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

Computers in Biology and Medicine 134 (2021) 104495

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Structural and conformational analysis of SARS CoV 2 N-CTD revealing monomeric and dimeric active sites during the RNA-binding and stabilization: Insights towards potential inhibitors for N-CTD

Arushi Chauhan [a], Pramod Avti [a,*], Nishant Shekhar [b], Manisha Prajapat [b], Phulen Sarma [b], Anusuya Bhattacharyya [c], Subodh Kumar [b], Hardeep Kaur [b], Ajay Prakash [b], Bikash Medhi [b,**]

[a] Department of Biophysics, Postgraduate Institute of Medical Education and Research, Chandigarh, India
[b] Department of Pharmacology, Postgraduate Institute of Medical Education and Research, Chandigarh, India
[c] Department of Ophthalmology, GMCH-32, Chandigarh, India

ARTICLE INFO

Keywords:
Dynamic cross-correlation matrix (DCCM)
Molecular mechanics/generalized born model and solvent accessibility (MM/GBSA)
Molecular dynamics (MD)
Nucleocapsid C terminal domain (N-CTD)
Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)

ABSTRACT

The advent of SARS-CoV-2 has become a universal health issue with no appropriate cure available to date. The coronavirus nucleocapsid (N) protein combines viral genomic RNA into a ribonucleoprotein and protects the viral genome from the host's nucleases. Structurally, the N protein comprises two independent domains: the N-terminal domain (NTD) for RNA-binding and C-terminal domain (CTD) involved in RNA-binding, protein dimerization, and nucleocapsid stabilization. The present study explains the structural aspects associated with the involvement of nucleocapsid C-terminal domain in the subunit assembly that helps the RNA binding and further stabilizing the virus assembly by protecting RNA from the hosts exonucleases degradation. The molecular dynamics (MD) simulations of the N-CTD and RNA complex suggests two active sites (site I: a monomer) and (site II: a dimer) with structural stability (RMSD: ~2 Å), Cα fluctuations (RMSF: ~3 Å) and strong protein-ligand interactions were estimated through the SiteMap module of Schrodinger. Virtual screening of 2456 FDA-approved drugs using structure-based docking identified top two leads distinctively against Site-I (monomer): Ceftaroline fosamil (MM-GBSA = −47.12 kcal/mol) and Cefoperazone (−45.84 kcal/mol); and against Site-II (dimer): Boceprevir, (an antiviral protease inhibitor, −106.78 kcal/mol) and Ceftaroline fosamil (−99.55 kcal/mol). The DCCM and PCA of drugs Ceftaroline fosamil (PC1+PC2 = 71.9%) and Boceprevir (PC1 +PC2 = 61.6%) show significant correlated residue motions which suggests highly induced conformational changes in the N-CTD dimer. Therefore, we propose N-CTD as a druggable target with two active binding sites (monomer and dimer) involved in specific RNA binding and stability. The RNA binding site with Ceftaroline fosamil binding can prevent viral assembly and can act as an antiviral for coronavirus.

1. Importance

The high incident rate and high risk of developing infectious mutant strains of SARS-CoV-2 has inspired researchers to explore extensively the therapeutically important target macromolecules. The nucleocapsid proteins (N and C-terminal domains) are necessary for the viral assembly, multiplication, and protecting the viral genome from the host's extracellular agents. So, targeting the nucleocapsid's C-terminal domain (N-CTD) will lead to improper assembly of its protein structure thus making the virus vulnerable to the host's extracellular agents. The present study shows N-CTD as a druggable target studied by understanding the structural stability after drug binding. The drug binding restricts the viral protein to multiply in huge numbers in the host. Therefore, this study points out the new direction of destabilizing the virus-protein complex structure and its multiplication mechanism due to drug binding. Thus, there is an urgent need to explore the possible therapeutic mechanisms for the effective treatment of SARS-CoV-2.

* Corresponding author.
** Corresponding author.
E-mail addresses: pramod.avti@gmail.com (P. Avti), drbikashus@yahoo.com (B. Medhi).

https://doi.org/10.1016/j.compbiomed.2021.104495
Received 10 February 2021; Received in revised form 11 May 2021; Accepted 11 May 2021
Available online 15 May 2021

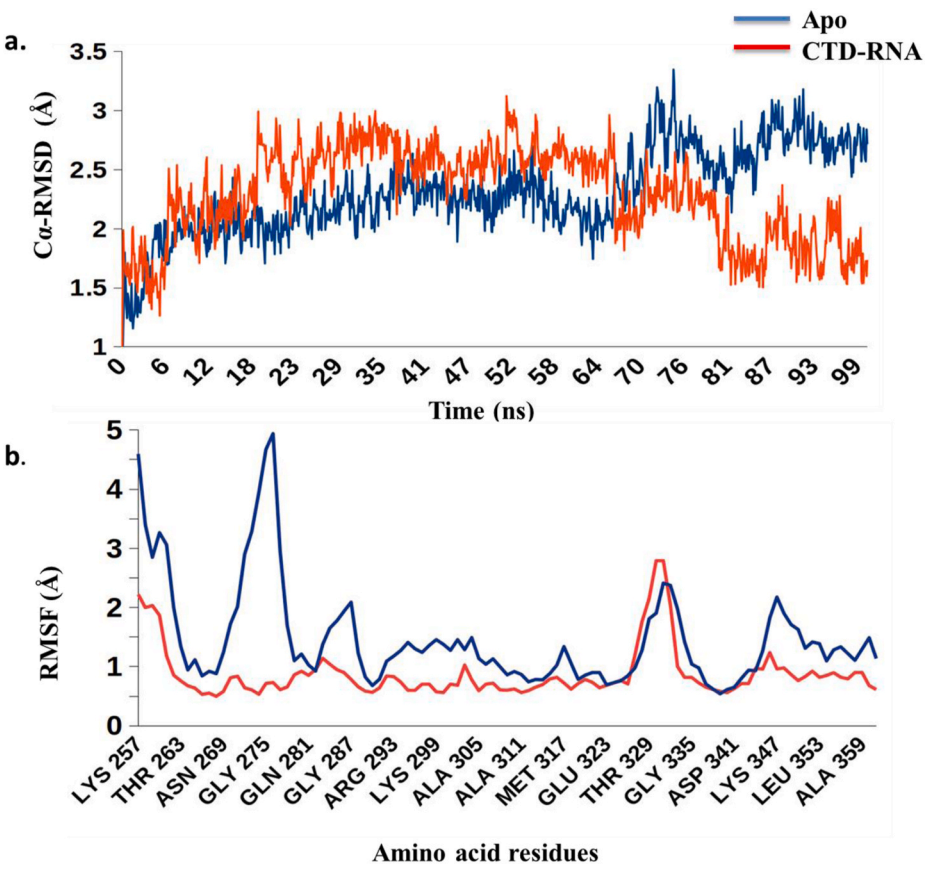

Fig. 1. The MD simulation profile for RNA-bound and unbound (apo) N-CTD dimer form. (a) Comparative C-alpha RMSD,RNA-bound CTD (red) with 2.29 Å and apo (blue) with 2.30 Å. (b) C alpha-RMSF comparison showing the mean local residue fluctuation throughout 100ns simulation.

2. Introduction

In December 2019, a highly virulent human virus, known as a coronavirus (COVID-19), was reported in Wuhan, China (epicenter of this virus). Majorly, six or more strains of human coronavirus are detected till date, the native strain that was first reported in 2019 is the L-strain which can mutate and convert into different strains. The second strain that appeared in early January 2020 is the S-strain, and towards the end of the month, new strains such as V and G have also emerged [1]. The coronavirus, caused by SARS CoV 2 (Severe acute respiratory syndrome coronavirus 2), is regarded as an enveloped "positive-stranded RNA virus" that belongs to the family of *Coronaviridae* [2,3]. It affects the lower respiratory system causing discomfort, dry cough, extreme fatigue, breathlessness, and pneumonia in some cases [4,5]. Usually, the elderly and patients having comorbidities such as diabetes, heart, and lung diseases are easily affected due to the compromised immunity [6].

Coronavirus is an encased particle that comprises spike (S), membrane (M), envelope (E), protease, and nucleocapsid (N) proteins. These four structural proteins are necessary for the formation of a well-organized structure of the infectious virions and can serve as an essential drug target for coronavirus [7–10]. The genome of coronavirus is a linear, positive-stranded mRNA of about 32 kb size is encapsulated via nucleocapsid (N) protein [11]. Besides all these structural proteins, SARS CoV 2 comprises of 16 non-structural proteins (NSPs) that get arranged into a replication-transcription complex. The complex is further packaged into a double membranous structure, thus serving as an initiator of replication [12]. The most plenteous protein in the viral cells in the N protein; it is the only protein that comes in contact with the transcriptase complex. The nucleocapsid protein's prime task is to bundle the 30-kb single-stranded RNA viral genome into a ribonucleoprotein complex (RNP) known as the capsid [12,13]. Some *in vitro* and *in vivo* observations have shown that N protein is critical in viral genome replication [14,15]. Due to its prominent role in regulating the synthesis of the viral RNA genome and high functional value, N protein can be considered a potential target for viral replication. Various studies reveal that the structure of N protein comprises three regions, two independent domains (N-terminal domain (NTD) and C-terminal domain (CTD)) and a linker region. The NTD was first described by Huang et al., 2014; it is the RNA binding domain (RBD) made up of five β-sheet (β4–β2–β3–β1–β5). The second and third beta-sheets are connected via a β-hairpin loop-like structure (β2′–β0) comprising positively charged amino acids, arginine and lysine [16]. Upon virus 3D packaging, a pocket-like structure is formed between the hairpin loop and the core structure of the virus, which serves as the RNA-binding site [17].

