High-Throughput Screening and Quantum Mechanics for Identifying Potent Inhibitors Against Mac1 Domain of SARS-CoV-2 Nsp3

Chandrabose Selvaraj, Dhurvas Chandrasekaran Dinesh, Umesh Panwar, Evzen Boura, and Sanjeev Kumar Singh

Abstract—SARS-CoV-2 encodes the Mac1 domain within the large nonstructural protein 3 (Nsp3), which has an ADP-ribosylhydrolase activity conserved in other coronaviruses. The enzymatic activity of Mac1 makes it an essential virulence factor for the pathogenicity of coronavirus (CoV). They have a regulatory role in counteracting host-mediated antiviral ADP-ribosylation, which is unique part of host response towards viral infections. Mac1 shows highly conserved residues in the binding pocket for the mono and poly ADP-ribose. Therefore, SARS-CoV-2 Mac1 enzyme is considered as an ideal drug target and inhibitors developed against them can possess a broad antiviral activity against CoV. ADP-ribose-1 phosphate bound closed form of Mac1 domain is considered for screening with large database of ZINC. XP docking and QPLD provides strong potential lead compounds, that perfectly fits inside the binding pocket. Quantum mechanical studies expose that, substrate and leads have similar electron donor ability in the head regions, that allocates tight binding inside the substrate-binding pocket. Molecular dynamics study confirms the substrate and new lead molecules presence of electron donor and acceptor makes the interactions tight inside the binding pocket. Overall binding phenomenon shows both substrate and lead molecules are well-adopt to bind with similar binding mode inside the closed form of Mac1.

Index Terms—COVID-19, Mac1, Macro X, Molecular Dynamics, Nsp3, Quantum Mechanics, SARS-CoV-2

1 Introduction

The Coronavirus Disease 2019 (COVID-19) pandemic outbreak, an infectious disease caused by a novel Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), is highly catastrophic and has a high impact on the global economy [1]. This coronavirus swiftly spread through communities and reached all continents except Antarctica in an incredibly limited time [2]. The current confirmed case numbers are more than five million with a mortality rate of 2.6 percent (according to https://ourworldindata.org/mortality-risk-covid). SARS-CoV-2 is an enveloped virus from family *Coronaviridae* and genus beta-coronavirus, comprising a large positive-strand single-strand RNA (+ssRNA) genome (~30 kb), which encodes four structural proteins (spike, envelope, membrane, and nucleocapsid protein) that are components of the virus particle, several nonstructural (Nsp) proteins mostly with enzymatic activities and accessory proteins [3]. The ORF1a and b encoded two large polyproteins which were later cleaved by viral proteases into 16 functional proteins (Nsp1-16), where Nsp11 alone is a short peptide. Nsp3 is the largest multidomain protein (~200kDa) in coronaviruses and is notable because of the presence of a key enzyme, papain-like cysteine protease (PL^{pro}), which is essential for viral replication and a target protein for drug discovery [4], [5]. The Nsp3 is found to be significantly different in two SARS-CoVs in comparison with other Nsps [6], [7]. Computational studies have shown that coronaviral Nsp3 may comprise 10-16 domains [8]. SARS-CoV-2 Nsp3 also includes multiple domains, an N-terminal ubiquitin-like (Ubl) domain followed by a highly variable and conserved macrodomain that binds to ADP-ribose [9], hereafter called as Mac1, previously called as Macro X domain because of its unknown structure and function, and multiple downstream SARS unique macrodomains, namely Mac2 and Mac3 (previously called as SUD-N and SUD-M) that binds to G-quadruplex, and a domain preceding Ubl2 and PL^{pro} (DPUP) domain (previously called as SUD-C) that binds to ssRNA [10], [11]. The macrodomains are followed by the Ubl2 domain, PL^{pro}, nucleic acid binding domain (NAB) and several other domains in the C-terminus which also includes transmembrane regions [12], [13], [14]. The macrodomains (~130-190 AAs) can be either stand-alone or a part of a multi-domain protein [15]. These domains are also widely present in viruses, bacteria, archaea, and several eukaryotes. They recognize ADP-ribose or its derivative or post-translational modification covalently bound to proteins, thereby playing a varied role in cellular processes (e.g., DNA repair, signal transduction, immune response) in all kingdoms of life [16]. The yeast macrodomain (Poa1) is found to bind ADP-ribose-1"-phosphate, which is a by-product of tRNA splicing [17], [18], [19]. In viruses, they play a role in virulence and suppressing the host innate immune response and

- *Chandrabose Selvaraj, Umesh Panwar, and Sanjeev Kumar are with the Department of Bioinformatics, Alagappa University, Karaikudi, Tamil Nadu 630003, India. E-mail: selnikraj@bioclues.org, {umesh.panwar07, skysanjeev}@gmail.com.*
- *Dhurvas Chandrasekaran Dinesh and Evzen Boura are with the Institute of Organic Chemistry and Biochemistry AS CR, 160 00 Prague, Czechia. E-mail: {dinesh.dhurvas, boura}@uochb.cas.cz.*

Manuscript received 1 June 2020; revised 6 Sept. 2020; accepted 26 Oct. 2020. Date of publication 11 Dec. 2020; date of current version 6 Aug. 2021. (Corresponding authors: Chandrabose Selvaraj and Sanjeev Kumar Singh.) Digital Object Identifier no. 10.1109/TCBB.2020.3037136

have shown potential roles in viral replication and pathogenesis [20], [21]. The macrodomains are found in three viral families, such as *Coronaviridae*, *Togoviridae*, and *Hepeviridae*, which are evolved to counteract the host defense mechanism [11], [22]. The coronaviral Mac1 domain specifically binds to ADP-ribose-1"-phosphate and has dephosphorylation or ADP-ribosylhydrolase activity, demonstrating their possible function in over-turning host antiviral ADP-ribosylation [23], [24]. Recent studies have shown that the SARS-CoV-2 Nsp3 Mac1 binds to ADP-ribose with a higher binding affinity compared to SARS-CoV *in vitro* and catalyzes the hydrolysis of ADP-ribose-1"-phosphate. Even though the Mac1 domain is found to be dispensable for viral RNA replication, several studies are available showing its regulatory role in counteracting host innate immunity [9], [11], [25]. Therefore, because of the emerging novel roles of macrodomains, they are considered as a potential therapeutic target for drug discovery in the viral world.

2 Materials and Methods

2.1 Sequence and Structure: Comparative Analysis

Pairwise protein structural comparison was performed between the Mac1 domain of SARS-CoV-2 (PDB ID:6WEY) and SARS-CoV (PDB ID:2FAV) using TM-align server [26], [27]. The structure-based sequence alignment of SARS-CoVs Mac1 domain sequences was prepared using ENDscript server 2.0 [28]. Secondary structural elements are displayed above the sequence and the residues are highlighted based on the residue conservation (red for identical and yellow for similar), with the lower bar indicating the solvent accessibility and hydropathy scale. All structure and interaction figures were rendered using PyMOL v2.3.5 [29].

