REVIEW

Biophysical analysis of SARS-CoV-2 transmission and theranostic development via N protein computational characterization

Godfred O. Sabbih[1] 🆔 | Maame A. Korsah[2] | Jaison Jeevanandam[3] |
Michael K. Danquah[1] 🆔

[1]Department of Chemical Engineering, University of Tennessee, Chattanooga, Tennessee

[2]Department of Mathematics, University of Tennessee, Chattanooga, Tennessee

[3]CQM - Centro de Química da Madeira, MMRG, Universidade da Madeira, Campus da Penteada, Funchal, Portugal

Correspondence
Michael K. Danquah, Department of Chemical Engineering, University of Tennessee, Chattanooga, TN 37403.
Email: michael-danquah@utc.edu

Abstract

Recently, SARS-CoV-2 has been identified as the causative factor of viral infection called COVID-19 that belongs to the zoonotic beta coronavirus family known to cause respiratory disorders or viral pneumonia, followed by an extensive attack on organs that express angiotensin-converting enzyme II (ACE2). Human transmission of this virus occurs via respiratory droplets from symptomatic and asymptomatic patients, which are released into the environment after sneezing or coughing. These droplets are capable of staying in the air as aerosols or surfaces and can be transmitted to persons through inhalation or contact with contaminated surfaces. Thus, there is an urgent need for advanced theranostic solutions to control the spread of COVID-19 infection. The development of such fit-for-purpose technologies hinges on a proper understanding of the transmission, incubation, and structural characteristics of the virus in the external environment and within the host. Hence, this article describes the development of an intrinsic model to describe the incubation characteristics of the virus under varying environmental factors. It also discusses on the evaluation of SARS-CoV-2 structural nucleocapsid protein properties via computational approaches to generate high-affinity binding probes for effective diagnosis and targeted treatment applications by specific targeting of viruses. In addition, this article provides useful insights on the transmission behavior of the virus and creates new opportunities for theranostics development.

KEYWORDS

COVID-19, machine learning, molecular dynamics, nucleocapsid proteins, SARS-CoV-2

1 | INTRODUCTION

In the first quarter of the year 2020, World Health Organization (WHO) declared COVID-19 infection as a global pandemic due to its rapid spread across countries, increasing 13-fold outside of China in the first 2 weeks before the announcement (WHO virtual press conference on COVID-19, March 11, 2020). COVID-19 infection is caused by SARS-CoV-2, which is a viral strain belonging to the family *Coronaviridae* and genus *Betacoronavirus*. Other subfamily members,

include alpha, gamma, and delta coronaviruses with the alpha and beta coronaviruses infecting only mammals, usually causing respiratory illness, gastroenteritis in other animals, and extensive attacks on organs that express angiotensin-converting enzyme II (ACE2), such as the heart, liver, testis, kidney, and intestines.[1-3] *Betacoronaviruses* have four distinct lineages that are identified and designated from A to D based on the amino acid (AA) sequence alignment analysis, and SARS-CoV-2 belongs to lineage C. Further, members of this family share common characteristics, such as a unimolecular, positive-stranded

RNA genome, an RNA-binding nucleocapsid protein (N), a homo-trimeric spike protein (S), an outer membrane glycoproteins (M), and a small pentameric membrane protein called the envelope protein (E).[4]

SARS-CoV-2 is the latest of the two previously identified zoonotic coronaviruses, such as SARS-CoV and MERS-CoV, that are belonging to the beta-subfamily. Recent phylogenetic analysis showed that SARS-CoV-2 is about 88% identical to two bat-derived SARS-like coronaviruses (bat-SLCoVZC45 and bat-SL-CoVZXC21), about 50% identical to MERS-CoV, and about 79% identical to SARS-CoV.[5,6] SARS-CoV-2 transmission occurs through the release of respiratory droplets from symptomatic and asymptomatic patients after sneezing or coughing into the environment. These droplets are aerosol-like and capable of staying in the air or surfaces for an extended period and can be transmitted to persons through inhalation or contact with contaminated surfaces.[7] The virus makes its way through the eyes, nose, or mouth and attaches itself to the mucous membrane for incubation, multiplies and then reaches the pulmonary system as well as other body organs, which have a high percentage of ACE2, especially in the lower respiratory system.[8] The virus begins its life cycle when the spike protein binds to the ACE2 cellular receptor, after which conformational changes in the spike protein promotes viral fusion through the endosomal pathway. Although the exact incubation period of the virus is unknown, current reported research indicates a range of about 2.1–24 days[9] upon contact with an infected person. Furthermore, SARS-CoV-2 releases its RNA into the host cell followed by translation of the genome RNA open reading frames 1ab (ORF1ab) into viral replicase polyproteins (pp1a and 1ab), which are then cleaved by viral proteinases.[4] The polymerase produces mRNAs, which are later translated into relevant proteins. These viral proteins and the genomic RNA are assembled into virions and released out of the host cell via vesicles.[2,10] The same pathway is observed for SARS-CoV.[6] However, SARS-CoV possesses a lower binding affinity to ACE2. This higher binding affinity of SARS-CoV-2 is attributed to a unique ACE2-interacting residue (Lys417), which increases the affinity via the formation of salt-bridge interactions, with ACE2 contributing to an overall positive surface potential of S protein which further increases the binding affinity.[11] Since the outbreak of SARS-CoV in 2011, and the recent emergence of SARS-CoV-2, there has been a significant interest in probing the transmission and structural characteristics of beta coronaviruses. While SARS-CoV-2 and SARS-CoV share similar structures, their characteristics differ in certain aspects and hence, there is a need for an in-depth study to better understand the unique transmission and infection characteristics of SARS-CoV-2. The high transmission and mortality rates of SARS-CoV-2 have catalyzed research interests in gaining the knowledge about dynamic nature of the virus, its spread among individuals, and replication in a host.

Mathematical models that characterize these viral attributes with high precision will aid in decision-making, such as better healthcare interventions. Li et al. developed a numerical model that describes the spatiotemporal elements of infections among 375 Chinese urban areas and explored measures, which slows down the spread of SARS-CoV-2.[12] Several transmission models, based on the basic susceptible-infected-recovered (SIR) compartmental model, have been proposed

to investigate and estimate the transmission dynamics of SARS-CoV-2.[12-14] For example, based on the model, initially reported estimates of the basic reproductive number, R_0 has been reviewed to acquire precise estimates of the spread.[10] They reported the estimated mean of the basic reproduction number, R_0 for SARS-CoV-2 as approximately 3.28, which is higher than the estimate provided by the WHO ($R_0 \sim 1.95$). Thus, the study concluded that the differences in the estimates may have resulted from insufficient data and the short onset time available for the calculation of the previous estimate. Also, a significant number of transmission models are based on SEIR compartmental model which represents the flow of individuals in the susceptible, exposed, infected, and recovered compartments. The SEIR model has been utilized to investigate the adequacy of quarantine and social distance mediations, while others have changed the model to discover analytical and numerical results to demonstrate that SARS-CoV-2 would stay endemic[14,15]. Most of the current models developed assume post-infection immunity, thus, the SEIR model is highly beneficial in the prediction of effective lockdown measures and immunity among patients after infection. Although, certain clinical diagnosis and model studies suggest that immunity gained after recovery may be short-lived and reinfection may occur within a year, there is currently limited evidence to support post-infection immunity[16]. In the analysis presented in this article, we assume the nonexistence of post-infection immunity and introduced the susceptible-exposed-infected-susceptible (SEIS) compartmental model. This model mimics the SEIR model except that individuals are moved back to the susceptible compartment after recovery, and this is important to understand the dynamics of transmission at full scale. Further, the article discusses the structural characteristics of the SARS-CoV-2 N protein with RNA-binding domains, and its role in interfering with the normal reproductive cycle of the host cell as well as participating in replication, transcription, and packaging of the viral genome.[6,17] Unlike the other structural proteins, the N protein is highly conserved and has been shown to be mostly expressed at the initial stages of the viral infection,[18] serving as a significant target for the development of theranostics to identify and treat the infection in the early stage.

2 | SEIS TRANSMISSION MODEL FOR COVID-19 INFECTION

Most epidemiological models are generated from the general deterministic SEIR model. This model comprises of four compartments, namely susceptible, exposed, infected, and recovered/removed. Other models can be obtained from SEIR under certain parametric restrictions. In the limiting case where recovery from infection confers no immunity, the R compartment is removed resulting in the SEI or SEIS model. Here, the infected individual either return to being susceptible. This model is also used, when the average period of immunity approaches zero. The basic form of the SEIS model is the SIS model. However, the SEIS model assumes that an individual who is susceptible would initially be latent before getting infectious, unlike the SIS model, as shown in Figure 1. The SIS model becomes an
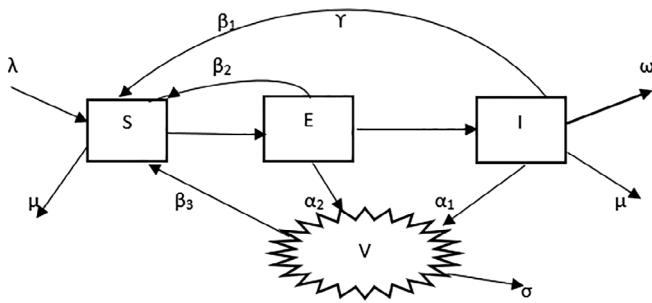
FIGURE 1   A COVID-19 SEIS (Susceptible-Exposed-Infected-Susceptible) transmission model

approximation of the SEIS model, when there is a short latent period. The SEIS framework is a well-known deterministic model as it considers the infectious force in the inert period between the phase of infection and the initiation of infectiousness. This transmission model has been used in the analysis of infectious diseases, such as gonorrhea, Nipa virus, and SARS.[13,19,20] The model describes the movement of individuals in a population within each of the aforementioned compartments that are interlinked by flows of different orders. Each compartment is well-defined, comprising of individuals, who flow into other compartments, following strict principles, that are set for each compartment. Individuals may move into the population under study by birth or immigration and can be removed from the population by death or emigration. In the model, each compartment is represented as a differential equation. The proposed SEIS COVID-19 transmission model with the virus compartment V is depicted by the accompanying system of differential equations as shown in the following equations:

$$\frac{dS}{dt} = \lambda - \beta_1 SI - \beta_2 SE - \beta_3 SV - \mu S + \gamma I \tag{1}$$

$$\frac{dE}{dt} = \beta_1 SI + \beta_2 SE + \beta_3 SV - (\epsilon + \mu)E \tag{2}$$

$$\frac{dI}{dt} = \epsilon E - (\omega + \gamma + \mu)I \tag{3}$$

$$\frac{dV}{dt} = \alpha_1 I + \alpha_2 E - \sigma V \tag{4}$$

where λ denotes the population influx, μ is the natural death rate of individuals, β_1, β_2, and β_3 are the constant transmission rates from infected persons (I), exposed individuals (E), and the concentration of virus in the environment (V), respectively. Further, Υ is the recovery rate, ϵ^{-1} represents the incubation period of the virus, ω is the rate of death induced by infection, α_1 and α_2 are the host shedding rates from infected and exposed individuals, and σ is the rate at which the virus is removed through activities, such as sanitizing of infected items and surfaces. The parameters utilized in the model are assumed to be positive. Similar to the experiment by Yang and Wang, we have determined the basic reproductive number, R_0, of the model utilizing the

strategy of the next generation matrix and further assess the transmission course functions.[15]