The CTD serves as a dimerization domain since it includes the residues that make homo-dimers and homo-oligomers through a domain-withstanding process for self-association [18,19]. Recently, various studies revealed that despite having a dimerization function, CTD might have a role in RNA binding. It is reported that the N-protein is depicted as a dimer formed via N-CTD interaction. A linker region (LKR) region that is highly disordered serves as the connection between the NTD and CTD, (it connects 2N-NTDs to N-CTD dimmers) [20]. Nucleocapsid protein resides at the CoV RNA synthesis sites called Replication Transcription Complexes (RTC). The N-terminal domain captures the RNA genome while CTD anchors non-structural protein (NSP-3) components of RTCs [22]. The NSP-3 serves as a vital framework component of the replication/transcription complex as it binds other essential proteins necessary for the lifecycle of SARS CoV 2. This emphasizes the role of CTD in RNA binding and makes CTD a critical druggable target [18–21]. The dimerization domain CTD comprises 6 alpha-helices and 4 beta-strands at both the ends [23]. The dimer is stabilized via intermolecular hydrogen bonds, and the residues ranging from 248 to 270

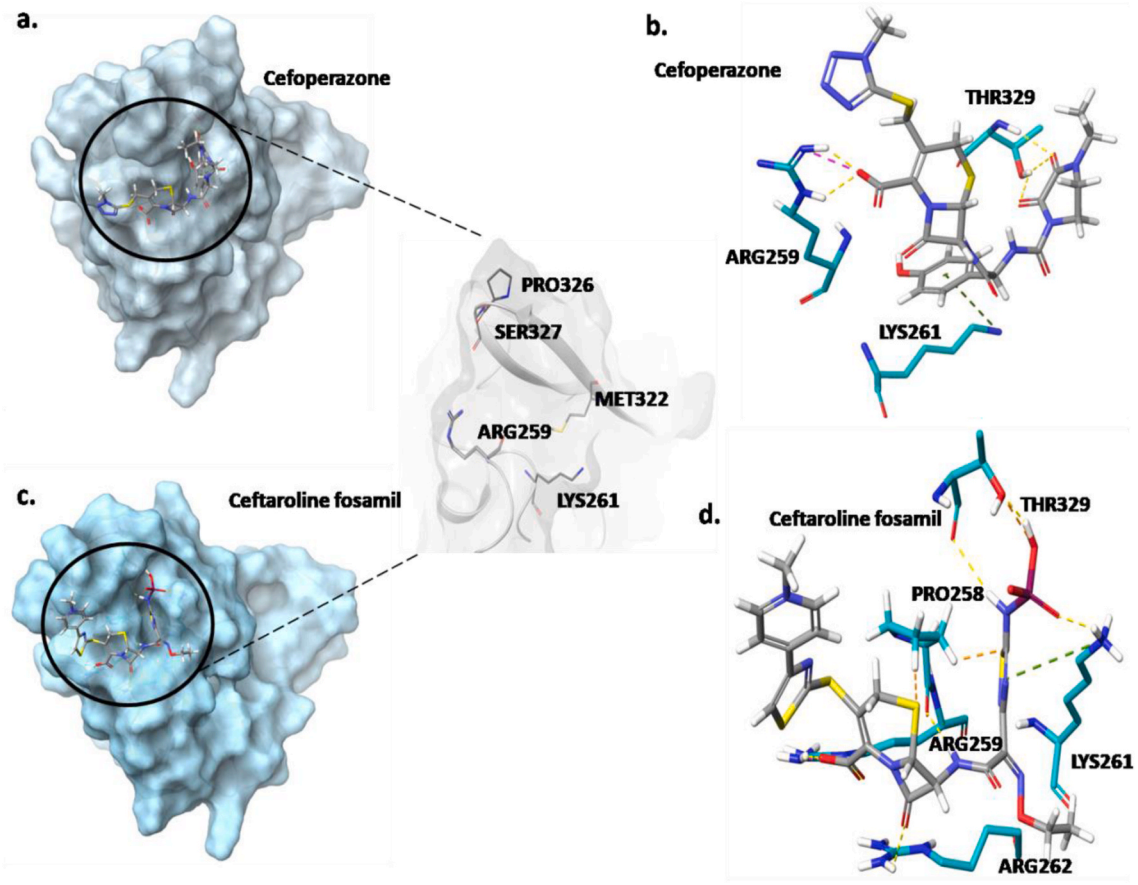

Fig. 2. Site I- a. Surface diagram of Cefoperazone-bound N-CTD, b. Interaction of the amino acid residues with Cefoperazone c. Surface diagram of Ceftaroline fosamil-bound N-CTD, d. Interaction of the amino acid residues with Ceftaroline fosamil. (Yellow dotted-hydrogen bonding; Green dotted-pi-cation bonding; Orange-weak van der Waals contacts).

also play an essential role in stabilizing the dimer. A previous study reported that the signal for the viral RNA packaging lies at the C-terminal half of the SARS-CoV [24]. Another interaction that the CTD is involved in is with the Membrane (M) protein. The N3 domain of the CTD interacts with the RNA and is displaced, which binds to the membrane protein [25]. Further, N-protein's structural study reveals that the residues of CTD are involved in RNA-binding and amino acid residues between 248 and 270 positions are essential for nucleotide-binding [20]. Computational aided drug designing has led to the identification of various targets and rapidly screening their possible inhibitors without much economic and resources burden in the shortest time possible [26,27]. Various *in vitro* clinically relevant models are also developed that can help in testing potential drug candidates against the novel targets to support the findings of the *in silico* studies [27,28].

Therefore, the present study explains the structural attributes and importance of CTD assembly domains and turns, facilitating stronger CTD binding. This forms a stable association with the RNA binding and improves the stability thereby protecting the RNA from the hosts nuclease degradation. We propose that the assembly of CTD is induced and gets stabilized concertedly, and targeting the protein-RNA assembly will inhibit the viral life cycle. This might show allosteric alterations in the conformationally active regions when introduced to a potent ligand on an inducible site. Another goal is to find out potential binding sites that may result in conformational changes on the ligand binding. Therefore, targeting the N-CTD can help inhibit the assembly without the proper capsid formation, the viral genome gets exposed to the host's nuclease machinery and limiting the propagation lifecycle of the virus. Hence, the NTD is the emerging target of focus for research groups [29,30]. The present study addresses some of the critical issues related to the structural and conformational analysis of N-CTD and addresses the monomer's role to dimer stabilization for the RNA binding and stabilization. Further, high throughput screening and identification of FDA-approved compounds and, therefore, the main target for developing potential inhibitors.

3. Material and methods

3.1. Protein preparation

The X-RAY crystallized protein structure was downloaded from RCSB with a PDB ID 7C22, possessing a resolution of 2.00 Å (https://www.rcsb.org/structure/7C22). The given PDB was selected as there are no bond length, angle, chirality, and planarity outliers. Other PDB structures like 6YUN, 6WJI, 6WZQ, and 2GIB are also available, but some are not published, and many of these structures show symmetric clashes, bond length, or bond angle outliers. The protein structures used in docking were first examined, and any ligand, metal ion, or other substances present in the structure were removed.

PDBePISA was used to explore the macromolecular interfaces and calculate protein interaction energy [31].

3.2. Molecular dynamics simulations

Molecular Dynamics (MD) simulation was performed to analyze the protein/ligand complexes by utilizing Desmond (D E Shaw Research (DESRES)). The 3-site model TIP3P was used to achieve high computational efficiency considering the actual geometry of water molecules. Equilibration was performed using NPT-isothermal–isobaric ensemble

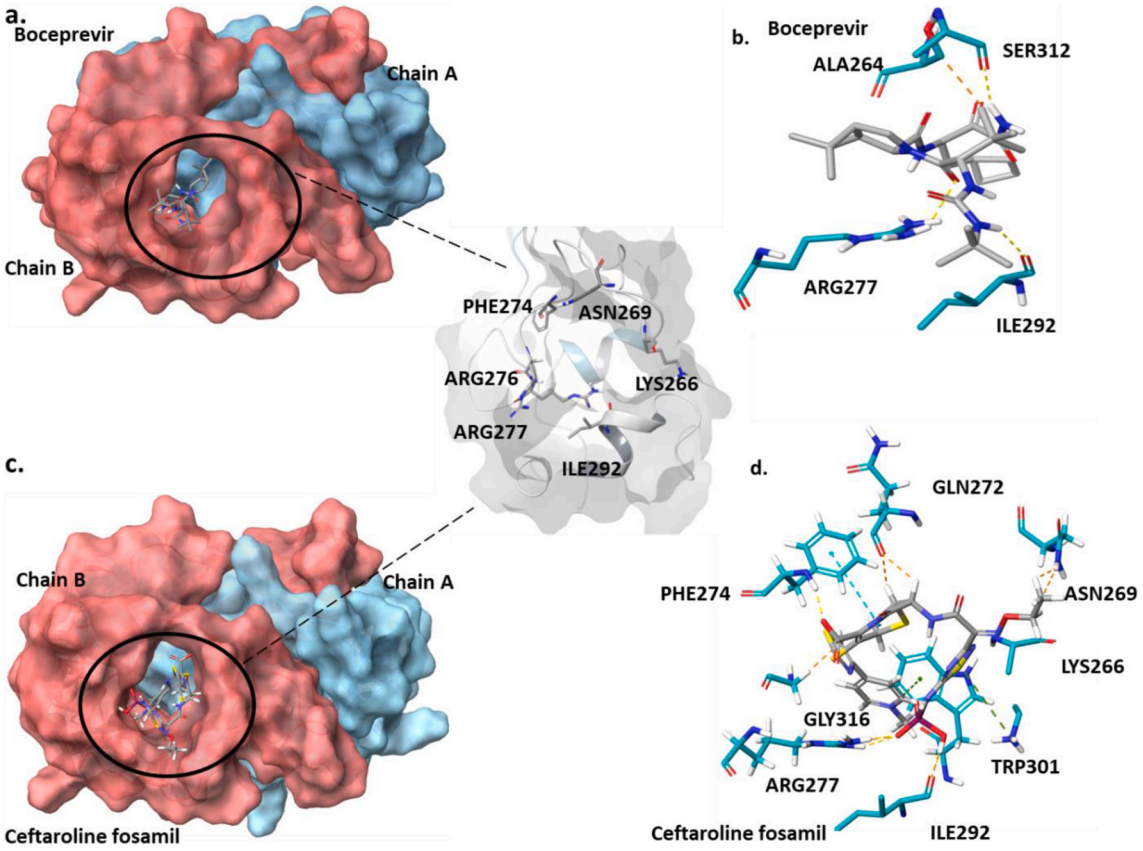

Fig. 3. Site II- a. Surface diagram of Boceprevir-bound N-CTD, b. Interaction of the amino acid residues with Boceprevir, c. Surface diagram of Ceftaroline fosamil-bound N-CTD, d. Interaction of the amino acid residues with Ceftaroline fosamil.(Yellow dotted-hydrogen bonding; Green dotted-pi-cation bonding; Orange-weak van der Waals contacts).

