



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.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



# Targeting the SARS-CoV2 nucleocapsid protein for potential therapeutics using immuno-informatics and structure-based drug discovery techniques

Alexander Kwarteng<sup>a,b,\*</sup>, Ebenezer Asiedu<sup>b</sup>, Samuel Amoah Sakyi<sup>c</sup>, Samuel Opoku Asiedu<sup>b</sup>

<sup>a</sup> Department of Biochemistry and Biotechnology, Kwame Nkrumah University of Science and Technology (KNUST), Kumasi, Ghana

<sup>b</sup> Kumasi Centre for Collaborative Research in Tropical Medicine (KCCR), Kumasi, Ghana

<sup>c</sup> Department of Molecular Medicine, Kwame Nkrumah University of Science and Technology (KNUST), Kumasi, Ghana

## ARTICLE INFO

### Keywords:

COVID-19  
Vaccine  
Immuno-informatics  
Epitope  
Molecular dynamics simulation  
Zidovudine

## ABSTRACT

The occurrence of the SARS-CoV2 infection has become a worldwide threat and the urgent need to discover therapeutic interventions remains paramount. The primary roles of the coronavirus nucleocapsid (N) protein are to interact with the viral genome and pack them into ribonucleoprotein complex. It also plays critical roles at many stages of the viral life cycle. Herein, we explore the N protein of SARS-CoV2 to identify promising epitope-based vaccine candidates and target the N-terminal domain of SARS-CoV2 N-protein for potential inhibitors using an integrative bioinformatics approach. We identified B-cell epitopes and T-cell epitopes that are non-toxic, non-allergenic, capable of inducing IFN- $\gamma$  and structurally stable with high global population coverage of response. The <sup>404</sup>SKQLQSMSSADS<sup>416</sup> and <sup>92</sup>RRIRGGDGKMKDL<sup>104</sup> sequences of N-protein were identified to induce B-cell immunity. We also identified <sup>79</sup>SPPDDQIGY<sup>87</sup> and <sup>305</sup>AQFAPSASAFFGMSR<sup>319</sup> as potential T-cell epitopes that form stable structures with human leucocyte antigens. We have also identified zidovudine triphosphate, an anti-HIV agent, as a potential inhibitor of the N-terminal domain of SARS-CoV2 N-protein based on docking and simulation analysis and should be considered for experimental validations. The findings of this study can help fast-track the discovery of therapeutic options to combat COVID-19.

## 1. Introduction

The global incidence of the coronavirus disease 2019 (COVID-19) has an estimated 20 million cases with over 740, 000 deaths as per current data available on 12th August 2020. The disease is caused by the severe acute respiratory syndrome coronavirus-2 (SARS-CoV2), the newest member of the family *Coronaviridae*. SARS-CoV2, SARS-CoV1 and Middle East Severe Acute Respiratory Syndrome Coronavirus (MERS-CoV) belong to the genus *Betacoronavirus*, thus, share particular genomic and proteomic similarities. It has been reported that SARS-CoV2 genome shares approximately 80 % and 50 % similarity with SARS-CoV1 and MERS-CoV, respectively [1,2]. The structural proteins and enzymes of SARS-CoV2 are 90 % homologous to those of SARS-CoV1 and MERS-CoV [3]. The clinical symptoms at the early stages of the disease include fever, fatigue and dry cough [4,5]. Progression of the disease to a severe case is manifested as pneumonia and is usually accompanied by dyspnea and dry cough [6]. There are several complications associated with the SARS disease which include acute respiratory distress syndrome (ARDS), organ dysfunction, cardiac injury

and systemic inflammation [4,6]. The structural proteins of SARS-CoV2 namely; spike (S) glycoprotein, envelope (E), membrane (M) and nucleocapsid (N) proteins, play crucial roles during entry and survival of the virus in the host cell. The primary roles of the coronavirus N protein are to interact with the viral genome and pack them into ribonucleoprotein (RNP) structure [7–9]. Notwithstanding, the coronavirus N protein is multifunctional, having critical roles at many stages of the viral life cycle [10]. Notable among them are viral budding [10], viral assembly [11], regulation of host cell cycle [12] and regulation of viral mRNA replication [13,14]. Considering the roles of the N protein and its high expression during viral infections, the N protein can be a promising target for drug and vaccine development. Targeting the N-protein provides opportunities to interfere with coronavirus assembly, transcription and replication. With the progression of the COVID-19 pandemic, search for pharmaceutical options are mainly driven by drug repurposing to reduce the drug discovery timeline [15].

Considering the novelty and timing of COVID-19, research groups are beginning to understand aspects of the immunological response to the SARS-CoV2 to aid in the discovery of vaccines. The quest for safe and

\* Corresponding author at: Department of Biochemistry and Biotechnology, KNUST, Kumasi, Ghana.

E-mail address: [akwarteng@knust.edu.gh](mailto:akwarteng@knust.edu.gh) (A. Kwarteng).

<https://doi.org/10.1016/j.bioph.2020.110914>

Received 13 August 2020; Received in revised form 29 September 2020; Accepted 18 October 2020

Available online 20 October 2020

0753-3322/© 2020 The Author(s).