At the unique disease-free equilibrium, $X_0 = (S_0, E_0, I_0, V_0) = \left(\frac{\lambda}{\mu}, 0, 0, 0\right)$, the infection (F) and transition (V) matrices can be obtained from the infection components of the model (E, I, V) as displayed in the following equations:

$$f = \begin{bmatrix} \beta_2 SE + \beta_1 SI + \beta_3 SV \\ 0 \\ 0 \end{bmatrix}, \quad \upsilon = \begin{bmatrix} (\epsilon + \mu)E \\ -\epsilon E + (\omega + \gamma + \mu)I \\ -\alpha_2 E - \alpha_1 I + \sigma V \end{bmatrix} \tag{5}$$

$$F = \begin{bmatrix} \beta_2 \frac{\lambda}{\mu} & \beta_1 \frac{\lambda}{\mu} & \beta_3 \frac{\lambda}{\mu} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad V = \begin{bmatrix} \epsilon + \mu & 0 & 0 \\ -\epsilon & (\omega + \gamma + \mu) & 0 \\ -\alpha_2 & -\alpha_1 & \sigma \end{bmatrix} \tag{6}$$

R_0 is evaluated as $\rho(FV^{-1})$ and indicates the spectral radius of FV^{-1}

$$R_0 = \rho\left(FV^{-1}\right) = S_0\beta_2 + \frac{S_0\epsilon}{(\omega + \gamma + \mu)}\beta_1 + \frac{[(\omega + \gamma + \mu)\alpha_2 + \epsilon\alpha_1]S_0}{\sigma(\omega + \gamma + \mu)}\beta_3 \tag{7}$$

Let,

$$R_a = S_0\beta_2 \tag{8}$$

$$R_b = \frac{S_0\epsilon}{(\omega + \gamma + \mu)}\beta_1 \tag{9}$$

$$R_c = \frac{[(\omega + \gamma + \mu)\alpha_2 + \epsilon\alpha_1]S_0}{\sigma(\omega + \gamma + \mu)}\beta_3 \tag{10}$$

where R_a represents the direct transmission route for the exposed (exposed to susceptible), R_b is the direct transmission route for the infected (infected to susceptible), and R_c comes from the contribution from the indirect transmission route (environment to susceptible).

The SEIS framework was first formulated in the study of the rabies population dynamics in fox.[21] Their results discussed certain quantitative measures for controlling rabies via culling or vaccinations. Further analysis and numerical simulations were conducted using Hopf bifurcation.[22] It is worthy to note that the SEIS models present periodic solutions from Hopf bifurcations, since the model has stable equilibrium points. Moreover, the SEIS transmission model has been an effective model in their mathematical analysis, for infections with transmitting features, such as SARS. Li and Zhen utilized the SEI framework in their study of such infections. Their outcomes gave conditions to the global asymptotic stability of the disease-free and epidemic equilibrium utilizing the Poincare-Bendixson property.[23] They also analyzed the global stability of the SEI transmission model. Their model has infectious force in both dormant and infection periods, similar to the present work. Their results exhibited global asymptotic stability in the disease-free and endemic equilibrium, thus showing when $R_0 \le 1$ (the disease-free equilibrium [DFE] is globally stable and the

malady vanishes in the end), but when $R_0 > 1$, the endemic equilibrium exists and the disease persists.[23] The transmission dynamics of the Nipah virus in bat and humans have also been analyzed based on the SEI model.[13] They studied the local and global stability conditions and performed numerical simulations that examined the flow of the Nipa virus infection in different compartments. Others have also analyzed the stochastic versions of the model and found significant results.[20,24] The effect of the presence of external noise on the disease transmission rate assessment of the SEI model has been studied.[25] The deterministic R_0 and stochastic R_0 were found, and the asymptotic stability of the disease-free equilibrium was also analyzed. It was demonstrated, that in any event, when the deterministic basic reproductive number is $R_0 < 1$, epidemic could in any case develop due to the existence of disparities in the stochastic SEI model.[25] In relation to coronaviruses, Elsheik et al. considered the SEIS model in studying its dynamic spread in Sudan.[20] They estimated the case detection proportion to be 22.7% and demonstrated that the passing pace of undetected cases was higher than the identified cases. Considering the varieties in the recognition pace of new cases in various parts of the world, and the distinctions in atmosphere, socioeconomics, and transmission dynamics of the coronavirus, the SEIS model will be highly beneficial to test the dynamics of the virus under explicit conditions.

3 | VIRAL REPLICATION MODEL

The procedure of viral replication happens via the stages, such as attachment, entrance and uncoating; transcription; synthesis of viral segments; and virion assembly and discharge. Adsorption is the initial phase of the viral replication process which occurs when proteins on the viral capsid attach themselves to the receptor proteins of the cell. Activities in this stage causes the two layers to stay as close as possible to facilitate further communications. Later, the virus proceeds to enter the target cell by breaching a phospholipid bilayer which serves as the cell's natural barrier. There are three means of entry into the target cell, such as membrane fusion, endocytosis, and virial penetration, depending on the type of virus. The SARS CoV-2, like other envelope viruses, is known to make an entry via endocytosis. Upon entry into the cell, viral contents are released by activities that cause the removal and degrading of the viral capsid. Viral contents then activate the formation of proteins to suppress the defense mechanisms and other cellular activities of the host cells, thereby gaining full control of host cellular activities. Further, the viral nuclei acids are

incorporated into the genetic material of the cell to induce replication of the viral genetic material. Furthermore, viral contents present in the cytoplasm take advantage of the host organelles to manufacture its viral components. For example, the virus mRNA can be translated on the host ribosomes into viral proteins. Later, newly created viral genome and proteins then assemble, forming virions. The virion assembling process takes place in either the cell nucleus, cytoplasm or the plasma membrane. The newly formed virions are discharged by budding off through the plasma membrane, by causing the cell to break apart or waiting for lysis. These virions are then able to infect other neighboring cells, thus repeating the entire viral replication cycle. Thus, it can be noted that the target cell-limited model is a popular framework in the field of viral dynamics. It has been used to study the viral replication of HIV, Influenza and Zika.[26-28] In our quest to evaluate the SARS-CoV-2 viral replication rate, we have introduced a modified target cell-limited model. Here, we have assumed that at the rate of k, a SARS-CoV-2 virus (V) infects a susceptible or target cell (T) and the infected cell (I) produces new viruses at the rate of ρ as shown in Figure 2. We also assume that infected cells could be cleared as a result of defense mechanisms from viral invasion as shown in the following equations:

$$\frac{dT}{dt} = \lambda - \mu T - kVT \tag{11}$$

$$\frac{dI}{dt} = kVT - (\delta + \mu)I \tag{12}$$

$$\frac{dV}{dt} = \rho I - cV \tag{13}$$

where λ represents the production rate of new cells, μ is the apoptosis rate, δ is the infection induced cell death rate, and c is the rate at which viruses are cleared resulting from cell activities and properties, such as cytopathic effect. The basic reproduction number is obtained as in the following equation:

$$R_0 = \frac{kT_0\rho}{c(\delta + \mu)} \tag{14}$$

where T_0 is the density of the pre-infected target cells. The target cell-limited framework was formulated from the knowledge that viral propagation is always constrained by the accessibility of target cells. Mathematical modeling of viral kinetics within host cells has broadened understanding of viral infection dynamics and has improved
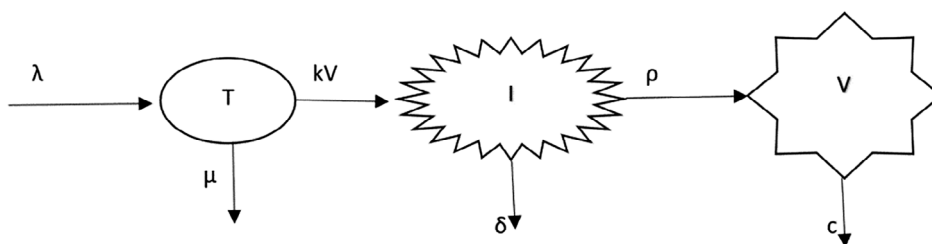

