In silico drug designing for COVID-19: an approach of highthroughput virtual screening, molecular and essential dynamics simulations

Rakesh Kumar
Dr.B.R.A.-Institute Rotary Cancer Hospital, All India Institute of Medical Sciences, New Delhi, INDIA-110029 https://orcid.org/0000-0002-1272-7560

Rahul Kumar
Dr.B.R.A.-Institute Rotary Cancer Hospital, All India Institute of Medical Sciences, New Delhi, INDIA-110029

Pranay Tanwar (pranaytanwar@gmail.com)
Dr.B.R.A.-Institute Rotary Cancer Hospital, All India Institute of Medical Sciences, New Delhi, INDIA-110029 https://orcid.org/0000-0002-2357-976X

Research Article

Keywords: SARS-CoV2, computational drug design, main protease, MD simulation

Posted Date: May 11th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-28221/v1

In silico drug designing for COVID-19: an approach of highthroughput virtual screening, molecular and essential dynamics simulations

Rakesh Kumar, Rahul Kumar and Pranay Tanwar*

Dr.B.R.A.-Institute Rotary Cancer Hospital, All India Institute of Medical Sciences, New Delhi, INDIA-110029

***Corresponding author:** Dr. Pranay Tanwar

Dr.B.R.A.-Institute Rotary Cancer Hospital, All India Institute of Medical Sciences, New Delhi, INDIA-110029

E-mail: pranaytanwar@gmail.com

ORCID Id

Rakesh Kumar: http://orcid.org/0000-0002-1272-7560

Pranay Tanwar: http://orcid.org/0000-0002-2357-976X

Total main figures : 10

Total main table : 1

Supplementary figures : 11

Supplementary tables : 7

Abstract

SARS-CoV2, a new coronavirus has emerged in Wuhan city of China, December last year causing pneumonia named COVID-19 which has now spread to entire world. By April 2020, number of confirmed cumulative cases crossed ~2.4 million worldwide, according to WHO. Till date, no effective treatment or drug is available for this virus. Availability of X-ray structures of SARS-CoV2 main protease (M^{pro}) provided the potential opportunity for structure based drug designing. Here, we have made an attempt to do computational drug design by targeting main protease of SARS-CoV2. Highthroughput virtual screening of million molecules and natural compounds databases was performed followed by docking. Six ligands showed better binding affinities which were further optimized by MD simulation and rescoring of binding energy was calculated through MM/PBSA method. In addition, conformational effect of various ligands on protein was examined through essential dynamics simulation. Three compounds namely ZINC14732869, ZINC19774413 and ZINC19774479 were finally filtered that displayed high binding free energies than N3 inhibitor and form conformationally stable complex. Hence, current study features the discovery of novel inhibitors for main protease of CoV2 which will provide effective therapeutic candidates against COVID19.

Introduction

Emergence of SARS (Severe acute respiratory syndrome coronavirus) in 2002-03 from Guangdong, China and MERS (Middle East respiratory syndrome coronavirus) in 2012-13 from Jeddah, Saudi Arabia toll up to ~10617 cases with a mortality of ~1640 as per WHO (World Health Organization) by January, 2020[1-2]. In March-2020, because of rapidly spreading of disease from Wuhan to the entire world, WHO declared pandemic and named this novel virus as 2019-novel coronavirus disease (COVID-19)[3]. According to WHO, by April 2020, number of confirmed cumulative cases crossed ~2.4 million with a fatality rate of 6.7% globally (Supplementary Table S1).

Coronavirus (CoV) is an enveloped virus, consist of a positive sense single stranded RNA (ssRNA) as a genetic material, comprising about 30kb genome size[4]. CoV is spherical in shape with spikes projecting on its outer surface, which help it to adhere on the host cell, followed by structural changes in the host cell that allow virus to enter into the host[5]. Upon entry, it started utilising host replication machinery for replication and accelerated its copy numbers in the cytoplasm. Its genome encodes various structural and accessory proteins which interfere innate immunity of the host[6]. Phylogenetic study reveals about its frequent natural and intermediate host alteration action[4]. The ability of coronavirus to adapt itself into new environment through mutation and recombination, fuel its evolution and constructed various strains that are highly pathogenic in nature[7]. These strains act as causative agent and lead to gastrointestinal, central nervous system and other diseases in addition to respiratory syndrome[8]. There are mainly 4 genera of coronavirus reported till now. Out of which, mammals are to be an ideal host for *Alphacoronavirus* and *Betacoronavirus* while an Aves hosts for *Gammacoronavirus* and *Deltacoronavirus*[9].

SARS-CoV2 is one such *Betacoronavirus* believed to be originated from their natural host, Bat. They are contagious and susceptible to infect people of all ages. Generally, the mode of transmission is through droplets or direct contact with infected person. But presence of SARS-CoV2 in anal swap and blood may indicates its transmission through multiple routes[10]. With the advancement of SARS-CoV2, millions of people from entire world are kept into quarantine. Current unparalleled trend in transmission suggests that influence of degree of climate as people living in cool and dry condition are more vulnerable to its exposure. High temperature (>15°C) forge less effective transmission but does not stop it[11] (Supplementary Table S2). Broadly, the incubation period of SARS-CoV2 infection is with a median of 14 days. However, this period depends upon the age and immune system of the patients. The

primary target of COVID-19 is lungs and cause severe respiratory disorders such as pneumonia, acute respiratory distress syndrome. In addition, it can also lead to systemic disorders such as diarrhoea, acute cardiac injury[12].

To combat with the family of deadly virus, there are many potential targets which are accessible for the sighting of an anti-viral drugs and one such target is main protease (M^{pro})[13]. M^{pro} or 3C-like protease, $3CL^{pro}$, is a highly conserved non-structural protein 5 (Nsp5) encoded by ORF1. It plays a crucial role in processing of polyproteins into mature Nsp's and this will in turn regulates viral life cycle by forming component of replication or transcription machinery. The essentiality of M^{pro} build them potential target for the discovery of antiviral drugs against coronavirus[14]. Recently, X-ray diffracted crystal structure of CoV2 main protease provide an excellent target for structure based drug designing against COVID19[15].

In the present study, highthroughput virtual screening targeting MD optimised tertiary structure of main protease (M^{pro}) of SARS-CoV2 was carried. Hit compounds were docked and high scored docking complexes were optimized and refined by MD simulations. Rescoring of binding free energy was done and conformational study was further proceeded. Hence, three ligands were filtered that showed better binding affinities and formed conformationally stable complex with M^{pro}, suggesting that these compounds act as potent inhibitors of main protease.

Results

Structure and dynamics of CoV2 main protease (M^{pro})

3D structure of CoV2 main protease (M^{pro}) was elucidated by X-crystallography at 2.1Å resolution[15]. M^{pro} consists of 3 domains wherein domain I (8-101 amino acids long) and domain II (102-184 amino acids long) have an antiparallel β-barrels. Domain III is 201-303 amino acids long segment, comprising an antiparallel globular cluster of five α-helices (Fig. 1A). Domains II and III is connected by large loop that provide site for substrate binding. Additionally, the protein has Cys-His (Cysteine-Histidine) catalytic dyad in between domains I and II which also formed binding pocket together with large loop. Tertiary structure taken from PDB was cleaned and validated by inspection of its stereochemical properties. About 99 and 0.4% residues were placed in favoured and disallowed regions of Ramachandran map, respectively and structure was having better 3D quality (Supplementary Fig. S1A and Table S3). Validated structure was subjected to MD simulation.

MD simulation was performed to investigate the stability and dynamics of protein. RMSD (root mean square deviation) was plotted with respect to time and showed stable behaviour

throughout the simulation with average value ~0.28nm (±0.04) except small deviation was found at 42ns time (Fig. 1B). RMSF or root mean square fluctuation with respect to amino acid was plotted and found no major fluctuation except residues at ~140-145 that showed moderate fluctuation with an average RMSF ~0.5nm (Fig. 1C and D). Protein compactness was estimated by measuring the Rg or radius of gyration, plotted against function of time and showed steady behaviour with an average Rg value ~2.23nm (±0.01) (Supplementary Fig. S1B). Overall structure of protein remained stable during MD simulation. But what will be the dynamics about protein local structures, such as secondary structure of protein? Next, we wanted to check the local stability of protein which was monitored through the inspection of secondary structure formation during MD simulation. Different moieties such as α-helices, β-sheets, coils, bends and turns were examined and found that, coil content was decreasing concomitantly with increase in β-sheets at the end of simulation period and α-helices remained stable (Fig. 2A and B). Additionally, the loop comprises ~15 residues (185-200), connecting domains II and III, interrupted with coils and bends in between and at the end of simulation periods (Fig. 2A). MD simulation results suggested that protein remained stable with minor changes in disorder regions implying that protein is ready for further downstream processes.

Virtual screening and docking

High throughput virtual screening was performed by RASPD tool and ~1 million compounds from million molecule and natural compound databases were screened. Top RASPD scored of ~100 compounds from each database were further docked individually by Autodock Vina tool (Supplementary Tables S4 and S5). X-rays structure of M^{pro} (protein-N3 inhibitor complex) was taken as a reference for binding site where all compounds were docked[15]. Binding pocket of protein was also predicted through computational methods. Binding site of protein was comprised about 38 residues with high Z score (14.2) and composed of both hydrophilic and hydrophobic residues (Supplementary Fig. S2A). Asparagine142 (present at turn between 4^{th} and 5^{th} β-sheets of domain II), Histidine164 (present at 6^{th} β-sheet of domain II), Proline168 (present at turn between 6^{th} and 7^{th} β-sheets of domain II) and Glutamine189 (present at loop connecting domains II and III) acts as key residues for ligand binding[15]. In addition to that, residues lied at cleft or surface of pocket also assisted in binding of ligand other than key residues as mentioned above. Docking results suggested that approximately 6 compounds namely ZINC14732869 (-8.5 kcal/mol), ZINC19774413 (-8.8 kcal/mol), ZINC12338080 (-8.6 kcal/mol), ZINC123845408 (-9.3 kcal/mol), ZINC19774479 (-8.7

kcal/mol) and ZINC31 (-7.3 kcal/mol) exhibited higher docking scores than N3 (known) inhibitor (-7.2 kcal/mol). We had also docked hydroxychloroquine that showed -5.6 kcal/mol docking score. Higher docking scores exhibited by novel compounds over N3 (known) inhibitor demonstrating the novel compounds have better binding affinities towards the M^{Pro} protein. Docking of novel inhibitors were superimposed with N3 inhibitor to check the similarities between common binding mode of all compounds (Supplementary Fig. S2B). Drug likeliness of novel compounds were examined through Lipinski rule of 5 and found that all novel compounds followed the rule, implying that newly screened compounds were having drug like properties (Supplementary Table S6). All known and novel compounds were bounded by both hydrophilic and hydrophobic residues, out of which N3 inhibitor formed 16 bonds (hydrophobic:10; hydrophilic:6), hydroxychloroquine formed 10 bonds (hydrophobic:9; hydrophilic:1), ZINC14732869 formed 16 bonds (hydrophobic:14; hydrophilic:2), ZINC19774413 and ZINC12338080 also formed 16 bonds (hydrophobic:15; hydrophilic:1), ZINC123845408 formed 12 bonds (hydrophobic:9 ; hydrophilic:3), ZINC19774479 formed 15 bonds (hydrophobic:14; hydrophilic:1) and ZINC31 formed 10 bonds (hydrophobic:9; hydrophilic:1) (Fig. 3A-H and Supplementary Table S6). Quantitative analysis of different bonding revealed that the compounds namely ZINC14732869, ZINC19774413, ZINC12338080 and ZINC19774479 formed fair numbers of bonds with receptor molecule. Top scored ligand compounds were further carried for ADMET (Absorption, distribution, metabolism, excretion and toxicity) analysis. ADMET properties were predicted by admetSAR server and found that almost all compounds were having better absorption and distribution, good metabolic and excretion profiles. Moreover, all compounds were nontoxic and non carcinogens (Supplementary Table S7). Top scored protein-ligand complexes along with known and hydroxychloroquine were further optimized by MD simulation.

