ORIGINAL PAPER

Potential antiviral peptides against the nucleoprotein of SARS-CoV-2

Iktedar Mahdi[1,2] · Humyara Yeasmin[1] · Imtiaz Hossain[1] · Raina Masnoon Badhan[1] · Md. Ackas Ali[1,4] ·
Md. Abdul Kaium[1] · Rajib Islam[1] · Md. Abu Sufian[3] · Mohammad A. Halim[4]

Abstract

Nucleoprotein is a conserved structural protein of SARS-CoV-2, which is involved in several functions, including replication, packaging, and transcription. In this research, 21 antiviral peptides that are known to have inhibitory function against nucleoprotein in several other viruses, were screened computationally against the nucleoprotein of SARS-CoV-2. The complexes of five best performing peptides (AVP1142, AVP1145, AVP1148, AVP1150, AVP1155) with nucleoprotein were selected for subsequent screening via 5 ns molecular dynamics (MD) simulation. Two peptides, namely AVP1145 and AVP1155, came out as promising candidates and hence were selected for 200 ns MD simulation for further validation, incorporating a DMPC-based membrane environment. In the long MD simulation, both AVP1155 and AVP1145 utilized multiple residues—mainly aromatic, acidic, and nonpolar residues—as interacting points to remain in contact with the nucleoprotein and formed predominantly hydrogen bonds along with hydrophobic and electrostatic interactions. However, AVP1155 proved to be superior to AVP1145 when its complex with nucleoprotein was analyzed in terms of root-mean-square deviation, root-mean-square fluctuation, radius of gyration, solvent accessible surface area and free energy landscape. In a nutshell, the findings of this research may guide future studies in the development of selective peptide inhibitors of SARS-CoV-2 nucleoprotein.

Keywords SARS-CoV-2 · Nucleoprotein · Antiviral peptides · Molecular dynamics simulation · Interacting residues

Introduction

A new disease named COVID-19 emerged in December, 2019 in Wuhan, China (Hasan et al. 2020; COVID Live - Coronavirus Statistics - Worldometer 14 Sep 2022). Since the beginning, it has managed to spread over 228 countries. Worldwide travel restrictions and continuous lockdown failed to halt the rapid transmission of this deadly disease. The number of affected cases mounted every day and reached 615,042,167 with a death toll of 6,521,147 as of September 14, 2022 (COVID Live - Coronavirus Statistics - Worldometer 14 Sep 2022).

COVID-19 is caused by severe Acute Respiratory Syndrome Corona Virus-2 (SARS-CoV-2). This deadly virus is a positive sense, single-stranded RNA virus, which encodes four main structural proteins including spike protein, membrane protein, envelope protein, and nucleoprotein (Ahmed et al. 2020a). Among the major structural proteins, nucleoprotein (NP) is functionally very diverse and plays crucial roles in transcription, replication as well as the packaging of the virions (Wang et al. 2010). Structurally, NP is a polypeptide of 419 amino acids, rich in glycine, serine, alanine, and glutamine. It consists of three conserved regions, namely N-terminal region (N-terminal arm and N-terminal domain), C-terminal region (C-terminal domain and C-terminal tail), and a central region enriched in Ser/Arg (SR) (Zhou et al. 2020). Functionally, these conserved regions of NP act as RNA-binding sites, and facilitates interaction of RNAs with NP in a newly assembled virion, thus protecting it in the early stage of its life cycle (Green and Xin Zhang 2021). Apart from protecting the genome, the RNA-NP complex

✉ Mohammad A. Halim
mhalim1@kennesaw.edu

1   Division of Infectious Diseases and Division
    of Computer-Aided Drug Design, The Red-Green Research
    Centre, BICCB, 16 Tejkunipara, Tejgaon, Dhaka 1215,
    Bangladesh

2   Department of Chemistry, Norwegian University of Science
    and Technology, 7491 Trondheim, Norway

3   School of Pharmacy, Temple University, Philadelphia,
    PA 19140, USA

4   Department of Chemistry and Biochemistry, Kennesaw State
    University, Kennesaw, GA 30144, USA

also facilitates the packaging of a virion, in several phases, by forming a wavy ribbon structure, or helical structure, both loosely and tightly coiled (Green and Xin Zhang 2021). Besides, thanks to its intrinsic multimerization capacity, NP forms the viral nucleocapsid core in a mature virus (Hurst et al. 2010; He et al. 2004). On the other hand, a body of literature suggests that NP interferes with the host cell cycle (Portela and Digard 2002; Almazán et al. 2004; Wurm et al. 2001) by interacting with cyclin dependent kinase complex (Zhou et al. 2020). Peng et al. reported that in animals NP can induce T cell specific immune responses (Peng et al. 2006). According to Wang et al., NP plays a major role in stimulating the host immune system, and may result in immunopathological effects (Wang et al. 2010). Thus, a highly conserved structure, along with high immunogenic potential, and other functional importance makes SARS-CoV-2 NP a potential therapeutic target.

Peptides represent a distinct class of pharmaceuticals, molecularly poised between proteins and small molecules, although they are different both therapeutically and biochemically (Lau and Dunn 2018). As potential inhibitors of SARS-CoV-2 NP, peptides are superior to small molecules because peptide-protein interaction has a large contact surface area (Smith and Gestwicki 2012). Peptide therapeutics also offer several crucial benefits over antibodies or proteins. The small size of peptides makes them easy to synthesize and allows them to diffuse easily through the cell membrane unlike proteins or antibodies (Marqus et al. 2017). Besides, peptides are less immunogenic as compared to recombinant antibodies or proteins, and there is no organ specific accumulation of peptides (e.g., accumulation in lungs or kidney), thus their toxic side effects are minimized (Marqus et al. 2017). However, peptides also have some disadvantages including oral bioavailability issues, high production costs, and short half-life (Haggag 2018). Despite all the challenges, several groups including ours have shown computationally as well as with experimental evidence that peptides are highly promising as a class of anti-SARS-CoV-2 therapeutics (Sakib et al. 2021; Chowdhury et al. 2020; Efaz et al. 2021). In several published works, we have shown that peptides that are experimentally proven to have inhibitory properties against various other viruses, offer high promise to inhibit SARS-CoV-2 as well (Sakib et al. 2021; Chowdhury et al. 2020; Efaz et al. 2021).

In this work, we focused on the nucleoprotein of SARS-CoV-2 as a potential target, as it has not been pursued thus far. Twenty one antiviral peptides that were experimentally tested against the nucleoprotein of other viruses have been selected and screened against nucleoprotein of SARS-CoV-2 using various computational tools (Qureshi et al. 2014) Upon analysis of binding interactions, binding affinity, as well as binding pose, best performing peptides were selected for further investigation by several developed techniques like molecular dynamics simulation, and principal component analysis. Here, the aim of this study was to provide a framework for the development of therapeutic peptides against the SARS-CoV-2 Nucleoprotein.