(constant temperature and constant pressure ensemble) class possessing Nose-Hoover thermostat of 310K and Martyna-Tobias-kleinbarostat at 1.0132Pa. A periodic cuboidal box of 10 × 10 × 20 Å was filled with protein-drug/ligand complexes, counterions, and salt solution. The simulation was planned for 100 ns, and Desmond was used to analyze the trajectory data.

3.3. MD trajectory analysis

MM/GBSA was calculated using trajectory analysis. The experiment was set up at an interval of 1 ns (step size = 10) for each protein-ligand complex using script *thermal_mmgbsa.py* available from schrodinger.com.

Using the Bio3d package in R studio, insights into the MD trajectories of protein-ligand complexes were obtained by examining the Dynamical Cross-Correlation Matrix (DCCM) and Principal Component Analysis (PCA) [32]. The DCCM, which depends on the degree to which the system's fluctuations are associated, depicts the correlation coefficient's magnitude. The key features of the DCCM include defining the correlation between the motions of various adjacent or distant domains. The PCA is a function that enables visualization of the variations in the dataset, i.e., it can be used in depicting the association between the significant global motions obtained through the protein dynamics trajectory.

3.4. Ligand preparation and virtual screening

A total of 2456 approved drugs were selected and modified using the ligprep module of Schrodinger; OPLS 2003 force field was adjusted to attain ligands with minimum energy.

For virtual screening, the binding sites were determined with the help of the SiteMap Schrodinger [33–35]. The virtual screening was performed using the "Glide" tool of "Maestro in the following order, High-Throughput Virtual Screening (HTVS) → Standard-Precision (SP) → Extra-precision (XP)." Top hits were selected for further analysis in case of both the monomer and the tetramer.

3.5. MM-GBSA: binding free energy calculation

The module, prime MM-GBSA Maestro, Schrödinger, LLC, New York, NY, USA, was used for binding free energy calculations between the ligand and receptor present in the docked complex. The ligand-bound pose-viewer complexes were subjected to MM-GBSA in VSGB solvation within 5.0 Å from flexible residues at the OPLS3 force field.

4. Results

4.1. Molecular dynamics simulations of N-protein CTD

Viral genome packaging is an essential step in viral replication. N-protein plays a vital role in stabilizing the protein-RNA assembly; therefore, targeting this stabilization can stop the viral replication. Performing MD simulation can help gather information about the interacting residues when the protein's active site is not well defined. Therefore, MD simulations of dimeric CTD (PDB: 7C22) apo form (unbound) and CTD dimer-RNA (6XEZ) bound form were performed for a span of 100ns to analyze the structural characteristics and protein folding. The trajectory analysis shows a stable RMSD value after 33ns in apo dimer and RNA-bound dimer (Fig. 1a). However, minor fluctuations were caught at the end, i.e., after 70ns in the dimer case with a 1.5 Å

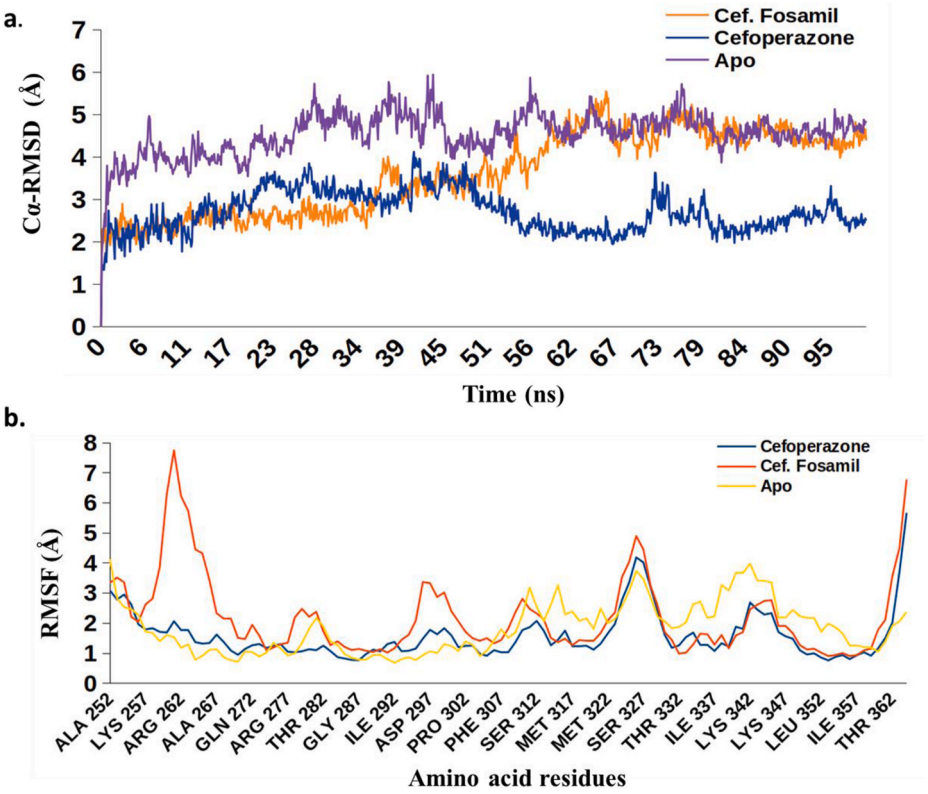

Fig. 4. The MD simulation comparative profile of N-CTD monomeric state highlighting the variation induced via inhibitory molecules in light of apo (unbound) CTD monomer. (a) Depicts C-alpha RMSD of CTD-inhibitor complexes i.e, Ceftarolinefosamil (3.62 Å**),** Cefoperazone (2.71 Å), and the unbound (apo) form (4.54 Å), and (b) C-alpha RMSF of residues.

RMSD value. The RMSF analysis of the above three protein structures depicts the amino acid residues involved in binding the nucleotide phosphate-backbone and dimer formation (Fig. 1b).

The results highlight several residues that help establish cross-linkages and form dimerization domains and aid in RNA-binding. As per Chang et al., 2005, CTD has a dimer developing capability with the cross-linking ability that includes the amino acids between positions 248–365 of SARS CoV-2 [36].

The crystal structure of N-CTD bound to a ligand was not available and the confirmation of the same was achieved with the MD analysis. Then the SiteMap module of Schrodinger was used to identify binding sites in the protein, which confirmed the same binding regions obtained from the MD analysis. Further, the targeted studies were designed with the same domain (Site I) residues obtained as obtained above (Fig. 2). Moreover, the quaternary structure reveals a peculiar binding groove when present in the dimerized-association, which is estimated to be the top rank binding site as per SiteMap output site-II (Fig. 3). Therefore, two studies were conducted simultaneously with distinct binding clefts (site I and site II) to observe different alterations in protein dynamics.

4.2. Ligand screening

After finalizing the binding sites through MD simulations and cross-checking them with SiteMap, virtual screening of approximately 2456 FDA-approved drugs was performed with the monomeric (site-I) and dimeric (site-II) binding sites. Top hits were selected based on the best docking scores and based on the MM/GBSA binding free energies the final scoring was done. The top hits were chosen based on the cut-off determined by the dimer assembly interaction energy, i.e., −45.1 kcal/mol.

The MD simulations of the protein-ligand complexes revealed essential characteristics of the domains participating in ligand binding in monomeric and dimeric forms. The N-CTD monomer unveils the dimerization domain requisite for stable dimer assembly, transforming into a binding grove for small molecules. The other binding site accessible in dimer-form attracts a different segment of ligands and exhibits substantial changes in N-CTD's conformational stability (Figs. 4 and 5).

4.3. The complexity and conformational dynamics of N-CTD

The protein assembly energetics revealed from EMBL-PISA tells us that chain A and chain B form complex with a free energy of $\Delta G^{int} = -39.1$ kcal/mol after the interaction. The same assembly's interaction with the RNA molecule bound at the polar interface is defined with −45.1 kcal/mol. The N-CTD structure experiences concerted stabilization upon RNA-binding. The C-alpha overlap of the apo N-CTD dimer and RNA-bound dimer in Fig. 6a & b depicts apo form (blue ribbons) and RNA-bound form (red ribbon), respectively. Fig. (6c) shows the stable N-CTD condition can be marked with uplifted beta-turn. The N-CTD structure experiences concerted stabilization upon RNA-binding. The rigorous residual-level analysis of the MD simulation trajectories was achieved with DCCM and PCA of the simulation entities. These figures provide a thorough understanding of the local structural changes that occur within major regions crucial for the dimer assembly when compared against the protein C-alpha RMSD, residue C-alpha fluctuations, and protein contact profiles respective ligands.