FIGURE 2 A SARS CoV-2 target cell-limited model showing the movement of target cells from the production stage to viral replication
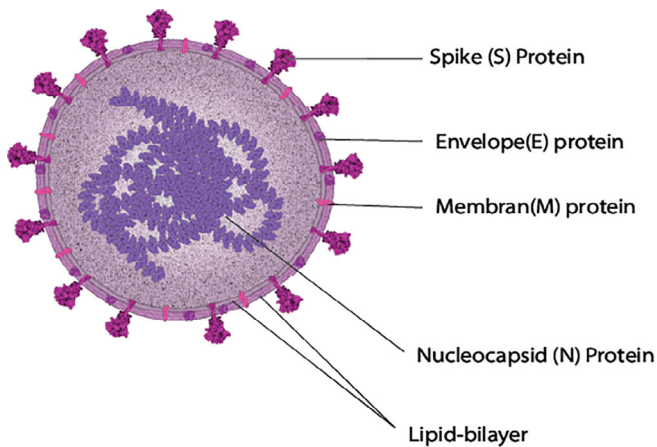
FIGURE 3 Pictorial view of SARS-CoV-2 showing its structural proteins such as spike (S), envelope (E), membrane (M), and nucleocapsid (N)

FIGURE 4 Image (a) is the trimeric S protein with the three protomers (A, B, and C) colored red, orange, and gray, respectively. S1 and S2 are the subunits of each monomer folded at the N and C terminal end as two independent domains. Image (b) is the overall structure of the S protein color coded for the different domains (RBD, NTD, CTD, FP, FPPR, HR1, CH, CD) (Figure generated with PBD ID:6VXX SARS-CoV-2 spike glycoprotein using VMD software[33])

healthcare interventions. In modeling, the "within-host" viral kinetics of influenza, modifications such as considering an eclipse phase of infected cells as well as ignoring the target cell production and death rates were integrated into the basic target cell-limited model. These modifications were based on the short duration of Influenza and the assumption that infected cells can support viral replication before they are cleared. The basic reproductive number, R_0, was estimated to be 22, when their modified model was also used to study the viral kinetics of H1N1 virus.[26] Several models have been developed to investigate SARS-CoV-2 within the replication dynamics of the host cell.[29] Further, they also compared the infection time of the coronavirus to other viral infections, such as Ebola and Influenza, and identified that the infecting time of SARS-CoV-2 is 3 times much slower than Ebola and 60 times slower than Influenza.[29] In deriving an analytical solution for the fundamental target cell-limited framework by utilizing a quasi-steady-state approximation without cytopathic effect, it was discovered that the deduction of the solution is critically subjected to the noncytopathic condition.[28] Depending on the innate immune controls and antiviral drug-induced defense mechanisms, certain target cells are cleared or inhibited as a result of viral invasion and are not able to synthesize viral components.[30,31] The SARS-CoV-2 target cell-limited model proposed in this article was formulated by introducing another parameter 'δ' into the infected cell compartment to account for the rate at which cells are cleared in the I-compartment for precise analytical results, since the analytical solutions of the viral kinetics model are dependent on the noncytopathic effects.[28] Thus, the model can be applied to the statistical data available to probe the dynamic behavior and transmission of COVID-19.

4 | SARS-COV-2 STRUCTURAL PROTEINS

The RNA of SARS-CoV-2 and other coronaviruses has approximately 30 kbs and expresses 16 nonstructural proteins that correspond to six open reading 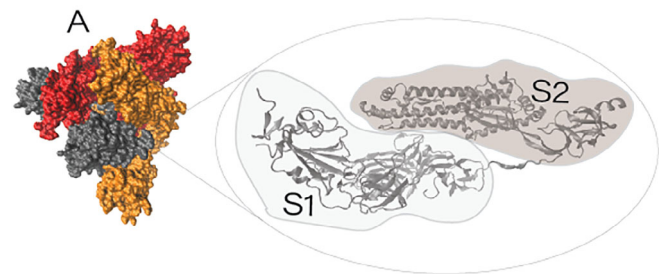frames (ORFs), with at least four major structural proteins, which are required to drive virus–host cell interaction, cytoplasmic viral assembly, and other accessory protein. The structural proteins of SARS-CoV-2 consists of the spike (S), nucleocapsid (N), membrane (M), and envelope (E) protein as shown in Figure 3.

The S protein is a large, homo-trimeric type I membrane glycoprotein of 1,128–1,472 amino acids. It is a fusion protein that mediates receptor binding and viral entry in the host cell, and is the main target for neutralization by the antibodies of the adaptive immune system in the host.[4] The S protein protrudes from the surface of the virus and interacts with the host cell via ACE2.[32] Each monomer of the S protein is about 180 kDa and have subunits, namely S1 and S2, folded as two separated units in the N- and C-terminal domains of the monomers as shown in Figure 4. The S1 subunit contains an NTD (residues 14–305), a RBD (residues 319–541), and two CTDs (residues 528–686), while the S2 subunit contains the FP (residues 788–806), FPPR (residues 834–910), HR1 (residues 912–984), CH (residues 985–1,035), CD (residues 1,036–1,068), HR2 (residues 1,163–1,213), TM domain (1213–1,237 residues), and the cytoplasmic domain (residues 1,237–1,273).[34,35] Either the N- and C-terminal domains can function as the receptor-binding region during interaction with the human ACE2 or the ACE2 of other mammals.[4] Hence, the S protein has gained significant attention in the antiviral drug development, due to its receptor-binding functionality.[4]

The viral membrane is composed of the lipid-bilayer, which is embedded in the M and E proteins. The M protein is a 23 kDa, highly conserved 232 amino acid nonglycosylated membrane protein, which possesses three transmembrane regions, as well as a $N^{exo}C^{endo}$ topology.[4] It is expressed by the 669 nucleotides (nt) long M gene of SARS-CoV-2 located after the 228 nt E gene, that encodes the E protein. The M protein is essential for virion assembly and has been identified to interact with the N protein during viral replication via its carboxy-terminal.[36] The E protein is a small pentameric protein of about 10 kDa, which spreads uniformly in the lipid bi-layer with about 20 copies per viral particle. Although its precise function is unknown, studies showed that it serves as a cation-selective channel and also plays an essential role in the virion assembly and morphogenesis.[4,37]

```
YP_009724397.2   1   MSDNGPQ-NQRNAPR-ITFGGPSDSTGSNQNGERSGARSKQRRPQGLPNNTASWFTALTQHGKEDLKFPRGQGVPINTNS   78
NP_828858.1      1   MSDNGPQsNQRSAPR-ITFGGPTDSTDNNQNGGRNGARPKQRRPQGLPNNTASWFTALTQHGKEELRFPRGQGVPINTNS   79
YP_007188586.1   1   MAS--PA-----APRaVSFADNNDITNTNLSRGRGRNPKPRAAP----NNTVSWYTGLTQHGKVPLTFPPGQGVPLNANS   69

YP_009724397.2   79  SPDDQIGYYRRATRRIRGGDGKMKDLSPRWYFYYLGTGPEAGLPYGANKDGIIWVATEGALNTPKDHIGTRNPANNAAIV  158
NP_828858.1      80  GPDDQIGYYRRATRRVRGGDGKMKELSPRWYFYYLGTGPEASLPYGANKEGIVWVATEGALNTPKDHIGTRNPNNNAATV  159
YP_007188586.1   70  TPAQNAGYWRRQDRKINTGNG-IKQLAPRWYFYYTGTGPEAALPFRAVKDGIVWVHEDGATDAPSTF-GTRNPNNDSAIV  147
```
Conserved region
```
YP_009724397.2   159 LQLPQGTTLPKGFYAEGSRGGSQASSRSSSR--SRNS--SRNSTPGSSRGTSPARMA---GNGGDAALALLLLDRLNQLE  231
NP_828858.1      160 LQLPQGTTLPKGFYAEGSRGGSQASSRSSSR--SRGN--SRNSTPGSSRGNSPARMA---SGGGETALALLLLDRLNQLE  232
YP_007188586.1   148 TQFAPGTKLPKNFHIEGTGG----NSQSSSRasSVSRnsSRSSSQGSRSGNSTRGTSpgpSGIGAVGGDLLYLDLLNRLQ  223

YP_009724397.2   232 SKMSGKGQQQQGQTVTKKSAAEASKKPRQKRTATKAYNVTQAFGRRGPEQTQGNFGDQELIRQGTDYKHWPQIAQFAPSA  311
NP_828858.1      233 SKVSGKGQQQQGQTVTKKSAAEASKKPRQKRTATKQYNVTQAFGRRGPEQTQGNFGDQDLIRQGTDYKHWPQIAQFAPSA  312
YP_007188586.1   224 ALESGKVKQSQPKVITKKDAAAAKNKMRHKRTSTKSFNMVQAFGLRGPGDLQGNFGDLQLNKLGTEDPRWPQIAELAPTA  303

YP_009724397.2   312 SAFFGMSRIGMEVTPS     GTWLTYTGAIKLDDKDPNFKDQVILLNKHIDAYKTFPPTEPKKDKKKKADETQ---ALPQR  385
NP_828858.1      313 SAFFGMSRIGMEVTPS     GTWLTYHGAIKLDDKDPQFKDNVILLNKHIDAYKTFPPTEPKKDKKKKTDEAQ---PLPQR  386
YP_007188586.1   304 SAFMGMSQFKLTHQNN[6]VYFLRYSGAIKLDPKNPNYNKWLELLEQNIDAYKTFPKKEKKQKAPKEESTDQmsePPKEH  386

YP_009724397.2   386 QKKQQTVTLLPAADLDDFSKQLQQSMS--SADSTQA   419
NP_828858.1      387 QKKQPTVTLLPAADMDDFSRQLQNSMSGASADSTQA   422
YP_007188586.1   387 RVQGTQRTRTRPSVQPG-------PMIDVNTD----   411
```

FIGURE 5 Multiple sequence analysis (MSA) comparing the N proteins of SARS-CoV-2 with MERS-CoV/SARS-CoV. Sequence identities are 50.8% and 92.5% for MERS-CoV and SARS-CoV, respectively
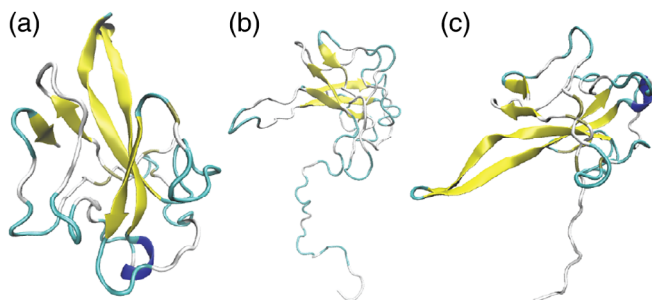
(a)   (b)   (c)

