAffinity of each residue substitution, which calculates the relative binding affinity changes upon substitution. The alanine scanning analysis suggested that five residues are preferable for substitution, which could impact the binding of the peptide. These residues include Thr2, Leu3, Ala5, and Ala7 and could be replaced by other residues to increase the binding affinity of the native reference peptide. Nineteen different amino acids replaced each selected residue, and the binding affinity was

calculated for each. This residue scan analysis was observed via dAffinity, which was measured as a difference between the binding energies of the wild type residue and the corresponding mutated amino acid residue. A mutational space of 95 mutations was designed and searched. By residue scanning, four different peptides were generated based on Affinity and dAffinity scores given in Table 1. The interface and the designed peptide library structure is given in Fig. 2.

Table 1: Final peptides library generated from residue scan approach based on the total affinity and dAffinity given in kcal/mol. The respective substitution is highlighted in red while the residues position is also given.

S. No	Peptide Sequence	Substation position	Affinity (kcal/mol)	dAffinity (kcal/mol)
1.	ATLQAIAS	Wild	-7.4	00
2.	A W LQAIAS	T2W	-11.70	-1.14
3.	A Y LQAIAS	T2Y	-11.56	-1.00
4.	AT R QAIAS	L3R	-11.64	-1.03
5.	ATLQ W IAS	A5W	-10.99	-0.62

Interaction analysis of the designed decoy peptides with 3CLpro

As given in Table 1, it can be seen that the substitution at position (**AWLQAIAS**) T2W increased the binding affinity from - 9.4 kcal/mol to -11.70 kcal/mol. The substitution of small polar threonine by large aromatic Tryptophan increased the binding affinity by forming an extra hydrogen bond with Gln189 of the 3CLpro. As given in Fig. 3(A) an extra h-bond is also formed with His41, which cannot be seen in the wild type. In the case of (**AYLQAIAS**) T2Y substitution the binding affinity was observed to be -11.56 kcal/mol while the dAffinity was reported to be -1.00 kcal/mol. This substitution of Tyrosine, an aromatic amino acid, favours an extra interaction with Thr190 along with those reported in T2W. The replacement of (**ATRQAIAS**) L3R shift the interaction paradigm and favors significant interactions through the binding energy is relatively lower than the first two peptides. Ten H-bonds were formed with the key active site residues, including the catalytic dyad His41 and Cys145. Furthermore, a single hydrogen bond was formed with Thr26, Asn142, and Glu166 was reported while two H-bonds with Meth165 and three H-Bonds with Gln189 were formed. The affinity for L3R was reported to be -11.64 kcal/mol. In the case of A5W (**ATLQWIAS**) substitution, extra interactions with the key residue His41 was observed. Here the main interacting residues include Thr25, Thr26, Asn142, Gly143, Glu166, and Gln189. Hence, the substitution makes a more favorable environment for interaction with the 3CLpro and suggests that the designed mutant peptides efficiently interact with the 3CLpro and block the key residues which are required for the function of the protein. The interaction pattern of the designed decoy peptides is given in Fig. 3.

Dynamics stability of 3CLpro-peptides complexes

To investigate the dynamic properties of these designed peptides in complex with 3CLpro molecular dynamics simulation of all the complexes was performed. The root mean square deviation as a function of time was calculated to quantify the stability of each system. Herein, in the case of the wild type, the system showed invariant behavior at a different interval. Initially, the RMSD was found to be ~ 1.5 Å until 40 ns, but then later on, the RMSD converged at different intervals until 100 ns. On the other hand, the T2W system relatively showed higher RMSD, but no major convergence was observed like the wild type. The average RMSD for the T2W was ~ 2.0 Å. In addition, the T2Y complex remained more stable when compared to the wild and T2W. The RMSD remained ~ 1.8 Å with a small convergence between 58–60 ns, but then until 100 ns, the system entered the equilibration state again. In the case of L3R, the average RMSD remained ~ 2.0 Å, with an acceptable convergence between 55–60 ns. On the other hand, the A5W system also remained stable, with a convergence between 87–88 ns. The average RMSD for A5W was also reported to be ~ 2.0 Å. These results show that the mutant systems relatively remained stable during the course of simulation than the wild type and imply that the mutant decoy peptides bind more tightly than the wild type and favours the stability of the whole complex. All the RMSD graphs are given in **Fig. 4**.

Residual Flexibility of the complexes

The RMSF graphs revealed that, in each formulated peptide, substitution has a specific impact on the residues flexibility. As given in Fig. 5, the wild type showed relatively high residual flexibility, which is in uniformity with the RMSD results. The wild type was reported to be relatively unstable; hence the residual flexibility also remained higher than the others. It can be seen that the other systems exhibited a more similar pattern of residual flexibility. In the case of the T2W and T2Y, the flexibility increased at the substitution point while the L3R and A5W flexibility decreased. These results reveal that the designed decoy peptides bind to the protein target more efficiently than the wild type and affect the residual dynamics by stabilizing the flexibility as compare to the wild type. Figure 5 is showing the flexibility of each system in a different colour.

The compactness of the peptide bound systems

The Gyration radius (RoG) was calculated to understand the compactness of each peptide bound system. Herein, the Rg calculation was performed to understand how the peptides remained intact with the protein target during the course of the simulation. This compactness also determines the stability of the system. As shown in Fig. 6, the wild type system showed highly converged Rg. Initially, the Rg value was 22.2 Å, but after reaching 20 ns, the Rg of the wild type converged all of a sudden, and the Rg value increased to 23 Å. The Rg value fell between 20–40 ns but remained higher until 100 ns. This shows that the wild type system is not that stable, and the peptide is probably released and the water entered the active pocket and lessened the peptide binding.

On the other hand, the R_g of the T2W remained stable, and the system showed more compactness. A little convergence between 75–80 ns was observed, but overall the system remained highly compact, and the average R_g value was reported to be 22.2 Å. In the case of T2Y, the system was comparatively more compact than the wild type and T2W. The average R_g value for T2Y was 22.2 Å. The R_g value for T2Y decreased after 50 ns, which shows the tight binding of the peptide to the receptor. In the case of L3R, the systems remained compact, but convergence between 80–100 ns was observed, which is in comparison with the RMSD, which shows a little unstable behaviour. The average R_g for L3R was observed to 22.2 Å. In addition, the R_g value for the A5W was reported to be 22.1 Å. No major convergence was observed except for a fluctuation at 40 ns and 80 ns. These results confirmed that the designed decoy peptides remained intact with the active site residues during the simulation time period. These results implied that the decoy peptides efficiently bind and inhibit 3CLpro. R_g (s) of all the systems are given in **Fig. 6**.