Methods

Protein, and peptide modeling and molecular docking

Homology model and probable binding sites of SARS CoV-2 nucleoprotein (NP) were collected from I-TASSER (Yang et al. 2014), without further validation. The following residues were mentioned and used as potential binding sites of NP: PRO46, THR49, ALA50, SER51, THR54, GLN83, ILE84, GLY85, ARG88, ALA90, ARG92, ARG93, LEU104, SER105, PRO106, ARG107, TYR109, TYR111, GLY114, THR115, THR148, ARG149, PRO151, ALA152, ALA156, GLU174, SER176 (Modeling of the SARS-COV-2 Genome using I-TASSER (Yang et al. 2014). PyMOL 2.3 was used to prepare, optimize, and model the NP structure (The PyMOL Molecular Graphics System 2017). Peptides sequences were collected from AVPdb database (Qureshi et al. 2014). Pepfold 3.5 (Lamiable et al. 2016) server was used for predicting de novo structures of selected peptides from the sequences. Haddock 2.4 (Zundert et al. 2016) was used for peptides-NP docking. The results from Haddock 2.4 (Zundert et al. 2016) were further confirmed and validated by PatchDock (Schneidman-Duhovny et al. 2005) and FireDock (Mashiach et al. 2008), Cluspro (Kozakov et al. 2017), and ZDOCK (Pierce et al. 2014).

Molecular dynamics simulation

Both short (5 ns) and long (200 ns) duration molecular dynamics (MD) simulation were performed. Peptide-NP complexes as well as Apo were subjected to either 5 ns or 200 ns MD simulations (Lyman et al. 2009). To mimic the physiological environment, a DMPC membrane was placed in the system during each MD simulation run (Figure S1) (Loschwitz et al. 2020). Before an MD run, the system was solvated by using an SPC water box with a buffer distance of 10 Å (Mark and Nilsson 2001), and the charge of the system was neutralized by adding 0.15 M NaCl. Each MD run employed OPLS2005 as the force field. The calculated system size for Apo, and AVP1145-NP, and AVP1155-NP complexes was 54,492 atoms, 55,618 atoms, and 56,376 atoms, respectively. The Desmond program (Bowers et al. 2006) was used to carry out the simulation. The NPT ensemble was used for minimization and relaxation of the system, where the number of particles was kept constant, and the temperature and pressure were set at 300 K, and

1 bar respectively. The M-SHAKE algorithm was applied in the system to satisfy geometric constraints of hydrogen bonds during the simulation period that allowed 2 fs time steps (Allouche 2012). For long-range electrostatic interactions, periodic boundary conditions were used based on the k-space Gaussian split Ewald method (Shan et al. 2005) and a cut-off distance of 9 Å was set for van der Waals interactions depending on a uniform density approximation. For short-range non-bonding interactions, 9.0 Å cut-off distance, calculated by a r-REPA integrator (Stuart et al. 1996) was considered. The required time for the update of short-range forces and the long-range forces were 2 fs and 6 fs, respectively. During the long (200 ns) MD simulation, snapshots were saved at every 100 ps to analyze the trajectory data. Root-mean-square deviation (RMSD), root-mean-square fluctuation (RMSF), radius of gyration (Rg), solvent accessible surface area (SASA) was calculated from the MD trajectory data to examine the structural parameters of the complexes during MD simulation.

Principal component analysis

Principal component analysis (PCA) was used to analyze several energy parameters to follow structural change in the complexes during MD simulation (Jong 1990; Jollife and Cadima 2016). The following parameters: bond angle, bond, dihedral angles, van der Waals energies, coulomb energies, and planarity obtained from the last 150 ns of MD trajectories of each peptide-protein complex were considered to generate training sets for PCA to identify any hidden difference in energy parameters of the test complexes (Islam et al. 2020; Ahmed et al. 2020b). The data was centered and scaled before the analysis. The different factors had been organized in X matrix and turned into two new matrices by following equation:

$$X = T_k P_k^T + E$$

here, T_k is score matrix signifying the relation of the samples among themselves, P_k is loading matrix concerning the relation of variables among themselves, k denotes the number of factors while E is an un-modeled variance. To perform all of the calculations, R(R Core Team (2014) R Core Team 2014), RStudio (RStudio Team 2015), and in-house generated codes were used; the package factoextra (Kassambara and Mundt 2017) was utilized for generating plots.

Binding free energy calculation

Binding Free Energy (BFE) was calculated for AVP1155-NP and AVP1145-NP by HawkDock (Weng et al. 2019) based on MM/GBSA method as well as by PRODIGY web server (Xue et al. 2016). BFE calculation based on PRODIGY employed Linear Interaction Energy (LIE) using the following equation:

$$\begin{aligned} \Delta G_{predicted} = &-0.09459\,IC_{s_{charged/charged}} - 0.10007\,IC_{s_{charged/apolar}} \\ &+ 0.19577\,IC_{s_{polar/polar}} - 0.22671\,IC_{s_{polar/apolar}} \\ &+ 0.18681\%NIS_{apolar} + 0.3810\%NIS_{charged} - 15.9433 \end{aligned}$$

where, ICs and NIS are interfacial contacts and non-interacting surfaces, respectively.

Statistical analysis

Student's t-test was used to compare structural and energy parameters of peptide-NP complexes.

Results and discussion

Peptides binding affinity and interactions in molecular docking

In this work, 21 peptides that are experimentally known to have inhibitory function against nucleoprotein of several other viruses were selected (Table S1). The selected peptides were screened against SARS-CoV-2 NP following the scheme outlined in Fig. 1. The selected peptides were modeled to generate 3D structures. The 3D models of the peptides were then docked against a homology model of SARS-CoV-2 NP using four different peptide-protein docking protocols. Haddock 2.4 was used as the primary docking tool to screen the peptides against NP, as it utilizes a residue focused docking of the peptides in the binding pocket of NP. However, other docking protocols, e.g., PatchDock and FireDock, Cluspro and ZDOCK were used to further confirm and validate the Haddock outputs. The results from all four docking protocols correlate relatively well (Table 1). In docking by Haddock, docking scores of the peptides ranged from −41 kcal/mol to −121 kcal/mol (Fig. 2a, Table 1). The five high frequency residues that have shown the highest interactions with the peptides during docking by Haddock were ARG92, ARG88, ARG149, ARG177, ARG 107 (Fig. 2b and d). The peptides-NP interactions were predominantly hydrogen bonds (55%), followed by hydrophobic (30%), and electrostatic (15%) interactions (Fig. 2c). From an initial screening of the

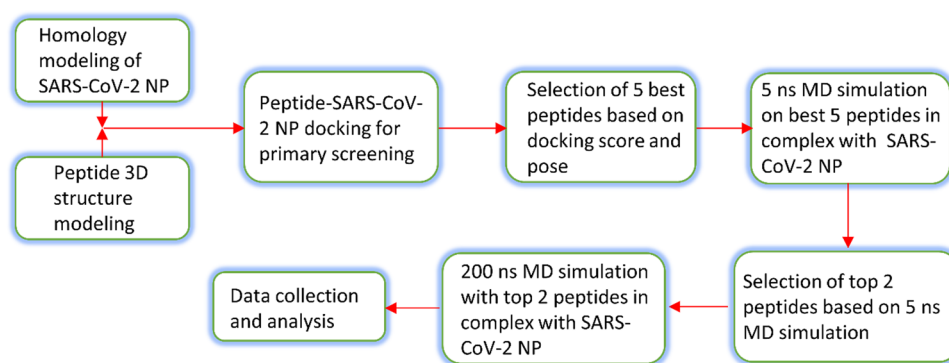**Fig. 1** Flow chart showing the research design and methodology