MD simulation of protein-ligand complexes and rescoring of binding free energy

MD simulation is advantageous for studying the stability of binding pose obtained from docking, to rescore the overall binding energy as well as energy contribution by different residues, to study the conformational changes occurring in protein due to ligand binding/unbinding etc[16-18]. Here, MD simulation of 50ns was carried to monitor the protein-ligand stability and rescore the binding energies of top docking complexes. RMSD of protein-ligand complexes were measured with respect to time. RMSD pattern of protein bound with N3 inhibitor was stabilised after 25ns with an average value 0.31nm (±0.04) (Fig. 4A and

Supplementary Fig. S3A). RMSD of protein bound with hydroxychloroquine was also stabilised after 25ns but showed minor deviations at the end of simulation period with average value 0.23nm (±0.03) (Fig. 4A and Supplementary Fig. S3A). On the other hand, the RMSD of protein bound with novel inhibitors such as ZINC14732869, ZINC19774413, ZINC12338080, ZINC123845408, ZINC19774479 and ZINC31 displayed consistent and stable behaviour with average values 0.25nm (±0.02), 0.24nm (±0.02), 0.23nm (±0.03), 0.22nm (±0.03), 0.21nm (±0.02) and 0.22nm (±0.02), respectively (Fig. 4A and B; Supplementary Fig. S3A). RMSD of whole protein bound with ligands remain stable and consistent. Since, the binding pocket of protein has large volume (~38 residues), composed of small turns and sheets, therefore we were interested to know the RMSD behaviour of pocket in bound and unbound form of proteins. RMSDs of pocket remain steady and slightly increased as compared to the RMSDs of whole protein (Supplementary Fig. S3B,E,F). Stability of ligands inside the binding pocket of protein was assessed by examining the ligand RMSD. N3 and hydroxychloroquine ligands showed maximum RMSD with average values 0.50nm (±0.07) and 0.22nm (±0.05), respectively and deviations during entire simulation period (Fig. 4C; Supplementary Fig. S3A). On the other hand, the RMSD of novel compounds ZINC14732869, ZINC19774413, ZINC12338080, ZINC123845408, ZINC19774479 and ZINC31 showed stable and consistence behaviour with minimum average RMSD values 0.11nm (±0.02), 0.13nm (±0.02), 0.15nm (±0.03), 0.13nm (±0.03), 0.11nm (±0.02) and 0.12nm (±0.02), respectively (Fig. 4A and B; Supplementary Fig. S3A). Large RMSD values of N3 and hydroxychloroquine ligands indicated that the ligands remain unstable in the binding pocket of the proteins, while smaller RMSD values of novel compounds implying that they are stable in the binding pocket.

To reveal the internal fluctuations of protein in presence of ligands, the RMSF with respect to residues was measured (Fig. 4E and F; Supplementary Fig. S4). Moderate fluctuations of Ser46 (Serine), Glu47 (Glutamate), Asp48 (Aspartate), Met49 (Methionine), Leu50 (Leucine), Tyr154 (Tyrosine), Asp197, Trp218 (Tryptophan), Arg222 (Arginine), Gly275 (Glycine), Met276, Asn277 (Asparagine), Gly278, and Arg279 were observed in almost all the protein-ligand complexes with average RMSF values ranging from 0.35 (minimum) to 0.59 (maximum). 3D structure inspection of these residues was investigated and observed that these residues occupied at loop or turn regions of protein. RMSF result suggested moderate fluctuations were found in loop and turn regions of protein indicating that these regions assist in ligand binding, as no such fluctuations were observed in apo protein.

Fluctuations of different ligands were also monitored at atomic level and found that N3 and hydroxychloroquine along with ZINC12338080 showed higher mobility consistent with RMSD result (Supplementary Fig. S5). Rest of compounds displayed fluctuations at flanking or extreme ends, which normally existed.

Effect of different ligands on protein compactness were gauged by radius of gyration (Rg) with function of time. Quantitative analysis of Rg of various ligands bound with protein showed no significant difference, but qualitative or different pattern of Rg was observed in different protein-ligands complexes throughout the simulation period (Fig. 4G and H). Average values of Rg of protein complexed with N3, hydroxychloroquine, ZINC14732869, ZINC19774413, ZINC12338080, ZINC123845408, ZINC19774479 and ZINC31 were 2.21nm (±0.01), 2.24nm (±0.05), 2.21nm (±0.01), 2.20nm (±0.01), 2.22nm (±0.01), 2.22nm (±0.01), 2.14nm (±0.01) and 2.22nm (±0.01), respectively (Fig. 4G and H; Supplementary Fig. S3C). Rg results suggested that ligands did not show any major effect on protein compactness. Surface area of protein available for solvent was examined through SASA (Solvent accessible surface area) analysis in presence or absence of ligands. SASA values for apo protein (M^{pro}) was 152nm^2 (±2.92) while for complexes of N3, hydroxychloroquine, ZINC14732869, ZINC19774413, ZINC12338080, ZINC123845408, ZINC19774479 and ZINC31 were 149nm^2 (±2.93), 149nm^2 (±3.21), 152nm^2 (±2.80), 150nm^2 (±2.86), 149nm^2 (±3.64), 151nm^2 (±2.68), 152nm^2 (±2.55) and 151nm^2 (±2.42), respectively (Supplementary Fig. S3D,G,H). SASA values of protein-N3 inhibitor, -hydroxychloroquine, -ZINC12338080 and -ZINC123845408 complexes were slightly reduced as compared to the rest of protein-ligand complexes and apo protein.

Protein secondary structural changes in presence of known and novel ligands were examined though DSSP method during entire simulation period. Compositions of β-sheets were slightly increased in case of protein bound with N3, ZINC19774413 and ZINC19774479, concomitantly with the reduction of bend contents (Fig. 5A,D,G; Supplementary Fig. S6A,D,G). Coil contents of protein bound with hydroxychloroquine, ZINC12338080, ZINC123845408, ZINC19774479 and ZINC31 were increased, while bend and turns were slightly decreased (Fig. 5B,E,G,H; Supplementary Fig. S6B,E,G,H). Secondary structures of Protein-ZINC14732869 complex remain unaffected (Fig. 5C; Supplementary Fig. 6C). α-helices remained stable and unaffected during binding of all ligands (Fig. 5 and Supplementary Fig. S6). Coils were mostly found in domain I and β-sheets were restricted to

domains I and II which formed binding cleft, therefore, variations in the coils and sheets were helping in providing the stable regime for proper ligand binding.

MD simulation approach is helpful in predicting the binding modes and calculate the binding free energy[19]. It generally used to refine the binding of docking complexes and helps in calculating the different binding energy components. Binding free energies of N3, hydroxychloroquine and novel compounds such as ZINC14732869, ZINC19774413, ZINC12338080, ZINC123845408, ZINC19774479 and ZINC31 were estimated from last stabled MD simulated trajectories by Molecular mechanics Poisson Boltzmann surface area approach (MM/PBSA). Protein-N3 and -hydroxychloroquine complexes exhibited -39.869 (±16.169) and -125.197 (±12.244) kJ/mol of binding energies, respectively (Table 1). Van der Waal energy was the major contributor for binding of both N3 (-62.074 ±21.728 kJ/mol) and hydroxychloroquine (-162.538 ±11.498 kJ/mol) compounds. On the other hand, the binding energies of novel compounds such as protein-ZINC14732869, -ZINC19774413, -ZINC12338080, -ZINC123845408, -ZINC19774479 and -ZINC31 were -151.532 (±3.999), -160.704 (±9.221), -50.282 (±16.805), -14.713 (±1.431), -136.809 (±7.656) and -105.356 (±2.44) kJ/mol, respectively. Again, the van der Waal energy was the major contributor for binding of ZINC14732869 (-206.956 ±3.646 kJ/mol), ZINC19774413(-225.656 ±11.387 kJ/mol), ZINC12338080 (-101.097 ±20.882 kJ/mol), ZINC123845408(-18.264 ±12.656 kJ/mol), ZINC19774479 (-184.18 ±13.735 kJ/mol) and ZINC31 (-191.574 ±1.447 kJ/mol) ligands. Results of binding energies of all compounds were similar with the docking results except ZINC123845408 compound which showed high docking score (-9.3kcal/mol) but very less binding energy -50.282kJ/mol (Table 1). Binding energies of hydroxychloroquine, ZINC14732869, ZINC19774413, ZINC19774479 and ZINC31 were higher than the N3 inhibitor, suggested that these novel compounds had strong binding affinities to main protease (M^{Pro}).

Energy decomposition and protein-ligand interaction studies

Binding energy results suggested that 4 novel compounds namely ZINC14732869, ZINC19774413, ZINC19774479 and ZINC31 along with hydroxychloroquine displayed higher binding affinity than N3 inhibitor. Here, we are interested to know how many residues contribute in binding energy (Figs. 6 and Supplementary Fig. S7). Energy distribution on residues basis has showed that protein bound with N3, ZINC19774479 and ZINC31 compounds exhibited a smaller number of residues contributed to total binding energy (Figs.

6A,E,F and Supplementary Fig. S7A,E,F). Also, the magnitude of energy contribution by these residues were lower which range in minimum (Thr25:-1.2kJ/mol) to maximum (Asp48:-1.7kJ/mol) in case of protein-N3 complex (Figs. 6A and Supplementary Fig. S7A), while protein complexed with ZINC19774479 (Met165:-8.1kJ/mol) and ZINC32 (Leu50:-4.8kJ/mol) compounds showed maximum value of energy contribution (Figs. 6E,F and Supplementary Fig. S7E,F). Rest of protein-ligand complexes along with hydroxyquinoline exhibited fair number of residues contributing in total energy (Figs. 6B,C,D and Supplementary Fig. S7B,C,D; Supplementary Table S4). Compounds ZINC14732869, ZINC19774413 bonded with maximum number of residues with higher energy values (Figs. 6C,D and Supplementary Fig. S7C,D; Supplementary Table S4). Energy decomposition results demonstrated that Leu27, Met49, Leu50 and Met165 were the key residues in ligand binding in all protein-ligand complexes.

Protein-ligand interactions were monitored through the inspection of morphology of binding pocket, hydrophobic interactions and number of hydrogen bonds formed during simulation. Above results indicating that hydroxychloroquine and 4 novel compounds such as ZINC14732869, ZINC19774413, ZINC19774479 and ZINC31 showed higher binding affinity than N3 inhibitor. From the above results it is also cleared that van der Waal energy is the major contributor for binding free energy (Table 1) of all ligands and hydrophobic residues (Methionine and Leucine) played a significant role in protein-ligand binding (Fig. 3 and Supplementary Table S4). Binding pocket analysis of protein complexed with known and novel ligands showed that all ligands were well accommodated in the binding pocket of protein and remained stable during MD simulation except known ligand that orient against surface of binding pocket (Supplementary Fig. S8A-F). Moreover, hydrogen bond (H-bond) formation was higher in known ligand with an average of 4-5 H-bonds were formed during MD simulation in contrast to 3-4 H-bonds were existed in novel ligands (Supplementary Fig. S9A-F). Similar results were also observed during docking as hydrophilic interactions were more in known compounds as compared to the hydrophobic interactions which were mainly found in novel ligands (Fig. 3 and Supplementary Table S4). Protein-ligand interactions study demonstrated that novel compounds were mainly stabilised by hydrophobic and partially by hydrophilic interactions. RMSD, RMSF, Rg, SASA and binding energy results suggested that novel compounds such as ZINC14732869, ZINC19774413, ZINC19774479 were most stable and potent inhibitors for M^{pro}.