2.2 Protein Preparation

The multi-domain Nsp3 is an essential component of the replication/transcription complex and from this the Mac1 domain structure is retrieved from the protein data bank (PDB ID: 6WEY). From the available Mac1 domains of SARS-CoV-2, the 6WEY has the ultra-high resolution of 0.95 Å and thus the apo structure is chosen for the study (Fig. 1a) along with the ADP-ribose from the PDB ID: 6YWL. The substrate co-ordinate, namely adenosine-5-diphosphoribose is taken from the X-ray structure of SARS-CoV-2 Mac1 domain in complex with ADP-ribose (Fig. 1c, PDB ID: 6YWL) and merged with high resolution structure using the Maestro (Schrödinger). After the complex formation, the correctness of the structure is obtained using the protein preparation wizard [30]. Here, the partial charges of the atoms that are filled along with bond orders, missing atoms and side chains are refined. Sidechain amino acids

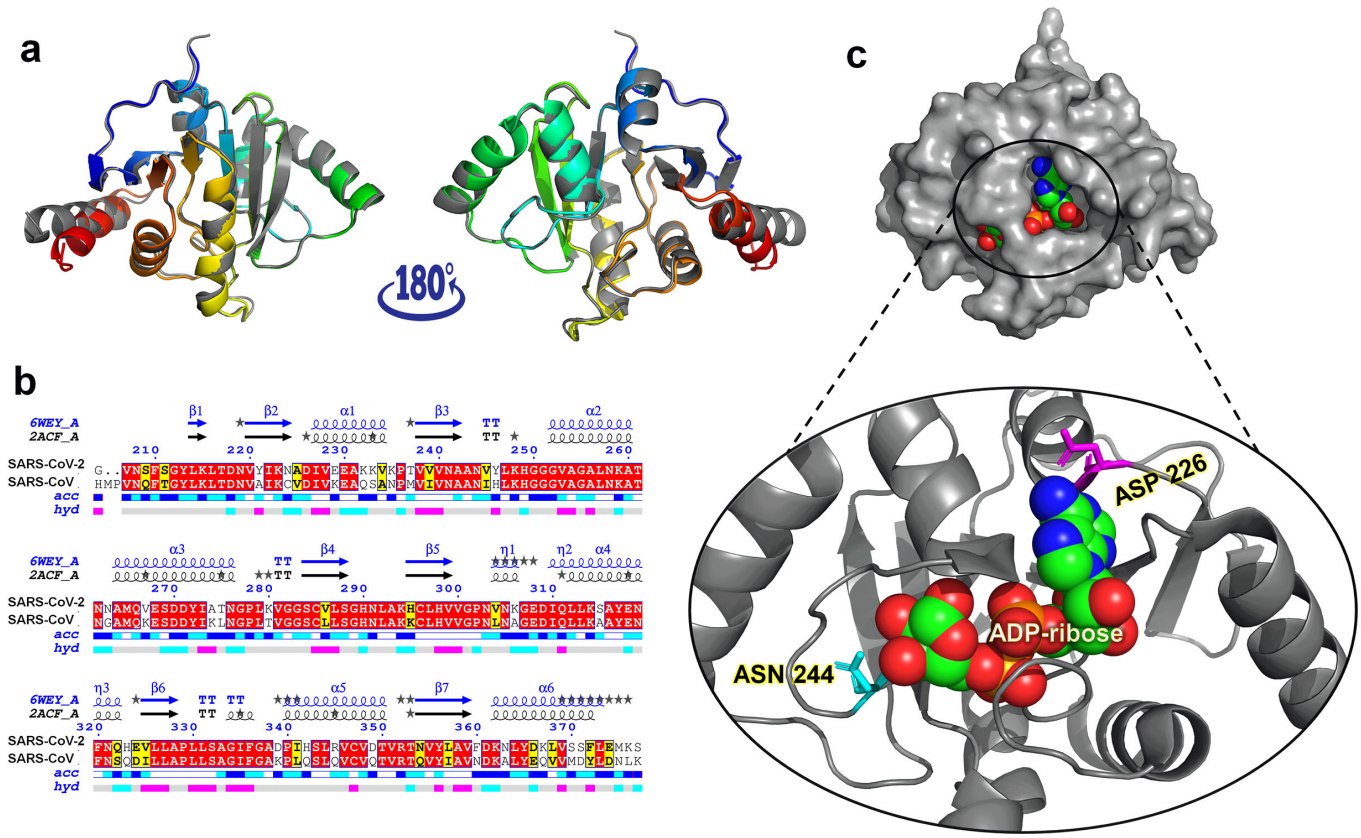

Fig. 1. Structure and sequence alignment of Mac1 domain from SARS-CoVs (a) Structural alignment of the Mac1 domain from experimental structures of SARS-CoV-2 (colored as the rainbow) with SARS-CoV (colored as gray) (b) Structure-based sequence alignment of Mac1 domain from SARS-CoVs highlighting identical residues in red and similar residues in yellow (c) Surface representation of closed ADP-ribose bound Mac1 domain structure from SARS-CoV-2 (PDB ID: 6YWL) and zoomed-in view of the binding site displaying crucial residues for ADP binding and catalytic activity.

Asn, Gln, and His angles are flipped to influence the H-bond formation and generate tautomer/ionized states [31]. The protein-substrate structure is applied with OPLS-3e force field (FF) for optimizing the intramolecular hydrogen bonds using the optimization method and minimized till the root-mean-square deviation (RMSD) levels reaches the 0.30 Å [32].

2.3 ZINC Database Preparation for Screening

The small molecule library of 230 million compounds is downloaded from ZINC database (https://zinc.docking.org/) [33]. The whole library is prepared by maintaining the pH values around 7.0 $\pm$ 2.0 for maintaining the ligand ionization state and ligands stereoisomer conformations are checked by LigPrep (Schrödinger) [34]. By applying the OPLS3e FF with all combinations applied, up to three conformations per ligand are generated.

2.4 Glide - Grid Generation

The prepared complex of Mac1-ADP-ribose complex is taken for grid generation using the Schrödinger's Glide–Grid generation method [35]. The ADP-ribose molecule is manually picked, and the grid box is customized around it. For softening the receptor non-polar regions, the vdW scaling factor is set to 1.0 kcal/mol and partial charge for atoms cut-off is set to default 0.25. The region around 2 Å from the ADP-ribose is set on focus for the molecular docking approach and the amino acids playing interactions with ADP-ribose are fixed as H-bond constrains [36]. Finally, the positional grid is generated around the ADP-ribose, where the new ligands will be replaced and from the best compounds will be reported using the molecular docking method.

2.5 Virtual Screening

The prepared ligand from the LigPrep and grid file from the receptor grid generation is input for the Schrödinger's Virtual Screening Workflow (VSW) [37]. Initial filtering is done by applying the Qikprop filter to pass the compounds having strong ADME profiles. Lipinski filter is avoided, as the substrate binding pocket is moderately bigger in size and able to hold large size molecules. Reactive functional group molecules are filtered out and from these filters, unnecessary compounds are filtered out and subjected to molecular docking approach. Initial level screening is performed using High-Throughput Virtual Screening (HTVS), to check for the apt compound that suits with ADP-ribose binding pocket. The top 10 percent of the successive compounds from HTVS is subject to pass to the standard precision (SP) docking method, and from that top 5 percent compounds will be examined for its interactions with Extra Precision (XP) docking method [38]. Pose validations will be done for the final compounds which are the best compounds from the glide docking score and Prime MM/GBSA ranking [39].