Table 1 Docking scores of selected 21 peptides against SARS-CoV-2 nucleoprotein

Peptide ID	Firedock score (kcal/mol)	ClusPro score	Zdock score	Haddock score (kcal/mol)
1137	− 71.75	− 942.6	1159.052	− 83.7 ± 0.4
1138	− 61.98	− 875.7	1128.526	− 46.8 ± 1.4
1139	− 86.26	− 862.4	1168.983	− 91.4 ± 0.6
1140	− 71.32	− 742.4	1086.002	− 69.6 ± 0.8
1141	− 53.13	− 839.8	999.878	− 66 ± 0.4
1142	− 82.97	− 1014.4	1319.405	− 108.8 ± 0.6
1143	− 70.83	− 876.4	993.179	− 68.7 ± 1.6
1144	− 42.21	− 1128.7	1360.77	− 76.8 ± 0.6
1145	− 56.39	− 1063.2	1180.404	− 106.2 ± 0.7
1146	− 56.57	− 966.8	1118.241	− 50.5 ± 5.8
1147	− 63.2	− 881.2	878.443	− 57.1 ± 3.2
1148	− 47.38	− 1021.9	1151.99	− 121 ± 0.7
1149	− 56.68	− 913	1229.801	− 88.7 ± 1.1
1150	− 48.22	− 1093.8	1361.696	− 119.7 ± 1.2
1151	− 47.8	− 1006.5	1055.465	− 70.4 ± 0.8
1152	− 61.81	− 964.3	1132.432	− 67.8 ± 0.5
1153	− 46.81	− 896.1	1328.602	− 87.8 ± 2.2
1154	− 68.03	− 1270.2	1440.125	− 77.7 ± 2.1
1155	− 69.71	− 1276.1	1387.157	− 92.8 ± 1.5
1156	− 57.95	− 749.6	775.674	− 53.1 ± 0.6
1157	− 47.06	− 941.3	913.843	− 72.5 ± 4.7

selected 21 peptides, based on docking scores and docking pose obtained from Haddock, five best performing peptides (AVP1142, AVP1145, AVP1148, AVP1150, AVP1155) were selected for further screening by a short duration MD simulation for 5 ns. The Haddock docking scores of the selected peptides were − 108.8 kcal/mol, − 106.2 kcal/mol, − 121 kcal/mol, − 119.7 kcal/mol, − 92.8 kcal/mol for AVP1142, AVP1145, AVP1148, AVP1150, AVP1155, respectively (Table 1).

Complex structural and energy parameters in MD simulation

In the beginning, a short duration (5 ns) MD simulation was performed with the complexes of five best performing peptides, i.e., AVP1148-NP, AVP1150-NP, AVP1142-NP, AVP1145-NP, and AVP1155-NP complexes. From 5 ns MD trajectories, RMSD, Rg and SASA were calculated for each complex and visualized against time (Figure S2). An analysis of Figure S2 revealed that the AVP1145-NP and AVP1155-NP complexes showed a more stable RMSD, Rg and SASA profile over time as compared to the other three complexes, and thus were selected for a long duration (200 ns) MD simulation to examine the structural and energy parameters of these complexes. In 200 ns MD simulation, Apo-NP was included along with AVP1145-NP and AVP1155-NP to observe how the binding of AVP1145 or AVP1155 influenced the structural parameters (RMSD, Rg, SASA, RMSF) and energy landscape (coulomb, bond, angle, vdw, dihedral and planarity) of the Apo-NP (Fig. 3).

RMSD profile of a peptide-protein complex is considered a crucial structural parameter. The relative dynamic stability of a set of peptide-protein complexes can be identified by plotting RMSD values against time (Shukla and Tripathi 2020). Additionally, snapshots of MD simulations at different time points can be used as a visual tool to explain key structural changes in a complex over the simulation period. Thus, RMSD values were calculated for AVP1145-NP, AVP1155-NP, and Apo-NP from an MD trajectory of 200 ns for each complex (Fig. 3a). Snapshots representing major structural changes at every 25 ns MD simulation were taken and presented in Fig. 4. An analysis of RMSD values of α-carbon atoms of the complexes, i.e., AVP1145-NP, AVP1155-NP and Apo-NP demonstrated that an initial hike at time zero was observed for both complexes including Apo-NP, which represents equilibration of the rigid structures in a simulated physiological environment. Following equilibration, the complexes took differential times to reach their average conformations, and subsequently experienced major or minor up or downward shifts from their average
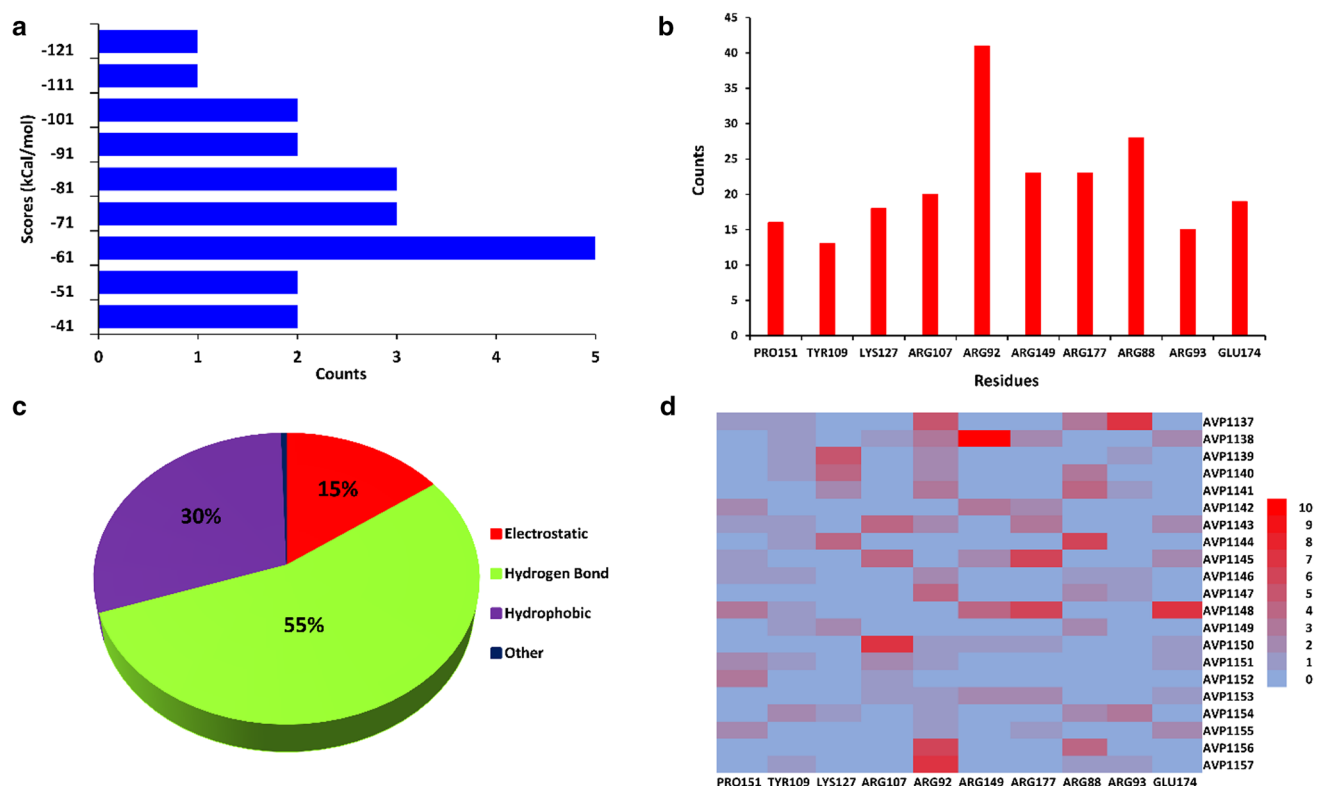
Fig. 2 Binding interactions **a** distribution of binding affinities **b** interacting nucleoprotein residues; **c** distribution of non-covalent interactions. **d** residue-residue contact of the peptides-NP complexes from Haddock results