Protein conformation and free energy landscape analyses

Ligand induced conformational changes were existed during MD simulation. The superimpositions of high scored binding energy complexes such as protein-ZINC14732869, -ZINC19774413, -ZINC19774479 along with N3 and hydroxychloroquine from the final snapshots of MD simulation showed that different ligands induced different conformational changes, even the starting conformation of protein was same during the MD simulation[20] (Fig. 7). Conformational changes were more pronounced in turn (between 4^{th} and 5^{th} β-sheets of domain II), loop (domains II and III) and α-helices (domain III). To further gain insight into the conformational changes induced during binding and unbinding of known and novel ligands, the principle component analysis (PCA) or essential dynamics was performed and stable conformation (minimum ΔG) of all complexes were obtained from free energy landscape study. Principle components analysis was performed to identify the dominant motions occurring in different complexes, where major dominant motions were achieved in first few eigenvectors (10) or PCs (Supplementary Fig. S10A). PCA indicated that first 3 eigenvectors (or PCs) accounted for 72.50, 77.68, 81.80, 76.79, 79.84 and 69.19% of the motions observed in protein (M^{pro}), protein-N3, -hydroxychloroquine, -ZINC14732869, -ZINC19774413 and -ZINC19774479 complexes, respectively (Supplementary Fig. S10A). Conformational behaviours of different protein-ligand complexes were monitored by plotting first two PCs in phase space (Supplementary Fig. S10B-F). Protein-N3 complex showed unequal space in phase space as compared to protein-hydroxychloroquine, -ZINC14732869, -ZINC19774413 and -ZINC19774479 complexes, indicating that protein-N3 complex is still exploring the conformation or large conformational changes may have existed. Conformational changes at structural level of different complexes were examined by assessing the low energy structures from PC1 and PC2 in free energy landscape (FEL) area. 3D and 2D plots of FEL were constructed to show the different energy barriers (Figs. 8,9 and Supplementary Fig. S11). FEL results suggested that protein complexed with N3 (RMSD:1.78Å) and hydroxychloroquine (RMSD:1.78Å) ligands undergo large conformational changes during MD simulations (Fig. 8B,E), as compared to protein complexed with -ZINC14732869 (RMSD:1.77Å), -ZINC19774413 (RMSD:1.44Å) and -ZINC19774479 (RMSD:1.59Å) ligands (Fig. 9B,E,H). Additionally, conformational changes were mostly restricted at the turn connecting 4^{th} and 5^{th} β-sheets of domain II, loop connecting domains II and III and α-helices in domain III as major motions was occurring in those regions.

Porcupine structures were generated from respective PCs to elucidate the nature of motions occurred in protein-ligand complexes and found that different rotational motions were existed in different protein-ligand complexes (Figs. 8 and 9). Arrows of porcupine structures indicated direction of motions and length of arrows denoted amplitude of motions. Interestingly, turns of domain II and α-helices of domain III face each other, indicating the expansion of binding cleft or increase in the volume of binding pocket that results instability of ligands in case of protein-N3 and -hydroxychloroquine complexes (Fig. 8C,F). In contrast, turns of domain II and α-helices of domain III moves against each other in protein-ZINC14732869, -ZINC19774413 and -ZINC19774479 complexes, indicating the contraction of binding cleft or restricted the volume of binding pocket, that provide stable regime to ligand binding (Fig. 9C,F,I). Asn142, His164, Pro168 and Gln189 are the key residues of protein which maintain morphology of binding cleft and also help in binding of the ligand[15]. Asp142 present at turn between 4^{th} and 5^{th} β-sheets of domain II, Hist164 lied at 6^{th} β-sheet of domain II, Proline168 occurred at turn connecting 6^{th} and 7^{th} β-sheets of domain II and Glutamine189 present at loop connecting domains II and III, assisted in the proper binding of ligands (Fig. 10). Conformational difference at structural level or relative alterations in binding pockets in various complexes were attributed by movement of Asn142, His164, Pro168 and Gln189 residues (Fig. 10G). Distance between Pro168-Gln189 and His164-Gln189 residues were changing in protein-ZINC14732869, -ZINC19774413 and -ZINC19774479 complexes as compared to protein-N3 and -hydroxychloroquine complexes and remain persistent during simulation period (Fig. 10B,C,E,F). The PCA results demonstrated that large conformational changes were existed in proteins when bound with N3 and hydroxychloroquine as compared to novel ligands. Moreover, residues near binding site showed large motions that might play an important role by providing proper orientation for ligand binding.

Discussion

In December, 2019 an outbreak of novel coronavirus was reported from Wuhan, China and its dissemination developed as epidemic[3]. In March, 2020, because of rapidly spreading of disease from Wuhan to the entire world, WHO (World Health Organization) declared pandemic and named this novel virus as COVID-19. In the absence of appropriate randomised controlled trial of any drug or vaccine, some measures suggested by various health institutions including WHO to curtail the infection of SARS-CoV2, such as self-quarantine, maintaining respiratory hygiene, social distancing etc. In the awake of this current

pandemic, preventive medicine on the basis of preliminary data like hydroxychloroquine, a derivative of chloroquine being used against COVID patients[21]. However, the mode of action of hydroxychloroquine is not fully understood, but it is known to increase the pH of endosome which arrest its maturation and thus prevents the entry of virus into the cytoplasm[22]. Recent study found that the medication of hydroxychloroquine causes several side effects such as cardiac arrest etc[23]. In the current scenario, the coarse and cost of drug development by conventional methods are the major hurdles for drug discovery against such type of fast emerging diseases. In short of time, the drug repurposing of already approved drugs is the only approach to identify the potential drug candidates against such disease. In this situation, *in silico* methods of drug designing play a fateful role to minimize the drug development hurdles. Few *in silico* studies have been come in last couple of months which identified various drug compounds targeting the viral proteins. One computational docking study has proposed few drugs such as Remdesivir, Sofosbuvir, Galidesivir, and Tenofovir that bind to RNA dependent RNA polymerase (RdRp) with high affinity[24]. Another deep docking study has screened thousands of hit compounds against viral main protease[25]. Recently, the X-ray structure of main protease of CoV2 virus has been submitted on PDB in complex with N3 inhibitor[15]. The inhibitor was identified through structure based virtual screening against the library consisting of clinical trial drug candidates, already approved drugs and natural compounds. None of the studies have provided the details of binding analysis of drugs and its effect on protein structure and conformation, a crucial parameter for precise or effective drug development.

Here, we have used atomistic computational microscope or molecular dynamics simulation technique to understand the nature and strength of binding of known drugs. We have also screened novel drug compounds that displayed high binding affinity and to reveal the conformational stability of drug binding to viral main protease (M^{pro}) along with known inhibitors. Present study was undertaken to *in silico* drug designing for COVID-19 by highthroughput virtual screening followed by docking. Drug binding modes were inspected through MD simulation and rescoring of binding energy was carried out through MM/PBSA method. Furthermore, effect of drugs on structures and conformation of protein were revealed by essential dynamics simulation and stability of protein structures were examined through free energy landscape study. For this, experimental resolved 3D structure of main protease (M^{pro}) was downloaded from PDB and stereochemical evaluation was done after energy minimisation. M^{pro} consists of 3 domains having mixtures of helices, sheets and loop with better stereochemical geometry. Protein structure was optimized through MD simulation and

different parameters such as RMSD, RMSF and Rg were measured. RMSD of structure showed consistent pattern and no major fluctuations were observed during RMSF analysis (Fig. 1). MD optimized structure was carried for highthroughput virtual screening and docking.

Nearly, 1 million compounds from million molecule and natural compound libraries were screened and top hit compounds were docked individually. Docking results demonstrated that novel compounds displayed high binding affinities than N3 and hydroxychloroquine. Physiochemical properties or drug likeliness of novel compounds showed drug like properties. Computational prediction of ADMET (Absorption distribution metabolism excretion toxicity) properties of newly drug are helpful to minimize the failure of drugs during clinical trials. Therefore, ADMET properties of novel compounds were examined and found that all drugs were having good absorption, distribution, metabolism and excretion properties and are less toxic with minimum side effects. Top binding modes obtained from docking were optimized and refined by MD simulation. Stabilities of all protein-ligands complexes were compared during 50ns MD simulation time. Low values and consistent pattern of RMSD of protein complexed with novel compounds were existed as compared to the RMSD of protein complexed with known (N3) and hydroxychloroquine compounds. Moreover, ligand RMSD of known and hydroxychloroquine compounds showed higher values and deviations as compared to all novel ligands, suggesting that novel compounds formed stable complex with M^{pro} in contrast to known inhibitor (Fig. 4). Effect of drugs on local protein structure were examined through RMSF analysis of protein and found that moderate fluctuations were observed in all protein-ligand complexes.

Rescoring of binding free energies of top scored docking complexes along with known and hydroxychloroquine were analysed using MD simulation trajectories and different binding energy components were measured. Binding free energy results revealed that ZINC14732869 (-151.532 ±3.999 kJ/mol), ZINC19774413 (-160.704 ±9.221kJ/mol), ZINC19774479 (-136.809 ±7.656kJ/mol) compounds interact with M^{pro} with highest free energies as compared to N3 (-39.869 ±16.169kJ/mol) and hydroxychloroquine (-162.538 ±11.498 kJ/mol) compounds, consistent with the docking results (Table 1). Energy decomposition and protein-ligand interaction studies demonstrated that Leu27, Met49, Leu50 and Met165 were the key residues in ligand binding in all protein-ligand complexes and both hydrophobic and hydrophilic forces played an important role in ligand binding (Fig. 6). Conformationally stable binding of ligand is necessary for effective drug development. Therefore, we elucidated the conformational changes induced by different ligands and examined the

stability of protein-ligand complexes through PCA and FES studies. PCA studies suggested that first 3 PCs accounted majority of protein motions. However, protein-N3 complex showed large conformational changes as compared to the rest of protein-ligand complexes (Figs. 8 and 9). Major motions were observed in loop and turns regions spanning near the binding pocket of proteins and α-helices of protein domain. Motions near the binding pocket was attributed by alterations in distance of key residues lied at the binding site which probably assist in stable binding of ligands (Fig. 10).

MD simulation technique is a promising approach for discovering the novel drug and help in understanding the dynamics and conformational changes induced by the ligands[26-28]. In the current study, our MD simulation was restricted to 50ns time scale which is however sufficient for discovering the new drug candidates for rapid growing disease such as COVID-19. We reckon that the newly discovered novel drugs from this computational study will provide the effective therapeutics against COVID-19.

Conclusion

The study documented here, has presented the computational findings of novel and potent inhibitors for CoV2 main protease. Highthroughput virtual screening coupled with docking was performed. MD simulation was applied to examined the stability and conformationally stable binding of ligands. Finally, 3 main compounds namely ZINC14732869, ZINC19774413 and ZINC19774479 were discovered with relatively better and promising binding affinities that may have huge impact in development of effective therapeutics against COVID-19 patients.

Materials and methods

Structure preparation and validation

Tertiary structure of Coronavirus2 (CoV2) main protease (M^{Pro}) was taken from Protein data bank (PDB:6LU7)[29]. X-ray crystallographic structure of CoV2 with N3 inhibitor complex was diffracted at a resolution of 2.1Å[15]. 3D structure of protein was extracted and cleaned in PyMOL. Structure was energy minimized in SwissPDB viewer. After that, stereochemical properties of minimized protein structure were assessed through SAVES (structure analysis and verification server) and ProSA web servers[30,31].

Molecular dynamics simulation protocol

Protein was subjected to atomistic molecular dynamics (MD) simulation by Gromacs 5.0 in conjunction with Gromos54a7 force field[32,33]. Initially, protein topology was prepared using pdb2gmx module and structure was placed in triclinic box with periodic distance 1.5nm between structure and edge of box. Protein was solvated using SPC/E (simple point charge/extended) water model and counterions were added to neutralize the system. After that, system was energy minimized by steepest descent method. NVT and NPT equilibration simulations were performed for 100 and 500ps respectively. All bonds were restrained using Lincs algorithm[34]. Temperature of 300K and pressure of 1 bar were maintained using v-rescale (modified Berendsen thermostat) and Parrinello-Rahman barostat, respectively. Particle Mesh Ewald (PME) method was used to process the electrostatic interactions. Finally, production simulation of 50ns was carried whereas 0.002ps of time step was selected and trajectories were recorded at every 10ps.

Highthroughput virtual screening and molecular docking

Highthroughput virtual screening was performed against million molecules and natural compounds databases of ZINC server by employing RASPD tool[35,36]. Top scored drugs from each library were further docked individually by Autodock Vina[37]. Protein and ligand structures preparation and different file format conversions were accomplished through MGL Tools1.5.6 and OpenBabel tools, respectively[38,39]. Drug likeliness of screened compounds were monitored through Lipinski filter[40]. During Autodock Vina, the receptor molecules were prepared in Auto Dock tools in which polar hydrogens were added and nonpolar hydrogen were merged. Ligand molecules were prepared with the addition of Gasteiger charges. Grid box of 28x28x28 dimension and size of x, y, z coordinates (-13.820, 21.037, 73.265) were prepared to cover the target binding site. Dimension and size of binding pocket was taken from the experimentally known binding site available at PDB structure[15]. Drug binding pocket was also verified through Metapocket server[41]. ADMET (Absorption, distribution, metabolism, excretion and toxicity) properties of top scored drugs were predicted through admetSAR server[42]. Binding energy was measured in kilocalorie per mol and top scored protein-ligand complexes were taken as input for MD simulation.