2.6 Quantum Polarized Ligand Docking

For obtaining the accurate interactions of the successive lead compounds, the redocking methodology using the Schrödinger's Quantum Polarized Ligand Docking (QPLD) is employed [40]. In this, the QSite and Jaguar enabled with Quantum Mechanics (QM) will be applied for the ligand and binding site residues. For the other regions of the proteins, the Molecular Mechanics (MM) will be applied using the Impact. QM region is assigned through DFT with the B3LYP using 6 -31G** basis set, and "Ultrafine" SCF accuracy level (iacc = 1, iacscf = 2) [41]. Through this, the pseudo charges of each protein-ligand atom will be refined and get accurate charges. QPLD methodology initially intakes normal Glide docking with XP; those poses will be subjected to charge refinement using the QM and MM methods and refined charges are employed for redocking method using the XP/QPLD [42], [43].

2.7 Density-Functional Theory (DFT) Calculations

The electronic charge transfer effects of drug like molecules that play insights in pharmacological effects are calculated using the DFT analysis [44]. For DFT analysis, the ADP-ribose substrate and the top five compounds are imported with jaguar workflow panel. Functional set is assigned to B3LYP with 6-31G**++ basis set. For calculating the multipole moments, the Coupled Perturbed Kohn-Sham (CPKS) equations are applied, and for elements solution, the Poisson-Boltzmann Finite (PBF) is applied. Among the QM properties, calculating the molecular orbitals (HOMO/LUMO/Band gap), electron density and molecular electrostatic potential (MESP) are marked for output [45], [46].

2.8 Molecular Dynamics Simulations

Molecular Dynamics (MD) simulations are performed for the ADP-ribose substrate bound with the Mac1 domain, and also for the top five lead compounds bound with the Mac1 domain using the GROMACS 5.1.4 (GROningen MAchine for Chemical Simulations, http://gromacs.org) [47], [48]. All the six protein-ligand complexes are simulated for the timescale of 50 ns for understanding the stability of ligand inside the protein binding pocket. For that, the complex files are solvated with TIP3P water model, periodic boundary set at 1.0 nm and systems prepared using OPLS-AA force field. Ligand charge adoption for the system is performed using the external PRODRG server (http://prodrg1.dyndns.org/) and the whole system is neutralized by adding suitable (Na+/Cl-) ions in required concentrations [49], [50]. Energy minimization step is performed for 1000 steps by applying steepest descent algorithm with tolerance of 10 kJ/mol/nm to avoid the steric clashes. Reference pressure is set to 1.0 bar using Parrinello-Rahman along with periodic boundary conditions with cut-offs for Lennard-Jones and reference temperature is set to 300 K using Berendsen thermostat. Successful minimized systems are subjected to NVT and NPT ensembles for initiating the MD simulation step for 50 ns of timescale [51].

3 Results and Discussion

3.1 Sequence and Structural Analysis of Mac1

The sequence comparison is done between the SARS-CoV-2 and SARS-CoV Mac1 domain. The Mac1 domain of SARS-CoV-2 has the sequence length of 172 amino acids, and the SARS-CoV has the sequence length of 182 amino acids. Both the SARS-CoV-2 and SARS-CoV share the sequence similarity of 81.7 percent (143/172) and identity of 70.9 percent (124/172) as shown in Fig. 1b. Even though the difference

between these sequences hold ~18 percent, the structural alignment between these two proteins shows similar structural and functional phenomenon with the RMSD of 1.2Å. The three-dimensional structure of Mac1 comprises a central six-stranded mixed beta sheet surrounded by helices that adopt a highly conserved globular three-layered $\alpha/\beta/\alpha$ sandwich fold with minor variations. The overall fold is quite similar to the human non-histone domain of macroH2A-like fold, classifying them as a member of MacroD-like family [16]. ADP-ribose binding cleft seems to undergo conformational changes upon binding or transition between apo and complex state. In SARS-CoV, key residues in loop 1 and 2 found between **β**2-**α**2 and **β**5-**α**5 respectively, contribute to ADP-ribose binding. Mutational studies on Asp23, Asn38, Asn41, His46, Gly131, Gly47+Gly48 and Phe133 from SARS-CoV have shown that they disrupt their catalytic function or virulence *in vivo* [11], [18], [52]. Protein topology network shows that SARS-CoV-2 Mac1 is arranged by 2 sheets, 1 beta alpha beta unit, 2 beta hairpins, 1 psi loop, 2 beta bulges, 7 strands, 9 helices, 6 helix-helix interacs, 13 beta turns and 2 gamma turns, while the SARS-CoV is arranged as 2 sheets, 1 beta alpha beta unit, 2 beta hairpins, 1 psi loop, 2 beta bulges, 7 strands, 9 helices, 7 helix-helix interacs, 11 beta turns and 2 gamma turns.

3.2 Molecular Interactions of ADP-Ribose With Mac1

The structure alignment provided in Fig. 1a shows the both proteins from the SARS-CoV-2 and SARS-CoV are structurally identical. Mac1 domain shows two distinct features in apo and holo forms by an open and closed binding pocket mechanism, respectively. Before the ADP-ribose binding, the protein substrate binding site is showing the open pocket, and while the ADP-ribose bound within the pocket, the loops come closer to each other and form the closed binding pocket. In SARS-CoV-2, the ADP-ribose bound to the substrate binding site (closed form) as shown in Fig. 1c, and the amino acids Asp226, Ile227, Asn244, Val253, Ser332 and Phe336 plays the role of holding the ADP-ribose through hydrogen bonding interactions. Phe360 also contributes to the binding of ADP-ribose through Π-Π interactions as shown in Figure S1, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TCBB.2020.3037136. In this the Asp226 and Asn244 plays the vital role in the catalytic reaction mechanism, and these two residues are highlighted in Fig. 1c.

3.3 Virtual Screening of Mac1 Domain

The bound form of ADP-ribose with Mac1 domain shows that the substrate binding pocket is holding the large regions of positive electron density with the core amino acid Asp226, Ile227, Asn244, Val253, Ser332, Phe336 and Phe360. Among these, Asp226 playing the role in holding ADP is negatively charged, where the other key residue Asn244 is a polar and non-charged amino acid (Fig. 1). Apart from these two amino acids, Ile227 and Val253 are non-polar categorized as aliphatic amino acid due to the availability of aliphatic side chain functional group. The other amino acid Ser332 is polar with non-charged amino acid, while Phe336 and Phe360 are the only aromatic amino acid in this pocket. Integrating the ligand and structure-based approach can determine a strong and potent compound that suitably bound towards these large regions of positive electron density pocket. Pointing the ADP-ribose focused ligands that retrieve few compounds through HTVS, SP and XP docking. Those compounds with docking score of $<$ -10 kcal/mol and the binding energy of $<$ -50 kcal/mol provides few compounds, that exactly matched the substrate binding pocket. For enhancing the accuracy of the docked complex, the redocking methodology using the Quantum Polarized Ligand Docking (QPLD) is utilized. This is to attain the select of correct and accurate binding poses by implementation of QM/MM docking. QPLD derived partial charges are playing the vital role in terms of enhancing the accuracy and provides the possible binding pose that is close to the experimental structures. From this screening, the compounds namely ZINC08765069, ZINC08792474, ZINC08879336, ZINC08879971 and ZINC00897592 arise as strong lead candidates, that actively bind with ADP-ribose binding pocket of Mac1 domain (Fig. 2), and all these compounds have the docking score of $<$ -10 kcal/mol and binding energy of $<$ -50 kcal/mol as shown in Table 1.