Binding Free Energy Calculation

A common technique, MM-GBSA, has been used to accurately predict the binding free energy of all the complexes. The Gibbs free energy determines the binding affinity between interacting molecules. Each energy term, including vdW, electrostatic energy, polar solvation, solvent accessible surface area, and total binding free energy of all the systems, are given in **Table 3**. As given in the total binding energy for the wild reference peptide is -59.13 kcal/mol. The vdW for wild reference peptide was reported to be -72.73 kcal/mol, electrostatic - 110.34 kcal/mol, while SASA - 8.56 kcal/mol. For T2W, the total binding energy was observed to be -70.04, while the vdW was reported to be -81.01 kcal/mol. The electrostatic energy for T2W was reported to be -111.12 kcal/mol, while SASA was observed to be -9.6 kcal/mol. For the decoy peptide T2Y the total binding energy was reported to be -66.47 kcal/mol. The vdW and electrostatic energies for T2Y were observed to be -79.08 kcal/mol and - 106.72 kcal/mol, respectively. The SASA for T2Y was reported to be -9.55 kcal/mol. Furthermore, the total binding energy for L3R was reported to be -67.47 kcal/mol with - 80.73 kcal/mol vdW, -225.82 kcal/mol electrostatic and - 9.52 kcal/mol SASA.

On the other hand, the total binding energy for A5W was reported to be -65.55 kcal/mol. In the case of A5W, both the vdW and electrostatic energies increased when compared to the wild type. The vdW for A5W was reported to be -76.45 kcal/mol, while electrostatic for A5W was observed to be -124.13 kcal/mol. These results significantly suggest that the designed decoy peptides possess strong inhibitory potential than the wild reference peptide. The contribution of the substituted residues increases the vdW and electrostatic energies significantly and eventually increases the total binding energy. The total binding energy and all the related terms are given in Table 2.

Table 2: The total binding energy of all the systems, including the wild type and the designed decoy peptides. All the energies are given in Kcal/mol.

S. No	Peptide Sequence	vdW	Elec	GB	SASA	G Total
1.	ATLQAIAS	-72.73	-110.34	132.51	-8.56	-59.13
2.	A W LQAIAS	-81.01	-111.12	131.69	-9.6	-70.04
3.	A Y LQAIAS	-79.08	-106.72	128.88	-9.55	-66.47
4.	A T RQAIAS	-80.73	-225.82	248.86	-9.52	-67.20
5.	ATLQ W IAS	-76.45	-124.13	144.31	-9.28	-65.55

Per-Residue Energy decomposition analysis

Furthermore, to understand the energetic contribution at residues level we used per-residues energy decomposition approach. This analysis further clarified the impact of these substitution and its contribution to the total energy. In case of the wild type Threonine at position 2, -1.19 kcal/mol. Upon mutation to Tryptophan the energy increased up to -7.52 kcal/mol. At the same position when Threonine was replaced by Tyrosine the energy observed was - 5.2 kcal/mol. With this substitution the energy of the Glutamine at position 4 also increased to -4.35 kcal/mol. In case of the L3R peptide the energy paradigm changed by at different residues. On the other hand, the substitution at position 5, replacing alanine by Tryptophan, the energy changed from - 0.1 kcal/mol to -2.92 kcal/mol. These results significantly justify the impact of the substitution on the total binding affinity of the designed decoy peptides. The bar graphs representing the residual contribution of each peptide are given in Fig. 7.

Principal Motions of the Wild Type and decoy designed Peptides

PCA was performed and plotted to pinpoint the principal structural variations exhibited by each system enacted by the binding of the designed peptides. The objective of the PCA analysis was to extract the knowledge on the conformational states of the Wild Type and the designed peptides complexes, using the trajectory of 100 ns. Significant motion in all the systems for the first three eigenvectors was observed. The first three eigenvectors indicated significant fluctuations, while the remaining eigenvectors showed localized fluctuations in each complex (Fig. 8). In the case of wild type peptide complex, the first three eigenvectors contributed 72% variance to the total observed motion, while in T2W 72%, in T2Y 42%, in L2R and A5W accounted for 33% variance in motion. This behavior may explain the structural rearrangement due to the peptides binding.

The first two eigenvectors were plotted against each other in order to gain conceivable attributed motions. The continuous representation of the red to blue color shows the switching along simulation time from one conformation to another. The dots represent each frame, beginning with red and ending in blue. In the Wild type and T2Y peptide complexes, more periodic jumps and continuous overlapping can be observed, while in the case of T2W, L3R and A5W showed localized fluctuations (Fig. 9). Together, all these observations imply that mutations significantly affect the structure and changes the internal dynamics of the complexes.

Discussion

The prolific spread of COVID-19 caused by a novel coronavirus (SARS-CoV-2) from its epicenter in Wuhan, China, to every nook and cranny of the world after December 2019, jeopardize the prevailing health system in the world and has raised serious concerns about human safety [1]. The World Health Organization (WHO), on March 20, 2020, declared COVID-19 as pandemic due to its worldwide penetration in a few months. The updates of August 07, 2020, have been reported 19.2 million infections and 717,754 death cases. The symptoms manifested by most patients with COVID-19 include fever, coughing, dyspnea, myalgia, shortness of breath, and radiological evidence of ground-glass lung opacities compatible with atypical pneumonia. However, some patients with asymptomatic or mildly symptomatic cases have also been reported[2–4]. The enormous number of confirmed COVID-19 positive (19.2 million) and death cases (717,754) as reported in updates of August 07, 2020 suggests the failure of the entire world to cope with this critical prevailing corona situation. The fast penetration of SARS-CoV-2 across the globe had adversely affected the human health, therefore researchers adopted various strategies to combat and control the havoc created by this pathogen. Computational biology approaches are very fast in identifying the possible molecular targets for novel vaccine and drug designing. Previous studies suggested the use of spike glycoprotein, proteases and RNA dependent RNA polymerases for controlling various diseases caused by diverse pathogens. The approaches such as novel vaccine designing, drug repurposing, novel drug development and protein–protein interactions (PPIs) were applied to find a workable solution to stop this deadliest pathogen and provide a feasible protection to the masses.

To date efforts are continuing to design small molecule inhibitor, vaccine and many other therapeutic options are practiced. Many companies claimed success in vaccine design but the recent re-infection and their long term effectivity is still questionable. Therefore designing such therapeutic small molecules or peptide inhibitors is strongly suggested which are of long term benefit. In this regard many studies reported thousands of small molecules and peptides but their final therapeutic potential is still to be tested. The notion of protein-protein interactions (PPIs) is gaining popularity day by day due to its applications in deciphering many biological processes. Thus, PPIs can be used as promising targets by incorporating new modulating entities into PPIs network. Using the old drug or vaccine or peptides could aid this process to avoid such long experimental procedure. Hence, here we have repurposed a small peptide from the previous study which reported the inhibitory effects of this peptide.