MD simulation and binding free energy calculation

MD simulation of top scored protein-ligand complexes were carried through Gomacs suite for 50ns time. Protein and ligand topologies were derived from Gromos54a7 and Prodrg server, respectively[43]. Procedure of MD simulation for complexes was same as used in case

of apo protein (M^{pro}). Molecular mechanics Poisson Boltzmann surface area (MM/PBSA) approach was used to calculate the binding free energies of protein-ligand complexes. Binding free energy of all protein-ligand complexes were computed using last 20 ns MD simulated trajectories by g_mmpbsa tool in Gromacs[44] as given in eq. 1.

$$\Delta G_{binding} = G_{complex} - (G_{protein} + G_{ligand}) \quad \ldots 1$$

where, $G_{complex}$ is the total free energy of the protein−ligand complex and $G_{protein}$ and G_{ligand} are total free energies of the isolated protein and ligand in solvent, respectively. Total binding free energy was computed through summing up the electrostatic, polar, van der Waals and solvent accessible surface area energies. In addition, energy decomposition per residues basis was also calculated as given in eq. 2.

$$\Delta R_X^{BE} = \sum_{i=0}^{n} \left(A_i^{bound} - A_i^{free}\right) \quad \ldots 2$$

where, A_i^{bound} and A_i^{free} are the energies of i^{th} atom from x residue in bound/unbound forms respectively. n is the total number of atoms in the residue. Energy contribution summed over all residues is equal to the binding energy, i.e. $\Delta G_{binding} = \sum_{x=0}^{m} \Delta R_X^{BE}$, where, m is the total number of residues in protein-ligand complexes[45]. Since we are comparing different ligands with the same receptor therefore, the entropy term was not included in our analyses. All energies were measured in kilojoule per mol.

Essential dynamics and free energy landscape studies

Essential dynamics (ED) approach was applied for both apo and complex protein systems to understand the overall motion of the protein that are significant to biological functions[46]. After removing the translational and rotational motions, the covariance matrix was constructed and diagonalize as given in equations 3 and 4.

$$\sigma_{ij} = \langle (q_i - \langle q_i \rangle)(q_j - \langle q_j \rangle) \rangle \quad \ldots 3$$

$$A^T C A = \lambda \quad \ldots 4$$

where, q_i and q_j represent mass-adjusted Cartesian coordinates of particles i and j, respectively, where $\langle \ldots \rangle$ is the ensemble average of all MD simulation structure samples over the course of the simulation. ED analysis was restricted to backbone atoms to avoid the statistical noise. In equation 4, C is symmetric matrix, A and λ represents the eigenvectors

and associated eigenvalues, respectively. Eigenvectors and eigenvalues were obtained through gmx covar and analyzed by gmx anaeig tools of Gromacs, respectively.

Free energy landscape (FEL) analysis was performed using first two principle components (PC1 and PC2) or eigenvectors as given in equation 5.

$$\Delta G_{(PC1,PC2)} = -k_b T ln P_{(PC1,PC2)} \quad \ldots 5$$

where, k_b is Boltzmann constant, T is temperature, and $P_{(PC1,PC2)}$ is the probability distribution of molecular system along principle component 1 and principle component 2. FEL was constructed by gmx sham module of Gromacs using the reaction coordinates of PC1 and PC2.

Analysis and data presentation

Root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (Rg), hydrogen bond (H-bond), solvent accessible surface area (SASA) and distance analysis were calculated in Gromacs. Secondary structure plots were calculated by do_dssp tool from MD simulation trajectories[47]. 2D plots of protein-ligand complexes were plotted in LigPlot+ tool[48]. All 3D figures of proteins and protein-ligand complexes for visualization were rendered in UCSF Chimera v1.7 (Computer Graphics Laboratory, University of California, San Francisco) and PyMOL (The PyMOL Molecular Graphics System, Version 1.3 Schrodinger, LLC). 2D graphs were plotted in Prism 6 (GraphPad Software, CA, USA, www.graphpad.com) and MS-Excel-2019 (Microsoft corporation 2018).

Data availability

All data generated or analysed during this study are included in this article/supplementary information.

References

1. Xu, R. H., He, J. F., Evans, M. R., Peng, G. W., Field, H. E., Yu, D. W., et.al. Epidemiologic clues to SARS origin in China. Emerg Infect Dis, 10(6), 1030 (2004).
2. Al-Osail, A. M., & Al-Wazzah, M. J. The history and epidemiology of Middle East respiratory syndrome corona virus. Multidiscip Respir Med, 12(1), 20 (2017).
3. Wang, C., Horby, P. W., Hayden, F. G., & Gao, G. F. A novel coronavirus outbreak of global health concern. The Lancet, 395(10223), 470-473 (2020).

5824. Marra, M. A., Jones, S. J., Astell, C. R., Holt, R. A., Brooks-Wilson, A., Butterfield, Y. S., et.al. The genome sequence of the SARS-associated coronavirus. Science, 300(5624), 1399-1404 (2003).
5855. Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C. L., Abiona, O., et.al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science, 367(6483), 1260-1263 (2020).
5886. Guo, Y. R., Cao, Q. D., Hong, Z. S., Tan, Y. Y., Chen, S. D., Jin, H. J., et.al. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak–an update on the status. Mil Med Res, 7(1), 1-10 (2020).
5917. Graham, R. L., & Baric, R. S. Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. J Virol, 84(7), 3134-3146 (2010).
5938. Weiss, S. R., & Navas-Martin, S. Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus. Microbiol. Mol Biol Rev, 69(4), 635-664 (2005).
5969. Woo, P. C., Lau, S. K., Lam, C. S., Lau, C. C., Tsang, A. K., Lau, J. H., et.al. Discovery of seven novel Mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus. J Virol, 86(7), 3995-4008 (2012).
60110. Zhang, W., Du, R. H., Li, B., Zheng, X. S., Yang, X. L., Hu, B., et.al. Molecular and serological investigation of 2019-nCoV infected patients: implication of multiple shedding routes. Emerg Microbes Infect, 9(1), 386-389 (2020).
60411. Araujo, M. B., & Naimi, B. Spread of SARS-CoV-2 Coronavirus likely to be constrained by climate. medRxiv (2020).
60612. Rothan, H. A., & Byrareddy, S. N. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. J Autoimmun, 102433 (2020).
60813. Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R., & Hilgenfeld, R. Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. Science, 300(5626), 1763-1767 (2003).
61114. Hilgenfeld, R. From SARS to MERS: crystallographic studies on coronaviral proteases enable antiviral drug design. The FEBS journal, 281(18), 4085-4096 (2014).
61315. Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., et.al. Structure of Mpro from COVID-19 virus and discovery of its inhibitors. Nature https://doi.org/10.1038/s41586-020-2223-y (2029).

16. Kumar, R., Maurya, R., & Saran, S. Identification of novel inhibitors of the translationally controlled tumor protein (TCTP): insights from molecular dynamics. Mol BioSyst, 13(3), 510-524 (2017).
17. Chen, Y. C. Beware of docking! Trends in Pharmacol Sci, 36(2), 78-95 (2015).
18. Zhao, H., & Caflisch, A. Molecular dynamics in drug design. Eur J Med Chem, 91, 4-14 (2015).
19. Kumar, R., & Saran, S. Structure, molecular dynamics simulation, and docking studies of *Dictyostelium discoideum* and human STRAPs. J Cell Biochem, 119(9), 7177-7191 (2018).
20. McNicholas, S., Potterton, E., Wilson, K. S., & Noble, M. E. M. Presenting your structures: the CCP4mg molecular-graphics software. Acta Crystallogr D Biol Crystallogr, 67(4), 386-394 (2011).
21. Sinha. N., Balayla. G., Hydroxychloroquine and covid-19. Postgrad Med J- 137785 (2020).
22. Liu, J., Cao, R., Xu, M., Wang, X., Zhang, H., et.al. Hydroxychloroquine, a less toxic derivative of chloroquine, is effective in inhibiting SARS-CoV-2 infection in vitro. Cell Discov, 6(1), 1-4 (2020).
23. Chorin, E., Dai, M., Shulman, E., Wadhwani, L., Bar-Cohen, R., Barbhaiya, C., et.al. The QT interval in patients with COVID-19 treated with hydroxychloroquine and azithromycin. Nature Med, 1-2 (2020).
24. Elfiky, A. A. Ribavirin, Remdesivir, Sofosbuvir, Galidesivir, and Tenofovir against SARS-CoV-2 RNA dependent RNA polymerase (RdRp): A molecular docking study. Life Sci, 117592 (2020).
25. Ton, A. T., Gentile, F., Hsing, M., Ban, F., & Cherkasov, A. Rapid Identification of Potential Inhibitors of SARS- CoV- 2 Main Protease by Deep Docking of 1.3 Billion Compounds. Mol Inform (2020).
26. Durrant, J. D., & McCammon, J. A. Molecular dynamics simulations and drug discovery. BMC Biol, 9(1), 71(2011).
27. Dong, Y. W., Liao, M. L., Meng, X. L., & Somero, G. N. Structural flexibility and protein adaptation to temperature: Molecular dynamics analysis of malate dehydrogenases of marine molluscs. Proc Natl Acad Sci U S A, 115(6), 1274-1279 (2018).
28. Simmerling, C., Strockbine, B., & Roitberg, A. E. All-atom structure prediction and folding simulations of a stable protein. J Am Chem Soc, 124(38), 11258-11259 (2002).
29. Rose, P. W., Beran, B., Bi, C., Bluhm, W. F., Dimitropoulos, D., Goodsell, D. S., et.al. The RCSB Protein Data Bank: redesigned web site and web services. Nucleic Acids Res, 39(suppl_1), D392-D401 (2010).

30. Luthy, R., Bowie, J. U., & Eisenberg, D. Assessment of protein models with three-dimensional profiles. Nature, 356(6364), 83-85 (1992).
31. Wiederstein, M., & Sippl, M. J. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res, 35(suppl_2), W407-W410 (2007).
32. Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. GROMACS: fast, flexible, and free. J Comput Chem, 26(16), 1701-1718 (2005).
33. Oostenbrink, C., Villa, A., Mark, A. E., & Van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force- field parameter sets 53A5 and 53A6. J Comput Chem, 25(13), 1656-1676 (2004).
34. Hess, B., Bekker, H., Berendsen, H. J., & Fraaije, J. G. LINCS: a linear constraint solver for molecular simulations. J Comput Chem, 18(12), 1463-1472 (1997).
35. Mukherjee, G., & Jayaram, B. A rapid identification of hit molecules for target proteins via physico-chemical descriptors. Phys Chem Chem Phys, 15(23), 9107-9116 (2013).
36. Irwin, J. J., & Shoichet, B. K. ZINC− a free database of commercially available compounds for virtual screening. J Chem Inform Model, 45(1), 177-182 (2005).
37. Trott, O., & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem, 31(2), 455-461 (2010).
38. Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem, 30(16), 2785-2791 (2009).
39. O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., & Hutchison, G. R. Open Babel: An open chemical toolbox. J Cheminform, 3(1), 33 (2011).
40. Lipinski, C. A. Lead-and drug-like compounds: the rule-of-five revolution. Drug Discov Today Technol, 1(4), 337-341 (2004).
41. Huang, B. MetaPocket: a meta approach to improve protein ligand binding site prediction. OMICS, 13(4), 325-330 (2009).
42. Shen, J., Cheng, F., Xu, Y., Li, W., & Tang, Y. Estimation of ADME properties with substructure pattern recognition. J Chem Inform Model, 50(6), 1034-1041 (2010).
43. Schuttelkopf, A. W., & Van Aalten, D. M. PRODRG: a tool for high-throughput crystallography of protein–ligand complexes. Acta Crystallogr D Biol Crystallogr, 60(8), 1355-1363 (2004).

344. Kumari, R., Kumar, R., Open Source Drug Discovery Consortium, & Lynn, A. g_mmpbsa-A GROMACS tool for high-throughput MM-PBSA calculations. J Chem inform Model, 54(7), 1951-1962 (2014).
645. Kumar, R., Maurya, R., & Saran, S. Introducing a simple model system for binding studies of known and novel inhibitors of AMPK: a therapeutic target for prostate cancer. J Biomol Struct Dyn, 37(3), 781-795 (2019).
946. Berendsen, H. J., & Hayward, S. Collective protein dynamics in relation to function. Curr Opin Struct Biol, 10(2), 165-169 (2000).
147. Kabsch, W., & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen- bonded and geometrical features. Biopolymers, 22(12), 2577-2637 (1983).
348. Laskowski, R. A., & Swindells, M. B. LigPlot+: multiple ligand–protein interaction diagrams for drug discovery. J Chem Inf Model, 51(10):2778-86 (2011).