Here, in this table, the scoring values are shown to hold the difference between the XP and QPLD docking. This is due to changes in atomic charges in the protein-ligand complex of XP docking is considered as pseudo charges, and the hybrid QM/MM incorporated in QPLD to replace the accurate charges that enhance the scoring and interaction pattern of the QPLD poses. These accurate charges derived protein ligand interactions are necessary, as the charge of each atom is replaced and thus, the interaction phenomenon of positive charge tend to negative charged atoms, while that of negative charge tend towards the positively charged atoms. The role of partial charges between the positive and negative charges tends to alter the hydrogen bonding in XP and QPLD and provides the slight variations in the binding pose. The RMSD variations of XP docked ligand and QPLD docked ligand are provided in Table 1, which shows the variations that occurred due to the replacement of atomic charges. From this, the role of QPLD in this screening has obtained higher benefit by the charge modulations, and explains that the XP docking provides pose conformations in binding pocket, and adding up the QPLD redocking method provides the change in angle conformations that bound more accurate inside the binding pocket.

3.4 HOMO/LUMO and MESP Calculations

The DFT calculations are performed to understand the key level atomic properties of the molecular structure and here, the QM techniques are applied to understand the charge transfer ability of atoms for the substrate and new lead compounds through HOMO/LUMO calculations. The exact conformations obtained from the QPLD docking pose are subject to optimize with B3LYP/3 - 6 -31G**++ level of theory, and by this the angles and bond lengths of the substrate and ligand optimized geometries are accommodated in realistic position. For understanding the solvation effect, geometry optimization, the base sets are performed with the 6 -31G**++ and PBF solvation. In the substrate and lead molecules, the electron donors are accessed by HOMO, the electron acceptors are accessed by LUMO, and the gap between the HOMO and LUMO is

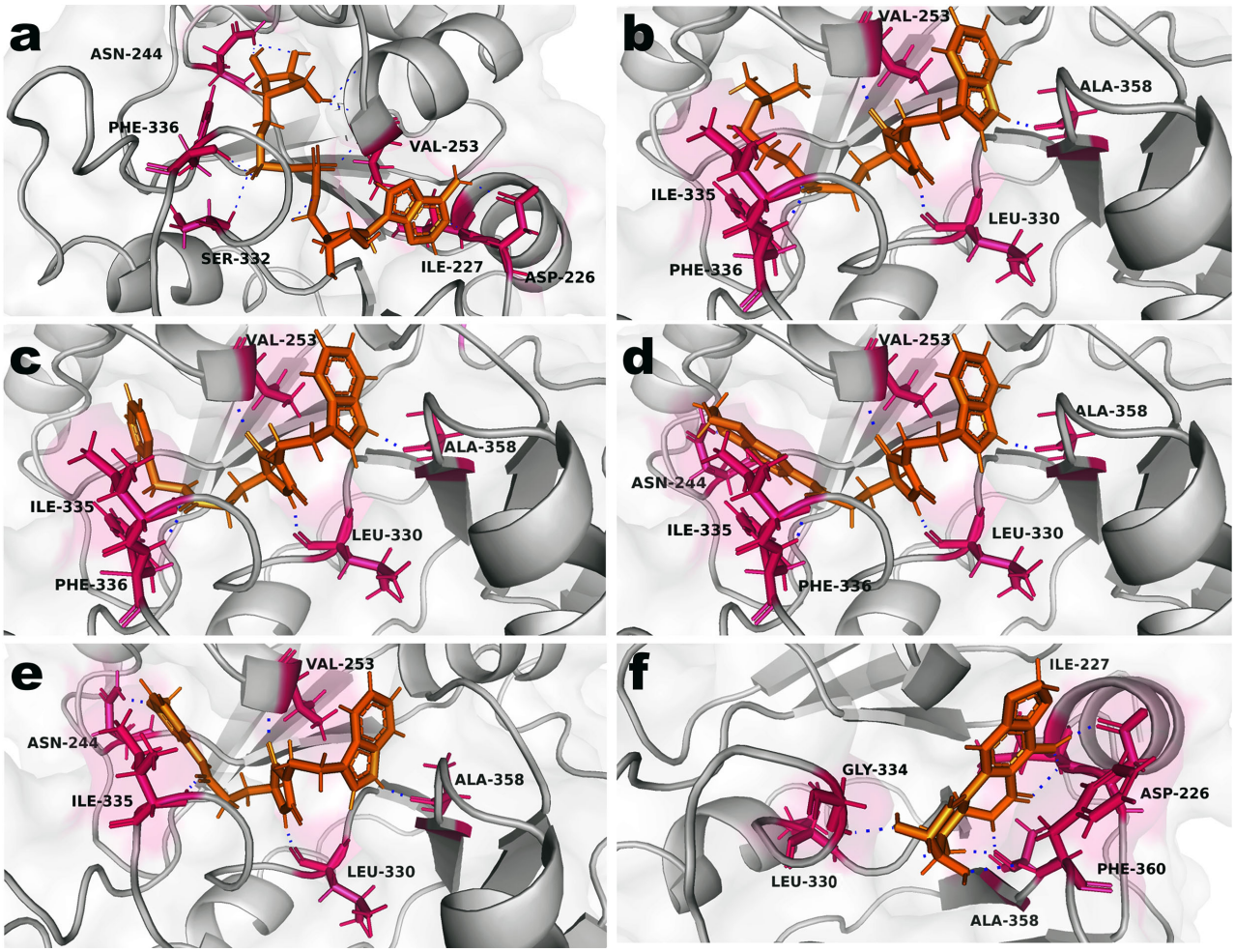

Fig. 2. Molecular docking interactions of ADP-ribose (a) and top lead compounds (ZINC08765069 (b), ZINC08792474 (c), ZINC08879336 (d), ZINC08879971 (e), ZINC00897592 (f)) with the Mac1 domain of SARS-CoV-2 using the Quantum Polarized Ligand Docking.

TABLE 1
Scoring Values of Substrate Binding and New Ligand Binding With Mac1 Domain of SARS-CoV-2

Compounds	Docking Score (XP) kcal/mol	Docking Score (QPLD) kcal/mol	MM/GBSA energy (XP) kcal/mol	MM/GBSA Energy (QPLD) kcal/mol	RMSD (Å)
ADP-ribose	−7.21	−7.49	−42.38	−51.31	0.058
ZINC08765069	−13.38	−12.86	−71.62	−70.33	0.106
ZINC08792474	−12.89	−13.20	−86.35	−85.55	0.092
ZINC08879336	−12.27	−12.83	−77.90	−81.37	0.068
ZINC08879971	−11.82	−11.88	−68.52	−70.30	0.061
ZINC00897592	−11.17	−11.55	−55.23	−58.69	0.023

called as band gap energy that elucidates the electron transfer swift between both energy levels. The representation of HOMO/LUMO and MESP is provided in Fig. 3 for the substrate and ligand by showing its quantum mechanical features of HOMO/LUMO and MESP for the ADP-ribose (a) and top lead compounds (b-f). When analyzing the structural features of substrate with new lead molecules, except the molecule ZINC00897592 all the other compounds show similar posture and while dissecting the electron donor region of the substrate and new leads, all the ligands show electron donor tendency in the head region, which is shown in Fig. 3 (HOMO). In terms of electron acceptor tendency except the compound ZINC08879336, all the other compounds show the LUMO regions in central part, and the ZINC08879336 shows the LUMO in the tail part. In terms of MESP, as shown in Fig. 3, red color surface indicates the most negative potential region, and the blue surface indicates the most positive potential region. This shows that, the substrate MESP surface is matched in the head and tail portions for all the new compounds except the ZINC00897592, and also the difference in central part is due to the presence of