Previous study reported that peptides could be a starting point for the drug discovery against SARS. Using molecular modeling and experimental approaches several peptides which are similar in structure and pharmacophoric features were designed and studies using molecular modeling and in vitro assays. Here utilizing the knowledge from previous study that a peptide (ATLQAIAS) significantly halt the replication of the SARS by targeting the 3CLpro. We repurposed the previous peptide against the SARS-CoV-2 by using computational pipeline. The peptide (ATLQAIAS) was docked against the 3CLpro from SARS-CoV-2. This analysis suggest that the peptide interacts with the catalytic dyad of the 3CLpro and blocks its important residues. Considering the initial knowledge, a useful residue scan approach was utilized to improve the native peptide based on in silico mutagenesis approach. The in silico mutagenesis replaced each amino acid by 19 substitutes and search for high affinity peptides based on the dAffinity

criteria. Our analysis yield only four peptides which possess stronger interactions and energy than the native wild type peptide. The activity of these rationally designed peptides was further confirmed by using molecular dynamics simulation and free energy calculations. Our analysis suggest that the designed novel peptides possess stable dynamics behaviour and compactness. On the other hand, the binding free energy revealed that all our designed peptides possess stronger binding affinity than the native wild type. This was also confirmed by the per-residue energy decomposition analysis which revealed that the mutated residues changed the energy level of each residue. In addition, the clustering of the MD trajectories also revealed the variation in the dynamics of each peptide.

Conclusion

In conclusion these analyses revealed a new paradigm for the development of peptide inhibitors against the SARs-CoV-2. Using such long computational pipeline from basic molecular modeling to free energy calculations confirmed the behaviour and inhibitory properties of the designed mutant peptides. Our wide-ranging analyses of binding affinities disclosed that our designed peptide owns the potential to hinder the SARS-CoV-2 and will reduce the progression of SARs-CoV-2-borne pneumonia. Our analysis strongly suggests the experimental and clinical validation of these peptides to curtail the recent corona outbreak.

Abbreviations

RMSD

Root mean square deviation

RMSF

Root mean square fluctuation

MD

molecular dynamics simulation

COVID-19

Coronavirus disease

Rg

Radius of gyration

PCA

principal component analysis

SARS

severe acute respiratory syndrome

WHO

World Health Organization

Declarations

Ethics approval and consent to participate

Not applicable

Consent for Publication

Not Applicable

Availability of data and materials

All the data will be provided on reasonable demand.

Competing Interest

The authors declare no conflict of interest.

Funding

Dong-Qing Wei is supported by the grants from the Key Research Area Grant 2016YFA0501703 of the Ministry of Science and Technology of China, the National Natural Science Foundation of China (Contract no. 61832019, 61503244), the Natural Science Foundation of Henan Province (162300410060) and Joint Research Funds for Medical and Engineering and Scientific Research at Shanghai Jiao Tong University (YG2017ZD14).

Acknowledgments

The computations were partially performed at the Center for High-Performance Computing, Shanghai Jiao Tong University. We acknowledge their help.

Authors' Contribution

AK, SU, AH, RF conceptualized the study. AK, ZB, MW, ZH did the analysis. AK, MR, NZ, SA, AS, LA and SS wrote the manuscript, revised, visualization and finalizing the final draft. DQW is an academic supervisor. He supervised the study.

References

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R: **A novel coronavirus from patients with pneumonia in China, 2019.***New England Journal of Medicine* 2020.
2. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X: **Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China.***The lancet* 2020, **395**:497-506.
3. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N: **Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding.***The Lancet* 2020, **395**:565-574.

4. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, Qiu Y, Wang J, Liu Y, Wei Y: **Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study.***The Lancet* 2020, **395**:507-513.
5. Guan W-j, Ni Z-y, Hu Y, Liang W-h, Ou C-q, He J-x, Liu L, Shan H, Lei C-l, Hui DS: **Clinical characteristics of 2019 novel coronavirus infection in China.***MedRxiv* 2020.
6. Weiss SR, Leibowitz JL: **Coronavirus pathogenesis.** In *Advances in virus research. Volume 81*: Elsevier; 2011: 85-164
7. Chan JF, Lau SK, To KK, Cheng VC, Woo PC, Yuen K-Y: **Middle East respiratory syndrome coronavirus: another zoonotic betacoronavirus causing SARS-like disease.***Clinical microbiology reviews* 2015, **28**:465-522.
8. Li F: **Structure, function, and evolution of coronavirus spike proteins.***Annual review of virology* 2016, **3**:237-261.
9. Tang Q, Song Y, Shi M, Cheng Y, Zhang W, Xia X-Q: **Inferring the hosts of coronavirus using dual statistical models based on nucleotide composition.***Scientific reports* 2015, **5**:17155.
10. Song Z, Xu Y, Bao L, Zhang L, Yu P, Qu Y, Zhu H, Zhao W, Han Y, Qin C: **From SARS to MERS, thrusting coronaviruses into the spotlight.***Viruses* 2019, **11**:59.
11. Khan A, Khan M, Saleem S, Babar Z, Ali A, Khan AA, Sardar Z, Hamayun F, Ali SS, Wei D-Q: **Phylogenetic analysis and structural perspectives of RNA-dependent RNA-polymerase inhibition from SARs-CoV-2 with natural products.***Interdisciplinary Sciences: Computational Life Sciences* 2020:1-14.
12. Khan A, Ali SS, Khan MT, Saleem S, Ali A, Suleman M, Babar Z, Shafiq A, Khan M, Wei D-Q: **Combined drug repurposing and virtual screening strategies with molecular dynamics simulation identified potent inhibitors for SARS-CoV-2 main protease (3CLpro).***Journal of Biomolecular Structure and Dynamics* 2020:1-12.
13. Durojaiye AB, Clarke J-RD, Stamatiades GA, Wang C: **Repurposing cefuroxime for treatment of COVID-19: a scoping review of in silico studies.***Journal of Biomolecular Structure and Dynamics* 2020:1-8.
14. Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G, Hu Y, Tao Z-W, Tian J-H, Pei Y-Y: **A new coronavirus associated with human respiratory disease in China.***Nature* 2020, **579**:265-269.
15. Perlman S, Netland J: **Coronaviruses post-SARS: update on replication and pathogenesis.***Nature reviews microbiology* 2009, **7**:439-450.
16. Hatada R, Okuwaki K, Mochizuki Y, Handa Y, Fukuzawa K, Komeiji Y, Okiyama Y, Tanaka S: **Fragment molecular orbital based interaction analyses on COVID-19 main protease-inhibitor N3 complex (PDB ID: 6LU7).***Journal of chemical information and modeling* 2020.
17. Needle D, Lountos GT, Waugh DS: **Structures of the Middle East respiratory syndrome coronavirus 3C-like protease reveal insights into substrate specificity.***Acta Crystallographica Section D: Biological Crystallography* 2015, **71**:1102-1111.
18. Rose PW, Prlić A, Altunkaya A, Bi C, Bradley AR, Christie CH, Costanzo LD, Duarte JM, Dutta S, Feng Z: **The RCSB protein data bank: integrative view of protein, gene and 3D structural information.***Nucleic*

acids research 2016:gkw1000.