Acknowledgments

Rakesh and Rahul thank Indian Council of Medical Research and University Grant Commission, respectively for their research fellowships.

Competing interests

The authors declare no competing interests.

Author contributions

Rakesh K. and P.T. conceived and designed the study. Rakesh K. and Rahul K. analysed the data and wrote manuscript. P.T. provided the laboratory infrastructure. All authors read and approved the final version of manuscript.

Legend to figures

Fig. 1 Structure and MD simulation of M^{pro}. **(A)** Tertiary structures of CoV2 main protease showing different structural units and domains, **(B)** RMSD in nanometer at y-axis with respect to time in nanosecond at x-axis, **(C)** 3D structure obtained from RMSF analysis showing mobile residues and (D) RMSF in nanometer at y-axis with respect to amino acids at x-axis. Structures were displayed in cartoon mode. Moderate mobile portion was highlighted in circular dotted line in (C) and shade region in RMSF plot indicated amino acid residues.

Fig. 2 Secondary structure formation during MD simulation. **(A)** 2D plot showing patten of different moieties like helices, sheets, tuns, loops, bends and bridge of protein formation during 50ns time and **(B)** 2D plot showing secondary structure variations at individually. Different secondary structure components labelled in different colour as indicated in figure.

Fig. 3 2D plots of protein-ligand complexes. **(A)** Protein-N3 complex, **(B)** Protein-hydroxychloroquine, **(C)** Protein-ZINC14732869, **(D)** Protein-ZINC19774413, **(E)** Protein-ZINC12338080, **(F)** Protein-ZINC123845408, **(G)** Protein-ZINC19774479 and **(H)** Protein-ZINC31. 2D plots were generated in LIGPLOT+ program and hydrophobic and hydrophilic residues denoted in red and black colours, respectively.

Fig. 4 MD simulations of protein-ligand complexes. **(A)** and **(B)** RMSDs in nanometer at y-axis with respect to time in nanosecond at x-axis of proteins in complex form, **(C)** and **(D)** RMSDs in nanometer at y-axis with respect to time in nanosecond at x-axis of ligands in complex form, **(E)** and **(F)** RMSFs in nanometer at y-axis with respect to amino acids at x-axis of proteins in complex form, **(G)** and **(H)** Rg in nanometer at y-axis with respect to time in nanosecond at x-axis of proteins in complex form. N3 (known), hydroxychloroquine, ZINC14732869, ZINC19774413, ZINC12338080, ZINC123845408, ZINC19774479 and ZINC31 were shown in black, red, mustard, green, cyan, yellow, blue and magenta colour, respectively.

Fig. 5 Secondary structure composition of protein in bound/unbound form. **(A)** Protein-N3 complex, **(B)** Protein-hydroxychloroquine, **(C)** Protein-ZINC14732869, **(D)** Protein-ZINC19774413, **(E)** Protein-ZINC12338080, **(F)** Protein-ZINC123845408, **(G)** Protein-ZINC19774479 and **(H)** Protein-ZINC31. M^{pro} (protein unbound) and bound with known, hydroxychloroquine, ZINC14732869, ZINC19774413, ZINC12338080, ZINC123845408, ZINC19774479 and ZINC31 were shown in white, black, red, mustard, green, cyan, yellow, blue and magenta colour, respectively.

Fig. 6 Energy contribution per residues basis. **(A)** Protein-N3 complex, **(B)** Protein-hydroxychloroquine, **(C)** Protein-ZINC14732869, **(D)** Protein-ZINC19774413, **(E)** Protein-ZINC19774479 and **(F)** Protein-ZINC31. Protein-N3, -hydroxychloroquine, -ZINC14732869, -ZINC19774413, -ZINC19774479 and -ZINC31 were shown in black, red, mustard, green, blue and magenta colour, respectively.

Fig. 7 Superimpositions of protein unbound and protein bound with N3, hydroxychloroquine, ZINC14732869, ZINC19774413, and ZINC19774479. M^{pro} (protein unbound) and bound with N3, hydroxychloroquine, ZINC14732869, ZINC19774413 and ZINC19774479 were shown in white, black, red, mustard, green and blue colour, respectively. Major structural differences were highlighted in black dotted circles.

Fig. 8 PCA and FEL studies of protein-N3 and -hydroxychloroquine complexes. **(A)** 2D plot of protein-N3 complexes, **(B)** Superimposed 3D structures of protein with or without complex form, **(C)** Porcupine structure of protein bound with known inhibitors, **(D)** 2D plot of protein-hydroxychloroquine complexes, **(E)** Superimposed 3D structures of protein with or without complex form and **(F)** Porcupine structure of protein bound with hydroxychloroquine. Structures of M^{pro} (protein unbound) and bound form were shown in white and black (N3), red (hydroxychloroquine) colour with cartoon mode, respectively. Arrow and length of cone denoted direction and magnitude of motions in porcupine structures. ΔG were measured in kilocalorie per mol.

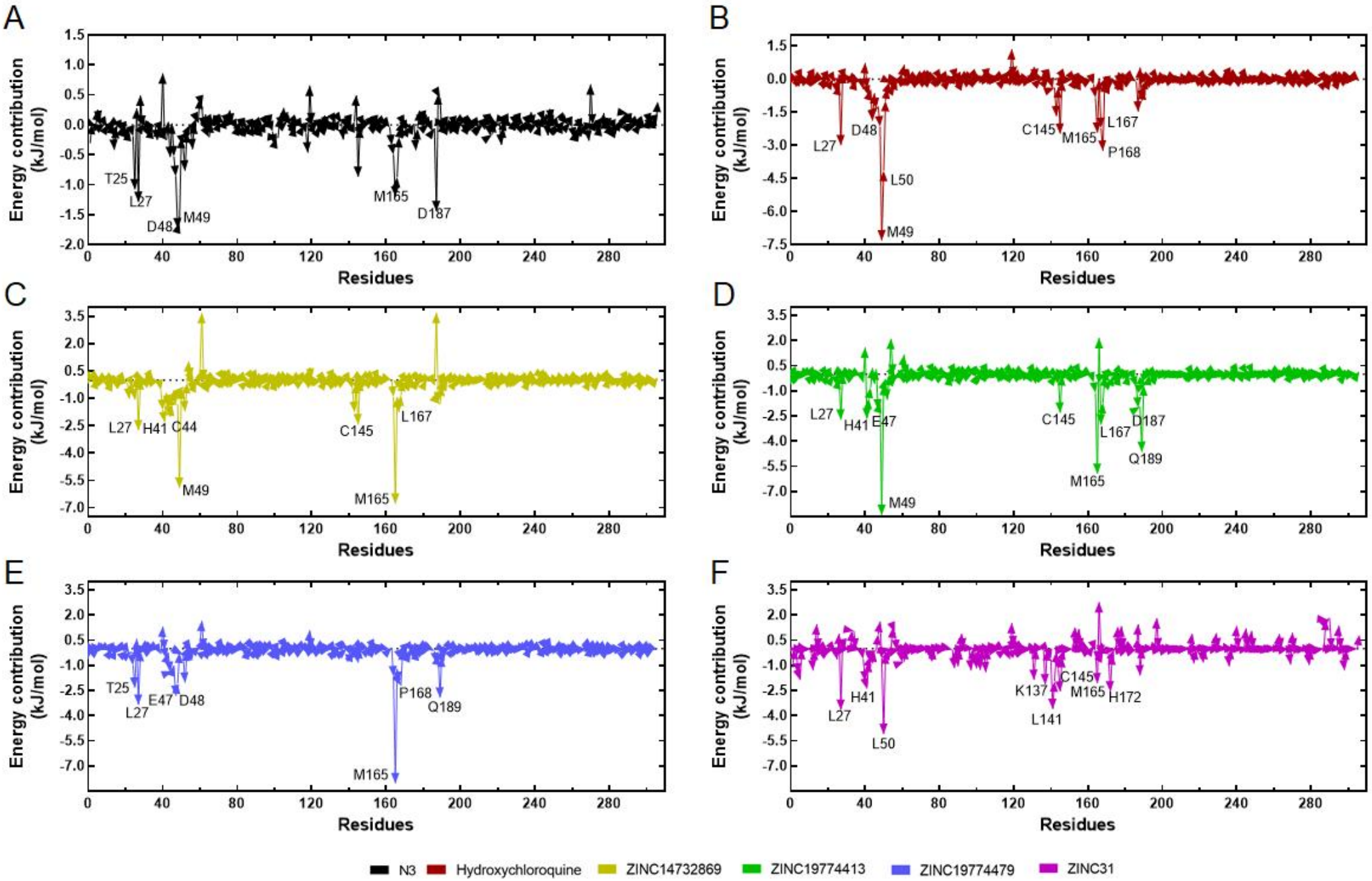
Fig. 9 PCA and FEL studies of protein-ZINC14732869, -ZINC19774413 and -ZINC19774479 complexes. **(A)** 2D plot of protein-ZINC14732869 complexes, **(B)** Superimposed 3D structures of protein with or without complex form, **(C)** Porcupine structure of protein bound

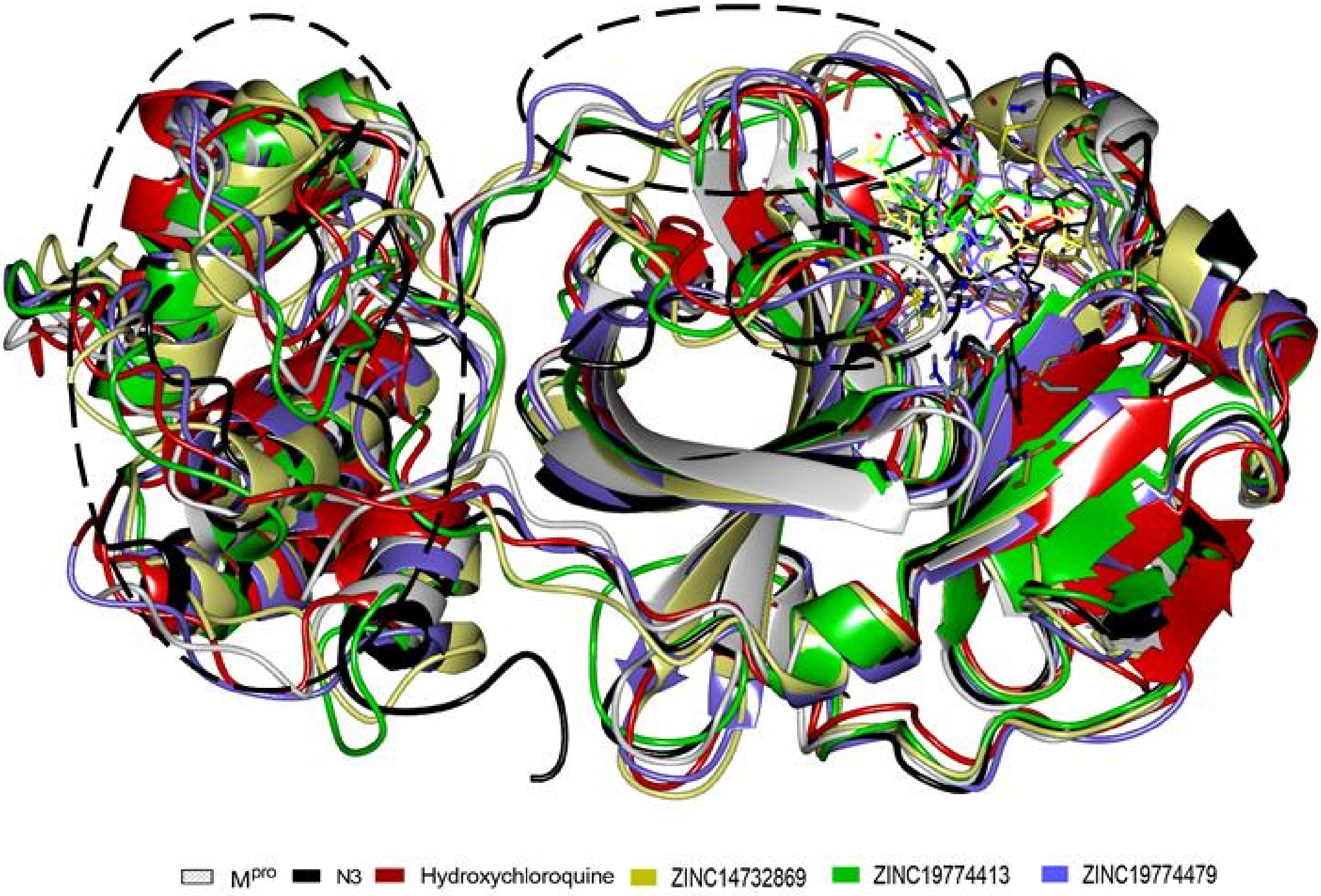
with ZINC14732869 inhibitors, **(D)** 2D plot of protein-ZINC19774413 complexes, **(E)** Superimposed 3D structures of protein with or without complex form, **(F)** Porcupine structure of protein bound with ZINC19774413, **(G)** 2D plot of protein-ZINC19774479 complexes, **(H)** Superimposed 3D structures of protein with or without complex form, **(I)** Porcupine structure of protein bound with ZINC19774479. Structures of M^{pro} (protein unbound) and bound with ligands were shown in white and mustard (ZINC14732869), green (ZINC19774413) and blue (ZINC19774479) colour with cartoon mode, respectively. Arrow and length of cone denoted direction and magnitude of motions in porcupine structures. ΔG were measured in kilocalorie per mol.