structures based on the complex stability (Fig. 3a). Time wise RMSD analysis for AVP1155-NP complex revealed minor up or downward shifts (Fig. 3a and 4b) and showed a stable RMSD (8.19 ± 0.89 Å) from the initial rigid complex. However, this RMSD was higher than that of Apo-NP (7.68 ± 0.79 Å), which suggests that AVP1155 binding to Apo-NP had a de-stabilization effect on the Apo structure. AVP1145-NP (8.56 ± 1.57 Å), on the other hand, experienced a higher average shift from its initial rigid structure (Fig. 3a and 4a). The RMSD of AVP1145 was higher than that of AVP1155-NP and Apo-NP, denoting stronger destabilizing effect on the Apo-NP structure upon AVP1145 binding as compared to AVP1155. A higher standard deviation of RMSD for AVP1145-NP (1.57 Å) indicates that this complex remained less stable as compared to AVP1155-NP over the 200 ns MD simulation period. When RMSD values of the peptides-NP complexes were compared in terms of change in RMSD values due to peptide binding to the Apo-NP structure, AVP1155-NP showed significantly lower average change relative to Apo-NP than that of AVP1145-NP (Figure S3a).

The time wise variation in SASA values were lower in AVP1155-NP complex ($23,018.96 \pm 365.55$) as compared to AVP-1145-NP complex ($22,638.7 \pm 427.44$), which was evident from the standard deviation values (Fig. 3b and S3b).

During the simulation period, its free Brownian movement allowed AVP1155-NP complex an average 26.3 ± 0.20 Å of radius of gyration (Rg), which was significantly lower than that of AVP1145-NP (Fig. 3c, Figure S3c). Notably, AVP1155 and AVP1145 are of similar in volume (calculated by Peptide Properties Calculator, northwestern.edu). Thus, lower Rg clearly points out better compactness of AVP1155-NP complex due to stronger binding of AVP1155 as compared to AVP1145. Residue wise fluctuations pattern for both AVP1155-NP and AVP1145-NP were roughly the same as that of Apo-NP (Fig. 3d). However, average residue wise fluctuation relative to apo-NP was significantly lower for AVP1155-NP complex as compared to that of AVP1145-NP complex (Figure S3d).

The root cause of manifest variance in structural parameters between AVP1145-NP and AVP1155-NP complexes, relative to Apo-NP may base in their inherent difference in the energy profiles. Thus, to investigate variation in energy profiles among the complexes, we developed a PCA model with the energy data obtained from their respective MD simulation trajectories. The generated PCA model was reliable in explaining energy variance among the test complexes as two PC axes, PC1 and PC2, explained 94.4% of total variance. The test complexes were clustered in a PCA scores plot to visualize the corresponding energy
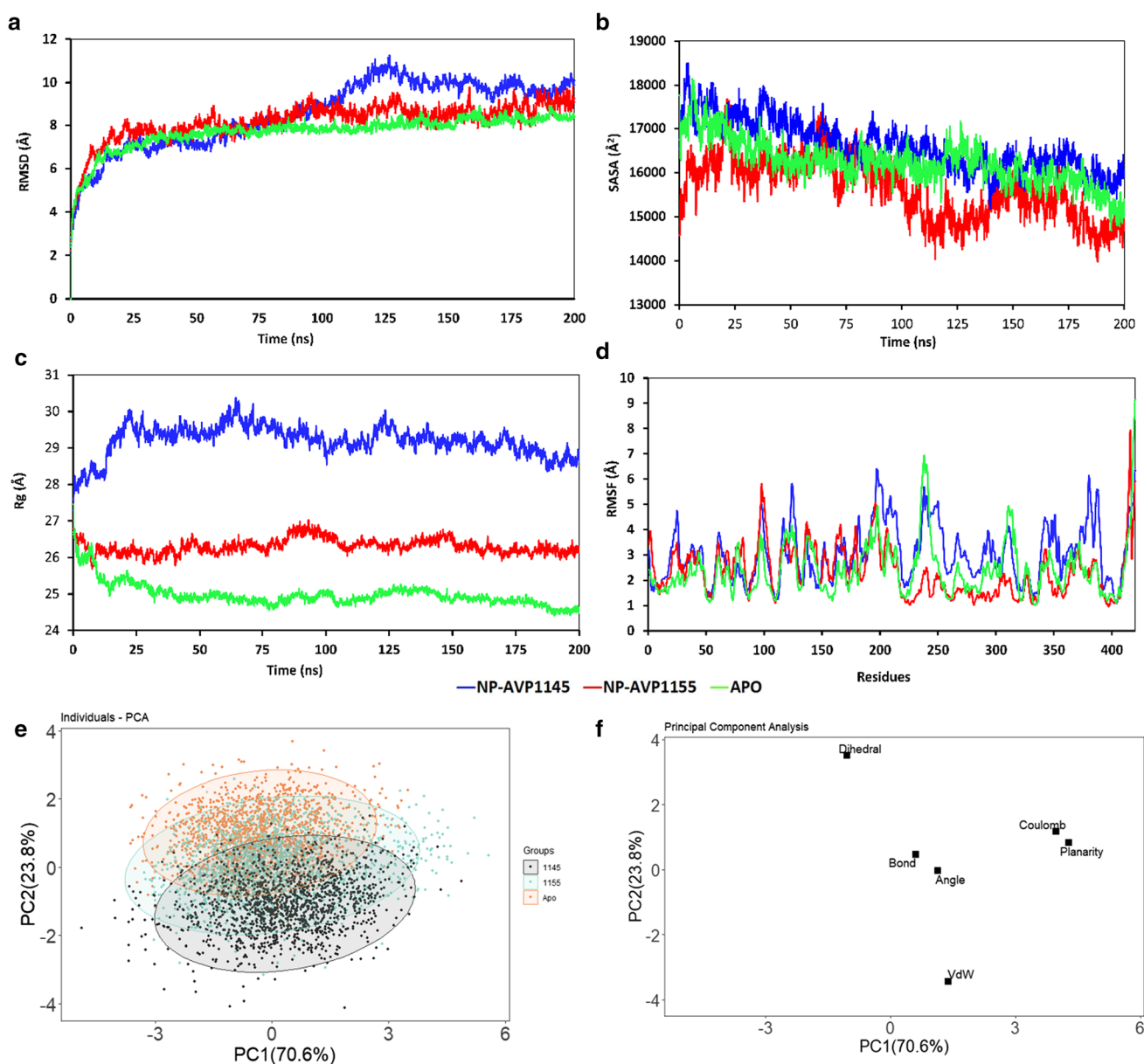
Fig. 3 Molecular dynamics simulation **a** root-mean-square deviation (RMSD); **b** solvent accessible surface area (SASA); **c** radius of gyration (Rg); and **d** root-mean-square fluctuation (RMSF); **e** scores plot and **f** loading plot of Apo-NP, and AVP1145-NP, AVP1155-NP complexes