19. Lamiable A, Thévenet P, Rey J, Vavrusa M, Derreumaux P, Tufféry P: **PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex.***Nucleic acids research* 2016, **44**:W449-W454.
20. Gan Y-R, Huang H, Huang Y-D, Rao C-M, Zhao Y, Liu J-S, Wu L, Wei D-Q: **Synthesis and activity of an octapeptide inhibitor designed for SARS coronavirus main proteinase.***Peptides* 2006, **27**:622-625.
21. Zhang R, Wei D-Q, Du Q-S, Chou K-C: **Molecular modeling studies of peptide drug candidates against SARS.***Medicinal Chemistry* 2006, **2**:309-314.
22. Pierce BG, Wiehe K, Hwang H, Kim B-H, Vreven T, Weng Z: **ZDOCK server: interactive docking prediction of protein–protein complexes and symmetric multimers.***Bioinformatics* 2014, **30**:1771-1773.
23. Vilar S, Cozza G, Moro S: **Medicinal chemistry and the molecular operating environment (MOE): application of QSAR and molecular docking to drug discovery.***Current topics in medicinal chemistry* 2008, **8**:1555-1572.
24. Junaid M, Shah M, Khan A, Li C-D, Khan MT, Kaushik AC, Ali A, Mehmood A, Nangraj AS, Choi S: **Structural-dynamic insights into the H. pylori cytotoxin-associated gene A (CagA) and its abrogation to interact with the tumor suppressor protein ASPP2 using decoy peptides.***Journal of Biomolecular Structure and Dynamics* 2019, **37**:4035-4050.
25. Salomon-Ferrer R, Case DA, Walker RC: **An overview of the Amber biomolecular simulation package.***Wiley Interdisciplinary Reviews: Computational Molecular Science* 2013, **3**:198-210.
26. Zwanzig R: **Nonlinear generalized Langevin equations.***Journal of Statistical Physics* 1973, **9**:215-220.
27. 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**:10089-10092.
28. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG: **A smooth particle mesh Ewald method.***The Journal of chemical physics* 1995, **103**:8577-8593.
29. Ryckaert J-P, Ciccotti G, Berendsen HJ: **Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes.***Journal of computational physics* 1977, **23**:327-341.
30. Roe DR, Cheatham III TE: **PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data.***Journal of chemical theory and computation* 2013, **9**:3084-3095.
31. Wold S, Esbensen K, Geladi P: **Principal component analysis.***Chemometrics and intelligent laboratory systems* 1987, **2**:37-52.
32. Pearson K: **LIII. On lines and planes of closest fit to systems of points in space.***The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 1901, **2**:559-572.
33. Balsera MA, Wriggers W, Oono Y, Schulten K: **Principal component analysis and long time protein dynamics.***The Journal of Physical Chemistry* 1996, **100**:2567-2572.

34. Ernst M, Sittel F, Stock G: **Contact-and distance-based principal component analysis of protein dynamics.***The Journal of chemical physics* 2015, **143**:12B640_641.
35. Ko J, Park H, Heo L, Seok C: **GalaxyWEB server for protein structure prediction and refinement.***Nucleic acids research* 2012, **40**:W294-W297.

Figures

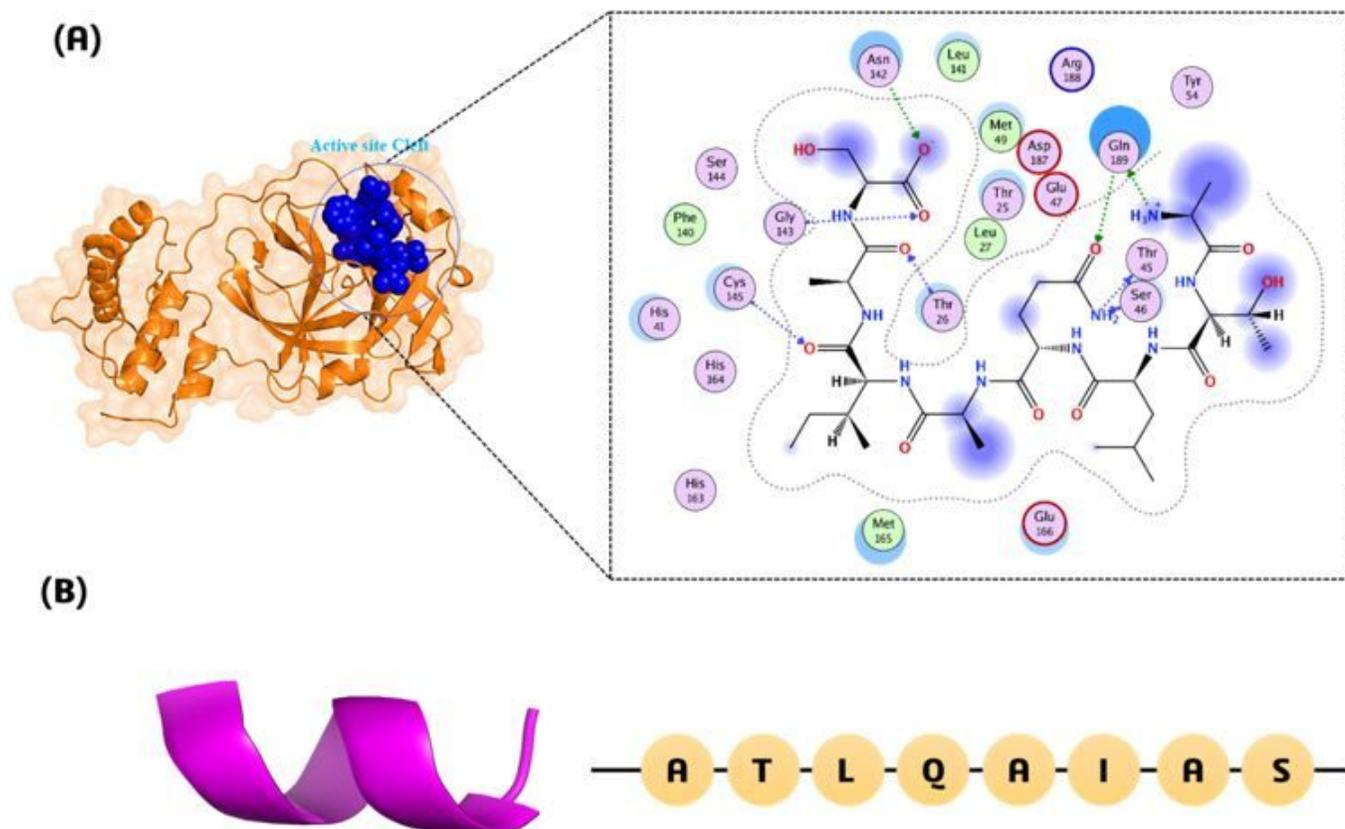


Figure 1

The 3D structure of 3CLpro in brown colour while the active site is shown in balls in blue colour. The Magenta colour shows the ribbon representation of the octapeptide (ATLQAIAS). The interaction of the octapeptide with the 3CLpro key residues such as Thr26, Thr45, Ser46, Asn142, Gly143, Cys145, and Gln189 is given in the panel (A) right side.