FIGURE 6 N-NTD of the three coronaviruses: (a) is SARS-CoV-2, (b) is SARS-CoV, and (c) is MERS-CoV. Both (a) and (b) share some similarities, having the same number of beta-sheets (5) while the C structural feature is a little distant

The nucleocapsid SARS-CoV-2 protein is an RNA-binding phosphoprotein that form a ribonucleoprotein complex with the viral RNA and is the core structure of the virus. It is expressed by a 1,260 nt N gene located next to ORF8 (which encodes the ORF8 non-structural protein). The N protein participates in the synthesis and translation of viral RNA, exhibits RNA chaperone activity, and also acts as a type I interferon antagonist,[4] making it immunogenic. It is structurally divided into N terminal RNA binding domain (NTD), C-terminal dimerization domain (CTD), and the intrinsically disordered central linker rich in serine and arginine.[18,38] It has been reported that the concentration of N protein in infected patients is usually higher than the other viral proteins,[18,39,40] implying that theranostic developmental efforts targeting the N protein are plausible. Currently, there exist abundant literature on the structural properties of the SARS-CoV N protein, but limited information on the structure of SARS-CoV-2 N protein, as expected. For instance, thermodynamic studies showed that the N protein of SARS-CoV is stable between pH 7 and 10, the maximum conformational stability is identified to be around pH 9, and the SARS-CoV N protein is observed to undergo irreversible thermal-induced denaturation.[18] Further, it has been reported that the N terminal domain (NTD) of SARS-CoV-2 N protein functions as its RNA binding domain.[39] Also, the authors predicted the druggable location of the NTD part of the protein using molecular dynamic simulations and revealed specific surface charge distributions that can aid in the discovery of drugs specifically targeting the RNA binding domain. Such therapeutic approaches will prevent the assembling of viral RNA by acting on the N protein of the virus to inhibit viral assembly and replication. Thus, this article discusses on the N protein as, besides being highly conserved and plays significant roles in RNA synthesis and translation, its concentrations in serum samples are high and detectable even after just a day of infection.[4,18] These traits present an opportunity to develop diagnostic solutions for early detection as well as synthesizing antiviral agents that will potentially intercept viral replication.

4.1 | RNA-binding domains of SARS-CoV-2 N proteins

The N terminal domain of the coronavirus N protein is responsible for binding to the viral RNA, resulting in a viral ribonucleoprotein (vRNA)

complex, that is essential for viral replication as shown in Figure 6. Additionally, the C terminal domain (CTD) is responsible for oligomerization (forming k-mers).[38,41] Certain studies focusing on the NTD of N protein for drug development and discovery has been performed using information deposited in public protein databases.[40,42] Recently, protein database (PDB) has received solved structures of the NTD of SARS-CoV-2, either in isolation or complexed with other molecules. Presently, there are three structures of the NTD protein of SARS-CoV-2 that have been solved by X-ray crystallography with IDs; 6M3M (2.7 Å),[39] 6WKP (2.67 Å),[43] 6VYO (1.7 Å).[44] 6M3M represents the structure of the N-Terminal RNA binding domain with the highest resolution of 2.7 Å above the 2.1 Å median resolution of X-ray diffraction structures deposited at PDB as of 2020 (https://www.rcsb.org/stats/distribution-resolution) and was reported along with the characterization of the RNA-binding N protein domains. 6VYO is currently the structure with the lowest resolution of 1.7 Å. However, unlike 6M3M, 6VYO and 6WKP have been deposited as a complex with two and four other smaller ligands respectively. Furthermore, 6VYO has the lowest R-factor of about 2.05, representing a high quality structure,[45,46] and thus may be preferred during molecular docking and molecular dynamics (MD) simulations studies.[39,40,47] Lastly, researchers have used nuclear magnetic resonance (NMR) spectroscopy to study the structure of N-NTD and its interaction with RNA and showed the existence of a net positive charge on the surface of the NTD as responsible for binding, confirming results also obtained by Zeng et al..[48,49] The SARS-CoV-2 N protein forms a dimer in solution via CTD–CTD interaction and binds to non-specific double-stranded deoxyribonucleic acid (dsDNA) through electrostatic interactions.[40,42] Similarly, PyMOL software-generated electrostatic surface potential map showed that the N protein has a net positive charge, both at the NTD and CTD sites.[50] Likewise, Kang et al. showed that the NTD tail residues (Asn 48, Asn 49, Thr 50, and Ala 51) possibly opens up the binding pocket to enable RNA binding, via atomic resolution. Further, it has been revealed that the N protein possesses a strong binding affinity to guanosine bases, by performing binding affinity experiments with ribonucleotide monophosphates (GMP, UMP, CMP, and AMP). Furthermore, the researchers predicted a drugability score of 0.66 on a scale of 0–1 (1 being the highest drugability value) with the binding pocket along its beta-sheet residues, via the pocket detection and analysis tool DoGSiteScorer. Moreover, the drugability score showed that the SARS-CoV-2 N-NTD bond is higher on the average of 15% than SAR-CoV, MERS-CoV, and HCoV-OC43 N-NTD.[39] A prior research showed that occupying the binding pocket of the mild type, homologous, coronavirus HCoV-OC43 N-NTD bond with a higher affinity ligand has hydrogen-bond-forming moieties, decreased the RNA-binding affinity of the N-protein, which is critical for the development of targeting agents.[51] In SARS-CoV-2, arginine residues, specifically Arg89, decrease the bond formation groups of the ligand core. The most abundant interactions of the N protein and RNA bases are the arginine–guanosine interactions, which will lead to a decrement in the effects of Arg89 on the aromatic core of the ligand. In summary, agents with high hydrogen bond-forming moieties may increase the binding affinity for theranostic applications, to increase the binding efficacy for SARS-COV-2.

4.2 | N proteins of SARS-CoV-2, MERS-CoV, and SARS-CoV

The structure of a protein determines its functionality. There are several ways to determine the functions of proteins based on the concept of annotation-by-homology, where annotations from well-characterized homologous proteins are used to predict the functions of new proteins.[52] Two of such methods that can be potentially applied to compare the N proteins of SARS-CoV-2, SARS-CoV, and MERS-CoV, are protein sequence alignments as shown in Figure 5 and structural alignments as displayed in Figure 6. However, it can be noted that the knowledge about structural proteins of SARS CoV-2, including the N protein, are based on the previous SARS-CoV and other human coronaviruses studies.

Protein sequence alignment is a relatively easy approach to compare proteins, where algorithms are used to align amino acid sequences to predict conserved regions and secondary structures. An example of these algorithms is BLAST, which was used to predict the percentage identity of the three coronaviruses in Table 1. Using BLASTp for protein sequence alignment analysis of the N proteins from the three zoonotic viral strains (SARS-CoV-2, SARS-CoV, and MERS-CoV) showed that the N protein of SARS-CoV-2 is much related to that of SARS-CoV, than MERS-CoV with 92.5% identity between SARS-CoV-2 and SARS-CoV as shown in Table 1 using BLAST. Further analysis of the conserved sequence regions can be used to predict the potential binding/active sites as well as the residues at the protein structure core for secondary structure prediction, since these sites are generally known to be highly conserved in homologous protein families. Although amino acid sequence alignments offer several benefits as discussed, it is not as accurate as structural alignment.[53] Structural alignment considers the spatial homologous protein evolution of close and/or distant family functional site predictions to identify the similarities and differences between them as well as searching for similar structures that have less sequence alignment identity.[54,55] Further, FATCAT server is an example of highly efficient tools, which are used for protein structure comparison. Its comparison algorithm showed hinges and internal arrangements in two protein structures.[55] It can be utilized after amino acid sequence alignment of the SARS-CoV-2 N protein, for instance, to identify proteins of similar structural characteristics and this can be beneficial in the design of inhibitors or binders for diagnostics.

5 | DIAGNOSIS OF SARS-COV-2 BASED ON IN SILICO ANALYSIS OF THE N PROTEIN

Current standard methods for the detection of viruses including polymerase chain reaction (PCR) and enzyme-linked immunosorbent assay (ELISA) require a considerable amount of processing time, slowing down the implementation of strategies to break the chain of transmission.[56,57] These methods are not suitable for mass screening of

Accession no	Viral strain	AA length	Percent identity (%)
YP_009724397.2	SARS-CoV-2	419	100
NP_828858.1	SARS-CoV	422	92.52
YP_007188586.1	MERS-CoV	411	50.82

TABLE 1 Comparison of the N protein sequences of SARS-CoV and MERS-CoV with SARS-CoV-2 by BLASTp