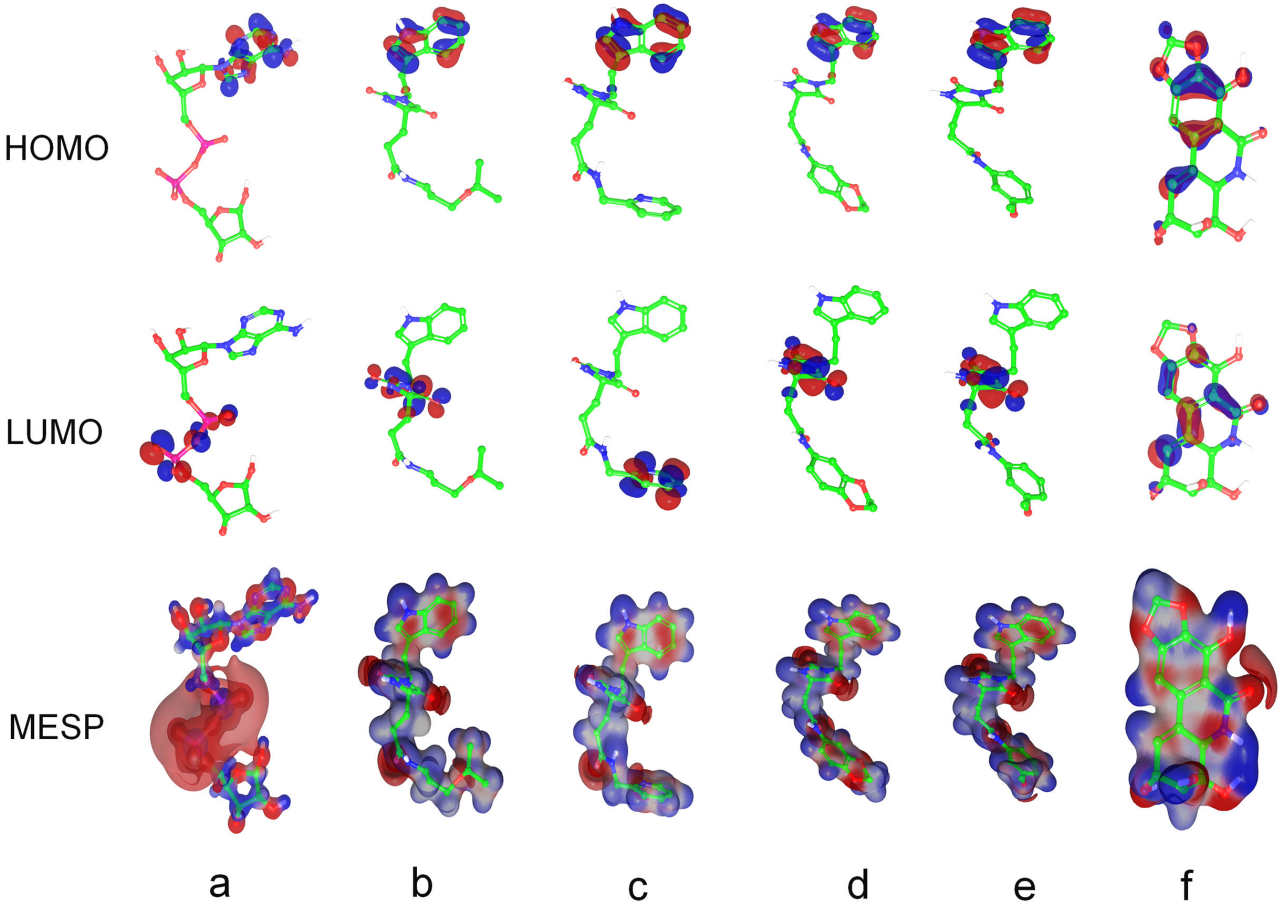

Fig. 3. DFT calculations for the substrate and ligand by showing its quantum mechanical features of HOMO/LUMO and MESP for the ADP-ribose (a) and top lead compounds (ZINC08765069 (b), ZINC08792474 (c), ZINC08879336 (d), ZINC08879971 (e), ZINC00897592 (f)).

phosphate atom in the substrates and those are more are less replaced by the nitrogen atoms in the ligands except ZINC00897592.

3.5 Molecular Dynamics Simulations

The protein ligand complex for the substrate and ligand bound complex is analyzed for the molecular dynamics simulations for the timescale of 50 ns. It is mandatory to compare the protein ligand bonds for the substrate and new lead molecules, and their exhibiting kinetic energy for their binding in the dynamic environment. Among the interactions, hydrogen bonding is core important to observe in dynamic environment, especially the amino acid contribution to hold the ligand in the aqueous environment, and so the average hydrogen bonds between the protein and ligands are shown in Fig. 4a. The average hydrogen bonds for the ADP-ribose show ~2.1, while the new lead compounds show ~2 and from that the compound ZINC00897592 shows the higher average number of hydrogen bonds by ~2.9. This is due to the compound ZINC00897592 allocation in binding pocket, which is comparatively smaller and able to hold with aromatic amino acids. The stableness of protein-ligand simulations is calculated using the RMSD, and RMSF fluctuations, provided in the supplementary information. The RMSD (Figure S2 available online) and RMSF (Figure S3 available online) favor the protein-ligand stability in the MD simulations on the dynamics aqueous state environment. In terms of kinetic energy, almost all the substrate and new lead molecules have the tendency to poseur ~80,000 kJ/mol and from this, only the compound ZINC08765069 can surpass the kinetic energy with the score of ~120,000 kJ/mol. Through this hydrogen bond analysis and kinetic energy analysis, we assume that the new compounds are having similar binding with ADP-ribose binding mechanism, and in dynamic environment, it can actively participate in tight binding with the Mac1 domain of SARS-CoV-2. As discussed earlier, the binding pocket has the unique feature by showing the open binding pocket in the apo form, and closed binding pocket in the holo form. The new lead compounds show similar HOMO/LUMO and MESP surfaces along with the substrate molecules shown in the DFT calculations. By that, the new leads are checked for its common binding mode with the substrate binding. The comparison is shown in Fig. 5, indicating that the new lead compounds have the tendency to bound in the similar binding pose with the substrate (green). Also, the new lead molecules binding did not alter the closed gate of the binding pocket, and this may be due to the new compounds having nitrogen atoms in the middle part that does not allow the binding pocket to open and remains stable in the MD simulations. We also assume that; these substrate binding can show open and closed binding pocket in the long-range timescale in the MD simulations.