Published by Elsevier Masson SAS. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

effective vaccines for COVID-19 has resulted in potential candidates with varying strategies and stages of developments. According to reports by the World Health Organization (WHO) on 10th August 2020, 28 vaccine candidates have been registered and are currently in clinical evaluations and 139 candidates in pre-clinical evaluations [16]. Many of these candidates are sub-unit vaccines, as they provide a more logistically workable approach [17] and are capable of inducing both cellular and humoral responses [18]. Majority of the candidates in clinical evaluations target the S protein, considering its immuno-dominant nature. However, there are reports that depict the N-protein as highly immunogenic with antigenic regions that are highly conserved across the coronaviruses [19–24]. The conventional approaches for epitope-based vaccine identification are resource-demanding, thus, designing epitope-based vaccines using bioinformatics offer a safer, inexpensive and more effective approach. In this paper, we integrate immuno-informatics, structure-based drug discovery and other computational methods to identify potential epitope-based vaccine and drug candidates that target the N-protein of SARS-CoV2. The results reveal potential N-protein epitopes that are highly immunogenic and elicit B-cell immunity, T-cell immunity, virus-specific antibody neutralization and natural killer cell-mediated immunity. The anti-HIV agent, zidovudine triphosphate was identified to potentially inhibit the N-terminal domain of SARS-CoV2 N-protein based on binding energy and simulation analysis.

## 2. Methods

### 2.1. Immuno-informatics approach

#### 2.1.1. Protein structure preparations

The SARS-CoV2 nucleocapsid protein (N-protein) sequence was fetched from the NCBI protein database (<https://www.ncbi.nlm.nih.gov/protein/QIC53221.1>) and human leucocyte antigen (HLA) protein sequences were retrieved from the UniProt database (<https://www.uniprot.org/>). We used the SWISS-MODEL server (<https://www.swissmodel.expasy.org/>) to generate structures of the SARS-CoV2 N-protein and the HLA proteins based on the co-ordinates of their respective homologous protein structures available at the protein data bank (<http://www.rcsb.org/>).

#### 2.1.2. Epitope prediction and assessment

**2.1.2.1. B-cell epitopes.** The linear and discontinuous B-cell epitopes were predicted using Bepipred 2.0 (<http://www.cbs.dtu.dk/services/BepiPred/index.php>) and ElliPro (<http://tools.iedb.org/ellipro/>), respectively. We used the interface provided by the IEBD web server (<http://www.iedb.org/>) for the predictions. Bepipred 2.0 uses a Random Forest regression algorithm to predict the probability that a residue of the antigenic protein is a member of an epitope [25]. The algorithm considers several properties of the antigenic residues including surface accessibility, hydrophobicity, secondary structure and polarity [25]. ElliPro combines three algorithms to predict discontinuous epitopes from a protein structure [26]. The first algorithm reconfigures the shape and surface of the query protein structure into an ellipsoid. Based on the ellipsoid shape, another algorithm estimates the protrusion index (PI) of each residue such that when a particular residue is seen outside the ellipsoid shape after 70 % of the protein atoms are captured within the ellipsoid shape, then the PI of that residue is 0.7. The last algorithm then determines the discontinuous epitopes by clustering neighboring residues based on a PI cut-off.

**2.1.2.2. T-cell epitopes.** All T-cell epitopes were predicted using the IEBD web server (<http://www.iedb.org/>). The T-cell epitope predictions were based on the peptides affinity to major histocompatibility complexes (MHC) or HLA. Both class I MHC-binding peptides and class II

MHC-binding peptides were predicted. For the class I MHC-binding peptides, a reference set consisting of several alleles each of HLA-A, HLA-B and HLA-C was used. Similarly, a reference set of MHC class II alleles was used for the prediction of class II MHC-binding peptides. The IEBD recommended prediction algorithms were employed in both cases.

**2.1.2.3. Epitope assessment.** The predicted epitopes were assessed based on toxicity, allergenicity and ability to induce IFN- $\gamma$ . The toxicity of the peptides was predicted using a support vector machine-based algorithm developed by Gupta et al. [27]. We used AllerTop [28] to predict the antigenicity of each peptide. The ability of the epitopes to induce IFN- $\gamma$  was predicted by IFNepitope [29] using the hybrid approach (motif and support vector machine).

**2.1.2.4. Epitope modeling and stability studies.** Structural configurations of the selected epitopes were modeled using PEP-FOLD 3 [30]. The peptide models were then simulated using GROMACS 2020.3, following the protocol described in section 2.1.7. A 20 ns run was performed for each epitope and their root mean square deviation was computed using the GROMACS analysis tool.

#### 2.1.3. Population coverage analysis

The population coverage analysis tool available at IEBD (<http://tools.iedb.org/population>) was used to predict the percentage of individuals that would respond to the epitope(s) based on the HLA allele genotypic frequency within their ethnicities/areas. The frequency of HLA genotype within a population is provided by the Allele frequency database (<http://www.allelefreqencies.net/>).