The RMSD and RMSF comparison plots of apo and RNA-bound form of N-CTD dimer highlights the local conformational changes like the structure is seen stabilizing after 23ns and further the RMSD declines after 76 ns (as low as 1.5 Å), whereas in case of APO higher values of RMSD are recorded at 76 ns (3–3.2 Å) towards the end. This depicts a generalized stabilization in the RNA bound protein form (Fig. 1). The RNA-bound form of N-CTD is essential to realize the primarily critical residues taking part in salt-bridge formation with the RNA; these

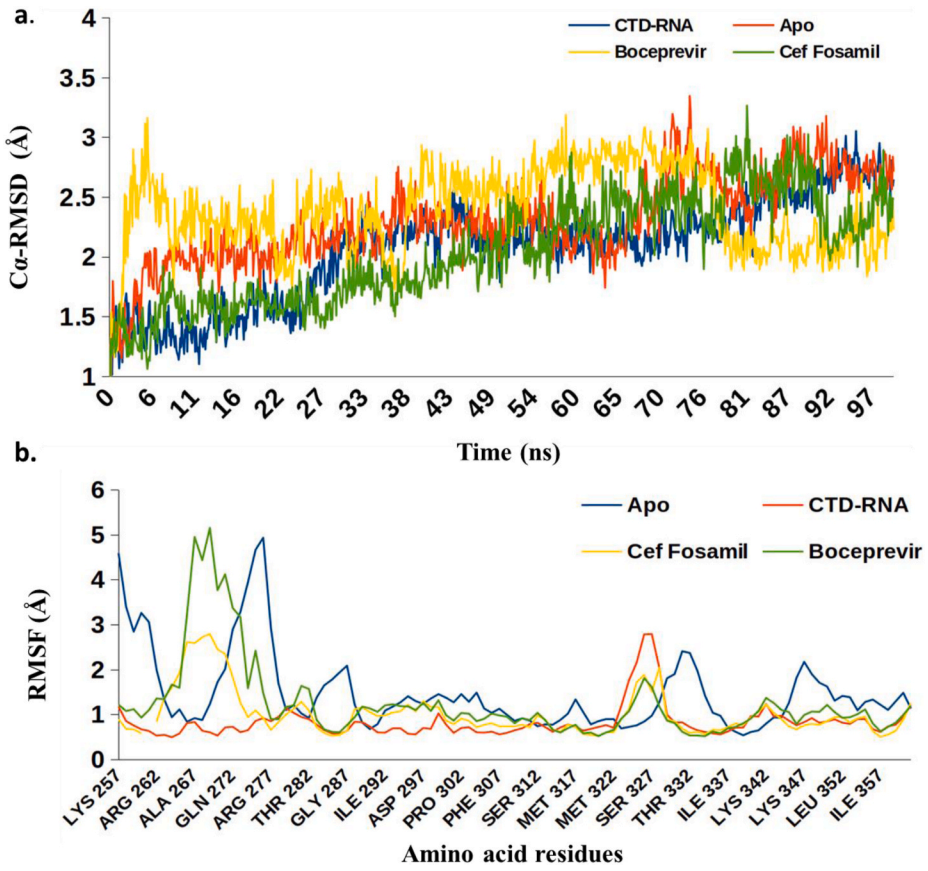

Fig. 5. The MD simulation comparative profile of N-CTD dimeric state highlighting the variation induced via inhibitory molecules in light of apo (unbound) CTD dimer. (a) Depicts C-alpha RMSD of CTD-inhibitor complexes i.e, Ceftarolinefosamil (2.05 Å**),** Boceprevir (2.39 Å), CTD-RNA complex (2.06 Å) and the unbound (apo) form (2.30 Å), and (b) C-alpha RMSF of residues.

residues are seen changing affinities for the posterior part of N-CTD upon RNA introduction (Fig. 7). The DCCM and PCA plots of apo and RNA-bound forms of N-CTD highlight the difference in correlated motions and the eigenvector values of the variation in N-CTD dynamics.

4.4. Docking/screening

The binding sites (Site-I and Site-II), obtained through MD analysis and SiteMap results were screened with an FDA-approved compound library. Targeting these active sites will lead to improper protein-RNA complex formation and stabilization, leading the reduced viral replication.

The site-I (monomeric) interacting residues include PRO326, SER327, ARG259, LYS261, and MET322. The ligands were screened with this active site, and the top six compounds were selected based on the docking score and MM/GBSA binding free energies (Table 1). The top two drugs of the 6 shortlisted candidates, Ceftaroline fosamil and Cefoperazone, were chosen as they possessed the highest binding free energies (−47.12 and −45.84 kcal/mol) (Table 1). MD analysis of The protein-drug complex was explored to get a clear insight into the structural alterations on inhibitor/drug binding such as their stability and protein-ligand contacts (Fig. 4). The enhanced side-chain interactions are seen in case of Ceftaroline fosamil at 7.8 Å and 5 Å (324-333a.a.); the RMSD plot stabilization corresponds to the apo form (60 ns, 5.0 Å). Cefoperazone is also observed interacting with the similar residues (324–333 a. a.) at 5 Å and it stabilizes after 50 ns (3.5 Å).

Similarly, in site-II (dimeric site), interacting residues include PRO326, SER327, THR325, VAL324, THR329, and GLY328 involved in the ligand-binding site are PHE274, ARG276, ARG277, ILE292, and LYS265. With this ligand-binding site, the FDA library was screened, and a total of seven compounds based on the docking score and MM/GBSA binding free energies were selected (Table 2). The top two drugs possessing the highest binding free energies, i.e., Boceprevir and Ceftaroline fosamil (−106.78 and −99.55 kcal/mol), were selected for further MD simulation study (Table 1 and 2). RMSF- The side chain involvement of Boceprevir and Ceftaroline fosamil shows enhanced contacts at the residue number LYS266 with further ligand contacts seen at ILE320, PRO326, SER327, GLY328, and VAL324. These similar residues are also seen participating in the dimer formation in Fig. 5. The Cα-RMSD is seen stabilizing after 50ns in all the protein states as compared to apo form in Fig. 5a. The highest structural fluctuations are recorded in the case of apo state with an average RMSD of 2.30 Å ± 0.36; a gradual structural instability is seen in the case of Ceftaroline fosamil and Boceprevir with an approximate RMSD of 2.05 Å ± 0.44 and 2.39 Å ± 0.34 respectively. Further, an overlap is also depicted between the 100ns and 0ns simulation structural change. Ceftaroline is seen interacting with protein residues; highly correlated regions were obtained via DCCM analysis (Fig. 8).

Ceftaroline fosamil and Cefoperazone at site-I showed the highest binding free energies (−47.12 and −45.84 kcal/mol); whereas, at site-II, Boceprevir and Ceftaroline fosamil (−106.78 and −99.55 kcal/mol) showed the highest binding free energies. In both the cases, Ceftaroline fosamil shows the highest binding free energy; the RMSF of Ceftaroline fosamil with the monomeric site is as high as 7.8 Å, and RMSD is seen stabilizing after 60 ns, whereas, in the case of the dimeric region, the RMSF peaks at 2.9 Å, and RMSD is seen fluctuating 2.05 Å ± 0.44 (Table 3).

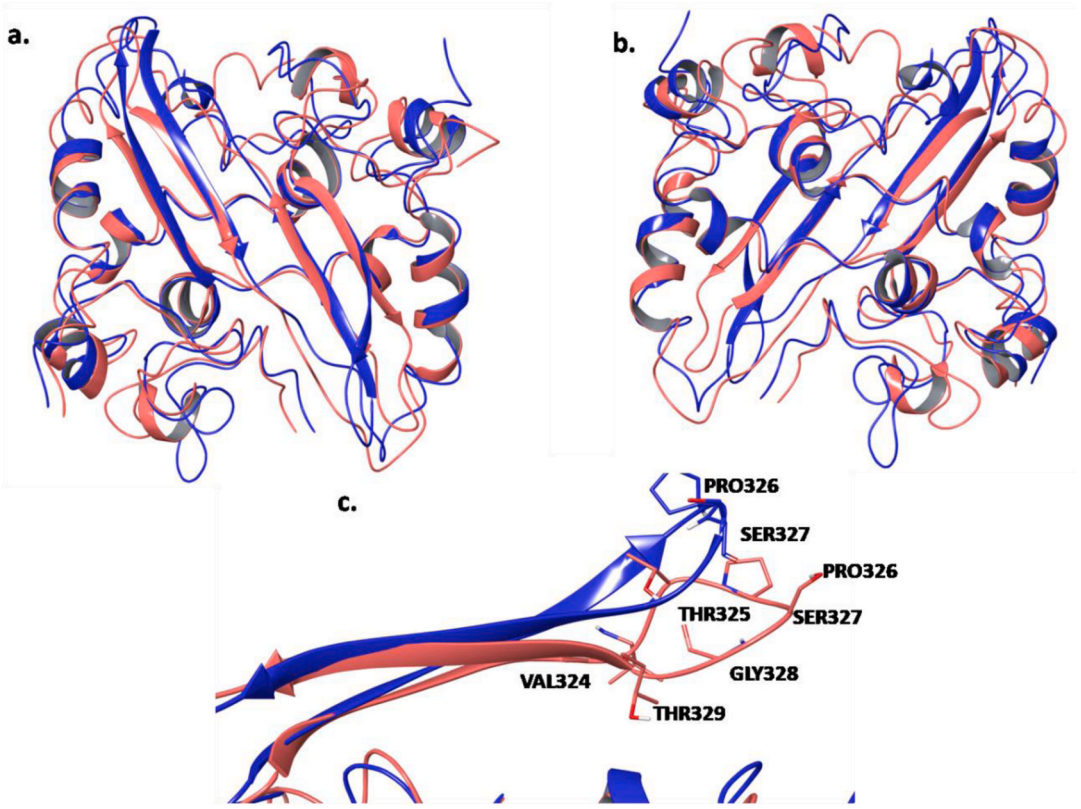

Fig. 6. The C-alpha overlay of N-CTD of SARS-CoV-2 in apo (blue) and RNA-bound (red) form. (a) Solvent-accessible domain overlap, (b) RNA-binding domain interface overlap (RMSD = 2.72 Å). (c) Deviation between the dimerization domain of RNA-bound (red) and unbound (blue) form of CTD (RMSD = 0.811 Å).