FIGURE 7 A flow diagram for the development of bio-probes to be used in diagnostic application via MD simulation and machine- learning techniques
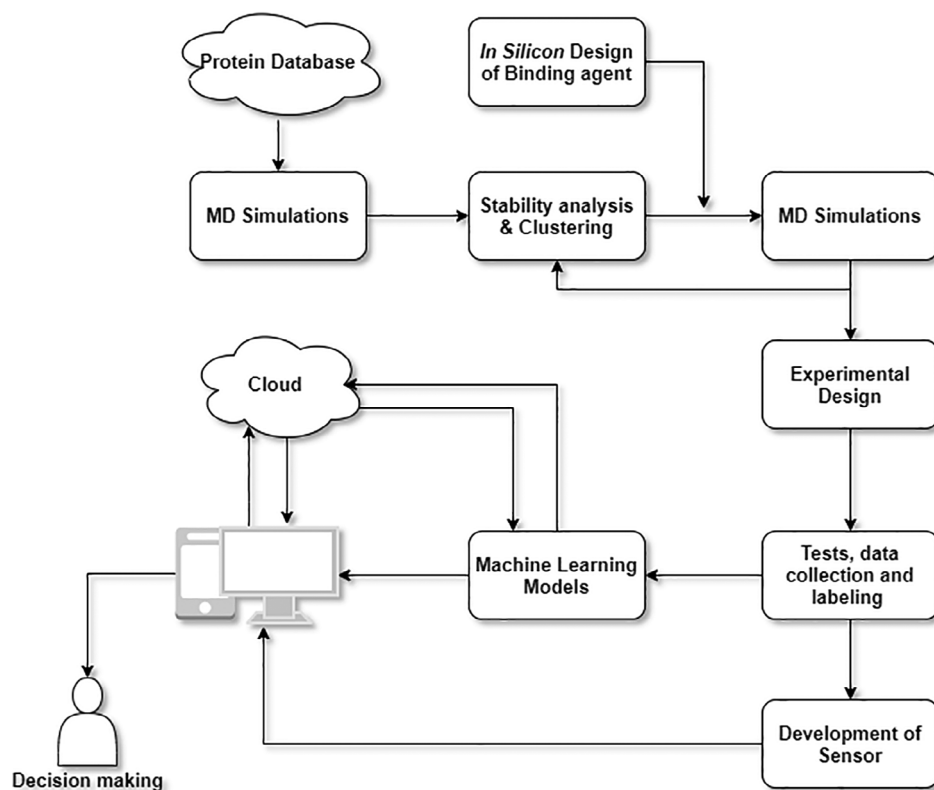
individuals in applications requiring almost instantaneous results, such as airport biosecurity programs, schools, and events. Further, certain emerging biosensing techniques that are promising and capable of alleviating some of the challenges associated with conventional assays, include lab-on-chip (LOC) and Field Effect transistor (FET) technologies. These technologies require the development of special biological probes that can bind specifically to the analyte of interest. Furthermore, experimental and in silico methods have been pursued by researchers toward the development of bio-probes. An example of an experimental approach is systematic evolution of ligands by exponential enrichment (SELEX), which is used for the generation of aptamers[58,59] and offer opportunities to develop aptasensors or FET biosensors[60-62] for enhanced SARS-CoV-2 detection. Another widely used experimental method is antibody-based-biosensors (immunosensors) that uses either polyclonal, monoclonal or recombinant antibodies as the bio-recognition elements as well as different materials as signal-carrying transduction technologies and can be either electrochemical, piezoelectric, or optical. These immunosensors are composed of antibodies that are immobilized onto the surface of the transducer and are connected to a control that reads the signals. For example, a piezoelectric immunosensor has been developed to detect SARS-CoV in sputum.[63] Other immunosensor-based technologies have also been reported in recent times.[64-66] Experimental methods augmented with in silico approaches can significantly speed up the process of bio-probe development for tailored biosensing applications. Figure 7 shows the use of in silico approaches to the development of bio-probes, such as aptamers for SARS-CoV-2 N protein detection. Although there are several in silico methods for studying macromolecules, MD simulations and machine learning approaches are considered in this article.

6 | MD SIMULATION

MD simulations predict the trajectory of every atom in a protein or other molecular systems to describe interatomic interactions known as force fields.[67] Such simulations will be helpful to capture a variety of essential bio-molecular processes that may not be possible, difficult and/or expensive to capture in wet-lab experiments. Some of the important processes, include but not limited to, conformational changes, ligand-binding activities, and protein folding, which are essential in the development of theranostics.[67,68] MD simulations can

be used to answer questions relating to the structural and conformational characteristics, and binding properties of the SARS-CoV-2 N protein to complement experimental results. The structural stability of the N protein or specific domains of the N protein can be investigated in isolation or in complex with designed ligands of interest using MD simulations. It can also be used to probe real-time conformational changes during intermolecular binding under varying biophysical and/or biochemical conditions, such as alterations in pH, ionic strength, redox, and molecular changes of binding motifs. Figure 8 shows 10 ns MD simulation, showing the root mean squared standard deviation (RMSD) of the SARS-CoV-2 N protein containing 419 residues. A high stability of the RNA-binding domain (RBD) residues (46–174) compared to the other domains of the N-Protein (CTD and Linker) is observed. This type of analysis facilitates further probing into the binding characteristics of the N protein with other ligands after conformational sampling[69] is performed to identify the most stable cluster of conformations.

Currently, MD simulation has been used together with other molecular modeling methods, such as molecular docking to understanding the binding and structural characteristics of the SARS-CoV-2.Tatar et al. reported molecular docking of 34 antiviral compounds with SARS-CoV 2 N protein and performed MD simulations to show that the N protein residues, such as Lys65, Gly69, Gln70, Pro67, Phe66, Lys123, Trp132, and Ala134 exhibit a strong binding affinity for Nafamostat (a synthetic serine protease inhibitor) primarily through hydrogen bond interactions. Root-mean-square-deviation (RMSD) analysis of the N protein also indicated its stability in the range of 0.1–0.3 nm in a 10 ns simulation. Further analysis based on the root mean square fluctuation (RMSF) and radius of gyration (Rg) analysis showed that the RNA-binding domain of the N protein is very stable and capable of forming stable complexes with ligands at its binding sites.[44] Similarly, Lin et al. used molecular docking to demonstrate that occupying the binding pockets of the homologous N-NTD of the mild type coronavirus HCoV-OC43 with a higher affinity ligand *N*-(6-oxo-5, 6-dihydrophenanthridin-2-yl) (*N,N*-dimethylamino) acetamide hydrochloride (PJ34) decreased the RNA-binding affinity of the N protein.[51] Although, there is currently not much literature on the application of MD simulations to probe the binding and structural characteristics of the SARS-CoV-2 N protein, MD simulations based on the N protein of other coronaviruses shows the capacity to develop and evaluate unique ligands for specific targeting of SARS-CoV-2 N proteins toward the development of advanced theranostics.

7 | MD SIMULATIONS FOR ENHANCED THERANOSTIC APPLICATIONS

MD simulations are highly helpful to obtain high-affinity protein–ligand structures for theranostic applications.[67] For example, the correlation between the binding affinity of wild-type protein–ligand and mutated protein–ligand complexes has been investigated by predicting the effects of mutation on the protein–ligand complexes using MD simulations and local geometrical features.[70] The authors measured the feature differences between the wild-type protein–ligand and mutated protein-ligand complexes as a way to evaluate changes in binding affinity using several machine learning models. Other researchers have also shown the importance of MD simulations in lead optimization for several protein structures.[71-73] Generally, MD simulation experiments take one of two forms; one that probes the structural properties of the proteins in question as shown in Figure 8 or one that tests nonstructural properties by mutating parts or whole structures.[70] Thus, it will be possible to discover new ligands with higher binding affinities or improve on existing ones for N protein binding by utilizing these two approaches. Since, the N protein is most expressed in the initial stage of SARS-CoV infection, it makes an attractive target for theranostics. MD probing of the N protein will not only be useful for theranostic applications but enable a better understanding of the SARS-CoV-2 activity. Currently, proposals for the use of the N protein as a diagnostic tool is primarily focused on two different strategies, such as development of antibodies for the N protein, and recombinant production of N protein for the detection of N-protein-specific antibodies.[18] These experimental approaches could be made more efficient through MD simulations. In a recent work, MD simulation was used as a part of four-step in silico procedure
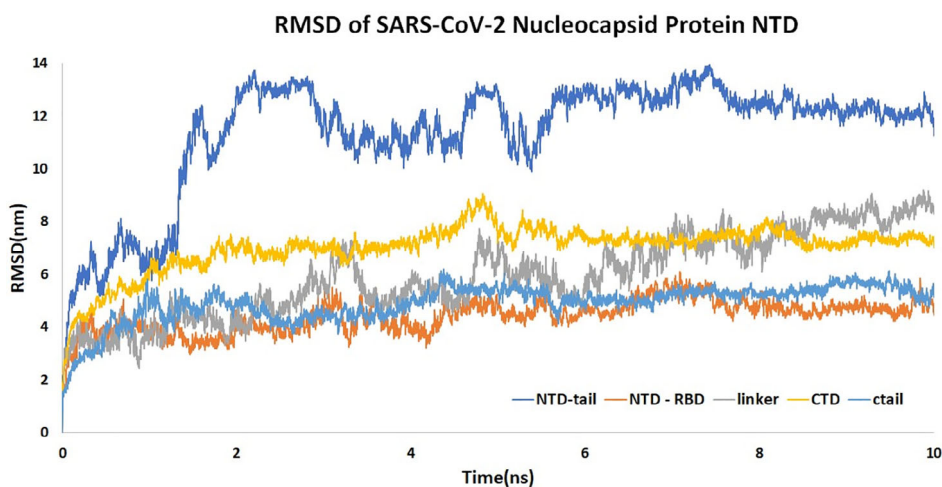

RMSD of SARS-CoV-2 Nucleocapsid Protein NTD

FIGURE 8 RMSD data of 10 ns MD simulation of SARS-CoV-2 N proteins showing the structural stability of the N terminal RBD compared other domains of the protein

(molecular docking, binding free-energy prediction, pharmacokinetics, and drug-likeness prediction as well as MD simulations) to determine the affinity binding characteristics of the drug candidate called ZINC00003118440 (8-(2-hydroxyethyl) aminophylline), which is a synthetic derivative of the Theophylline drug known to act as a bronchodilator with antiviral properties as confirmed in the literature.[74] Further, ZINC0000146942 (ethyl (4S)-4methyl-2-oxo-6-[(1S)-1-phenylethyl]-3,4-dihydro-1H-pyrimidine5-carboxylate) is a pyrimidone derivatives, that are used against viral infections and are recently employed to inhibit SARS-CoV-2. In the study, a 100 ns simulation of the two candidates in complex is performed with the N-NTD protein structures, which are obtained from the PDB (PDBID: 2OFZ) and used RMSD as well as RSMF analysis to determine the protein–ligand stability.[75] Furthermore, since ZINC00003118440 is a bronchodilator, the binding affinity of all approved bronchodilators have been investigated against SARS-CoV-2, where N-NTD showed that other bronchodilators including formeterol, terbutaline, ipratropium bromide, Tiotropium bromide and salbutamol can be used a potential SARS-CoV-2 N-NTD blockers. Moreover, ZINC000000146942 confirmed that pyrimidone derivatives, which are already used as major components in antiretroviral drugs,[76-79] can also be used in the design of drugs to inhibit SARS-CoV-2 by targeting N-NTD. In conclusion, in vitro assays using these drug candidates will be required to demonstrate the effectiveness of the methods and serve as a deciding factor to select specific bronchodilator for the treatment of COVID-19 infection.