#### 2.1.4. Epitope conservation analysis

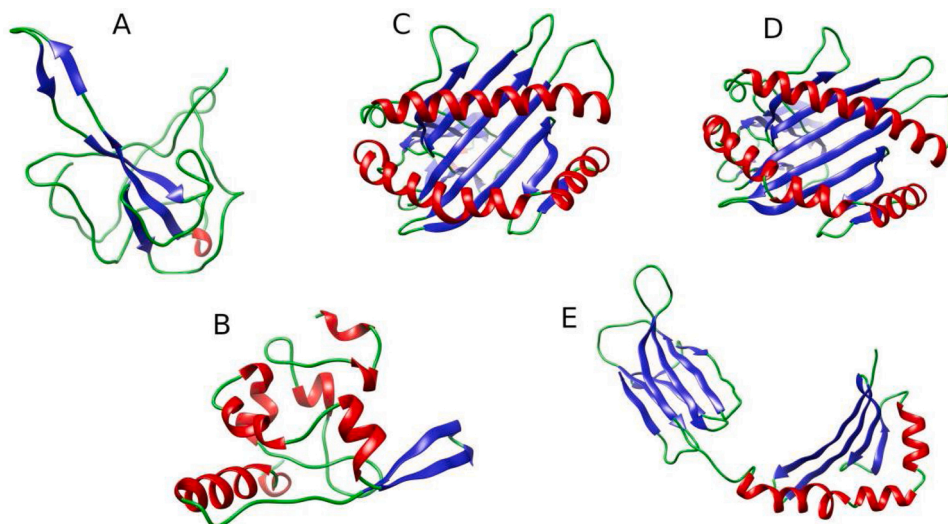
The occurrence of the epitope sequence in other related coronaviruses was analyzed using Clustal omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>). The N-protein of SARS-CoV1, MERS-CoV, Human coronaviruses (hCoV) OC43 and NL63 were considered for the homology analysis, noting the conservation of the epitope sequences across the related coronaviruses.

#### 2.1.5. Molecular docking of epitopes with HLA proteins

The best binding conformation of the epitope-HLA complex was determined with Cluspro 2.0 server [31] using the antibody model [32]. The Cluspro algorithm performs three steps in the protein-protein binding prediction [31]. The algorithm performs a rigid-body docking on a very large set of conformations using a Fast Fourier Transform (FFT)-dependent algorithm called PIPER [33,34]. The docked conformations are clustered based on root mean square deviation to select representative conformations of lowest energy [31]. The last computational step is energy minimization of the top conformations from step 2 using CHARMM force field [35]. The server is available at <https://cluspro.bu.edu/home.php>.

#### 2.1.6. Molecular dynamics simulation of complexes

The proteins were subjected to molecular dynamics simulations using GROMACS 2020.3 software suite [36] and CHARMM 2019 force field. The structures were centered in a dodecahedron box, at least 1 nm from the edges and solvated with spc216 water model. The system was electrostatically neutralized with 0.154 M NaCl. In order to get rid of inappropriate geometry and steric hindrances, energy minimization was performed on the system using the steepest descent algorithm for 10,000 steps with a maximum force threshold of 100 kJ/mol/nm. Van der-Waals interactions were treated with a single cut-off of 1.216 nm and long-range electrostatics were treated with the Particle-Mesh Ewald (PME) method with 0.16 FFT grid spacing and 4th order B-spline interpolation. Neighbor search was performed every 20 steps using the grid method with Verlet cut-off scheme. Protein and non-protein components of the system were independently coupled to v-rescale



**Fig. 1.** Structural models of N-protein and HLAs. (A) N-terminal domain of SARS-CoV2 N-protein (modeled based on PDB ID: 6M3M) (B) C-terminal domain of SARS-CoV2 N-protein based on PDB ID: 7C22 (C) HLA-A\*03:01 based on PDB ID: 6O9B (D) HLA-B\*35:01 based on PDB ID: 6AVG (E) HLA-DRB1\*09:01 based on PDB ID: 1A6A. The  $\beta$ -sheets (blue),  $\alpha$ -helix (red) and loops (green) are also shown.

thermostat and an isotropic Berendsen algorithm for pressure coupling. The pressure was maintained by weak coupling to a reference pressure of 1 bar, with an isothermal compressibility of  $4.6 \times 10^{-5} \text{ bar}^{-1}$ . The bond lengths within the protein were constrained using the LINCS algorithm. The system was subjected to a 400 ps NVT and NPT equilibration, followed by 200 ns simulation using the leap-frog integrator. The trajectories were analyzed with built-in GROMACS commands and all graphic data were generated with the GRACE tool (<http://plasma-gate.weizmann.ac.il/Grace>).

#### 2.1.7. Immune simulation

The immune response that would be elicited by the predicted epitopes was simulated using the C-ImmSim server (<http://150.146.2.1/C-IMMSIM/index.php>). C-ImmSim is based on the models developed by Celada-Seiden [37–39] and operates as an agent-based model that uses machine learning techniques to simulate the immune response towards a particular pathogen. The players of the immune system are represented as agents and the interactions between the agents are guided by set of rules that emanate from present understanding of immunological principles. These principles that are incorporated into the C-ImmSim algorithm include MHC restriction [40], development of memory [41,42], clonal deletion theory [43], processing and presentation of antigens [44–46] among others [47]. The algorithm does well to mimic the immune system as it simultaneously simulates immunological processes in the bone marrow (production of lymphoid and myeloid cells), the thymus (T-cell education), and the lymphatic system (immune surveillance) [47]. Three injections were given according to the dose series of 0, 2 and 6 months. All simulation parameters were set at default with time steps set at 1, 180, and 360 (each time step represent 8 h in reality).