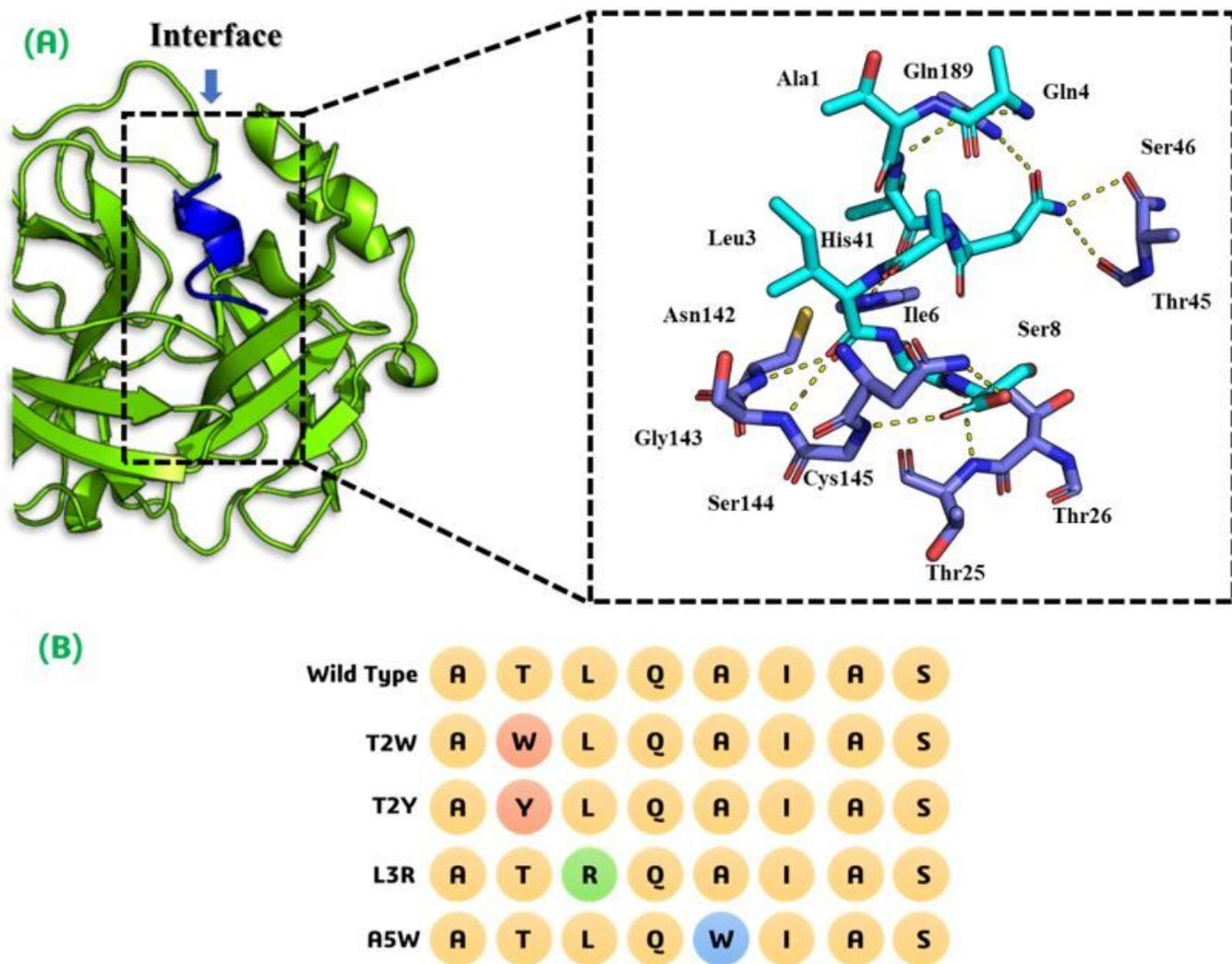


Figure 2

(A) The interface analysis of the wild peptide and the 3CLpro. (B) showing the designed peptide library, and each amino acid replacement is coloured differently.

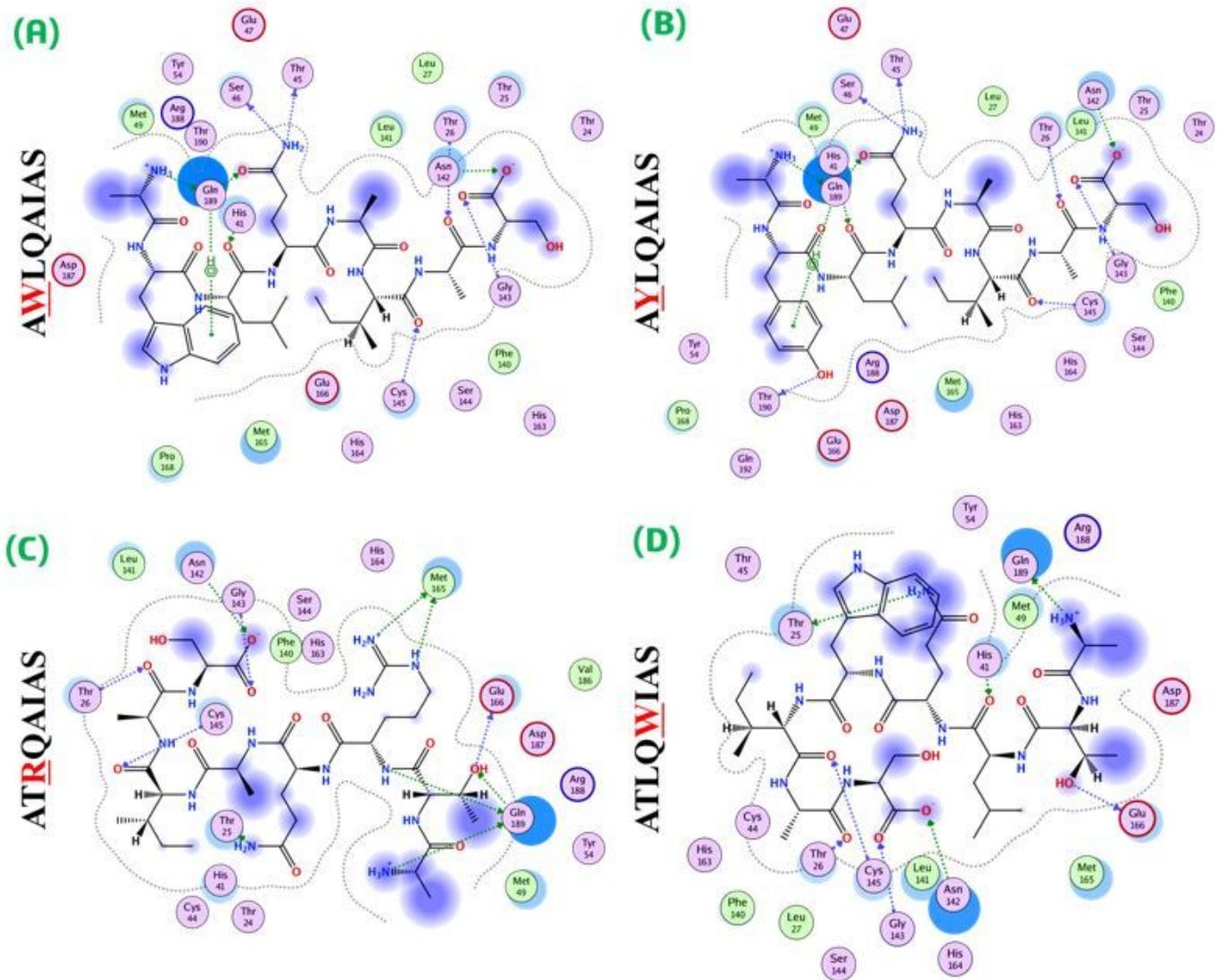


Figure 3

Interaction pattern of the designed decoy inhibitory peptides of SARS-CoV-2 3CLpro. (A) AWLQAIAS, (B) AYLQAIAS, (C) ATRQAIAS, and (D) ATLQWIAS.