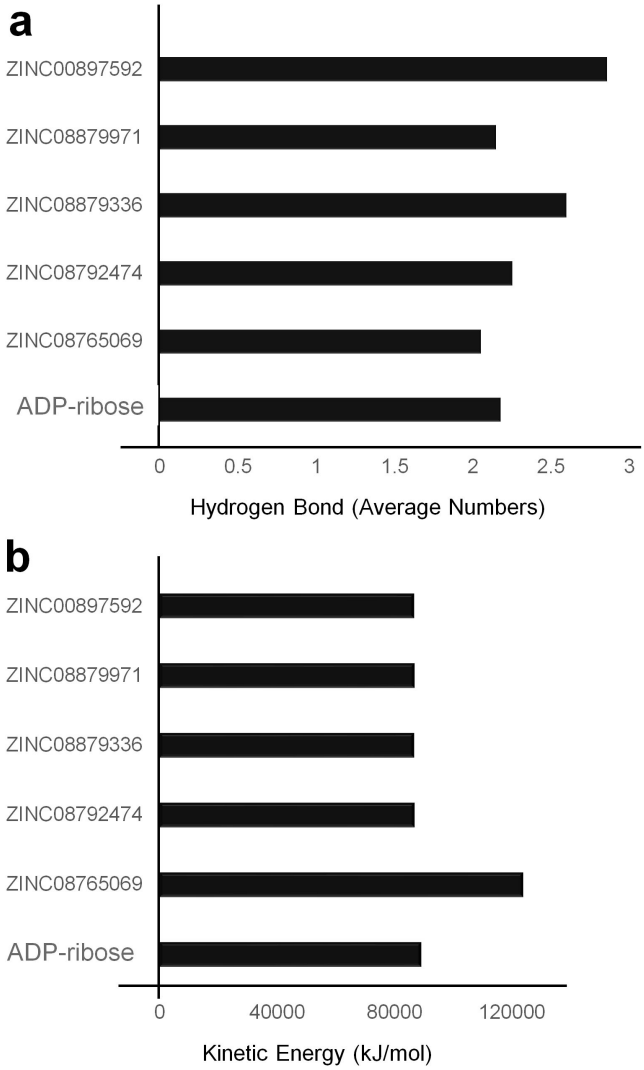

Fig. 4. Molecular dynamic simulations of the protein ligand complex for the substrate and new ligands analyzed for its average hydrogen bonds (a) and the average kinetic energy (b).

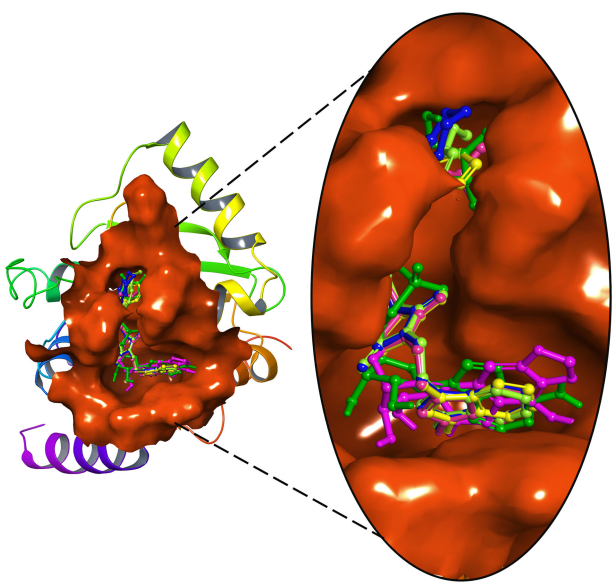

Fig. 5. Molecular interactions of substrate (green) and other new lead compounds showing similar binding inside the Mac1 domain of SARS-CoV-2.

4 Conclusion

In summary of this work, we come up with deep understanding of ADP-ribose substrate binding with Mac1 domain and based on the structural insights, we found few strong potential lead molecules that can regulate the SARS-CoV-2 pathogenesis. Through this work, we report five new lead molecules from the ZINC database namely ZINC08765069, ZINC08792474, ZINC08879336, ZINC08879971 and ZINC00 897592. These compounds are the outcome of strong filtering criteria of docking score of $<$ -10 kcal/mol and binding energy of $<$ -50 kcal/mol. The interacting amino acids and the ligand orientations are making accurate by applying the QM charges towards the ligand binding environment. The QM-based DFT calculations strongly says that the electron donor atoms are functioning in head part of the substrate and new lead molecules, while the electron acceptor is located in the middle part of the substrate and new lead molecules, except ZINC08879971 molecule. This unique feature of the presence of electron donor and acceptor in the head and middle part of the ligands is making these ligands to adopt well inside the binding pocket with the similar binding mode. Interestingly, in substrate-binding and lead molecule binding does not affect the closed binding pocket of the Mac1. This substrate and lead molecule binding is validated using the molecular dynamics in aqueous environment, showing that the substrate and lead molecules are stable by their tendency of forming two or more average hydrogen bonds between the protein-ligand complex involving crucial residues Asp226 and Asn244 which are essential for ADP-ribose binding and catalytic activity in SARS-CoV Mac1 domain. This complex structure also verified with kinetic energy shows all the complex shows lower 80000 kJ/mol that indicates, strong and stable binding of substrate and new lead molecules. We also report that, the closed pocket of Mac1, may have open and closed gated mechanism in their enzyme reaction active states. Through this work, we recommend that, these five lead compounds on further experimental validations along with strong international effort can make these compounds as strong inhibitors against the SARS-CoV-2 Mac1 domain ADP-ribosylhydrolase activity.

Acknowledgments

CS and SKS thankfully acknowledge RUSA-Phase 2.0 Policy (TNmulti-Gen), Dept. of Edn, Govt. of India (Grant No: F.24-51/2014-U). DCD and EB acknowledge European Regional Development Fund; OP RDE; Project: "Chemical biology for drugging undruggable targets (ChemBioDrug)" (No. CZ.02.1.01/0.0/0.0/16_019/0000729), Programu podpory perspektivních lidských zdrojů (PPLZ) of the Czech academy of sciences - postdoctoral fellowship (No.L200551951) and the Academy of Sciences of Czech Republic (RVO: 61388963).

Reference

[1] C. Contini *et al.*, "The novel zoonotic COVID-19 pandemic: An expected global health concern," *J. Infect. Dev. Countries*, vol. 14, no. 3, pp. 254–264, Mar. 31, 2020.

[2] M. G. Hemida and M. M. Ba Abduallah, "The SARS-CoV-2 outbreak from a one health perspective," *One Health*, vol. 10, Mar. 16, 2020, Art. no. 100127.