landscape, whereas a PCA loading plot was generated to explain the shapes of each assembly and their relative positions in the scores plot (Fig. 3e and f). Evidently, all the clusters were elliptical in shape with roughly similar major and minor axes lengths. It implies that the time wise energy distribution within the complexes followed a similar pattern. Within a complex, bond, angle, planarity and coulomb energy landscapes changed the most with respect to time as the ellipse major axis lined in parallel to PC1. To focus on variance between groups, the cluster of AVP1145-NP showed a downward shift as compared to

Apo-NP or AVP1155-NP. Looking at the loading plot, the downward shift of this complex relative to others was contributed by major changes in van der Waals energy values. Considering shifts along both PC1 and PC2, AVP1155-NP was closer to apo-NP as compared to AVP1145-NP (Fig. 3e). As evidenced from both scores plot and loading plot, although binding of AVP1145 and AVP1155 to NP impacted the energy landscape of NP, AVP1155 introduced minimal changes in its energy profile and hence stabilized the complex more as compared to AVP1145 (Fig. 3e–f). Besides, when the change in energy landscape of NP because of AVP1155 or AVP1145 binding
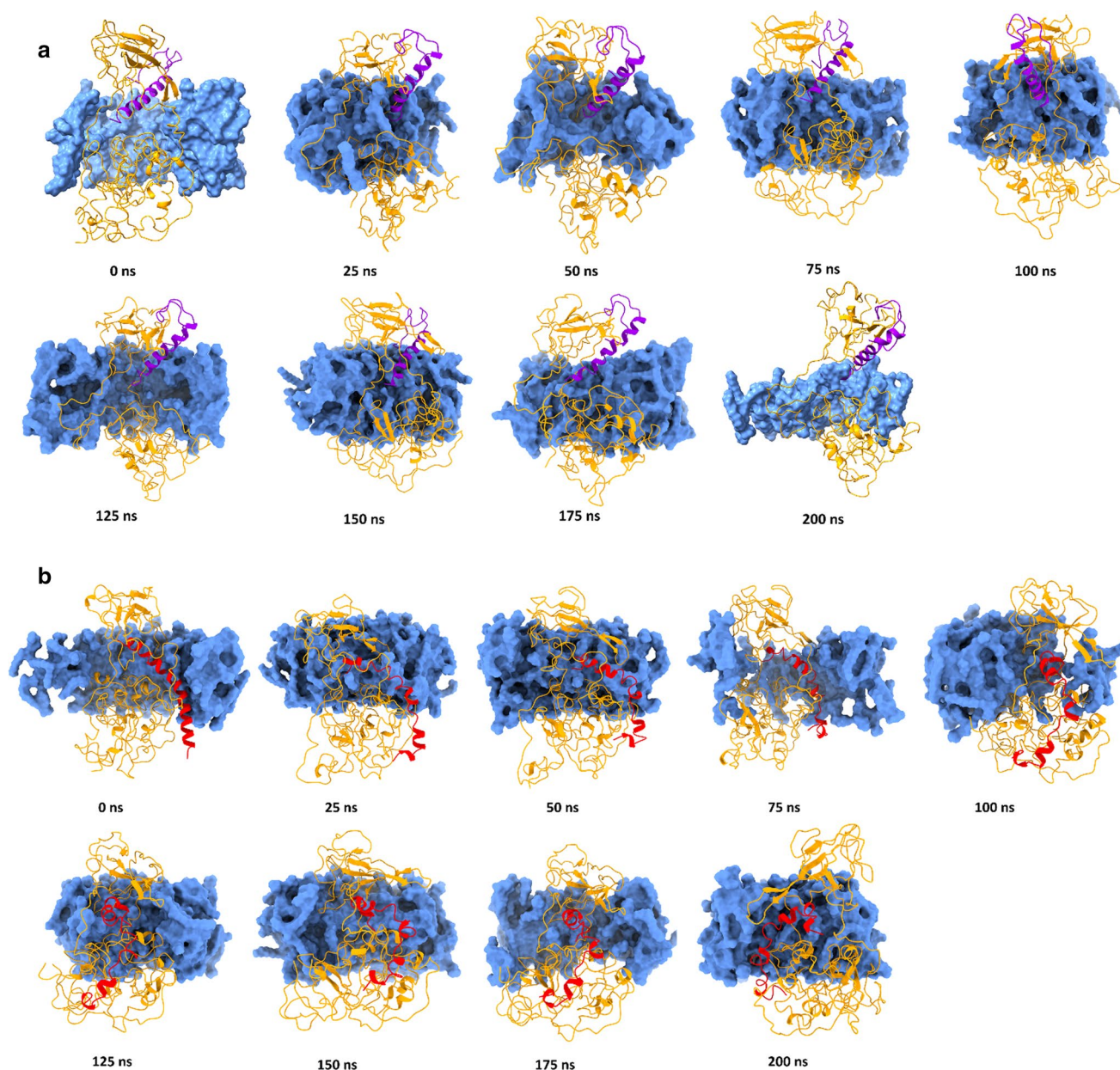
Fig. 4 Representative snapshots **a** AVP1145-NP (Violet); **b** AVP1155-NP (Red) during 200 ns MD simulation. Membrane is highlighted in corn flower blue.

was statistically compared, a significantly greater change was observed due to AVP1145 binding to NP as compared to AVP1155 binding (Figure S4).