## 2.2. Structure-based drug discovery

### 2.2.1. Structure preparation

In our structure-based drug discovery approach, we targeted the N-terminal domain of SARS-CoV2 N-protein because the domain is reported to have binding pockets for ribonucleoside 5'-monophosphates such as adenosine 5'-monophosphate (AMP) [8]. We then retrieved an experimental structure of a related coronavirus in complex with AMP to aid in our analysis of the ligand's binding site in SARS-CoV2 N-protein. The receptor (SARS-CoV2 N-protein) was energy minimized for 10,000 steps using GROMACS 2020.3 and the low-energy structure was used for

the analysis.

### 2.2.2. AMP binding site analysis

The AMP binding site in the N-terminal domain of SARS-CoV2 N-protein was elucidated through alignment of protein structures using PyMol 2.2 [48]. The algorithm performs sequence alignment followed by a structural superposition, and then cycle refinement in order to reject structural outliers found during the fitting. Since we were also interested in the all-atom root mean square deviation, we turned off the refinement process.

### 2.2.3. Discovery of potential N-protein NTD inhibitors

The N-terminal domain of the SARS-CoV2 nucleocapsid protein was targeted for potential inhibitors using a library of approved antiviral drugs. The active site of the N-protein NTD accommodates nucleic acid substrates when packing the viral genome into a ribonucleoprotein complex. Targeting the N-protein can provide means to interrupt critical stages of the virus life cycle like virus assembly, transcription and replication processes. Thus, existing antiviral agents known to interfere with viral processes such as reverse transcription, virus integration, viral transcription and translation were considered because their mechanisms involve viral genetic materials. A set of 21 antiviral agents were randomly selected for this study. The selected ligands have a wide inhibitory spectrum including nucleoside analogues, nucleoside reverse transcriptase inhibitors, integrase inhibitors and acyclic guanosine analogues. The 3D structures of the drug candidates were generated with UCSF Chimera 1.14 [49] using their SMILE strings obtained from PubChem (<https://pubchem.ncbi.nlm.nih.gov/>). Ligands and receptor preparations were performed with AutoDock 4.2.6 tools [50]. AutoDock compatible structure files (pdbqt formats) were generated for both ligands and receptor after addition of hydrogen atoms and computation of Gasteiger charges. The Lamarckian genetic algorithm (LGA) was used and search parameters were set as follows; 50 GA runs, population size of 150, a maximum number of 25,000,000 energy evaluations, 27,000 number of generations, mutation rate of 0.02, and a crossover rate of 0.8. Grid was centered on the binding pocket of AMP and the size configured with dimensions 54, 52 and 48 along the x, y, and z-axis, respectively and 0.375 Å spacing. RMS cluster tolerance was set to 2.0 Å. We clustered the docking conformations based on root mean square deviation and a tolerance of 2.0 Å. The docked conformations were evaluated on the basis of binding energy and frequency of pose (cluster).

**Table 1**

Evaluation and characterization of epitopes. The predicted HLA allele that binds well to the MHC-restricted epitopes are also provided.