8 | MACHINE-LEARNING APPROACHES

Machine-learning algorithms are either supervised or unsupervised, where the supervised learning refers to algorithms, that use labeled data sets to make prediction/inferences and focuses more on data classification by approximating with high accuracy. Contrarily, unsupervised learning uses unlabeled data sets for clustering techniques.[80] Some of the algorithmic methods that have been developed and used in machine learning, includes neural networks, naive Bayes, instance-based learning, principal component analysis, and logistic regression. Machine learning methods have been used in biotechnology to recognize biosensing specific signal features. It has also been used to analyze MD trajectory data for clustering of ensembles with similar conformations to be used in docking and virtual screening studies, as well as for dimensionality reductions to identify most relevant features in trajectories, that reduces noise, and to develop empirical MD force fields for simulations.[81-84] The relevance of these two applicational areas is significant in the search for diagnostic methods to target N protein in SARS-CoV-2.

Biomolecular simulation trajectories are inherently of high dimensions and therefore present a challenge in developing insights from datasets, which are generated by MD simulations. By applying deep learning methods, it is possible to reduce the dimensionality, making it easier to predict essential underlying motions of complex biological processes. Bhowmik et al. developed a deep convolutional variational auto-encoder (CVAE) neural network that reduces high dimensions of protein folding pathways into a reduced number of conformational forms with similar structural properties. The method primarily utilized a type of auto-encoder architecture known as variational auto-encoder (VAE), which represents the output latent attributes of the encoder section as a probability distribution, allowing for the latent space sampling at any point to generate new outcomes that corresponds to the patterns found in the initial data. The method involves in the generation of contact matrices as shown in Figure 9, where atoms are separated by ~8 Å were considered, and the output (contact matrices) were fed as inputs in the CVAE to generate the VAE embeddings, that has a lower dimensional representation.[85] ML methods can also be used to develop empirical molecular dynamic force fields for MD simulations,[86-89] where deep potential for molecular dynamics (DeePMD), which is a deep neural network model. This
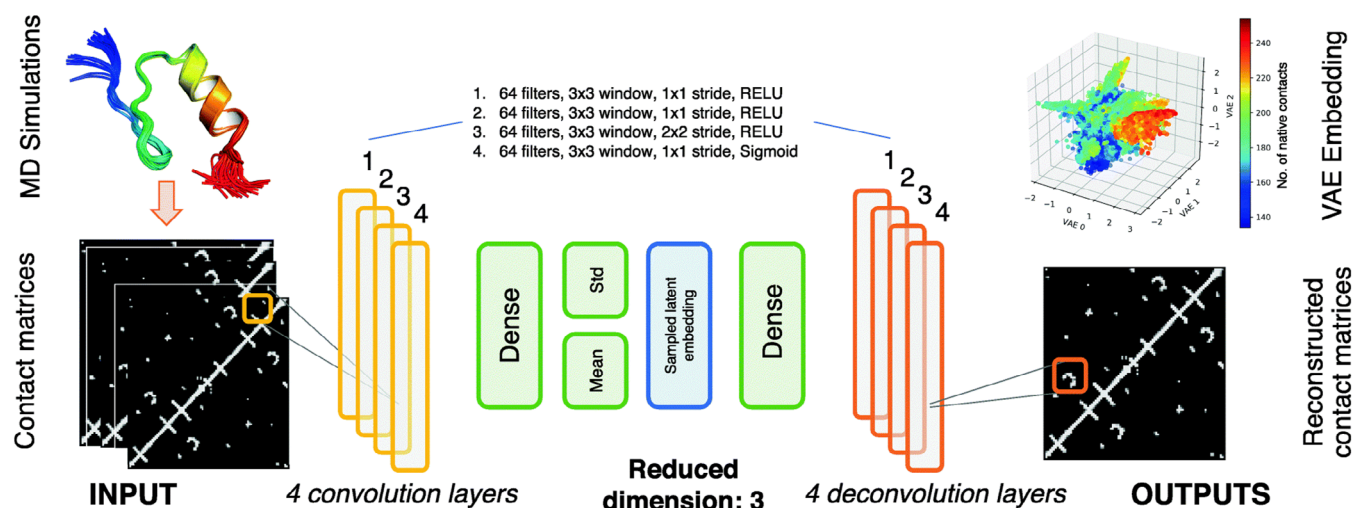

FIGURE 9 Convolutional variational auto encoder architecture. The deep learning network that takes contact maps (2D images) as inputs and outputs VAE embedding (low dimensionality) and reconstruct the contacts maps based on the learned embedding[85]

FIGURE 10 Potential in silico pipeline for the design of high-affinity ligands for use in diagnostics strategies for SARS-CoV-2
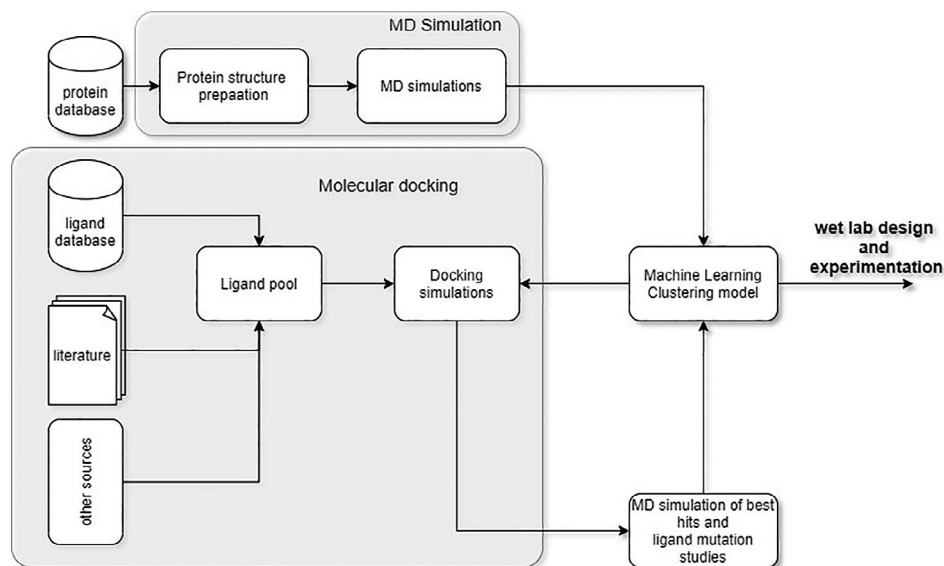
MD simulation model is based on a many-body potential and interatomic forces trained with ab initio data is of particular interest[90] in the development of DeePMD kit, which is a software library that simplifies the development of energy fields for MD simulations.

Specific applications of ML in the development of theranostics, include extraction of informative structures from raw data, classification of different high dimensional biomedical data forms in an organized structure, prediction of protein structure and functions, as well as recovery of clinically significant biomarkers.[91,92] Thus, specific ML methods that have been used for these applications, include principal component analysis (PCA), single valued decomposition (SVD), support vector machines (SVM), and deep learning models. Ge et al. applied machine learning and statistical approaches to identify SARS-CoV-2 drug candidates by integrating large-scale available coronavirus-related data from over 6,000 drug candidates. Their ML approach, together with experimental methods, helped to discover poly-ADP-ribose polymerase 1 (PARP1) inhibitor called CVL218, as a potential drug candidate to treat COVID-19 infection. Additionally, the researchers performed molecular docking studies and showed that CVL218 binds favorably to the N-NTD protein of SARS-CoV-2, thus providing a possible mode of antiviral activity against SARS-CoV-2.[93] Besides, Google DeepMind's ALPHAFOLD deep learning system has also been utilized to predict the related protein structures of SARS-CoV-2,[94] providing valuable insights to develop vaccine and drug to combat COVID-19 infection.

9 | PROPOSED IN SILICO PIPELINE FOR SARS-COV-2 N PROTEIN DIAGNOSTICS

A computational pipeline for the development of a highly efficient bio-probe via RNA-binding domain of SARS-CoV-2 N protein is proposed as displayed in Figure 10. The pipeline is primarily divided into three sections, such as MD simulation, clustering via machine learning,

and molecular docking approach. In a typical process, the protein structure is retrieved from widely used databases, such as the Protein Databank and prepared for simulation and stability analysis using tools, namely Chemistry at Harvard Molecular Mechanics (CHARMM) force field,[95] PropKa (protonation state prediction),[96] visual molecular dynamics (VMD) (for visualization and analysis)[33] and Nanoscale molecular dynamics (NAMD) (to perform the simulation).[97] In the next stage, machine-learning models are used to perform unsupervised learning or clustering of the trajectory information obtained from the MD simulations in order to select the most stable conformation of the proteins for molecular docking experiments. The molecular docking experiments also begin with the development of a potential ligand pool followed by the docking simulations. It can be noted that the ligand pool is built with molecules that may be sourced from databases, such as ZINC and PubChem. After the docking process, high affinity ligand–protein complexes are selected for further MD simulation and clustering, followed by wet laboratory experiments after successful simulations and modifications are performed in silico.