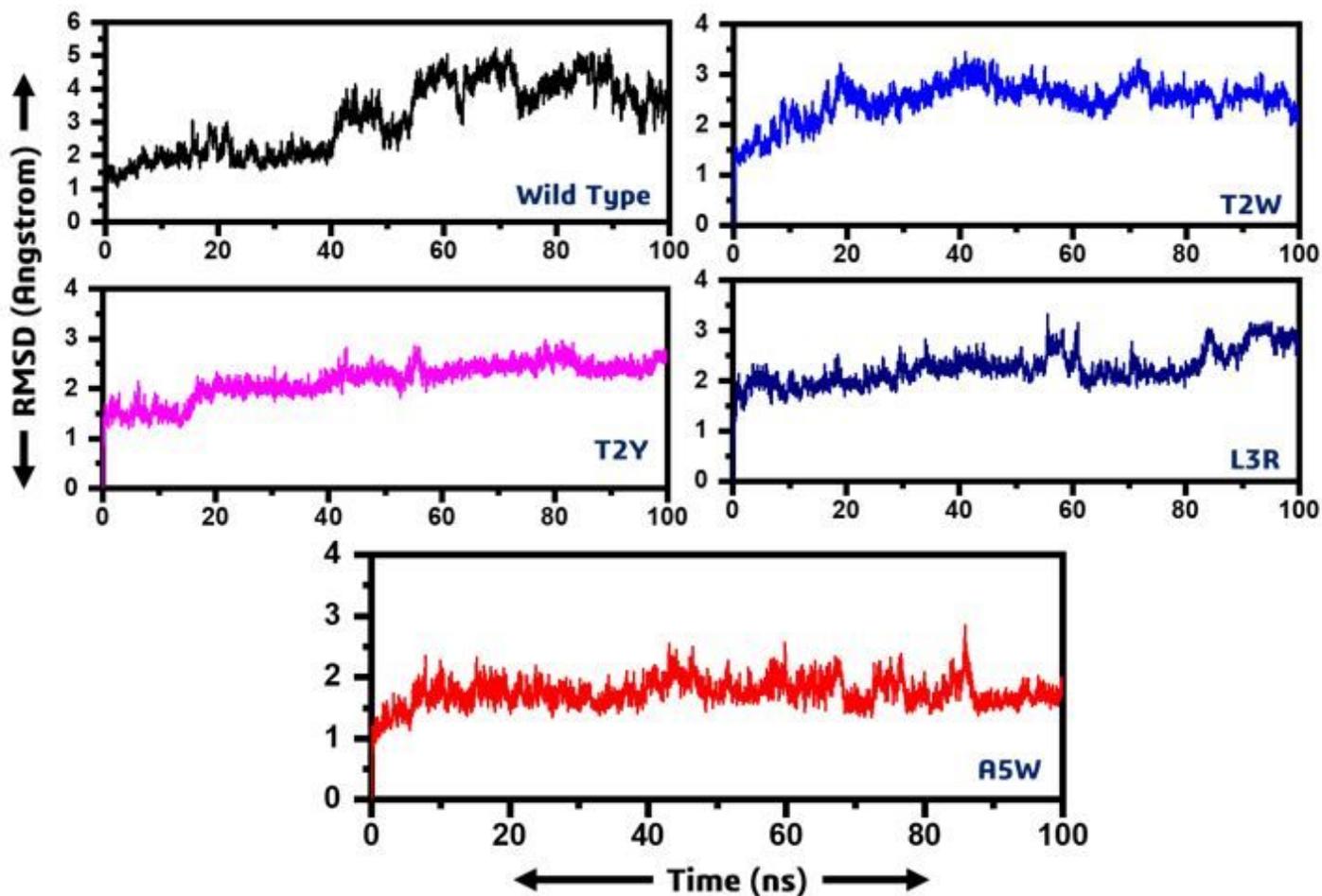


Figure 4

Showing the dynamic stability of each system as RMSD. A different colour represents each complex. The x-axis is showing time in nanoseconds while the y-axis is showing RMSD in Angstrom.

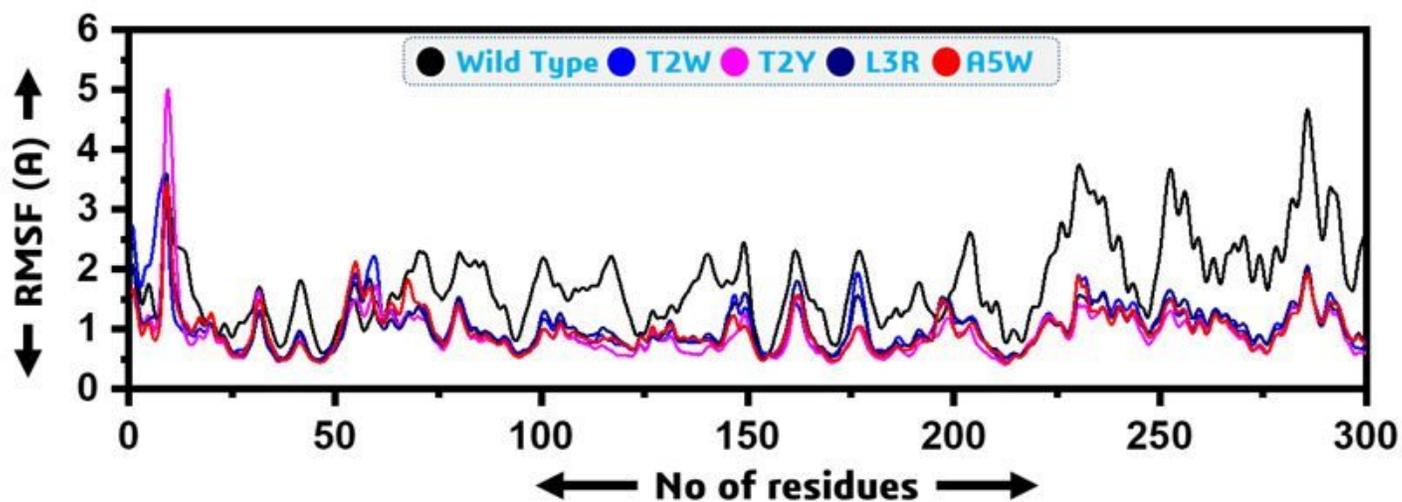


Figure 5

The residual flexibility of each system. The x-axis is showing the number of residues while the y-axis is showing RMSF in Angstrom.