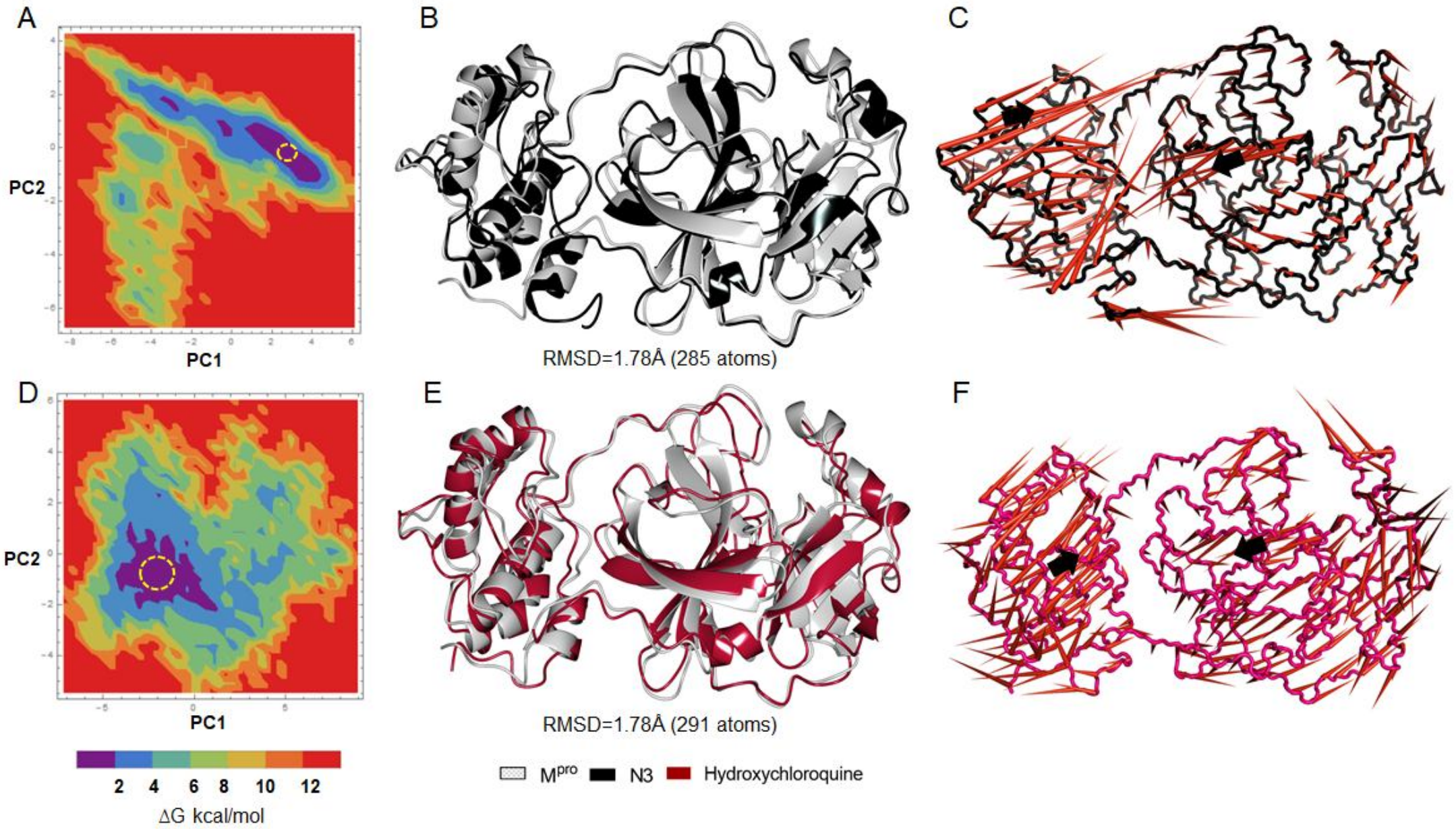
Fig 10 Distance analysis of key residues in binding pocket. **(A)** Distance between Asn142 to Gln189 in nanometer at y-axis with respect to time in nanosecond at x-axis, **(B)** Distance between His164 to Gln189 in nanometer at y-axis with respect to time in nanosecond at x-axis, **(C)** Distance between Pro168 to Gln189 in nanometer at y-axis with respect to time in nanosecond at x-axis, **(D)** Bar diagram showing average values and deviations obtained from graph (A), **(E)** Bar diagram showing average values and deviations obtained from graph (B), **(F)** Bar diagram showing average values and deviations obtained from graph (C) and **(G)** 3D pose of protein showing bond length in nanometer between Asn142-Gln189 (magenta dotted line), His164-Gln189 (black dotted line) and Pro168-Gln189 (green dotted line). Protein 3D structure was displayed in cartoon mode and residues were labelled in stick mode. M^{pro} (protein unbound) and protein bound with N3, hydroxychloroquine, ZINC14732869, ZINC19774413 and ZINC19774479 in 2D graphs were shown in white, black, red, mustard, green and blue colour, respectively.

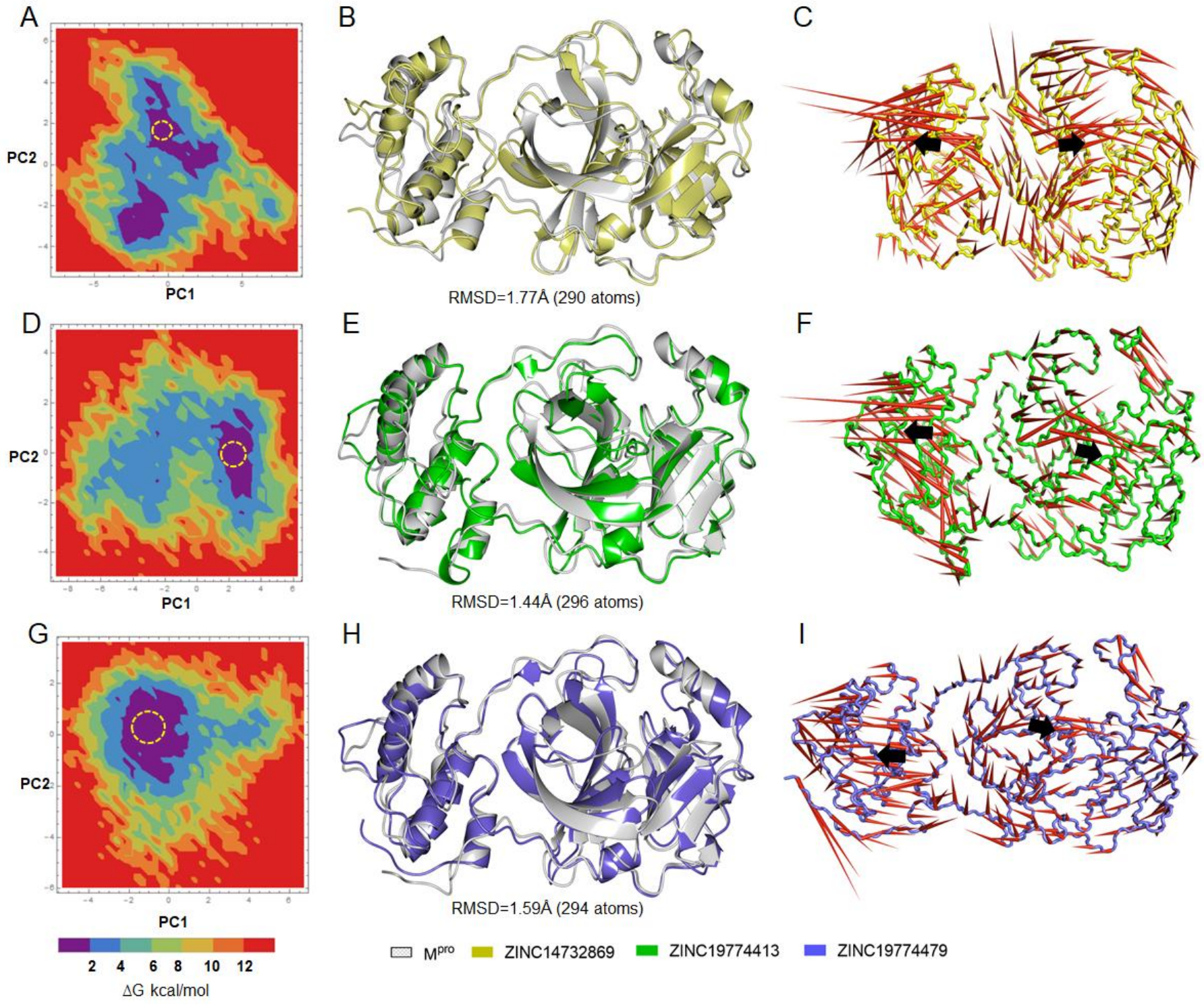
Table 1. Binding free energies calculation by MM/PBSA methods (All energies in kJ/mol) (Bold rows highlighted the ligands with higher binding energies).