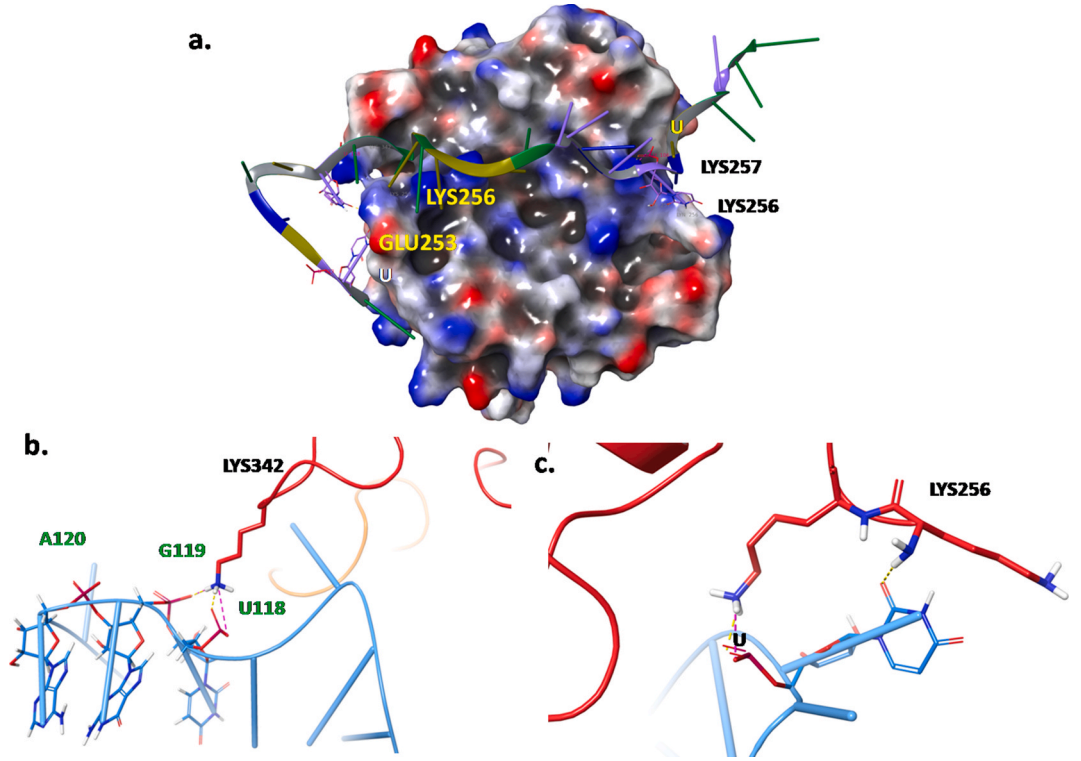

Fig. 7. RNA-bound N-CTD a. Surface diagram showing the residue property in form of electric potential exposing the polar RNA binding positive surface with spread out blue region, b & c. Highlight the extrinsic contacts between the N-CTD and RNA genome backbone. (Yellow-hydrogen bonding; Orange- van der Waals contact; Pink- Salt bridge).

Table 1
Docking score and MM/GBSA of drugs bound at Site-I.

Drugs	Docking score	MM/GBSA (kcal/mol)
Ceftaroline fosamil	−5.07	−47.12
Cefoperazone	−6.86	−45.84
Deferoxamine	−5.45	−42.33
Polymyxin B	−6.47	−31.99
Acyclovir	−5.89	−31.77
Imidurea	−5.05	−32.21

Table 2
Docking score and MM/GBSA of drugs bound at Site-II.

Drugs	Docking score	MM/GBSA (kcal/mol)
Boceprevir	−8.83	−106.78
Ceftaroline fosamil	−9.68	−99.55
Glycerol phenylbutyrate	−8.60	−95.69
Ioxilan	−8.66	−91.38
Pantethine	−8.37	−90.06
Riboflavin	−7.95	−85.80
Dinoprost	−8.78	−82.89
Fenoterol	−7.89	−60.01

5. Discussion

The N-CTD complexes into a homotetramer, which binds and surrounds the single-stranded RNA backbone, thus, protect the viral RNA from degradation by extracellular materials. The monomers assemble to form 2 pairs of dimers; if the monomer assembly is targeted, the dimer won't be created, and this can cause instability in the structure. These dimers are formed to possess a highly polar surface determined for nucleic acid-binding; therefore, if the dimer is targeted, the RNA assembly to the N-protein will be hindered. Various studies are conducted on the SARS CoV 2-N terminal domain; as earlier, it was assumed that only NTD has a crucial role in assembling the virion. Still, lately, researchers have explored the other terminal too [17,24], the mechanistic study on the molecular understanding of SARS CoV 2 pathogenesis trails behind. This study improvises the idea of simulating the nucleic acid genome of the virion particle with the N-CTD protein to observe and study the mechanical variations in the complex. This is encouraged by the idea that nucleic acid binding brings out communicative changes in the protein structure, driving it positively towards the stable complex formation. Targeting the distinct ribonucleotide-binding pocket will lead to an unstable complex making it prone to be attacked by the cellular nucleases.

The N-CTD has four chains, A, B, C, and D, of which A-B and C-D occur as a dimer and are joined by non-covalent and hydrophobic interactions to make a series of N-CTD capsomere unit. The RNA was aligned at the RNA binding interface, where the N-CTD binds to RNA phosphodiester bonded backbone with polar residues such as LYS342, LYS256, LYS261, and LYS257. This binding of the RNA stiffens the

Table-3
Comparison of Docking score, MM/GBSA, and RMSD±SD of the potential drugs according to Site-I and Site-II.

Protein state	Drugs	MM/GBSA (kcal/mol)	Dock score	RMSD±SD
Monomer	Ceftaroline fosamil	−62.312	−5.077	3.62 ± 0.96
	Cefoperazone	−52.421	−6.86	2.71 ± 0.48
Dimer	Ceftaroline fosamil	−106.16	−9.68	2.05 ± 0.44
	Boceprevir	−102.03	−8.83	2.39 ± 0.34

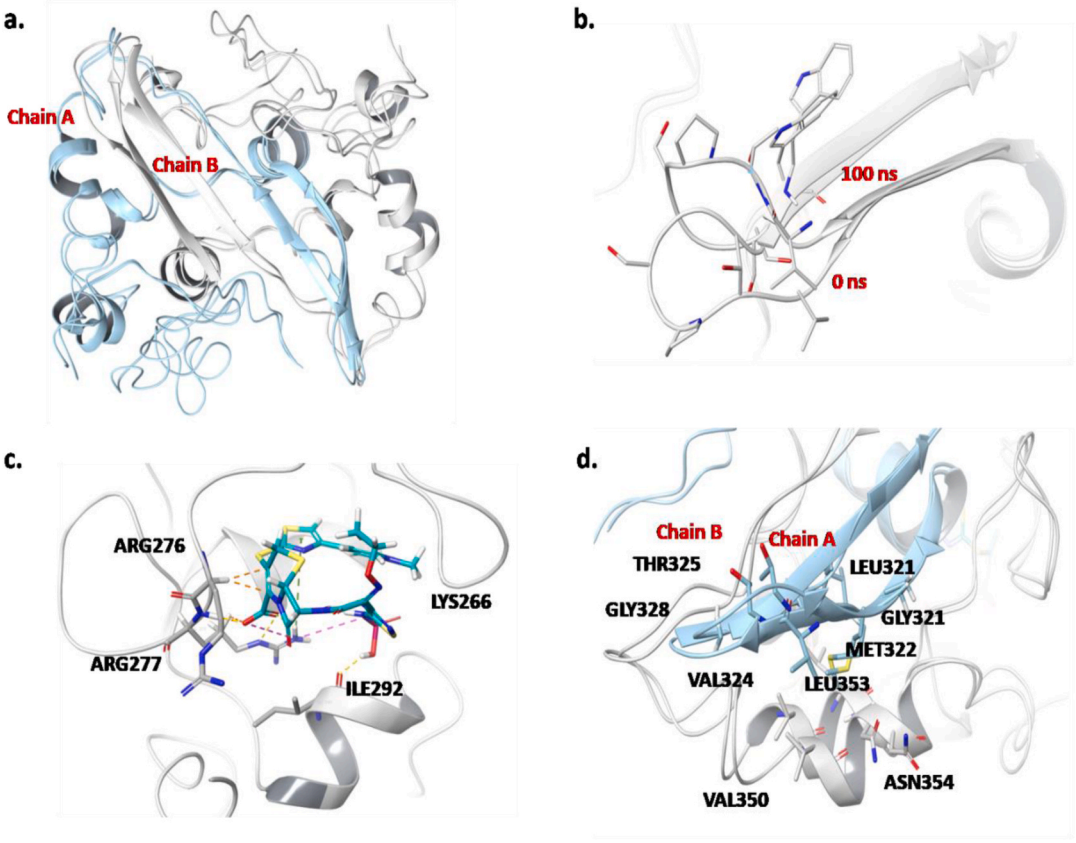

Fig. 8. Ceftaroline fosamil (DCCM) interactions with the protein residues at the site-II- a. Chain A& B overlap at 0 &100 ns, b. Zoomed in view, c & d. Residue interaction with Ceftaroline fosamil.(Yellow-hydrogen bonding; Orange- van der Waals contact; Pink- Salt bridge; Green dotted-pi-cation bonding).