10 | CONCLUSIONS

Understanding the transmission dynamics of SARS-CoV-2 is one of the initial steps toward the deployment of intervention strategies for COVID-19 mitigation. The present article has discussed the development of an intrinsic model to describe the growth characteristics of the virus, and this provide relevant information to probe the transmission rate of the virus under various conditions. Computational approaches in the search for theranostic methods to combat against SARS-CoV-2 are highly useful. MD simulations and machine-learning methods are amongst the efficient tools that facilitate the development of theranostics for new pathogens, such as SARS-CoV-2, providing opportunities to develop new and tailored biomedical technologies. Thus, the in silico methods for

probing ligand–target interactions will play a key role in identifying and designing diagnostic and therapeutic strategies to mitigate viral infections in the future.

AUTHOR CONTRIBUTIONS

Godfred Sabbih: Data curation; formal analysis; investigation; methodology. **Maame Korsah:** Conceptualization; data curation; formal analysis; methodology. **Jaison Jeevanandam:** Writing-review and editing. **Michael Danquah:** Supervision; validation; writing-review and editing.

CONFLICT OF INTEREST

None.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1002/btpr.3096.

DATA AVAILABILITY STATEMENT

Data available on request due to privacy/ethical restrictions

ORCID

Godfred O. Sabbih https://orcid.org/0000-0002-9141-1715
Michael K. Danquah https://orcid.org/0000-0002-4232-7138

REFERENCES

1. Luan J, Yue L, Jin X, Zhang L. Spike protein recognition of mammalian ACE2 predicts the host range and an optimized ACE2 for SARS-CoV-2 infection. *Biochem Biophys Res Commun*. 2020;526(1):165-169.
2. Rabi FA, Al Zoubi MS, Kasasbeh GA, Salameh DM, Al-Nasser AD. SARS-CoV-2 and coronavirus disease 2019: what we know so far. *Pathogens*. 2020;9(3):231–244.
3. Jeevanandam J, Banerjee S, Paul R. Challenges and opportunities to develop diagnostics and therapeutic interventions for severe acute respiratory syndrome- Corona virus 2 (SARS-CoV-2). *J Biomed Res Environ Sci*. 2020;6(10):219-232.
4. King AM, Adams MJ, Carstens EB, Lefkowitz EJ, Eds. Family - Coronaviridae. *Virus Taxonomy*. Ninth Report of the International Committee on Taxonomy of Viruses, Amsterdam: Elsevier; 2012:806-828. https://www.sciencedirect.com/book/9780123846846/virus-taxonomy.
5. Lu R, Li J, Niu P, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet*. 2020;395(10224):565-574.
6. Li X, Peng Y, Meng L, Lu S. Molecular immune pathogenesis and diagnosis of COVID-19. *J Pharm Anal*. 2020;10(2):102-108.
7. Morawska L, Cao J. Airborne transmission of SARS-CoV-2: the world should face the reality. *Environ Int*. 2020;139:105730.
8. Cao W, Li T. COVID-19: towards understanding of pathogenesis. *Cell Res*. 2020;30(5):367-369.
9. Li Q, Guan X, Wu P, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *N Engl J Med*. 2020;382(13):1199-1207.
10. Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *J Travel Med*. 2020;27(2):1-4.
11. Lan J, Ge J, Yu J, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*. 2020;581(7807):215-220.
12. Li R, Pei S, Chen B, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science*. 2020;368(6490):489-493.
13. Shah NH, Suthar AH, Thakkar FA, Satia MH. SEI-model for transmission of Nipah virus. *J Math Comput Sci*. 2018;8(6):714-730.
14. Yang Z, Zeng Z, Wang K, et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *J Thorac Dis*. 2020;12(3):165-174.
15. Yang C, Wang J. A mathematical model for the novel coronavirus epidemic in Wuhan, China. *Math Biosci Eng*. 2020;17(3):2708-2724.
16. Kirkcaldy RD, King BA, Brooks JT. COVID-19 and Postinfection immunity: limited evidence, many remaining questions. *JAMA*. 2020;323(22):2245-2246.
17. Tilocca B, Soggiu A, Sanguinetti M, et al. Comparative computational analysis of SARS-CoV-2 nucleocapsid protein epitopes in taxonomically related coronaviruses. *Microbes Infect*. 2020;22(4-5):188-194.
18. Surjit M, Lal SK. The sars-cov nucleocapsid protein: a protein with multifarious activities, infection, genetics and evolution. *J Mol Epidemiol Evolution Genet Infect Dis*. 2008;8(4):397-405.
19. Xue Y, Yuan X, Liu M. Global stability of a multi-group SEI model. *Appl Math Comput*. 2014;226:51-60.
20. Elsheikh S, Abbas M, Bakheet M, Degoot A. A Mathematical Model for the Transmission of Corona Virus Disease (COVID-19) in Sudan. *Preprint*. 2020. https://www.researchgate.net/profile/Abdo_Degoot/publication/341804444_A_Mathematical_Model_for_the_Transmission_of_Corona_Virus_Disease_COVID-19_in_Sudan/links/5ed93fb7299bf1c67d3c9a71/A-Mathematical-Model-for-the-Transmission-of-Corona-Virus-Disease-COVID-19-in-Sudan.pdf.
21. Anderson RM, Jackson HC, May RM, Smith AM. Population dynamics of fox rabies in Europe. *Nature*. 1981;289:765-771.
22. Swart JH. Hopf bifurcation and stable limit cycle behavior in the spread of infectious disease, with special application to fox rabies. *Math Biosci*. 1989;95(2):199-207.
23. Li G, Zhen J. Global stability of an SEI epidemic model with general contact rate. *Chaos Solitons Fract*. 2005;23:997-1004.
24. Chen Q. A new idea on density function and covariance matrix analysis of a stochastic SEIS epidemic model with degenerate diffusion. *Appl Math Lett*. 2020;103:106200.
25. Otunuga OM. Global stability of nonlinear stochastic SEI epidemic model with fluctuations in transmission rate of disease. *Int J Stochast Anal*. 2017;2017:6313620.
26. Canini L, Perelson AS. Viral kinetic modeling: state of the art. *J Pharmacokinet Pharmacodyn*. 2014;41(5):431-443.
27. Zitzmann C, Kaderali L. Mathematical analysis of viral replication dynamics and antiviral treatment strategies: from basic models to age-based multi-scale Modeling. *Front Microbiol*. 2018;9:1546-1546.
28. Cangelosi RA, Schwartz EJ, Wollkind DJ. A quasi-steady-state approximation to the basic target-cell-limited viral dynamics model with a non-cytopathic effect. *Front Microbiol*. 2018;9:54-54.
29. Hernandez Vargas EA, Velasco-Hernandez JX. In-host modelling of COVID-19 kinetics in humans. *medRxiv*. 2020;20044487. https://europepmc.org/article/ppr/ppr151196.
30. Jamieson AM. Host resilience to emerging coronaviruses. *Future Virol*. 2016;11(7):529-534.
31. Ashida H, Mimuro H, Ogawa M, et al. Cell death and infection: a double-edged sword for host and pathogen survival. *J Cell Biol*. 2011;195(6):931-942.

32. Zhou P, Yang X-L, Wang X-G, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020;579: 270-273.

33. Humphrey W, Dalke A, Schulten K. VMD - visual molecular dynamics. *J Mol Graph*. 1996;14:33-38.

34. Cai Y, Zhang J, Xiao T, et al. Distinct conformational states of SARS-CoV-2 spike protein. *Science*. 2020;369(6511):1586-1592. https://science.sciencemag.org/content/369/6511/1586.

35. Huang Y, Yang C, Xu X-f, Xu W, Liu S-w. Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. *Acta Pharmacol Sin*. 2020;41: 1141-1149.

36. Fang X, Ye L, Khalid Amine T, et al. Peptide domain involved in the interaction between membrane protein and Nucleocapsid protein of SARS-associated coronavirus. *Korean Soc Biochem Mol Biol - BMB Reports*. 2005;36(6):381-385.

37. Lauren W, Carolyn M, Peter G, Gary E. SARS coronavirus E protein forms cation-selective ion channels. *Virology*. 2004;330(1):322-331.

38. Chun-Yuan C, Chung-ke C, Yi-Wei C, et al. Structure of the SARS coronavirus nucleocapsid protein RNA-binding dimerization domain suggests a mechanism for helical packaging of viral RNA. *J Mol Biol*. 2007;368(4):1075-1086.

39. Kang S, Yang M, Hong Z, et al. Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. *Acta Pharm Sin B*. 2020;10:1228-1238.

40. Zeng W, Liu G, Ma H, et al. Biochemical characterization of SARS-CoV-2 nucleocapsid protein. *Biochem Biophys Res Commun*. 2020;527 (3):618-623.

41. McBride R, van Zy M, Fielding BC. The coronavirus nucleocapsid is a multifunctional protein. *Viruses*. 2014;6(8):2991-3018.

42. Chang C-K, Sue S-C, Yu T-H, et al. Modular organization of SARS coronavirus nucleocapsid protein. *J Biomed Sci*. 2006;13(1):59-72.

43. Madej T, Lanczycki CJ, Zhang D, et al. MMDB and VAST+: tracking structural similarities between macromolecular complexes. *Nucleic Acids Res*. 2014;42(D1):D297-D303.

44. Tatar G, Turhan K. Investigation of N terminal domain of SARS CoV 2 Nucleocapsid protein with antiviral compounds based on molecular Modeling approach. *ScienceOpen Preprints*. 2020. https://www.scienceopen.com/hosted-document?doi=10.14293/S2199-1006.1.SOR-.PPPT99I.v1.

45. Paula FR, Fernandes MS, da Silva FS, Freitas ACSG, de Melo EB, Trossini GHG. Insights on 3D structures of potential drug-targeting proteins of SARS-CoV-2: application of cavity search and molecular docking. *Mol Informs*. 2020. https://onlinelibrary.wiley.com/doi/full/10.1002/minf.202000096.

46. Yadav R, Imran M, Dhamija P, Suchal K, Handu S. Virtual screening and dynamics of potential inhibitors targeting RNA binding domain of nucleocapsid phosphoprotein from SARS-CoV-2. *J Biomol Struct Dyn*. 2020;1-16. https://www.tandfonline.com/doi/full/10.1080/07391102.2020.1778536.