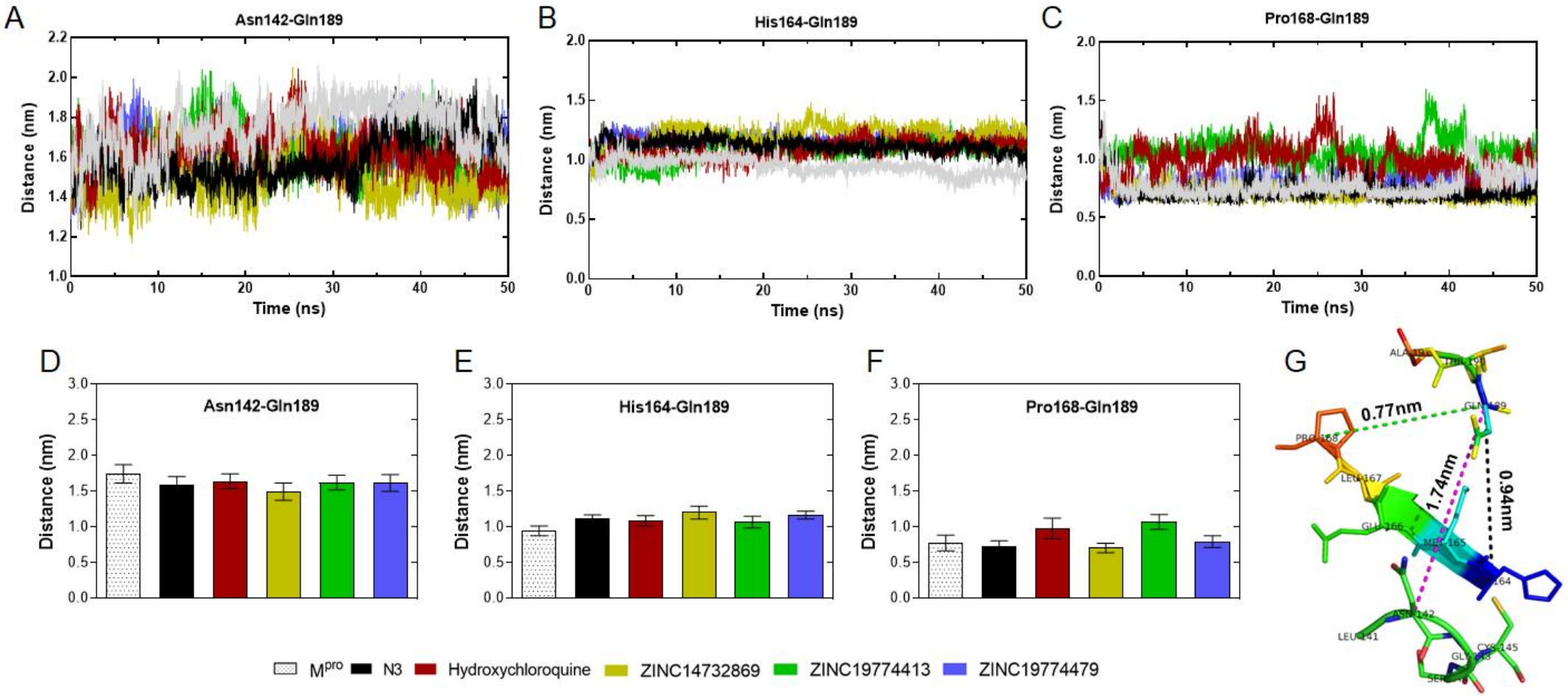
Ligands name	vander Waal energy	Stdev	Electrostatic energy	Stdev	Polar solvation energy	Stdev	SASA energy	Stdev	Binding energy	Stdev
N3	-62.074	±21.728	-8.083	±2.996	37.189	±13.473	-6.372	±2.326	-39.869	±16.169
Hydroxychloroquine	-162.538	±11.498	-2.083	±0.773	54.411	±2.892	-15.303	±1.273	-125.197	±12.244
ZINC14732869	**-206.956**	**±3.646**	**-10.707**	**±1.13**	**85.494**	**±2.176**	**-19.259**	**±0.291**	**-151.532**	**±3.299**
ZINC19774413	**-225.657**	**±11.387**	**-6.869**	**±1.052**	**91.834**	**±4.599**	**-19.744**	**±1.039**	**-160.704**	**±9.221**
ZINC12338080	-101.097	±20.882	11.535	±1.224	49.063	±7.553	-10.094	±1.895	-50.282	±16.805
ZINC12845408	-18.264	±12.656	-1.889	±1.225	7.721	±8.811	-2.011	±1.258	-14.713	±1.431
ZINC19774479	**-184.18**	**±13.735**	**-13.17**	**±1.448**	**78.154**	**±8.947**	**-17.517**	**±1.357**	**-136.809**	**±7.656**
ZINC31	-191.574	±1.447	-15.921	±1.677	118.993	±2.529	-16.884	±0.208	**-105.356**	±2.44

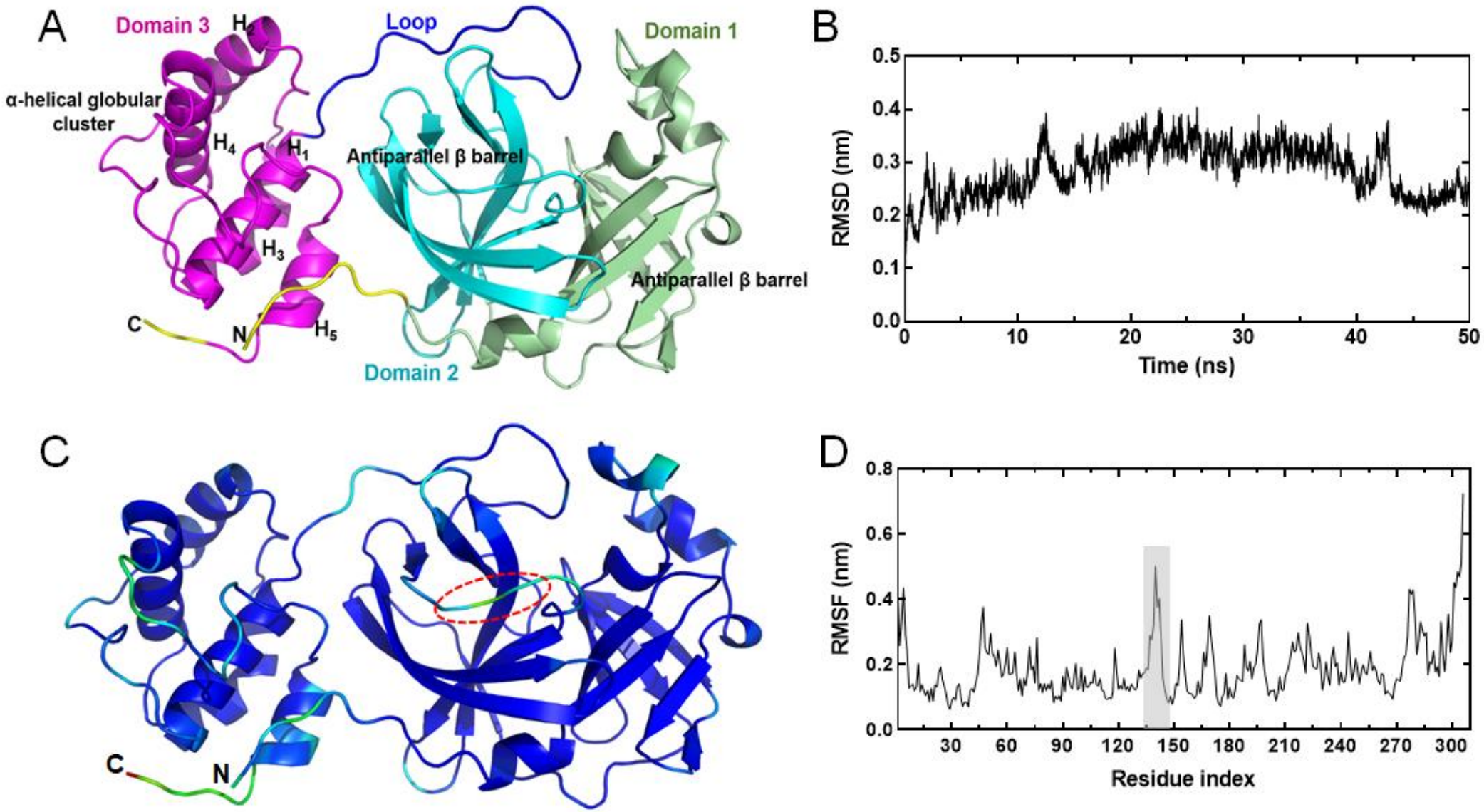

A
B
C
D
E
F
Energy contribution (kJ/mol)
Residues
T25
L27
D48
M49
M165
D187
L27
D48
L50
M49
C145
M165
L167
P168
L27
H41
C44
M49
C145
M165
L167
L27
H41
E47
M49
C145
M165
L167
D187
Q189
T25
L27
E47
D48
M165
P168
Q189
L27
H41
L50
K137
L141
C145
M165
H172
N3
Hydroxychloroquine
ZINC14732869
ZINC19774413
ZINC19774479
ZINC31

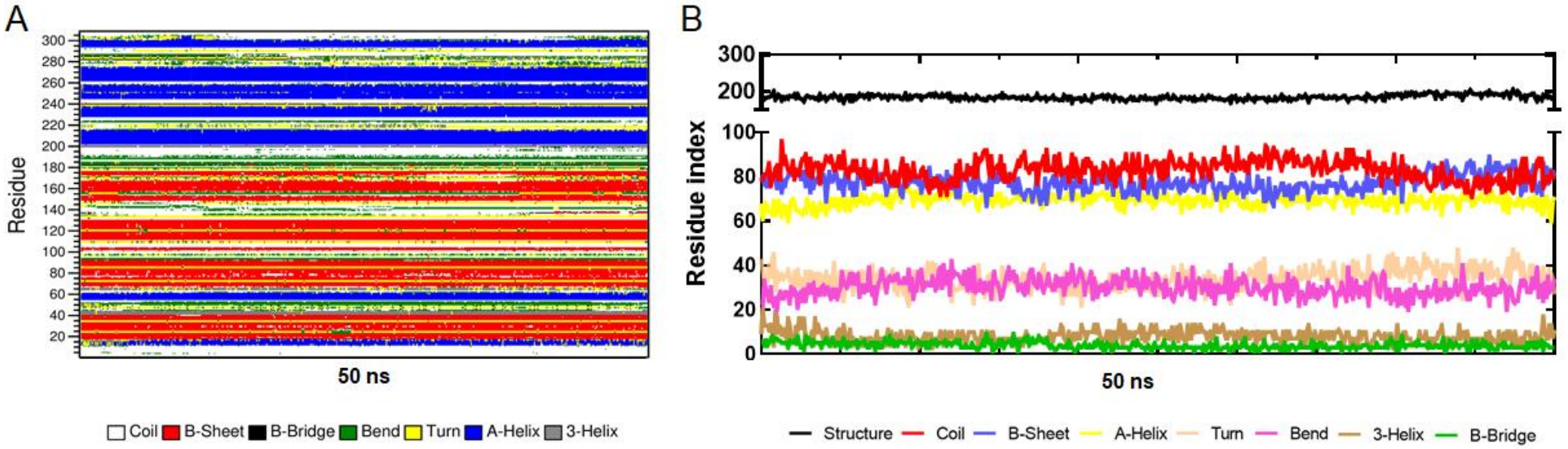

M^pro
N3
Hydroxychloroquine
ZINC14732869
ZINC19774413
ZINC19774479

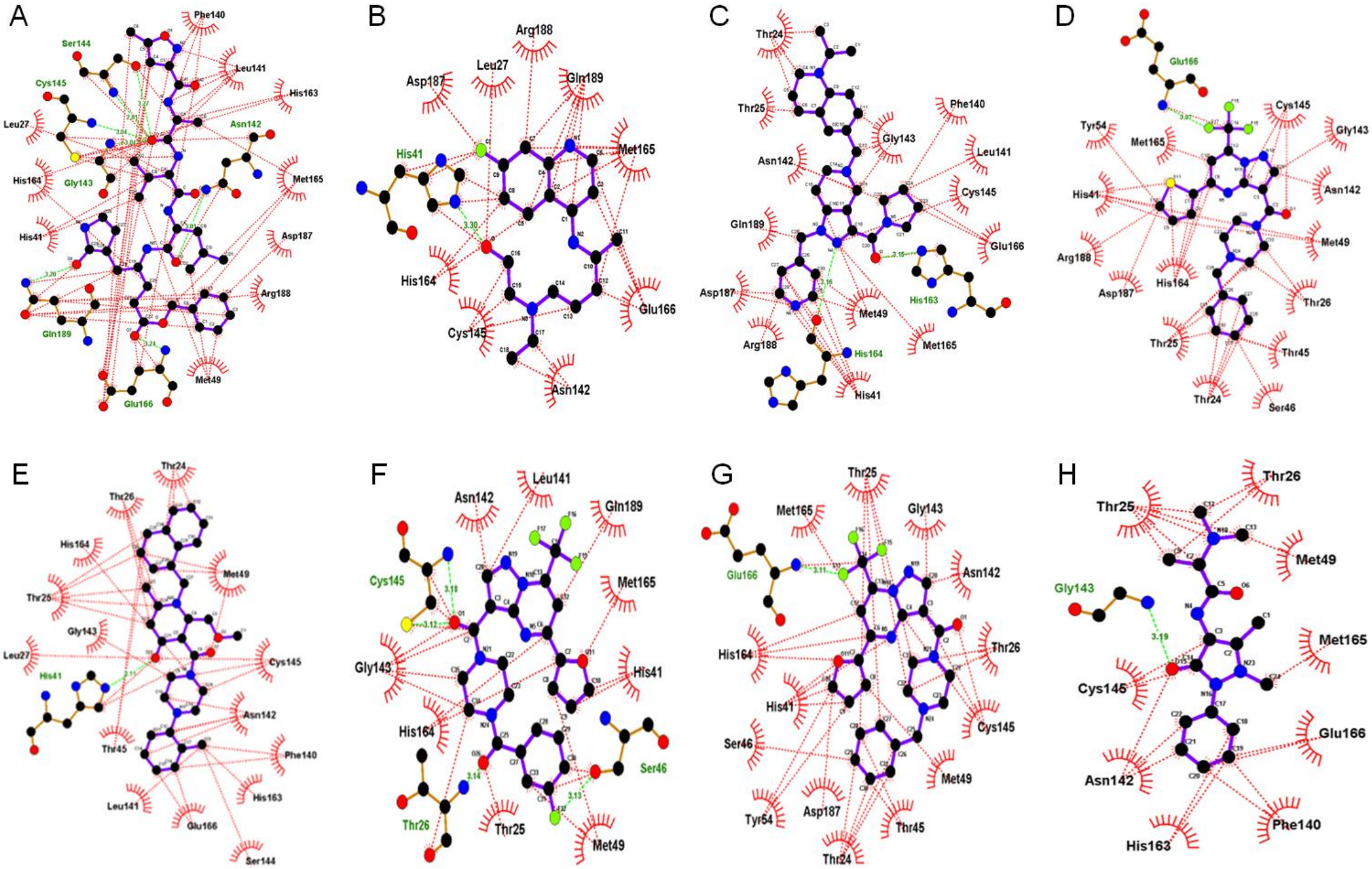

A
PC2
PC1
B
RMSD=1.78Å (285 atoms)
C
D
PC2
PC1
2 4 6 8 10 12
ΔG kcal/mol
E
RMSD=1.78Å (291 atoms)
M^pro N3 Hydroxychloroquine
F