Non-covalent interactions in AVP1155-NP and AVP1145-NP obtained from MD simulation

During 200 ns MD simulation, conformers were generated at 1 ns interval for both AVP1145-NP and AVP1155-NP complexes. Meticulous analysis of these conformers revealed interacting peptide and protein residues at each time point,

participation time of a residue in its corresponding bond, distribution of non-covalent interactions, time wise frequency of different types of interactions (hydrogen, hydrophobic or electrostatic), and binding free energy landscape for each of these complexes. Upon analysis of AVP1155-NP dynamics, it was observed that ARG277, ASP22 and PRO279 were the major interacting points of NP, whereas ASP22 and PRO279 (100% participation time) anchored the peptide in its binding groove during 200 ns MD simulation. On the other hand, TRP15, ARG40 and GLU24 served as the main interaction as well as anchoring points for AVP1155 (Fig. 5a and b).
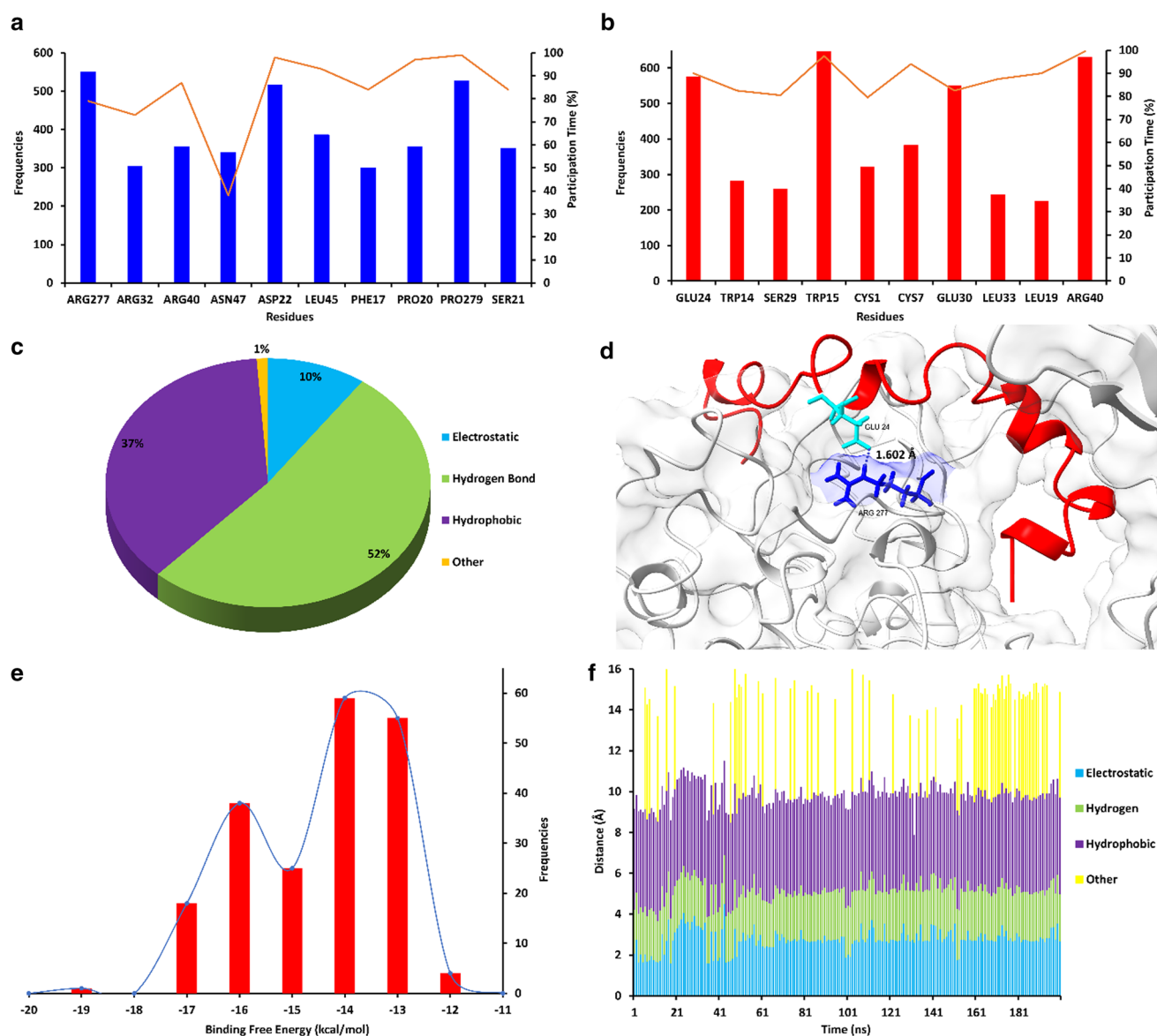
Fig. 5 AVP1155-NP complex **a** interacting NP residues with percent participation time; **b** interacting AVP1155 residues with percent participation time; **c** distribution of non-covalent interactions; **d** representative snapshot (at 200th ns): AVP1155 (dark red) and nucleo-protein (grey); **e** distribution of binding free energies (kcal/mol); **f** average distance of interactions (hydrogen, hydrophobic, electrostatic, and other) in AVP1155-NP complexes over 200 ns

Hydrogen bonds, hydrophobic interactions and electrostatic interactions contributed to 52%, 37% and 10% of total interactions of AVP1155 with NP and over the 200 ns MD simulation, this interaction pattern remained stable (Fig. 5c and f). Thus, AVP1155 was strongly held in the binding pocket of NP, which was also evidenced from its free energy of interaction (-14.485 ± 1.393 kcal/mol) (Fig. 5d, 5f and Table S2).

In case of AVP1145-NP complex, ARG107, ARG92 and HIS59 of NP and GLU33, GLU37 and TYR44 of AVP1145 were the major interacting residues in terms of both participation time and frequency (Fig. 6a and b). Similar to AVP1155,