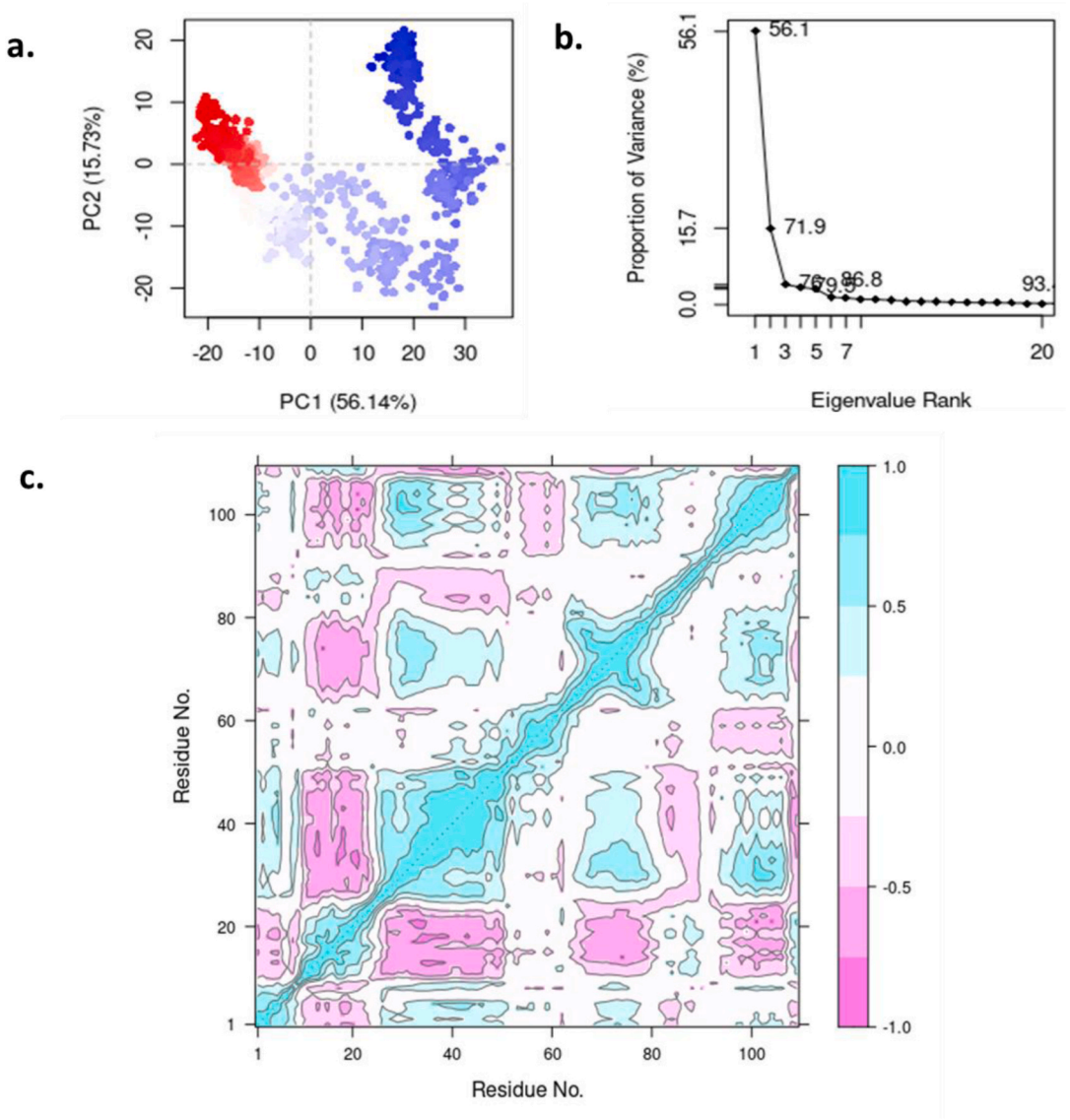

Fig. 9. Ceftaroline fosamil- (a) PCA plot showing the covariance shared between principal component 1 and 2 (PC1 & PC2) attributing the overall movement of N-CTD atom coordinates, where (b) shows the eigenvalues of these principal components. There are more regions with highly correlated motions (c) between the binding site residues 287–301, 267–277 and the dimer interaction region (a.a. 322–331) which is evident for the induced disassemble caused by Ceftaroline fosamil.

highly exposed structure of CTD like a closed wing structure. It is attributed to reduced RMSF of the RNA binding region, primarily ASN269-GLY278, leading to stronger binding affinity within the dimer assembly, chain A and B (−45.1 kcal/mol); and with the packed solvent accessible surface, with increased intermolecular non-covalent interactions.

The molecular dynamics and sitemap results of the N-CTD un-bound (apo) and bound (RNA & drug) form reveal the sites involved in dimer association and nucleotide-binding. Almost similar residues were seen interacting in previous studies; a structural survey of SARS CoV 2 N protein-CTD claims that 248–365 amino acids are involved in the dimer formation [36]. Further, considering the global confidence of accurate binding prediction, SiteMap shows 86% efficacy in large-scale validations [37]. Eliciting the subject, it is evident that the monomeric binding site's site score obtained in our experiment was 0.878, and the druggability score was 1.00 at 0 ns. In contrast, at 100 ns, the site score was 0.911, and the druggability score was also enhanced to 1.026. On the other hand, the dimeric site's site score was 0.881 and druggability score 1.003 at 0ns after simulation of 100 ns, the site score was 0.903, and druggability score also improved to 1.011. This shows that at the stage of protein structure equilibration i.e., after running an MD simulation of 100 ns, the binding site score and druggability score were significantly improved. The same is reflected by the critical profile of the candidate molecules above.

The dimer interface is made up of two consistent anti-parallel beta-sheets side-lined by one helix. This beta-sheet flap interacts with the extended random loop region of the complementary chain to make lid-like anchorage. Keeping this simple organization, to sum up, N-CTD's structure, it can be added that the N-CTD's anterior face has an enormous polar surface area (PSA). That determines its binding to the negatively charged backbone of the RNA molecule, and the posterior part is solvent accessible. It provides a barrier from direct exposure of nucleic acid to the adjacent solvating environment.

Therefore, two simultaneous screenings were conducted to observe the changing assembly of the protein and based on docking scores, MM/GBSA energy, and protein-interaction energy, the ligands were selected for further analysis.

The monomeric state revealing the dimer interlocking interface was also intensely occupied by Cefoperazone and Ceftaroline fosamil; both are cephalosporin antibiotics. The extended loop region from ASN354 to

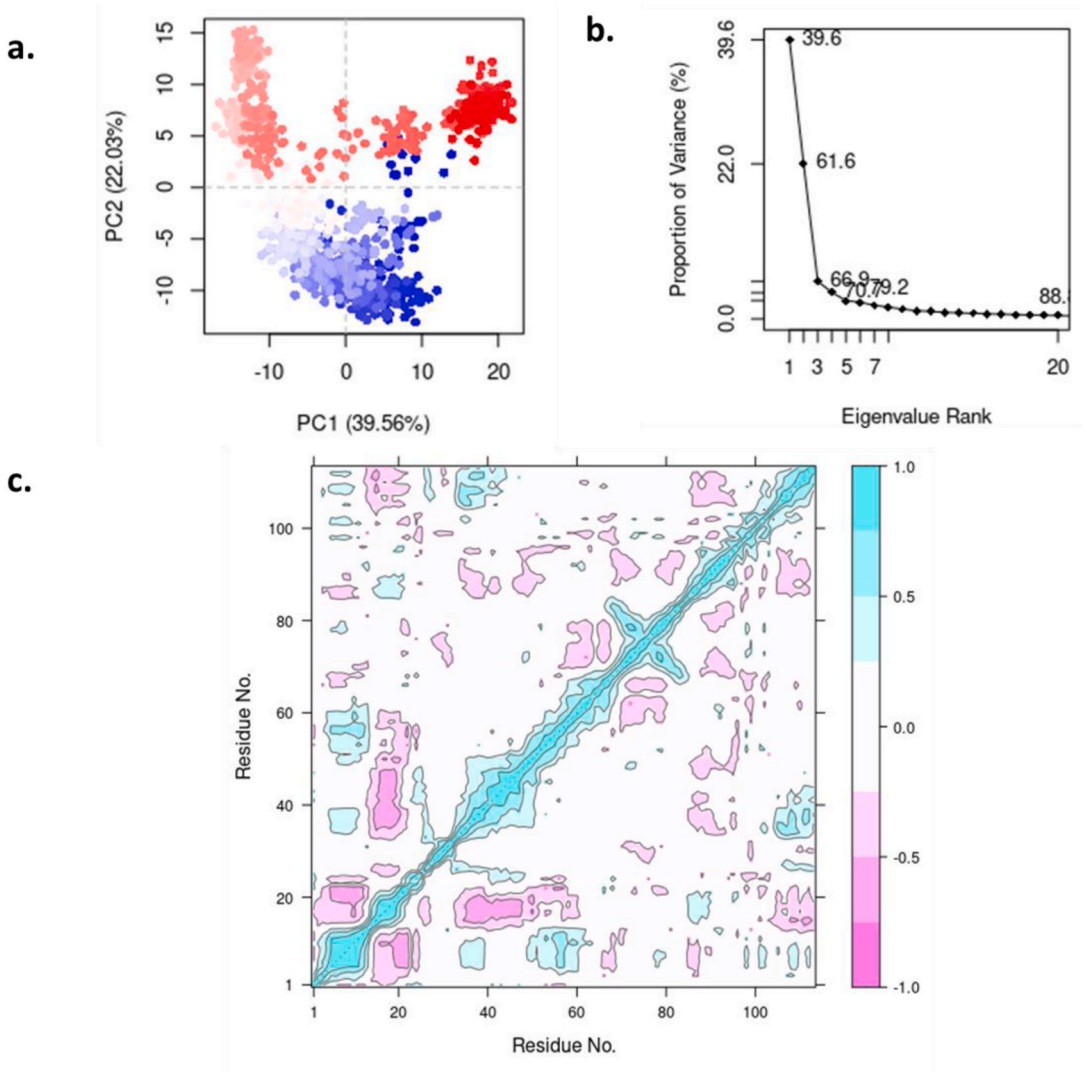

Fig. 10. Boceprevir- The PCA plot (a) and (b) shows poor covariance clustering and lower value eigenvectors. It can be seen in the DCCM plot (c) there are few regions with fairly correlated motions but are not strong enough to extrapolate the induced dynamics.