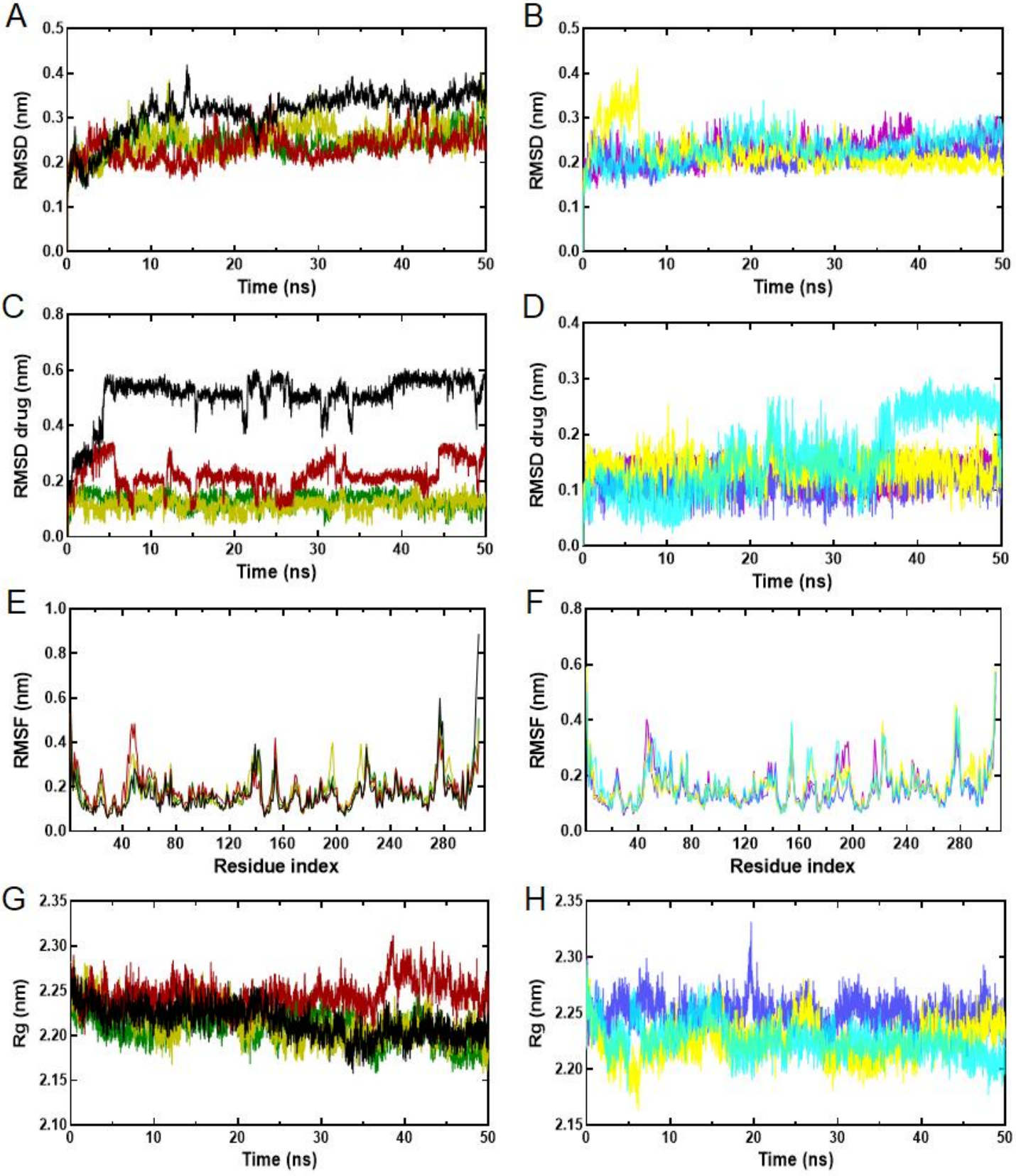

A
B
C
PC2
PC1
RMSD=1.77Å (290 atoms)
D
E
F
PC2
PC1
RMSD=1.44Å (296 atoms)
G
H
I
PC2
PC1
2 4 6 8 10 12
ΔG kcal/mol
RMSD=1.59Å (294 atoms)
M^pro
ZINC14732869
ZINC19774413
ZINC19774479

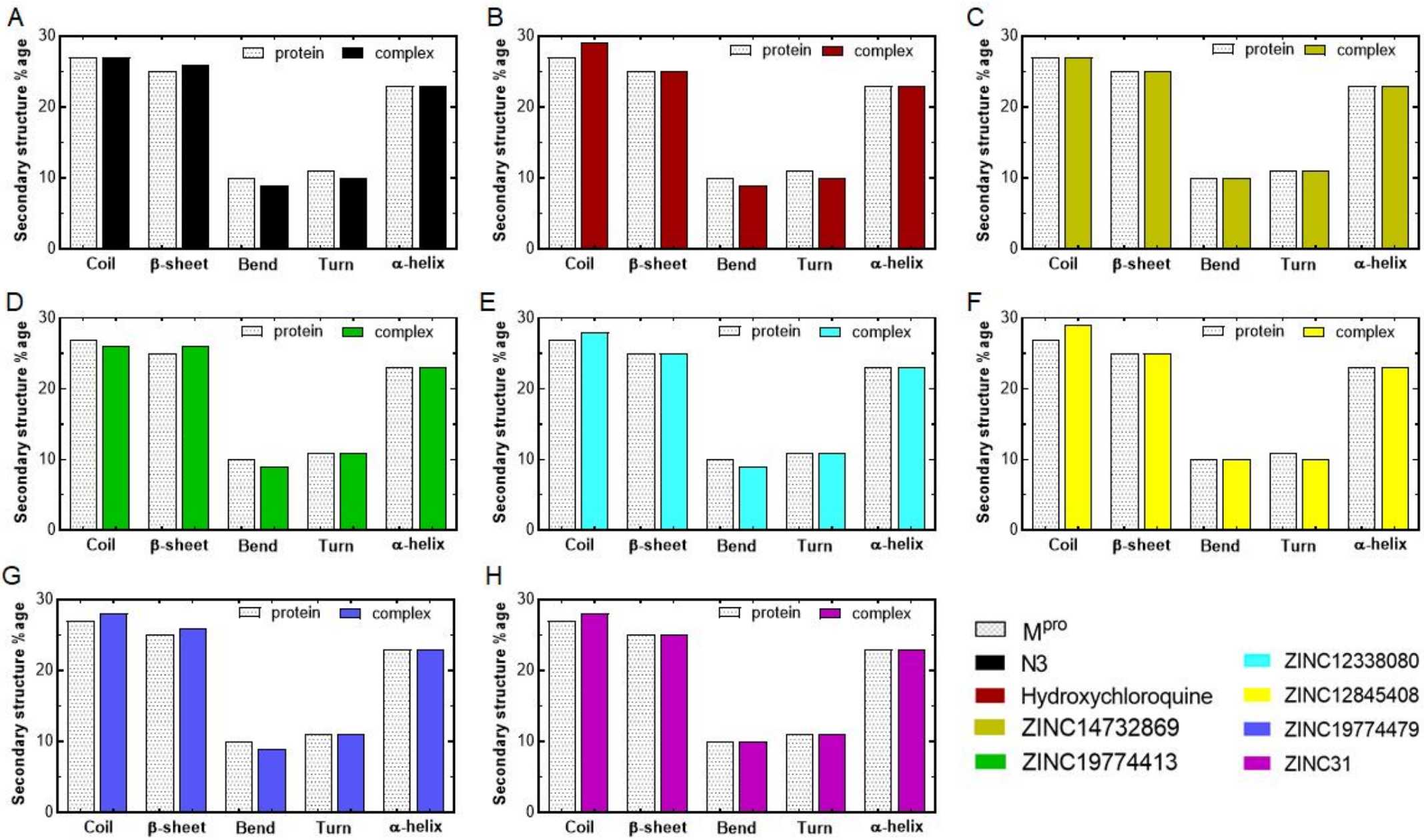

A
Asn142-Gln189
Distance (nm)
Time (ns)
B
His164-Gln189
Distance (nm)
Time (ns)
C
Pro168-Gln189
Distance (nm)
Time (ns)
D
Asn142-Gln189
Distance (nm)
E
His164-Gln189
Distance (nm)
F
Pro168-Gln189
Distance (nm)
G
0.77nm
1.74nm
0.94nm
M^pro
N3
Hydroxychloroquine
ZINC14732869
ZINC19774413
ZINC19774479


A
Domain 3
H6
Loop
Domain 1
α-helical globular cluster
H4
H1
Antiparallel β barrel
H3
Antiparallel β barrel
C
N
H5
Domain 2
B
RMSD (nm)
0.0
0.1
0.2
0.3
0.4
0.5
0
10
20
30
40
50
Time (ns)
C
C
N
D
RMSF (nm)
0.0
0.2
0.4
0.6
0.8
30
60
90
120
150
180
210
240
270
300
Residue index


A
Residue
300
280
260
240
220
200
180
160
140
120
100
80
60
40
20
50 ns
Coil
B-Sheet
B-Bridge
Bend
Turn
A-Helix
3-Helix
B
Residue index
300
200
100
80
60
40
20
0
50 ns
Structure
Coil
B-Sheet
A-Helix
Turn
Bend
3-Helix
B-Bridge


A
Phe140
Ser144
Leu141
Cys145
His163
Leu27
Asn142
His164
Gly143
Met165
His41
Asp187
Arg188
Gln189
Met49
Glu166
B
Arg188
Leu27
Asp187
Gln189
His41
Met165
His164
Glu166
Cys145
Asn142
C
Thr24
Thr25
Phe140
Gly143
Leu141
Asn142
Cys145
Gln189
Glu166
His163
Asp187
Met49
Arg188
Met165
His164
His41
D
Glu166
Cys145
Tyr54
Gly143
Met165
Asn142
His41
Met49
Arg188
His164
Asp187
Thr26
Thr25
Thr45
Thr24
Ser46
E
Thr24
Thr26
His164
Met49
Thr25
Gly143
Leu27
Cys145
His41
Asn142
Thr45
Phe140
His163
Leu141
Glu166
Ser144
F
Asn142
Leu141
Gln189
Cys145
Met165
Gly143
His41
His164
Ser46
Thr26
Thr25
Met49
G
Thr25
Met165
Gly143
Glu166
Asn142
His164
Thr26
His41
Cys145
Ser46
Met49
Tyr54
Asp187
Thr45
Thr24
H
Thr26
Thr25
Met49
Gly143
Met165
Cys145
Glu166
Asn142
Phe140
His163


A
RMSD (nm)
Time (ns)
B
RMSD (nm)
Time (ns)
C
RMSD drug (nm)
Time (ns)
D
RMSD drug (nm)
Time (ns)
E
RMSF (nm)
Residue index
F
RMSF (nm)
Residue index
G
Rg (nm)
Time (ns)
H
Rg (nm)
Time (ns)
N3
Hydroxychloroquine
ZINC14732869
ZINC19774413
ZINC12338080
ZINC12845408
ZINC19774479
ZINC31


A
B
C
D
E
F
G
H
Secondary structure % age
protein
complex
Coil
β-sheet
Bend
Turn
α-helix
0
10
20
30
M^pro
N3
Hydroxychloroquine
ZINC14732869
ZINC19774413
ZINC12338080
ZINC12845408
ZINC19774479
ZINC31

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementaryinformation.pdf