Candidate	<i>B-cell epitope</i> Peptide	Length	Toxicity	Allergenicity	RMSD (nm)		
1	AGLPYGANK	9	Non-toxin	Probable non-allergen	0.47		
2	SKQLQQSMSSADS	13	<b>Non-toxin</b>	<b>Probable non-allergen</b>	<b>0.14</b>		
3	RRIRGGDGKMKDL	13	<b>Non-toxin</b>	<b>Probable non-allergen</b>	<b>0.24</b>		
4	LPQGTTLPKGF	11	Non-toxin	Probable non-allergen	0.39		
<i>MHC I</i>							
	Peptide	length	Toxicity	Allergenicity	IFN- $\gamma$ induction	RMSD (nm)	HLA allele
1	KTFPTEPK	9	Non-toxin	Probable non-allergen	Positive	0.52	HLA-A*11:01, HLA-A*03:01, HLA-A*30:01, HLA-A*31:01, HLA-A*68:01, HLA-A*32:01, HLA-A*30:02, HLA-A*33:01, HLA-B*57:01, HLA-B*58:01, HLA-A*26:01, HLA-A*01:01, HLA-B*15:01, HLA-B*07:02
2	SSPDDQIGY	9	<b>Non-toxin</b>	<b>Probable non-allergen</b>	<b>Positive</b>	<b>0.26</b>	<b>HLA-A*01:01, HLA-A*26:01, HLA-A*30:02, HLA-B*35:01, HLA-B*58:01, HLA-B*15:01, HLA-B*57:01, HLA-A*68:01, HLA-B*53:01, HLA-A*32:01, HLA-A*11:01, HLA-A*03:01, HLA-B*44:03, HLA-B*44:02, HLA-A*68:02, HLA-A*33:01</b>
3	SSPDDQIGYY	10	Non-toxin	Probable non-allergen	Positive	0.35	HLA-A*01:01, HLA-A*30:02, HLA-B*35:01, HLA-A*26:01, HLA-B*53:01, HLA-B*57:01, HLA-B*58:01, HLA-B*15:01, HLA-A*11:01, HLA-A*68:01, HLA-B*07:02, HLA-B*44:03, HLA-A*03:01, HLA-B*44:02, HLA-A*32:01, HLA-B*51:01
4	VTPSGTWLTY	10	Non-toxin	Probable non-allergen	Positive	0.3	HLA-A*01:01, HLA-A*30:02, HLA-B*35:01, HLA-A*26:01, HLA-B*53:01, HLA-B*15:01, HLA-B*58:01, HLA-B*57:01, HLA-A*11:01, HLA-A*32:01, HLA-B*07:02, HLA-A*03:01, HLA-A*68:01, HLA-A*24:02, HLA-B*51:01, HLA-A*23:01, HLA-A*30:01, HLA-A*33:01, HLA-A*31:01, HLA-B*44:03, HLA-B*44:02
<i>MHC II</i>							
1	AQFAPSASAFFGMSR	15	<b>Non-toxin</b>	<b>Probable non-allergen</b>	<b>Positive</b>	<b>0.23</b>	<b>HLA-DRB5*01:01, HLA-DRB5*01:01, HLA-DRB4*01:01, HLA-DRB4*01:01, HLA-DRB1*11:01, HLA-DRB1*12:01, HLA-DRB1*01:01</b>
2	IAQFAPSASAFFGMS	15	<b>Non-toxin</b>	<b>Probable non-allergen</b>	<b>Positive</b>	<b>0.24</b>	<b>HLA-DQA1*01:01/DQB1*05:01, HLA-DQA1*03:01/DQB1*03:02, HLA-DRB1*11:01, HLA-DRB1*01:01, HLA-DRB3*01:01, HLA-DPA1*02:01/DPB1*14:01, HLA-DRB1*09:01, HLA-DRB1*12:01</b>
3	PQIAQFAPSASAFFG	15	<b>Non-toxin</b>	<b>Probable non-allergen</b>	<b>Positive</b>	<b>0.24</b>	<b>HLA-DRB1*04:05, HLA-DRB1*04:01, HLA-DRB4*01:01, HLA-DQA1*04:01/DQB1*04:02, HLA-DRB3*01:01, HLA-DQA1*04:01/DQB1*04:02, HLA-DRB3*02:02, HLA-DRB1*15:01, HLA-DPA1*02:01/DPB1*05:01</b>
4	QIAQFAPSASAFFGM	15	Non-toxin	Probable non-allergen	Positive	0.37	HLA-DPA1*01:03/DPB1*02:01, HLA-DQA1*01:02/DQB1*06:02, HLA-DPA1*02:01/DPB1*01:01, HLA-DPA1*02:01/DPB1*01:01, HLA-DPA1*03:01/DPB1*04:02, HLA-DQA1*03:01/DQB1*03:02, HLA-DRB5*01:01, HLA-DPA1*03:01/DPB1*04:02, HLA-DRB5*01:01
5	WPQIAQFAPSASAFF	15	Non-toxin	Probable non-allergen	Positive	0.29	HLA-DRB4*01:01, HLA-DPA1*02:01/DPB1*14:01, HLA-DRB4*01:01, HLA-DRB1*11:01, HLA-DRB1*04:01, HLA-DPA1*01:03/DPB1*02:01, HLA-DPA1*02:01/DPB1*01:01, HLA-DRB5*01:01

RMSD was calculated after least square fitting of C $\alpha$  atoms. The epitope candidates considered for further analysis are in bold.

#### 2.2.4. Binding energy analysis: MM/PBSA approach

The energy terms governing the protein-ligand interactions were analyzed using the MM-PBSA (Molecular Mechanics with Poisson-Boltzmann Surface Area) approach [51]. All energy terms were calculated by the g\_mmpbsa 5.1.2 tool [52] based on simulation trajectories. We used GROMACS 5.1.5 (compatible with the g\_mmpbsa tool) to run all the simulations following the protocol discussed in section 2.1.7 with slight modifications. We performed NVT and NPT equilibrations for 200 ps each under constraint conditions followed by 5 independent simulations each lasting for 1 ns with a sampling frequency of 10 ps. Frames from the last 500 ps of each run were used for the binding energy calculations.

#### 2.2.5. Simulation of protein-ligand complexes

The behavior of the protein-ligand complexes at the atomistic level was elucidated with molecular dynamics simulation using the method described in section 2.1.7 with some modifications. Ligand topologies were generated by the Swiss Param server (<http://www.swissparam.ch/>