[3] F. K. Yoshimoto, "The proteins of severe acute respiratory syndrome coronavirus-2 (SARS coV-2 or n-COV19), the cause of COVID-19," *Protein J.*, vol. 39, pp. 198–216, May 23, 2020.
[4] I. Imbert, E. J. Snijder, M. Dimitrova, J. C. Guillemot, P. Lecine, and B. Canard, "The SARS-Coronavirus PLnc domain of nsp3 as a replication/transcription scaffolding protein," *Virus Res.*, vol. 133, no. 2, pp. 136–148, May 2008.
[5] B. H. Harcourt *et al.*, "Identification of severe acute respiratory syndrome coronavirus replicase products and characterization of papain-like protease activity," *J Virol.*, vol. 78, no. 24, pp. 13600–13612, Dec. 2004.
[6] Y. R. Guo *et al.*, "The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak - an update on the status," *Mil. Med. Res.*, vol. 7, no. 1, Mar. 13, 2020, Art. no. 11.
[7] A. Wu *et al.*, "Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in china," *Cell Host Microbe*, vol. 27, no. 3, pp. 325–328, Mar. 11, 2020.
[8] B. W. Neuman, "Bioinformatics and functional analyses of coronavirus nonstructural proteins involved in the formation of replicative organelles," *Antiviral Res.*, vol. 135, pp. 97–107, Nov. 2016.
[9] J. Lei, Y. Kusov, and R. Hilgenfeld, "Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein," *Antiviral Res.*, vol. 149, pp. 58–74, Jan. 2018.
[10] K. S. Saikatendu *et al.*, "Structural basis of severe acute respiratory syndrome coronavirus ADP-ribose-1"-phosphate dephosphorylation by a conserved domain of nsP3," *Structure*, vol. 13, no. 11, pp. 1665–1675, Nov. 2005.
[11] A. R. Fehr, G. Jankevicius, I. Ahel, and S. Perlman, "Viral macrodomains: Unique mediators of viral replication and pathogenesis," *Trends Microbiol.*, vol. 26, no. 7, pp. 598–610, Jul. 2018.
[12] X. Deng, R. C. Mettelman, A. O'Brien, J. A. Thompson, T. E. O'Brien, and S. C. Baker, "Analysis of coronavirus temperature-sensitive mutants reveals an interplay between the macrodomain and papain-like protease impacting replication and pathogenesis," *J. Virol.*, vol. 93, no. 12, pp. 1–16, Jun. 15, 2019.
[13] Y. Kusov, J. Tan, E. Alvarez, L. Enjuanes, and R. Hilgenfeld, "A G-quadruplex-binding macrodomain within the "SARS-unique domain" is essential for the activity of the SARS-coronavirus replication-transcription complex," *Virology*, vol. 484, pp. 313–322, Oct. 2015.
[14] Y. M. Baez-Santos, S. E. St John, and A. D. Mesecar, "The SARS-coronavirus papain-like protease: Structure, function and inhibition by designed antiviral compounds," *Antiviral Res.*, vol. 115, pp. 21–38, Mar. 2015.
[15] A. K. L. Leung, R. L. McPherson, and D. E. Griffin, "Macrodomain ADP-ribosylhydrolase and the pathogenesis of infectious diseases," *PLoS Pathogens*, vol. 14, no. 3, Mar. 2018, Art. no. e1006864.
[16] J. G. Rack, D. Perina, and I. Ahel, "Macrodomains: Structure, function, evolution, and catalytic activities," *Annu. Rev. Biochem.*, vol. 85, pp. 431–454, Jun. 2, 2016.
[17] A. Putics, W. Filipowicz, J. Hall, A. E. Gorbalenya, and J. Ziebuhr, "ADP-ribose-1"-monophosphatase: A conserved coronavirus enzyme that is dispensable for viral replication in tissue culture," *J. Virol.*, vol. 79, no. 20, pp. 12721–12731, Oct. 2005.
[18] M. P. Egloff *et al.*, "Structural and functional basis for ADP-ribose and poly(ADP-ribose) binding by viral macro domains," *J. Virol.*, vol. 80, no. 17, pp. 8493–8502, Sep. 2006.
[19] N. P. Shull, S. L. Spinelli, and E. M. Phizicky, "A highly specific phosphatase that acts on ADP-ribose 1"-phosphate, a metabolite of tRNA splicing in saccharomyces cerevisiae," *Nucleic Acids Res.*, vol. 33, no. 2, pp. 650–660, 2005.
[20] T. Nelemans and M. Kikkert, "Viral innate immune evasion and the pathogenesis of emerging RNA virus infections," *Viruses*, vol. 11, no. 10, Oct. 18, 2019, Art. no. 961.
[21] M. Kikkert, "Innate immune evasion by human respiratory RNA viruses," *J. Innate Immun.*, vol. 12, no. 1, pp. 4–20, 2020.
[22] Y. X. Lim, Y. L. Ng, J. P. Tam, and D. X. Liu, "Human coronaviruses: A review of virus-host interactions," *Diseases*, vol. 4, no. 3, Jul. 25, 2016, Art. no. 26.
[23] M. E. Grunewald *et al.*, "The coronavirus macrodomain is required to prevent PARP-mediated inhibition of virus replication and enhancement of IFN expression," *PLoS Pathogens*, vol. 15, no. 5, May 2019, Art. no. e1007756.
[24] Y. M. O. Alhammad and A. R. Fehr, "The viral macrodomain counters host antiviral ADP-Ribosylation," *Viruses*, vol. 12, no. 4, Mar. 31, 2020, Art. no. 384.
[25] J. Lei and R. Hilgenfeld, "RNA-virus proteases counteracting host innate immunity," *FEBS Lett.*, vol. 591, no. 20, pp. 3190–3210, Oct. 2017.
[26] C. C. S. De Oliveira, G. R. C. Pereira, J. Y. S. de Alcantara, D. Antunes, E. R. Caffarena, and J. F. De Mesquita, "In silico analysis of the V66M variant of human BDNF in psychiatric disorders: An approach to precision medicine," *PLoS One*, vol. 14, no. 4, 2019, Art. no. e0215508.
[27] S. B. Pandit and J. Skolnick, "Fr-TM-align: A new protein structural alignment method based on fragment alignments and the TM-score," *BMC Bioinf.*, vol. 9, Dec. 12, 2008, Art. no. 531.
[28] X. Robert and P. Gouet, "Deciphering key features in protein structures with the new ENDscript server," *Nucleic Acids Res.*, vol. 42, no. Web Server issue, pp. W320–W324, Jul. 2014.
[29] M. A. Lill and M. L. Danielson, "Computer-aided drug design platform using PyMOL," *J. Comput. Aided Mol. Des.*, vol. 25, no. 1, pp. 13–19, Jan. 2011.
[30] R. Yadav *et al.*, "Investigating into the molecular interactions of flavonoids targeting NS2B-NS3 protease from ZIKA virus through in-silico approaches," *J. Biomolecular Struct. Dyn.*, pp. 1–13, Jan. 10, 2020.
[31] A. R. Muralidharan, C. Selvaraj, S. K. Singh, J. R. Sheu, P. A. Thomas, and P. Geraldine, "Structure-Based virtual screening and biological evaluation of a calpain inhibitor for prevention of selenite-induced cataractogenesis in an in vitro system," *J. Chem. Inf. Model.*, vol. 55, no. 8, pp. 1686–1697, Aug. 24, 2015.
[32] C. Selvaraj and S. K. Singh, "Validation of potential inhibitors for SrtA against bacillus anthracis by combined approach of ligand-based and molecular dynamics simulation," *J. Biomolecular Struct. Dyn.*, vol. 32, no. 8, pp. 1333–1349, 2014.
[33] Z. Chen, X. Cen, J. Yang, X. Tang, K. Cui, and K. Cheng, "Structure-based discovery of a specific TLR1-TLR2 small molecule agonist from the ZINC drug library database," *Chem. Commun. (Camb)*, vol. 54, no. 81, pp. 11411–11414, Oct. 9, 2018.
[34] I. J. Chen and N. Foloppe, "Drug-like bioactive structures and conformational coverage with the ligprep/confgen suite: Comparison to programs MOE and catalyst," *J. Chem. Inf. Model*, vol. 50, no. 5, pp. 822–839, May 24, 2010.
[35] C. Selvaraj, J. Sivakamavalli, B. Vaseeharan, P. Singh, and S. K. Singh, "Examine the characterization of biofilm formation and inhibition by targeting SrtA mechanism in bacillus subtilis: A combined experimental and theoretical study," *J. Mol. Model*, vol. 20, no. 8, Aug. 2014, Art. no. 2364.
[36] C. Selvaraj, J. Sivakamavalli, V. Baskaralingam, and S. K. Singh, "Virtual screening of LPXTG competitive SrtA inhibitors targeting signal transduction mechanism in bacillus anthracis: A combined experimental and theoretical study," *J Receptors Signal Transduction Res.*, vol. 34, no. 3, pp. 221–232, Jun. 2014.
[37] H. Le-Thi-Thu, G. M. Casanola-Martin, Y. Marrero-Ponce, A. Rescigno, C. Abad, and M. T. Khan, "A rational workflow for sequential virtual screening of chemical libraries on searching for new tyrosinase inhibitors," *Curr. Top Med. Chem.*, vol. 14, no. 12, pp. 1473–1485, 2014.
[38] C. Selvaraj, J. Sivakamavalli, B. Vaseeharan, P. Singh, and S. K. Singh, "Structural elucidation of SrtA enzyme in enterococcus faecalis: An emphasis on screening of potential inhibitors against the biofilm formation," *Mol. BioSyst.*, vol. 10, no. 7, pp. 1775–1789, Jul, 2014.
[39] G. N. S. Hema Sree, S. G. Rajalekshmi, M. Murahari, and R. R. Burri, "Reappraisal of FDA approved drugs against alzheimer's disease based on differential gene expression and protein interaction network analysis: An in silico approach," *J. Biomolecular Struct. Dyn.*, vol. 38, pp. 3972–3989, Oct. 14, 2019.
[40] K. K. Reddy, S. K. Singh, S. K. Tripathi, and C. Selvaraj, "Identification of potential HIV-1 integrase strand transfer inhibitors: In silico virtual screening and QM/MM docking studies," *SAR QSAR Environ. Res.*, vol. 24, no. 7, pp. 581–595, 2013.
[41] P. Vijayalakshmi, C. Selvaraj, S. K. Singh, J. Nisha, K. Saipriya, and P. Daisy, "Exploration of the binding of DNA binding ligands to staphylococcal DNA through QM/MM docking and molecular dynamics simulation," *J. Biomolecular Struct. Dyn.*, vol. 31, no. 6, pp. 561–571, 2013.
[42] C. Selvaraj, P. Singh, and S. K. Singh, "Molecular insights on analogs of HIV PR inhibitors toward HTLV-1 PR through QM/MM interactions and molecular dynamics studies: Comparative structure analysis of wild and mutant HTLV-1 pR," *J. Mol. Recognit.*, vol. 27, no. 12, pp. 696–706, Dec, 2014.