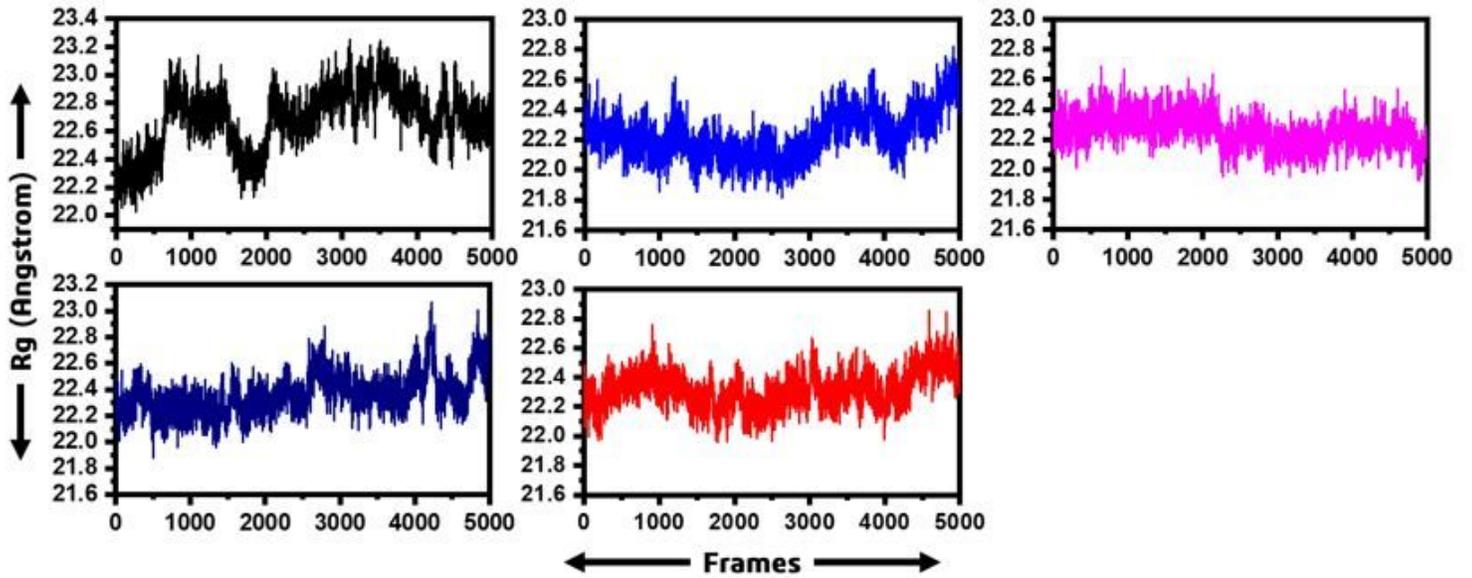


Figure 6

Rg of all the systems. The x-axis is showing the number of frames while the y-axis is showing Rg in Angstrom. Wild (black), T2W (blue), T2Y (Magenta), L3R (navy) while the A5W is shown in red colour.

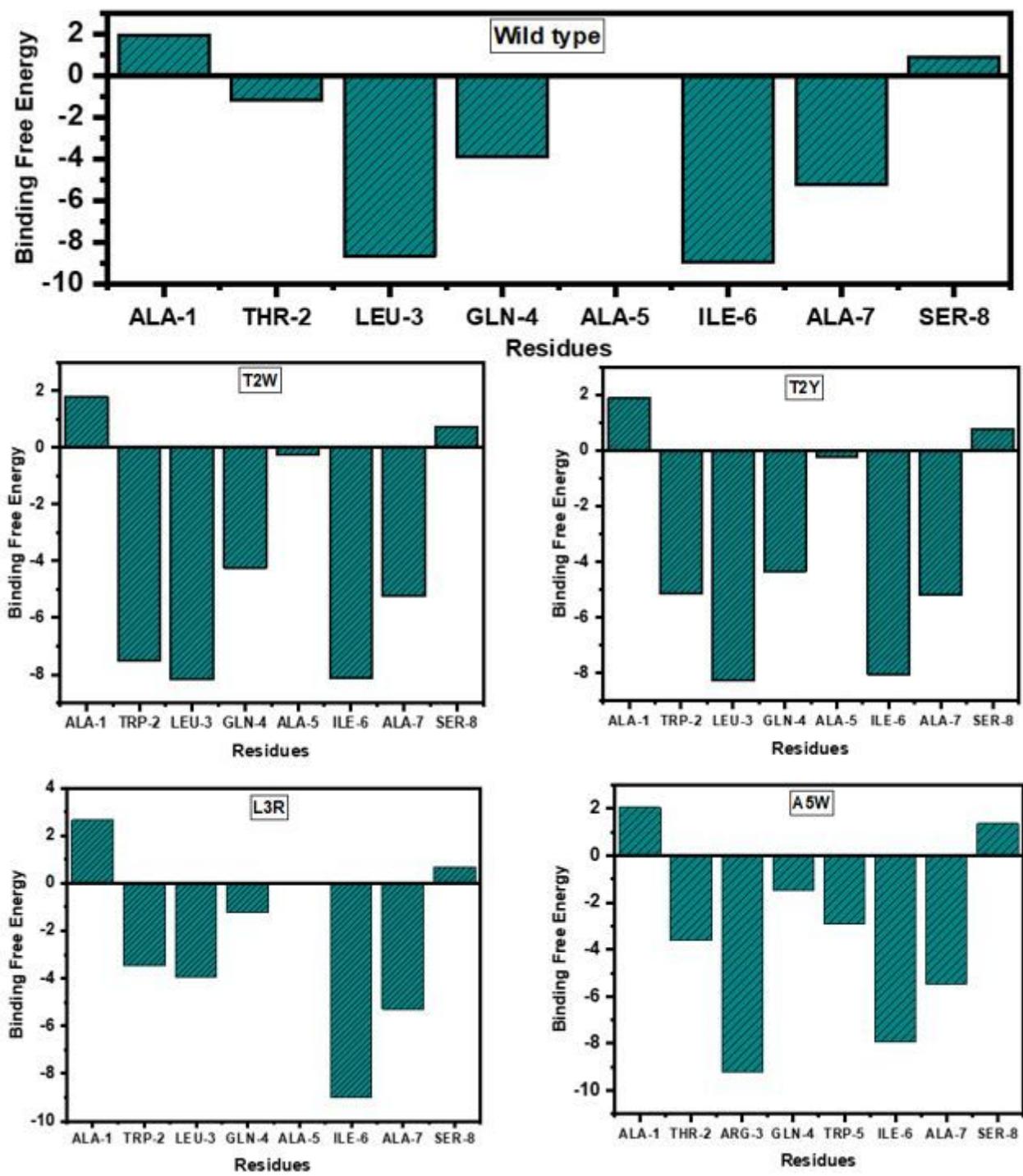


Figure 7

Per-residue energy decomposition analysis of all the designed peptides. The graphs are given in Wild type, T2W, T2Y, L3R and A5W sequence.