THR362, which partly constitutes the RNA binding residues and dimerization beta-ribbon from GLY321 to LEU327, making up the dimer interface 'site-I.' This site was seen well blocked by Ceftaroline fosamil & Cefoperazone with a binding affinity of −47.12 and −45.84 kcal/mol respectively. The binding to this site was attributed to stabilized complex formation Cefoperazone (2.6 Å) and Ceftaroline fosamil (4.5 Å) and stronger binding to crucial residues such as ARG259, MET322, and SER327, which are necessary for dimer association.

Ligand binding at site II alters the protein impeding a distorted and loose dimerization upon Boceprevir binding, an antiviral drug used to treat chronic hepatitis C virus. The free energy of assembly formation between chains A & B was −38.0 kcal/mol and exhibited higher energy at site II. This can be deduced from Fig. (5a); Boceprevir significantly accentuates the fluctuation of RNA binding region residues, i.e., LYS266 to GLY278. Contrastingly, Ceftaroline fosamil distorts the dimer and increases its binding energy to −37.1 kcal/mol, with more allosteric changes within the CTD protein, leading to loose dimer formation.

The covariance matrix analysis of the N-CTD protein dynamics trajectory displays correlated motions of residues behaving in a dramatically different manner at various conditions. While Boceprevir shows stronger occupancy of RNA binding residue, it exhibits weak CTD modulation. On the contrary, Ceftaroline fosamil-bound CTD shows highly correlated and distant regions. The residues 268 to 278 (site II) indirectly affect the movements of residue number 321 to 336, which forms the interlocking dimerization domain in Fig. 8c and d. Additionally, a notable correlation was observed for site II with C-terminal residues actively taking part in RNA binding; hence unfolding the close association affirms the high binding affinity of Ceftaroline fosamil with CTD site II residues. The drastic difference of conformations can be spotted in the PCA plot (Figs. 9 and 10) in the form of distinct clusters as a function of time. The first two PCs (PC1 & PC2) express 71.9% of the variation in the protein's conformation ensembles upon binding with Ceftaroline fosamil. Boceprevir also constitutes brilliant 61.6% variation in CTD confirmation ensembles in PC1 & PC2. Ceftaroline fosamil brings the structural modulation in Chain B where it is bound and attenuates the beta-ribbon interlocking region in Chain-A Fig. (8d). Appreciate how the beta-sheet loses its complexity and turns into a loop, losing its native secondary structure, highlighting its weakened interchain tertiary structure. This study highlights N-CTD's functional importance in the virulence of CTD and its structural complexity, which harmonizes the effective binding of CTD with the viral genome. Thus, it is postulated that targeted drug treatment with appropriate drug molecules like Cefoperazone, Boceprevir, and Ceftaroline fosamil can decrease the N-CTD dimerization rate.

Moreover, this study is directed towards discovering particular binding sites that bring noteworthy changes affecting the whole protein

dynamics. As in the case of site II, hence in the present study, such allosteric changes must be studied with denoted concern as these sites are highly influencing when it comes to protein's overall packaging and assembly. Earlier, various experimental studies are conducted that justify the binding sites or clarify the protein structure. Takeda et al., 2008 performed the study to explore the solution structure of the SARS-CoV NP CTD by NMR using the stereo-array isotope labeling (SAIL) method. They also employed NMR, mutation analyses, and electrophoretic mobility shift assays (EMSAs) to determine the RNA-binding site, but they did not specifically address druggable active sites. The amino acid residues that were mutated at the CTD include R320A and H335A, significant chemical shift changes were observed leading to loss of binding affinity between the protein and nucleotide and thus concluding that the N-CTD- R320, H335, and A337 residues were involved in binding RNA particularly [18]. Therefore, our study can be further analyzed by various experimental techniques and can be proved as a boon in targeting the deadly coronavirus.

6. Conclusion

The outbreak of COVID-19 has posed a severe threat to public health. Therefore, new targets and their inhibitors are necessary to combat this pandemic. We propose a new target, i.e., the C-terminal domain of the nucleocapsid protein of SARS CoV-2. It plays a significant role in stabilizing the protein-nucleotide assembly and protecting the viral genome from the host's machinery; thus, targeting this can stop viral replication and spread. In this study, the computer-aided analysis was exploited to detect two active sites, Site-I or the monomeric site involved in the initial assembly of the protein complex, and Site-II or the dimeric site engaged in RNA-binding and protein-nucleotide stabilization. Secondly, potential FDA-approved inhibitors were screened and identified against these two sites. According to Site-I, Ceftaroline fosamil, and Cefopaerazone (cephalosporin antibiotic) are the drugs of choice. And as per Site-II or to target the RNA association with protein, Ceftaroline fosamil and Boceprevir (hepatitis C virus protease inhibitor) are the top compounds possessing the highest binding free energies and positively related dynamic cross-correlation. However, Ceftaroline fosamil, an antibacterial drug, might have the ability to target the assembly of protein initially and at the RNA-binding stage since it remained at the top when screened with both the active sites. Thus, it can be concluded that Ceftaroline fosamil and Boceprevir can bind specifically to the dimeric site. These can play vital role and can act as potential antivirals in inhibiting the RNA-binding, preventing further the virus replication and translation.

Data and software availability

The X-RAY crystallized structure of the nucleocapsid protein chosen was PDB-ID (7C22) (https://www.rcsb.org/structure/7C22). The protein preparation, modeling, molecular docking simulations operations were performed on Maestro (release 2020-4), Schrödinger, LLC, New York, NY, the USA. The FDA-approved ligand library was retrieved from https://go.drugbank.com/releases/5-1-8/downloads/all-full-database. MM/GBSA was calculated using trajectory analysis with the script *thermal_mmgbsa.py* available from https://www.schrodinger.com/scriptcenter. PDBePISA v.1.48 (open source) was used to explore the macromolecular interfaces and calculate the protein assembly interaction energyhttps://www.ebi.ac.uk/pdbe/pisa/. Using the Bio3d v.2.4–1.9000 (http://thegrantlab.org/bio3d/) package in R studio v1.4.1103, MD trajectories of protein-ligand complexes was obtained by examining the Dynamical Cross-Correlation Matrix (DCCM) and Principal Component Analysis (PCA) (https://rstudio.com/products/rstudio/download-commercial-desktop/).

Declaration of competing interest

All the authors declare no conflict of interest.

Acknowledgement

The authors thank Dr. Prajwal Nandekar and Mr. Vinod Devaraji of Schrodinger Corporation for their kind help and Mr. Nripendra Bhatta for his logistic help.