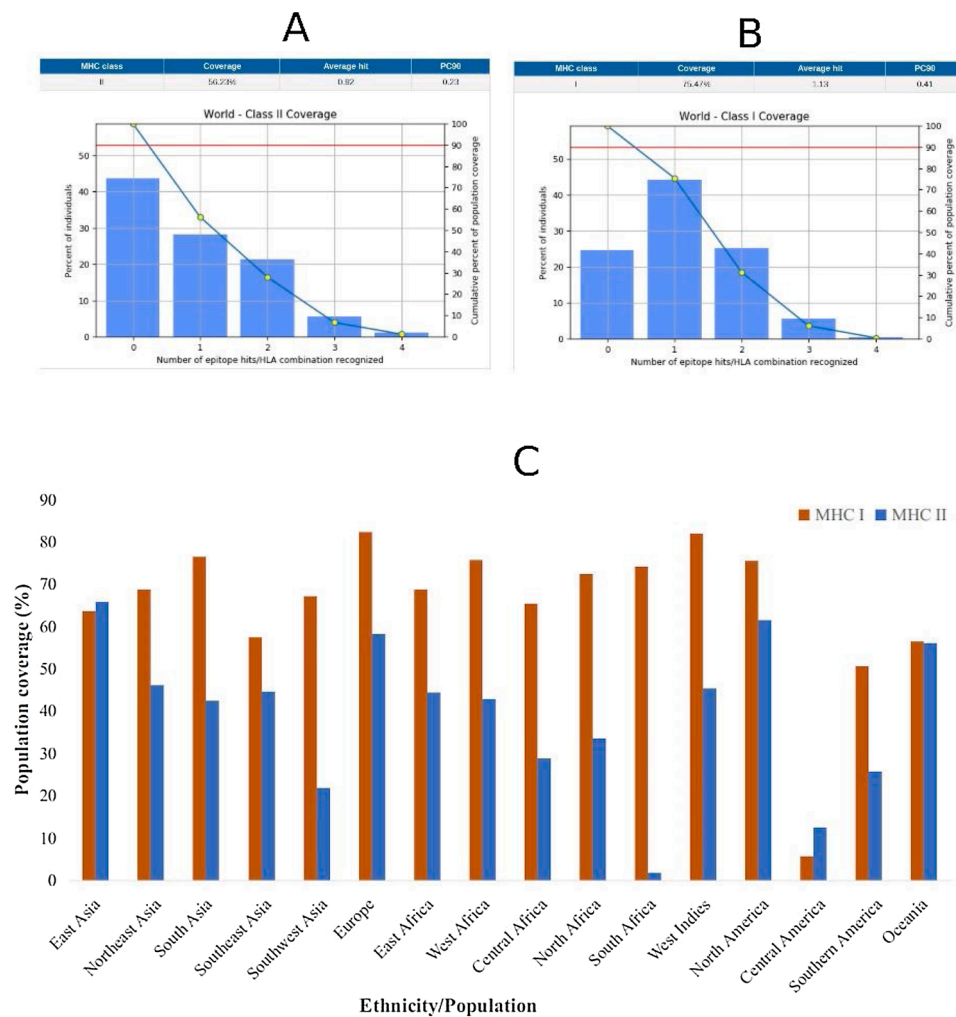
). Each complex was simulated for 100 ns. RMSD calculation was performed with the 100 ns simulation trajectory while the last 80 ns simulation trajectory was used for all other protein-ligand complex analysis.

### 3. Results

#### 3.1. Immuno-informatics approach

##### 3.1.1. Prediction of B-cell epitopes

To determine the regions of SARS-CoV2 N-protein that may induce B-cell immunity, we predicted both linear and discontinuous B-cell epitopes using BepiPred 2.0 and Ellipro, respectively. The N-protein sequence obtained from the NCBI protein database (Accession number: QIC53221.1) was used for the prediction of the linear epitopes. The discontinuous epitopes, on the other hand, were predicted using a 3D structural conformation of the SARS-CoV2 N-protein. Structures of N-protein C-terminal domain and N-terminal domain were modeled as



**Fig. 2.** Population coverage of MHC restricted epitope response. (A) Global population coverage of MHC I-restricted epitopes (B) Global population coverage of MHC II-restricted epitopes (C) Population coverage of MHC I and MHC II epitopes in particular ethnicities across the world.

separate molecules using SWISS-MODEL (Fig. 1). The built structures were used to predict the discontinuous epitopes. Using a prediction cut-off of 0.5, the Bepipred 2.0 server predicted 11 candidates of varying residual length ranging from 6 to 52. The Ellipro server predicted 5 and 7 epitope candidates mapped to the C-terminal and N-terminal domains, respectively.

### 3.1.2. Prediction of T-cell epitopes

We also predicted epitopes that may induce T-cell response against SARS-CoV2, based on ability to interact with specific HLA alleles. With over 3400 predicted MHC class I restricted epitopes, consideration was given to those below a percentile rank of 0.5%. The percentile rank represents a randomly selected number of epitopes that are graded better than the candidate of interest, hence, a lower percentile rank is preferred.

### 3.1.3. Assessment of epitopes

As potential vaccine candidates, the predicted epitopes were profiled according to their toxicity, allergenicity and ability to induce IFN- $\gamma$ . For the B-cell epitopes, preference was given to the non-toxins and non-allergen candidates with 8–11 amino acid residues. This resulted in 4 B-cell epitope candidates. The MHC-restricted epitopes that were predicted to be non-toxic, non-allergic and capable of inducing IFN- $\gamma$  were considered for further analysis. This resulted in a total of 9 MHC-restricted epitopes. The sequences of the selected epitopes after the evaluation process are shown in Table 1.

### 3.1.4. Epitope stability studies

We performed molecular dynamics simulation for all the selected epitopes in order to inspect the stability of the peptides at the atomistic level. Prior to that, we built structural models for the epitopes using the PEP-FOLD3.5 server. The root mean square deviation (RMSD) from a 20 ns simulation was computed for each peptide and the results are shown in Table 1. The RMSD tool compares the simulating model to the initial structural configuration (reference structure) and computes the deviation between the two structures. The epitopes with low RMSD values were selected for further analysis.