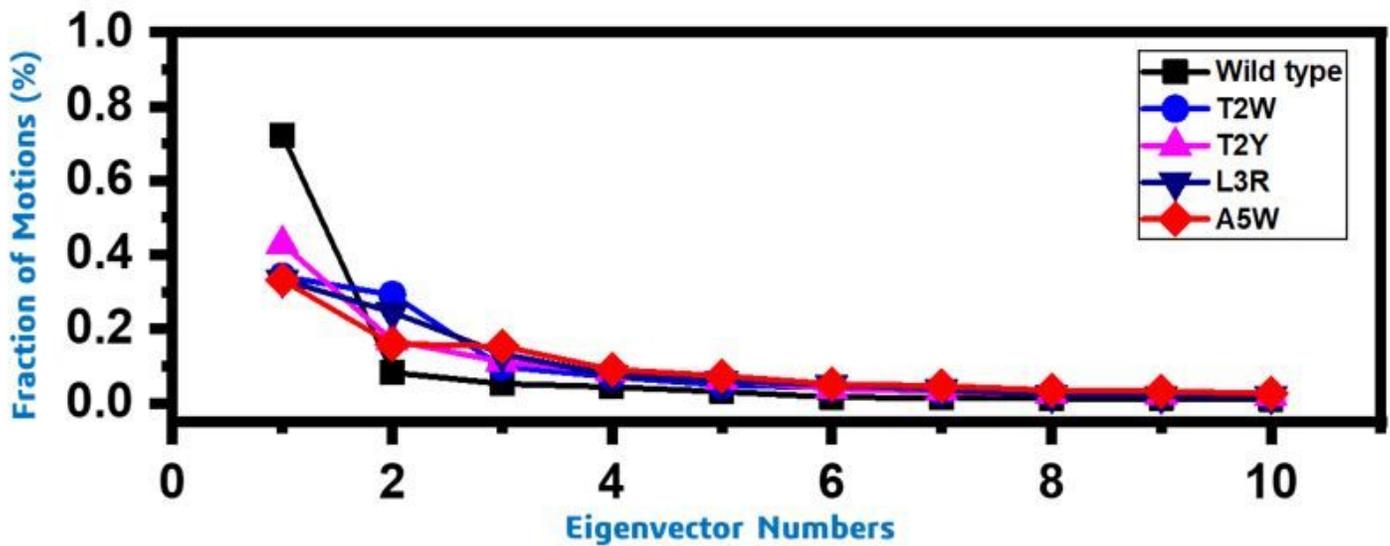


Figure 8

Fraction of the first ten eigenvectors. The (%) contribution of each eigenvector obtained from the covariance matrix plotted against the corresponding eigenvector indices constructed from the MD trajectory.

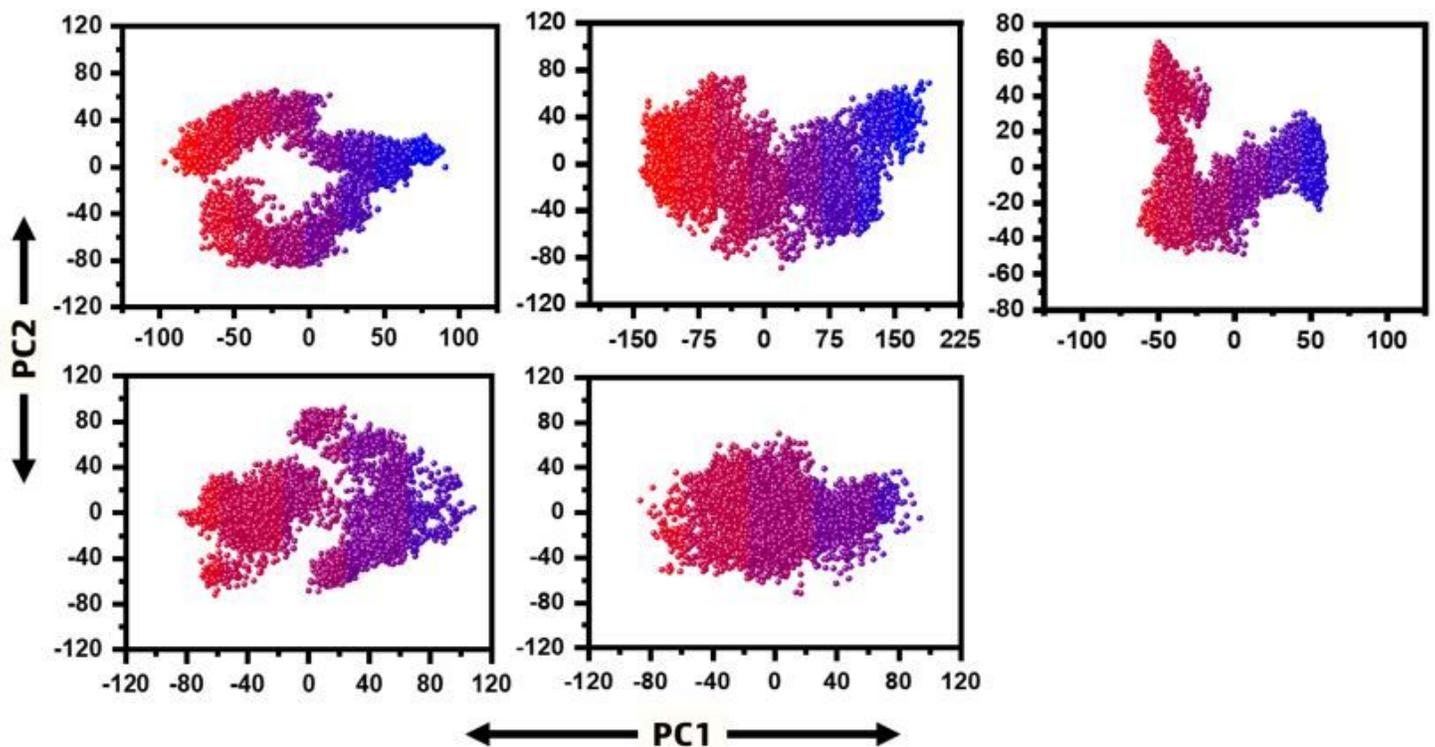


Figure 9

Principal component analysis (PCA) of wild and designed decoy peptides bound to the 3CLpro. The first PC1 and second PC2 from the PCA of the backbone carbon were used. The sequence is Wild type, T2W, T2Y, L3R, and A5W.