References

[1] D. Marcatelli, F.M. Giorgi, Geographic and genomic distribution of SARS-CoV-2 mutations, Front. Microbiol. 11 (2020) 1800.
[2] S.G. Siddell, P.J. Walker, E.J. Lefkowitz, A.R. Mushegian, M.J. Adams, B.E. Dutilh, GorbalenyaAE, B. Harrach, R.L. Harrison, S. Junglen, N.J. Knowles, Additional changes to taxonomy ratified in a special vote by the International Committee on Taxonomy of Viruses (October 2018), Arch. Virol. 164 (3) (2019 Mar 1) 943–946.
[3] J.D. Berry, S. Jones, M.A. Drebot, A. Andonov, M. Sabara, X.Y. Yuan, H. Weingartl, L. Fernando, P. Marszal, J. Gren, B. Nicolas, Development and characterisation of neutralising monoclonal antibody to the SARS-coronavirus, J. Virol Methods 120 (1) (2004 Sep 1) 87–96.
[4] A.A. Elfiky, Ribavirin, Remdesivir, Sofosbuvir, Galidesivir, and Tenofovir against SARS-CoV-2 RNA dependent RNA polymerase (RdRp): a molecular docking study, Life Sci. 253 (2020 Jul 15) 117592.
[5] C. Liu, Q. Zhou, Y. Li, L.V. Garner, S.P. Watkins, L.J. Carter, J. Smoot, A.C. Gregg, A.D. Daniels, S. Jervey, D. Albaiu, Research and development on therapeutic agents and vaccines for COVID-19 and related human coronavirus diseases, ACS Cent. Sci. 6 (3) (2020) 315–331, https://doi.org/10.1021/acscentsci.0c00272.
[6] T.K. Boehmer, J. DeVies, E. Caruso, K.L. van Santen, S. Tang, C.L. Black, K. P. Hartnett, A. Kite-Powell, S. Dietz, M. Lozier, A.V. Gundlapalli, Changing age distribution of the COVID-19 pandemic—United States, May–August 2020, MMWR (Morb. Mortal. Wkly. Rep.) 69 (39) (2020 Oct 2) 1404.
[7] M. Prajapat, P. Sarma, N. Shekhar, P. Avti, S. Sinha, H. Kaur, S. Kumar, A. Bhattacharyya, H. Kumar, S. Bansal, B. Medhi, Drug targets for corona virus: a systematic review, Indian J. Pharmacol. 52 (1) (2020 Jan) 56.
[8] P. Sarma, M. Prajapat, P. Avti, H. Kaur, S. Kumar, B. Medhi, Therapeutic options for the treatment of 2019-novel coronavirus: an evidence-based approach, Indian J. Pharmacol. 52 (1) (2020 Jan) 1.
[9] M. Prajapat, P. Sarma, N. Shekhar, A. Prakash, P. Avti, A. Bhattacharyya, H. Kaur, S. Kumar, S. Bansal, A.R. Sharma, B. Medhi, Update on the target structures of SARS-nCoV-2: a systematic review, Indian J. Pharmacol. 52 (2) (2020 Mar 1) 142.
[10] Pramod Avti, Arushi Chauhan, Nishant Shekhar, Manisha Prajapat, Phulen Sarma, Hardeep Kaur, Anusuya Bhattacharyya, Subodh Kumar, Ajay Prakash, Saurabh Sharma, Bikash Medhi, Computational basis of SARS-CoV 2 main protease inhibition: an insight from molecular dynamics simulation based findings, Journal of Biomolecular Structure and Dynamics (2021), https://doi.org/10.1080/07391102.2021.1922310.
[11] C. Drosten, S. Günther, W. Preiser, S. Van Der Werf, H.R. Brodt, S. Becker, H. Rabenau, M. Panning, L. Kolesnikova, R.A. Fouchier, A. Berger, Identification of a novel coronavirus in patients with severe acute respiratory syndrome, N. Engl. J. Med. 348 (20) (2003 May 15) 1967–1976.
[12] P.S. Masters, L.S. Sturman, Background paper functions of the coronavirus nucleocapsid protein, in: Coronaviruses and Their Diseases, Springer, Boston, MA, 1990, pp. 235–238.
[13] Y. Cong, F. Kriegenburg, C.A. De Haan, F. Reggiori, Coronavirus nucleocapsid proteins assemble constitutively in high molecular oligomers, Sci. Rep. 7 (1) (2017 Jul 18), 1-0.
[14] G.W. Nelson, S.A. Stohlman, S.M. Tahara, High affinity interaction between nucleocapsid protein and leader/intergenic sequence of mouse hepatitis virus RNA, J. Gen. Virol. 81 (1) (2000 Jan 1) 181–188, https://doi.org/10.1099/0022-1317-81-1-181.
[15] J. Mu, XuJ, L. Zhang, T. Shu, D. Wu, M. Huang, Y. Ren, X. Li, Q. Geng, Y. Xu, Y. Qiu, SARS-CoV-2-encoded nucleocapsid protein acts as a viral suppressor of RNA interference in cells, Sci. China Life Sci. (2020 Apr 10) 1–4, https://doi.org/10.3390/v6082991.
[16] Q. Huang, L. Yu, A.M. Petros, A. Gunasekera, Z. Liu, N. Xu, P. Hajduk, J. Mack, S. W. Fesik, E.T. Olejniczak, Structure of the N-terminal RNA-binding domain of the SARS CoV nucleocapsid protein, Biochemistry 43 (20) (2004 May 25) 6059–6063.
[17] C.K. Chang, M.H. Hou, C.F. Chang, C.D. Hsiao, T.H. Huang, The SARS coronavirus nucleocapsid protein–forms and functions, Antivir. Res. 103 (2014 Mar 1) 39–50, https://doi.org/10.1016/j.antiviral.2013.12.009.
[18] M. Takeda, C.K. Chang, T. Ikeya, P. Güntert, Y.H. Chang, Y.L. Hsu, T.H. Huang, M. Kainosho, Solution structure of the c-terminal dimerization domain of SARS coronavirus nucleocapsid protein solved by the SAIL-NMR method, J. Mol. Biol. 380 (2008 Jul 18) 608–622.
[19] I.M. Yu, M.L. Oldham, J. Zhang, J. Chen, Crystal structure of the severe acute respiratory syndrome (SARS) coronavirus nucleocapsid protein dimerization domain reveals evolutionary linkage between corona-and arteriviridae, J. Biol. Chem. 281 (25) (2006 Jun 23) 17134–17139.
[20] C.Y. Chen, C.K. Chang, Y.W. Chang, S.C. Sue, H.I. Bai, L. Riang, C.D. Hsiao, T. H. Huang, Structure of the SARS coronavirus nucleocapsid protein RNA-binding

dimerization domain suggests a mechanism for helical packaging of viral RNA, J. Mol. Biol. 368 (4) (2007 May 11) 1075–1086.
[21] X. Hu, Z. Zhou, F. Li, Y. Xiao, Z. Wang, J. Xu, F. Dong, H. Zheng, R. Yu, The study of antiviral drugs targeting SARS-CoV-2 nucleocapsid and spike proteins through large-scale compound repurposing, Heliyon (2021 Mar 1), e06387.
[22] B.W. Neuman, Bioinformatics and functional analyses of coronavirus non-structural proteins involved in the formation of replicative organelles, Antivir. Res. 135 (2016 Nov 1) 97–107.
[23] C.K. Chang, S.C. Sue, T.H. Yu, C.M. Hsieh, C.K. Tsai, Y.C. Chiang, S.J. Lee, H. H. Hsiao, W.J. Wu, C.F. Chang, T.H. Huang, The dimer interface of the SARS coronavirus nucleocapsid protein adapts a porcine respiratory and reproductive syndrome virus-like structure, FEBS Lett. 579 (25) (2005 Oct 24) 5663–5668.
[24] P.K. Hsieh, S.C. Chang, C.C. Huang, T.T. Lee, C.W. Hsiao, Y.H. Kou, I.Y. Chen, C. K. Chang, T.H. Huang, M.F. Chang, Assembly of severe acute respiratory syndrome coronavirus RNA packaging signal into virus-like particles is nucleocapsid dependent, J. Virol. 79 (22) (2005 Nov 15) 13848–13855.
[25] L. Kuo, C.A. Koetzner, P.S. Masters, A key role for the carboxy-terminal tail of the murine coronavirus nucleocapsid protein in coordination of genome packaging, Virology 494 (2016 Jul 1) 100–107.
[26] M. Prajapat, N. Shekhar, P. Sarma, P. Avti, S. Singh, H. Kaur, A. Bhattacharyya, S. Kumar, S. Sharma, A. Prakash, B. Medhi, Virtual screening and molecular dynamics study of approved drugs as inhibitors of spike protein S1 domain and ACE2 interaction in SARS-CoV-2, J. Mol. Graph. Model. 101 (2020 Dec 1) 107716.
[27] N. Shekhar, P. Sarma, M. Prajapat, P. Avti, H. Kaur, A. Raja, H. Singh, A. Bhattacharya, S. Sharma, S. Kumar, A. Prakash, In silico structure-based repositioning of approved drugs for spike glycoprotein S2 domain fusion peptide of SARS-CoV-2: rationale from molecular dynamics and binding free energy calculations, mSystems 5 (5) (2020 Oct 27).
[28] S. Kumar, P. Sarma, H. Kaur, M. Prajapat, A. Bhattacharyya, P. Avti, N. Sehkhar, H. Kaur, S. Bansal, S. Mahendiratta, V.M. Mahalmani, Clinically relevant cell culture models and their significance in isolation, pathogenesis, vaccine development, repurposing and screening of new drugs for SARS-CoV-2: a systematic review, Tissue Cell (2021 Jan 26) 101497.
[29] S. Kang, M. Yang, Z. Hong, L. Zhang, Z. Huang, X. Chen, S. He, Z. Zhou, Z. Zhou, Q. Chen, Y. Yan, Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites, Acta Pharm. Sin. B (2020 Apr 20).
[30] P. Sarma, N. Shekhar, M. Prajapat, P. Avti, H. Kaur, S. Kumar, S. Singh, H. Kumar, A. Prakash, D.P. Dhibar, B. Medhi, In-silico homology assisted identification of inhibitor of RNA binding against 2019-nCoV N-protein (N terminal domain), J. Biomol. Struct. Dyn. (2020 May 16) 1–9.
[31] E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state, J. Mol. Biol. 372 (3) (2007 Sep 21) 774–797.
[32] B.J. Grant, A.P. Rodrigues, K.M. ElSawy, J.A. McCammon, L.S. Caves, Bio3d: an R package for the comparative analysis of protein structures, Bioinformatics 22 (21) (2006 Nov 1) 2695–2696, https://doi.org/10.1093/bioinformatics/btl461.
[33] SiteMap SR. 4: Version 3.4, Schrödinger, LLC, New York, NY, 2015.
[34] T. Halgren, New method for fast and accurate binding-site identification and analysis, Chem. Biol. Drug Des. 69 (2) (2007 Feb) 146–148.
[35] T.A. Halgren, Identifying and characterizing binding sites and assessing druggability, J. Chem. Inf. Model. 49 (2) (2009 Feb 23) 377–389.
[36] C.K. Chang, S.C. Sue, T.H. Yu, C.M. Hsieh, C.K. Tsai, Y.C. Chiang, S.J. Lee, H. H. Hsiao, W.J. Wu, C.F. Chang, T.H. Huang, The dimer interface of the SARS coronavirus nucleocapsid protein adapts a porcine respiratory and reproductive syndrome virus-like structure, FEBS Lett. 579 (25) (2005 Oct 24) 5663–5668.
[37] T.A. Halgren, Identifying applications, and recent advances of protein-ligand docking in structure-based drug design, J. Chem. Inf. Model. 49 (2009) 377–389.