in AVP1145 binding to NP, hydrogen bonds contributed to 53% of total interactions. However, AVP1145-NP and AVP1155-NP differed in electrostatic and hydrophobic interactions (Fig. 5c and 6c). In AVP1145-NP complex, hydrophobic and electrostatic bonds accounted for 28% and 19% of total interactions, respectively (Fig. 6c). The time wise interaction pattern of AVP1145 with NP was relatively less stable as compared to that of AVP1155 (Fig. 5f and 6f). The binding free energy of AVP1145 with NP was -11.3225 ± 0.7069 kcal/mol, which implies that AVP1155 was more securely held in the binding groove of NP, as compared to AVP1145 (Fig. 5e, 6e and Table S2).
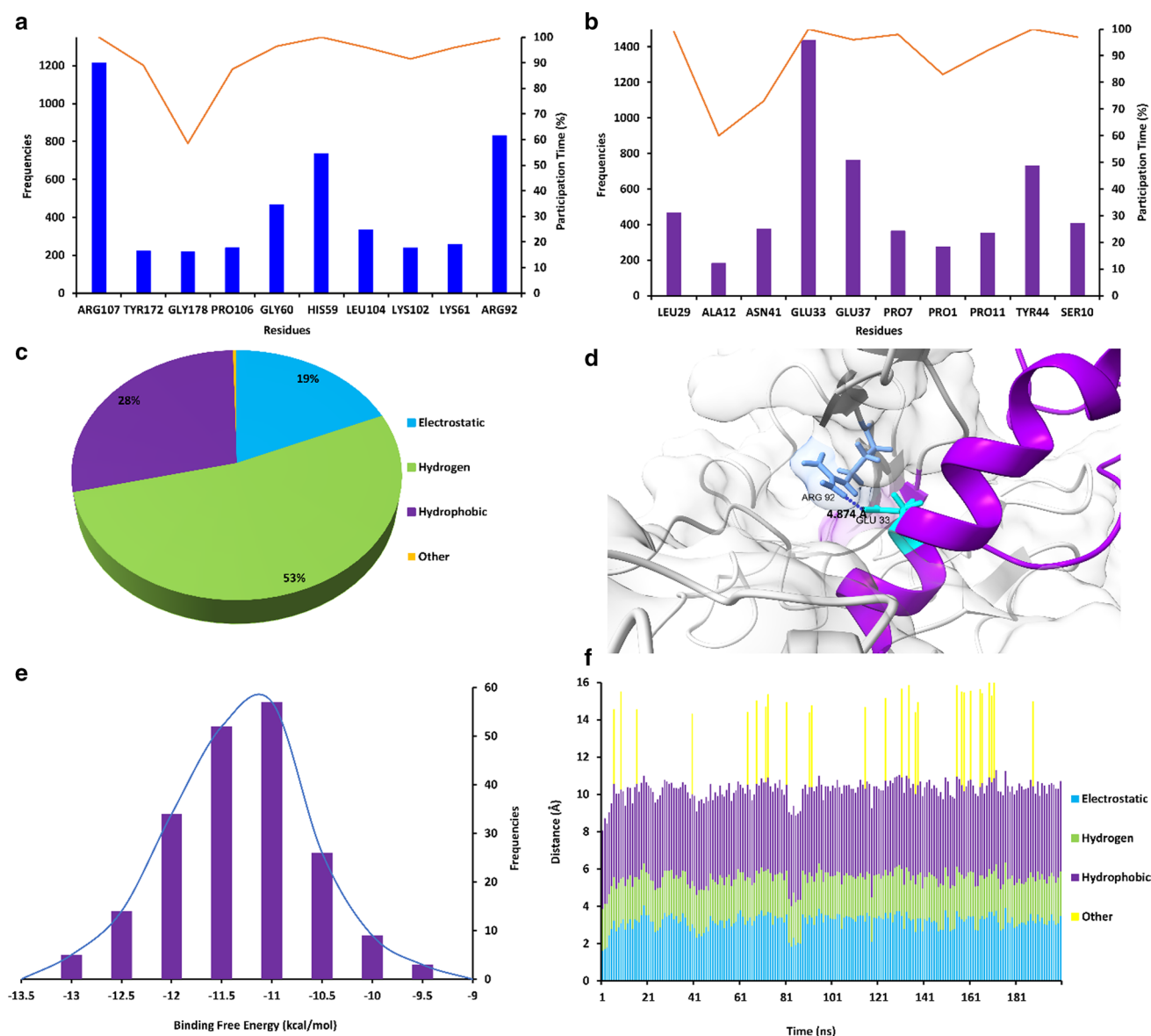
Fig. 6 AVP1145-NP complex **a** interacting NP residues with percent participation time; **b** interacting AVP1145 residues with percent participation time; **c** distribution of non-covalent interactions; **d** representative snapshot (at 200th ns): AVP1145 (dark violet) and nucleo- protein (grey); **e** distribution of binding free energies (kcal/mol); **f** average distance of interactions (hydrogen, hydrophobic, electrostatic, and other) in AVP1145-NP complexes over 200 ns

Conclusions

Nucleoprotein of SARS-CoV-2 is a very promising target to design and develop anti-COVID therapeutics. In this research, 21 antiviral peptides, well-known to have experimental evidence against nucleoprotein of other viruses were attempted to repurpose against nucleoprotein of SARS-CoV-2 as potential inhibitors. Two peptides namely, AVP1145 and AVP1155 were identified as very promising candidates, which showed strong hydrogen, hydrophobic and electrostatic interactions with the intended target. Key interaction points of the peptides as well as

the binding pocket in the target were demonstrated. The dynamic behavior of the peptides-protein complexes was pursued and changes in structural parameters and energy landscape during a long duration simulation in a physiologically relevant environment were disclosed. Altogether, we set forth a framework to design peptide therapeutics against a suitable target and we believe that our research framework and preliminary computational data may serve as a starting point to rationally design peptide therapeutics with strong inhibitory potential against the nucleoprotein of SARS-CoV-2.

Declarations

Conflict of interest All the researchers involved in this project are enlisted in the author's list. Every author confirms that this work is original and has not been published elsewhere, nor it is currently under consideration for publication elsewhere. We have no conflict of interest to disclose.

References

Ahmed SF, Quadeer AA, McKay MR (2020a) Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. Viruses. https://doi.org/10.3390/v12030254

Ahmed S, Mahtarin R, Ahmed SS et al (2020b) Investigating the binding affinity, interaction, and structure-activity-relationship of 76 prescription antiviral drugs targeting RdRp and Mpro of SARS-CoV-2. J Biomol Struct Dyn 0:1–16. https://doi.org/10.1080/07391102.2020.1796804

Allouche A (2012) Software news and updates Gabedit — a graphical user interface for computational chemistry softwares. J Comput Chem 32:174–182. https://doi.org/10.1002/jcc

Almazán F, Galán C, Enjuanes L (2004) The nucleoprotein is required for efficient coronavirus genome replication. J Virol 78:12683–12688. https://doi.org/10.1128/jvi.78.22.12683-12688.2004

Bowers KJ, Chow DE, Xu H, et al (2006) Scalable algorithms for molecular dynamics simulations on commodity clusters. In: SC '06: proceedings of the 2006 ACM/IEEE conference on supercomputing. p 43

Chowdhury SM, Talukder SA, Khan AM et al (2020) Antiviral peptides as promising therapeutics against SARS-CoV-2. J Phys Chem B 124:9785–9792. https://doi.org/10.1021/acs.jpcb.0c05621

COVID Live - Coronavirus Statistics - Worldometer. https://www.worldometers.info/coronavirus/. Accessed 14 Sep 2022

De Jong S (1990) Multivariate calibration, H. Martens and T. Naes, Wiley, New York, 1989. ISBN 0 471 90979 3. Price: £75.00, US$138.00. No. of pages: 504. J Chemom 4:441–441. https://doi.org/10.1002/cem.1180040607

Efaz FM, Islam S, Talukder SA et al (2021) Repurposing fusion inhibitor peptide against SARS-CoV2. J Comput Chem 42:2283–2293. https://doi.org/10.1002/jcc.26758

Green TJ, Xin Zhang G (2021) Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. Am Assoc Adv Sci 313:357–360

Haggag YA (2018) Peptides as drug candidates: limitations and recent development perspectives. Biomed J Sci Tech Res 8:6659–6662. https://doi.org/10.26717/bjstr.2018.08.001694

Hasan A, Paray A, Hussain A et al (2020) A review on the cleavage priming of the spike protein on coronavirus by angiotensin-converting enzyme-2 and furin. J Biomol Struct Dyn. https://doi.org/10.1080/07391102.2020.1754293