### 3.1.5. Population coverage analysis of MHC restricted epitopes

The aim of the population coverage analysis was to predict the percentage of a population that would respond to the epitopes, considering the frequency of HLA genotypic traits in the population. In general, a higher fraction of the population was predicted to respond to the MHC I-restricted epitope than the MHC II-restricted epitopes. Considering the global population, 75.47% of individuals would respond to the MHC I-restricted epitope while 56.23% of the individuals would respond to the MHC II-restricted epitopes (Fig. 2A and B). The lowest population coverage for the epitopes (both MHC I and II) was recorded in Central America and the highest recorded in East Asia (Fig. 2C). The South African population did not respond very well to the MHC II-restricted epitopes, although there was a significant population coverage in other regions of Africa. The best coverage of the MHC I-restricted epitope was observed in Europe, South Asia, Africa, North and Southern America

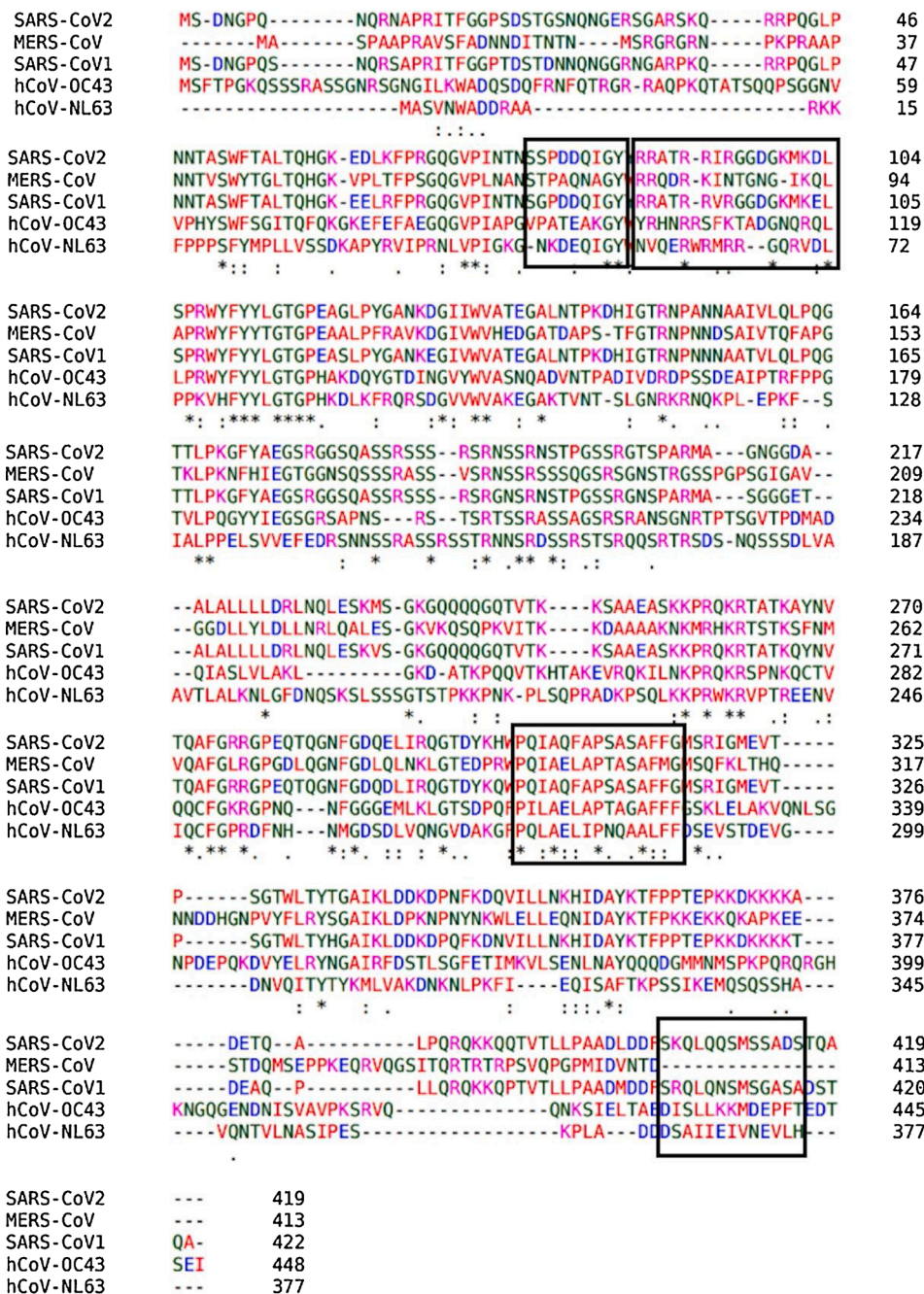


Fig. 3. Multiple sequence alignment of SARS-CoV2 N-protein with other related coronaviruses. The sequences in the rectangular shape are peptide sequence of epitopes.

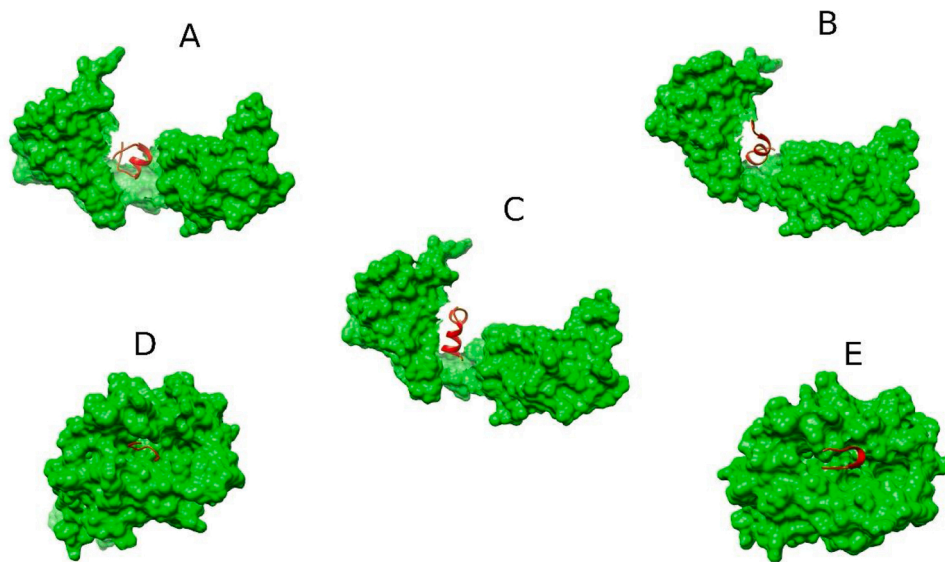
and the Oceania.