47. Durrant JD, McCammon JA. Molecular dynamics simulations and drug discovery. *BMC Biol*. 2011;9:71.

48. Dinesh DC, Chalupska D, Silhan J, Veverka V, Boura E. Structural basis of RNA recognition by the SARS-CoV-2 nucleocapsid phospho-protein. *bioRxiv*. 2020. https://www.biorxiv.org/content/10.1101/2020.04.02.022194v1.abstract.

49. Zeng W, Liu G, Ma H, et al. Biochemical characterization of SARS-CoV-2 nucleocapsid protein. *Biochem Biophys Res Commun*. 2020;527 (3):618-623.

50. Schr¨odinger, L., The PyMOL molecular graphics system, version 1.8 (November 2015); 2015.

51. Lin S-Y, Liu C-L, Chang Y-M, Zhao J, Perlman S, Hou M-H. Structural basis for the identification of the N-terminal domain of coronavirus Nucleocapsid protein as an antiviral target. *J Med Chem*. 2014;57(6): 2247-2257.

52. Loewenstein Y, Raimondo D, Redfern OC, et al. Protein function annotation by homology-based inference. *Genome Biol*. 2009;10 (2):207.

53. Florencio P, Jung-Wook B. Computational prediction of functionally important regions in proteins. *Curr Bioinform*. 2006;1(1):15-23.

54. Modi V, Dunbrack RL. A structurally-validated multiple sequence alignment of 497 human protein kinase domains. *Sci Rep*. 2019;9(1): 19790.

55. Ye Y, Godzik A. FATCAT: a web server for flexible structure comparison and structure similarity searching. *Nucl Acids Res*. 2004;32 (suppl_2):W582-W585.

56. Zhu H, Fohlerov Z, Pekárek J, Basov E, Neužil P. Recent advances in lab-on-a-chip technologies for viral diagnosis. *Biosens Bioelectron*. 2020;153:112041.

57. Carter LJ, Garner LV, Smoot JW, et al. Assay techniques and test development for COVID-19 diagnosis. *ACS Cent Sci*. 2020;6(5): 591-605.

58. Acquah C, Chan YW, Pan S, et al. Characterisation of aptamer-anchored poly(EDMA-co-GMA) monolith for high throughput affinity binding. *Sci Rep*. 2019;9(1):14501.

59. Jeevanandam J, Tan KX, Danquah MK, Guo H, Turgeson A. Advancing aptamers as molecular probes for cancer Theranostic applications—the role of molecular dynamics simulation. *Biotechnol J*. 2020;15(3):1900368.

60. Ocaña C, del Valle M. Chapter 8 - Impedimetric Aptasensors using nanomaterials. In: Nikolelis DP, Nikoleli G-P, eds. *Nanotechnology and biosensors*. A volume in Advanced Nanomaterials, Amsterdam: Elsevier; 2018:233-267.

61. Pereira AC, Sales MGF, Rodrigues LR. Chapter 3 - biosensors for rapid detection of breast cancer biomarkers. In: Khan R, Mohammad A, Asiri AM, eds. *Advanced Biosensors for Health Care Applications*, Inamuddin. Amsterdam: Elsevier; 2019:71-103.

62. Di Pietrantonio F, Cannatà D, Benetti M. Chapter 8 - biosensor technologies based on nanomaterials. In: Dinca V, Suchea MP, eds. *Functional Nanostructured Interfaces for Environmental and Biomedical Applications*. Amsterdam: Elsevier; 2019:181-242.

63. Zuo B, Li S, Guo Z, Zhang J, Chen C. Piezoelectric Immunosensor for SARS-associated coronavirus in sputum. *Anal Chem*. 2004;76(13): 3536-3540.

64. Gogola JL, Martins G, Caetano FR, et al. Label-free electrochemical immunosensor for quick detection of anti-hantavirus antibody. *J Electroanal Chem*. 2019;842:140-145.

65. Haji-Hashemi H, Safarnejad MR, Norouzi P, Ebrahimi M, Shahmirzaie M, Ganjali MR. Simple and effective label free electrochemical immunosensor for fig mosaic virus detection. *Anal Biochem*. 2019;566:102-106.

66. Viter R, Savchuk M, Starodub N, et al. Photoluminescence immunosensor based on bovine leukemia virus proteins immobilized on the ZnO nanorods. *Sens Actuators B*. 2019;285:601-606.

67. Scott AH, Ron OD. Molecular dynamics simulation for all. *Neuron*. 2018;99(6):1129-1143.

68. Rafael CB, Marcelo CRM, Klaus S. Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochim Biophys Acta*. 2015;1850(5):872-877.

69. Hiqmet K. Faster protein folding using enhanced conformational sampling of molecular dynamics simulation. *J Mol Graph Model*. 2018;81: 32-49.

70. Debby DW, Le O-Y, Haoran X, Mengxu Z, Hong Y. Predicting the impacts of mutations on protein-ligand binding affinity based on molecular dynamics simulations and machine learning methods. *Comput Struct Biotechnol J*. 2020;18:439-454.

71. Nair PC, Miners JO. Molecular dynamics simulations: from structure function relationships to drug discovery. *In Silico Pharmacol*. 2014;2 (1):4.

72. Liu X, Shi D, Zhou S, Liu H, Liu H, Yao X. Molecular dynamics simulations and novel drug discovery. *Expert Opin Drug Discovy*. 2018;13(1):23-37.

73. Okimoto N, Futatsugi N, Fuji H, et al. High-performance drug discovery: computational screening by combining docking and molecular dynamics simulations. *PLOS Computat Biol*. 2009;5(10):e1000528.

74. Zheng Z, Li J, Sun J, et al. Inhibition of HBV replication by theophylline. *Antiviral Res*. 2011;89(2):149-155.

75. Sarma P, Shekhar N, Prajapat M, et al. In-silico homology assisted identification of inhibitor of RNA binding against 2019-nCoV N-protein (N terminal domain). *J Biomol Struct Dyn*. 2020;1-9. https://www.tandfonline.com/doi/full/10.1080/07391102.2020.1753580.

76. Wierenga W. Antiviral and other bioactivities of pyrimidinones. *Pharmacol Ther*. 1985;30(1):67-89.

77. Skulnick HI, Weed SD, Eidson EE, Renis HE, Stringfellow DA, Wierenga W. Pyrimidinones. 1. 2-Amino-5-halo-6-aryl-4(3H)-pyrimidinones. Interferon-inducing antiviral agents. *J Med Chem*. 1985;28(12):1864-1869.

78. Sharma V, Chitranshi N, Agarwal AK. Significance and biological importance of pyrimidine in the microbial world. *Int J Med Chem*. 2014;2014:202784.

79. Seley-Radtke KL, Yates MK. The evolution of nucleoside analogue antivirals: a review for chemists and non-chemists. Part 1: early structural modifications to the nucleoside scaffold. *Antiviral Res*. 2018;154:66-86.

80. Deo RC. Machine learning in medicine. *Circulation*. 2015;132(20):1920-1930.

81. Wang H, Zhang L, Han J, Weinan E. DeePMD-kit: a deep learning package for many-body potential energy representation and molecular dynamics. *Comput Phys Commun*. 2018;228:178-184.

82. Abdullah WFH, Othman M, Ali MAM, Islam MS. Improving ion-sensitive field-effect transistor selectivity with backpropagation neural network. *WSEAS Trans Circuits SystArchiv*. 2010;9:700-712.

83. Shriyaa M, Diwakar S. Recruiting machine learning methods for molecular simulations of proteins. *Mol Simul*. 2018;44(11):891-904.

84. Wehmeyer C, Noé F. Time-lagged autoencoders: deep learning of slow collective variables for molecular kinetics. *J Chem Phys*. 2018;148(24):241703.

85. Bhowmik D, Gao S, Young MT, Ramanathan A. Deep clustering of protein folding simulations. *BMC Bioinform*. 2018;19(18):484.

86. Wang J, Olsson S, Wehmeyer C, et al. Machine learning of coarse-grained molecular dynamics force fields. *ACS Cent Sci*. 2019;5(5):755-767.

87. Li Y, Li H, Pickard FC, et al. Machine learning force field parameters from ab initio data. *J Chem Theor Comput*. 2017;13(9):4492-4503.

88. Jinnouchi R, Karsai F, Kresse G. On-the-fly machine learning force field generation: application to melting points. *Phys Rev B*. 2019;100(1):014105.

89. Li Z, Kermode JR, De Vita A. Molecular dynamics with on-the-Fly machine learning of quantum-mechanical forces. *Phys Rev Lett*. 2015;114(9):096405.

90. Zhang L, Han J, Wang H, Car R, Weina E. Deep potential molecular dynamics: a scalable model with the accuracy of quantum mechanics. *Phys Rev Lett*. 2018;120:143001-1-143001-6.

91. Alimadadi A, Aryal S, Manandhar I, Munroe PB, Joe B, Cheng X. Artificial intelligence and machine learning to fight COVID-19. *Physiol Genomics*. 2020;52(4):200-202.

92. Torrisi M, Pollastri G, Le Q. Deep learning methods in protein structure prediction. *Comput Struct Biotechnol J*. 2020;18:1301-1310.

93. Ge Y, Tian T, Huang S, et al. A data-driven drug repositioning framework discovered a potential therapeutic agent targeting COVID-19. *bioRxiv*. 2020;2020.986836. https://www.biorxiv.org/content/10.1101/2020.03.11.986836v1.abstract.

94. Senior AW, Evans R, Jumper J, et al. Improved protein structure prediction using potentials from deep learning. *Nature*. 2020;577(7792):706-710.

95. MacKerell AD Jr, Banavali N, Foloppe N. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers*. 2000;56(4):257-265.

96. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci*. 2001;98(18):10037.

97. Kalé Laxmikant SR, Milind B, Robert B, et al. NAMD2: greater scalability for parallel molecular dynamics. *Journal of Computational Physics*. 1999;151(1):283-312.