He R, Leeson A, Ballantine M et al (2004) Characterization of protein-protein interactions between the nucleocapsid protein and membrane protein of the SARS coronavirus. Virus Res 105:121–125. https://doi.org/10.1016/j.virusres.2004.05.002

Hurst KR, Ye R, Goebel SJ et al (2010) An interaction between the nucleocapsid protein and a component of the replicase-transcriptase complex is crucial for the infectivity of coronavirus genomic RNA. J Virol 84:10276–10288. https://doi.org/10.1128/JVI.01287-10

Islam R, Parves MR, Paul AS et al (2020) A molecular modeling approach to identify effective antiviral phytochemicals against the main protease of SARS-CoV-2. J Biomol Struct Dyn 0:1–12. https://doi.org/10.1080/07391102.2020.1761883

Jollife IT, Cadima J (2016) Principal component analysis: a review and recent developments. Philos Trans R Soc A Math Phys Eng Sci. https://doi.org/10.1098/rsta.2015.0202

Kassambara A, Mundt F (2017) Factoextra: extract and visualize the results of multivariate data analyses. R Packag Version 1:337–354

Kozakov D, Hall DR, Xia B et al (2017) The ClusPro web server for protein-protein docking. Nat Protoc 12:255–278. https://doi.org/10.1038/nprot.2016.169

Lamiable A, Thévenet P, Rey J et al (2016) PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. Nucleic Acids Res 44:W449–W454. https://doi.org/10.1093/nar/gkw329

Lau JL, Dunn MK (2018) Therapeutic peptides: historical perspectives, current development trends, and future directions. Bioorganic Med Chem 26:2700–2707. https://doi.org/10.1016/j.bmc.2017.06.052

Loschwitz J, Olubiyi OO, Hub JS et al (2020) Computer simulations of protein–membrane systems, 1st edn. Elsevier Inc

Lyman E, Higgs C, Kim B et al (2009) A role for a specific cholesterol interaction in stabilizing the apo configuration of the human A2A adenosine receptor. Structure 17:1660–1668. https://doi.org/10.1016/j.str.2009.10.010

Mark P, Nilsson L (2001) Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. J Phys Chem A 105:9954–9960. https://doi.org/10.1021/jp003020w

Marqus S, Pirogova E, Piva TJ (2017) Evaluation of the use of therapeutic peptides for cancer treatment. J Biomed Sci 24:1–15. https://doi.org/10.1186/s12929-017-0328-x

Mashiach E, Schneidman-Duhovny D, Andrusier N et al (2008) FireDock: a web server for fast interaction refinement in molecular docking. Nucleic Acids Res 36:229–232. https://doi.org/10.1093/nar/gkn186

Modeling of the SARS-COV-2 Genome using I-TASSER. https://zhanglab.ccmb.med.umich.edu/COVID-19/. Accessed 1 May 2020

Peng H, Yang L, tao, Wang L yun, et al (2006) Long-lived memory T lymphocyte responses against SARS coronavirus nucleocapsid protein in SARS-recovered patients. Virology 351:466–475. https://doi.org/10.1016/j.virol.2006.03.036

Pierce BG, Wiehe K, Hwang H et al (2014) ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers. Bioinformatics 30:1771–1773. https://doi.org/10.1093/bioinformatics/btu097

Portela A, Digard P (2002) The influenza virus nucleoprotein: a multifunctional RNA-binding protein pivotal to virus replication. J Gen Virol 83:723–734. https://doi.org/10.1099/0022-1317-83-4-723

Qureshi A, Thakur N, Tandon H, Kumar M (2013) AVPdb: a database of experimentally validated antiviral peptides targeting medically important viruses. Nucleic Acids Res 42:1147–1153. https://doi.org/10.1093/nar/gkt1191

R Core Team (2014) R Core Team (2014). R: a language and environment for statistical computing. R Found Stat Comput Vienna, Austria URL http//wwwR-project.org/

RStudio Team (2015) RStudio: integrated development for R. RStudio, Inc., Boston, MA. URL http://www.rstudio.com/. RStudio, Inc 10

Sakib MMH, Nishat AA, Islam MT et al (2021) Computational screening of 645 antiviral peptides against the receptor-binding

domain of the spike protein in SARS-CoV-2. Comput Biol Med 136:104759. https://doi.org/10.1016/j.compbiomed.2021.104759

Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson H J (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. Nucleic Acids Research 33(Web Server):W363–W367. https://doi.org/10.1093/nar/gki481

Schrodinger LLC (2017). The PyMOL molecular graphics systems, 2.0, https://www.pymol.org/2/

Shan Y, Klepeis JL, Eastwood MP et al (2005) Gaussian split Ewald: a fast Ewald mesh method for molecular simulation. J Chem Phys 122:13. https://doi.org/10.1063/1.1839571

Shukla R, Tripathi T (2020) Molecular dynamics simulation of protein and protein-ligand complexes. Comput Drug Des. https://doi.org/10.1007/978-981-15-6815-2_7

Smith MC, Gestwicki JE (2012) Features of protein-protein interactions that translate into potent inhibitors: topology, surface area and affinity. Expert Rev Mol Med 14:1–20. https://doi.org/10.1017/erm.2012.10

Stuart SJ, Zhou R, Berne BJ (1996) Molecular dynamics with multiple time scales: the selection of efficient reference system propagators. J Chem Phys 105:1426–1436. https://doi.org/10.1063/1.472005

Van Zundert GCP, Rodrigues JPGLM, Trellet M et al (2016) The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes. J Mol Biol 428:720–725. https://doi.org/10.1016/j.jmb.2015.09.014

Wang Q, Li C, Zhang Q et al (2010) Interactions of SARS coronavirus nucleocapsid protein with the host cell proteasome subunit p42. Virol J 7:1–8. https://doi.org/10.1186/1743-422X-7-99

Weng G, Wang E, Wang Z, Liu H, Zhu F, Li D, Hou T (2019) HawkDock: a web server to predict and analyze the protein–protein complex based on computational docking and MM/GBSA. Nucleic Acids Res 47(W1):W322–W330. https://doi.org/10.1093/nar/gkz397

Wurm T, Chen H, Hodgson T et al (2001) Localization to the nucleolus is a common feature of coronavirus nucleoproteins, and the protein may disrupt host cell division. J Virol 75:9345–9356. https://doi.org/10.1128/jvi.75.19.9345-9356.2001

Xue LC, Rodrigues JP, Kastritis PL et al (2016) PRODIGY: a web server for predicting the binding affinity of protein–protein complexes. Bioinformatics 32:3676–3678. https://doi.org/10.1093/bioinformatics/btw514

Yang J, Yan R, Roy A et al (2014) The I-TASSER suite: protein structure and function prediction. Nat Methods 12:7–8. https://doi.org/10.1038/nmeth.3213

Zhou R, Zeng R, von Brunn A, Lei J (2020) Structural characterization of the C-terminal domain of SARS-CoV-2 nucleocapsid protein. Mol Biomed 1:1–11. https://doi.org/10.1186/s43556-020-00001-4