[43] C. Selvaraj, P. Singh, and S. K. Singh, "Molecular modeling studies and comparative analysis on structurally similar HTLV and HIV protease using HIV-PR inhibitors," *J. Receptors Signal Transduction Res.*, vol. 34, no. 5, pp. 361–371, Oct. 2014.
[44] V. Suryanarayanan, S. K. Singh, S. K. Tripathi, C. Selvaraj, K. K. Reddy, and A. Karthiga, "A three-dimensional chemical phase pharmacophore mapping, QSAR modelling and electronic feature analysis of benzofuran salicylic acid derivatives as LYP inhibitors," *SAR QSAR Environ. Res.*, vol. 24, no. 12, pp. 1025–1040, 2013.
[45] S. Chinnasamy *et al.*, "Combining in silico and in vitro approaches to identification of potent inhibitor against phospholipase A2 (PLA2)," *Int. J. Biol. Macromol.*, vol. 144, pp. 53–66, Feb. 1, 2020.
[46] S. Chinnasamy *et al.*, "Molecular docking and molecular dynamics simulation studies to identify potent AURKA inhibitors: Assessing the performance of density functional theory, MM-GBSA and mass action kinetics calculations," *J. Biomol. Struct. Dyn.*, pp. 1–11, Oct. 14, 2019.
[47] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. Berendsen, "GROMACS: Fast, flexible, and free," *J. Comput. Chem.*, vol. 26, no. 16, pp. 1701–1718, Dec. 2005.
[48] C. Selvaraj, S. Sakkiah, W. Tong, and H. Hong, "Molecular dynamics simulations and applications in computational toxicology and nanotoxicology," *Food Chem. Toxicol.*, vol. 112, pp. 495–506, Feb. 2018.
[49] D. Paschek, R. Day, and A. E. Garcia, "Influence of water-protein hydrogen bonding on the stability of Trp-cage miniprotein. A comparison between the TIP3P and TIP4P-Ew water models," *Phys. Chem. Chem. Phys.*, vol. 13, no. 44, pp. 19840–19847, Nov. 28, 2011.
[50] D. M. van Aalten, R. Bywater, J. B. Findlay, M. Hendlich, R. W. Hooft, and G. Vriend, "PRODRG, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules," *J. Comput. Aided Mol. Des.*, vol. 10, no. 3, pp. 255–262, Jun. 1996.
[51] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, "GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation," *J. Chem. Theory Comput.*, vol. 4, no. 3, pp. 435–547, Mar. 2008.
[52] H. R. Jonsdottir and R. Dijkman, "Coronaviruses and the human airway: A universal system for virus-host interaction studies," *J. Virol.*, vol. 13, Feb. 6, 2016, Art. no. 24.

Chandrabose Selvaraj received the PhD degree in bioinformatics from Alagappa University, India. He carried out his postdoctoral research in International esteemed institutes that include FDA-USA, CEITEC-Brno, Konkuk University -Seoul, IIT-Mandi and currently he is working as RUSA-SPDF with the Department of Bioinformatics, Alagappa University, India. His works are mostly dedicated to understanding the molecular level interactions between the pathogenic microbes and humans. Presently he is working on SARS-CoV-2 related projects analyzing drug targets for their interactions and identifying potential drug targets to combat against COVID- 19.

Dhurvas Chandrasekaran Dinesh received the PhD degree in biochemistry from Martin-Luther University, Germany. He has also received an advance diploma in bioinformatics from Madurai Kamaraj University and worked as a junior researcher in Indian Institute of Sciences, India. Currently, he is currently working as a postdoctoral fellow with the Institute of Organic Chemistry and Biochemistry of the CAS, Czech Republic. His research interest includes structural biology, biochemistry, and bioinformatics.

Umesh Panwar is currently working toward the doctoral degree in the Department of Bioinformatics, Alagappa University, India. His research involves the application of QM and MD simulation techniques in viral RNA/DNA binding with host proteins.

Evzen Boura received the PhD degree from Charles University, CR. He continued his post-doctoral training at the NICHD, NIH, USA. He is a senior group leader with the Department of Molecular Biology and Biochemistry, Institute of Organic Chemistry and Biochemistry-CAS, Czech Republic. His current group mainly focuses on +RNA virus replication, virus-host factor interaction, and antiviral research. He has contributed more than 50 macromolecular structures to the protein data bank (PDB) and several high-impact publications.

Sanjeev Kumar Singh is currently working as the group leader of Structural Bioinformatics and CADD lab at the Department of Bioinformatics, Alagappa University, India. He has extensive exposure in the research towards HIV and other viruses like ZIKV, DENV, HPV and HTLV. His research focuses mainly on the identification of novel inhibitors towards the inhibition of viruses viral interaction with the human host. He has tremendous experience in the field of the CADD and structural bioinformatics and he has published more than 140 publications in peer reviewed journals. He works in close with experimental biologists and has contributed to research in several fundamental subjects including host-pathogen interactions, drug discovery towards viruses, molecular modeling, and structural bioinformatics.