### 3.1.6. Epitope sequence conservation analysis

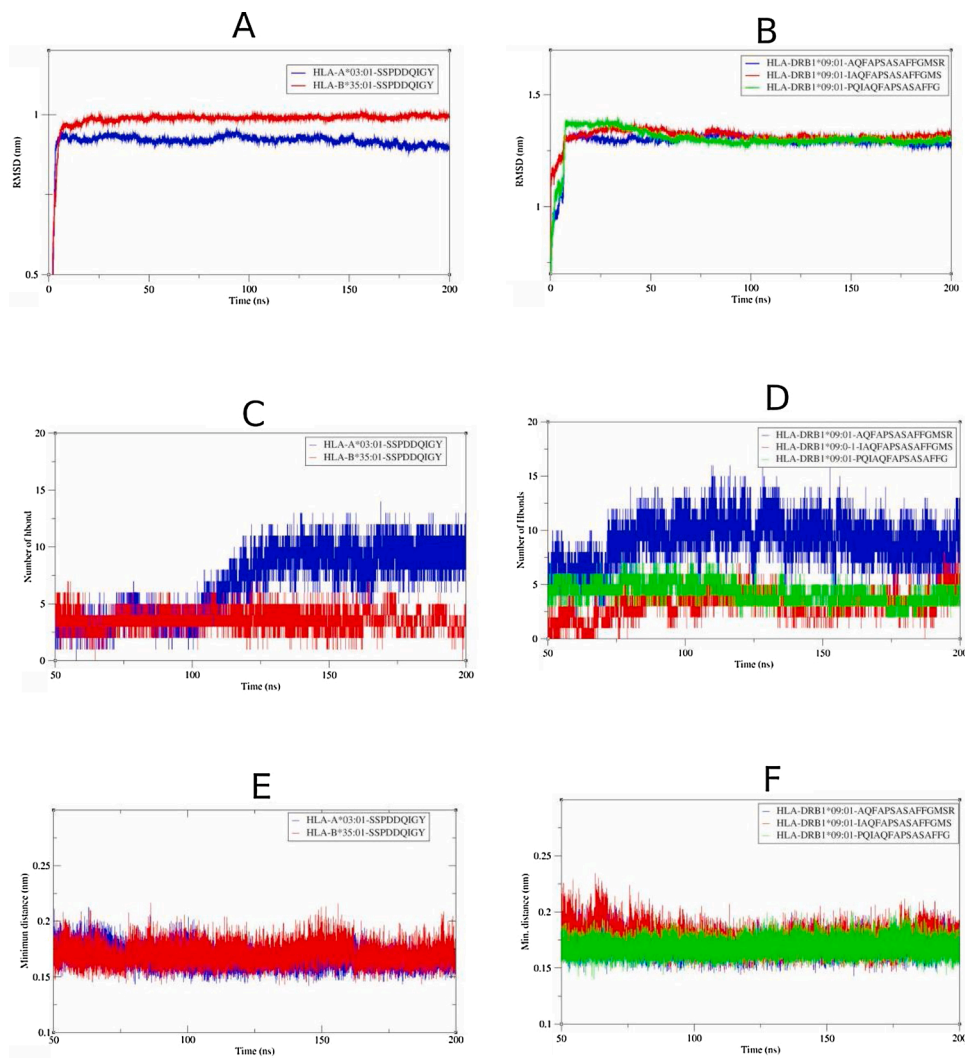
In order to check the cross-reactivity of the predicted epitopes across other related coronaviruses, a multiple sequence alignment study was performed using Clustal omega server. The N-protein sequence of SARS-CoV2 was aligned with N-protein sequences of SARS-CoV1, MERS-CoV, hCoV-OC43 and hCoV-NL63. The alignments of the protein sequences are shown Fig. 3. Interestingly, the sequences of the MHCII-restricted epitopes were highly homologous to sequences of all the related coronaviruses except hCoV-NL63 and the MHCI-restricted epitope sequence was highly homologous to only the SARS-CoV1 N protein. The linear B-cell epitope, on other hand, shared low homology to N protein of SARS-CoV1 but none to the other related coronaviruses.

### 3.1.7. Structural studies of HLA-epitope complexes

The interaction between the selected epitopes and their respective HLA proteins were studied using docking and simulation analysis. We predicted the binding conformations of the T-cell epitopes to HLA alleles using the Cluspro 2 server. The best binding conformations of the T-cell epitopes to their respective HLA alleles are shown in Fig. 4. We used HLA-A\*03:01 and HLA-B\*35:01 as the representative MHC class I proteins and HLA-DRB1\*09:01 as the representative MHC class II protein. To further characterize the HLA-epitope interactions, we evaluated the structural properties and behavior of the complexes from a simulation trajectory using root mean square deviation, number of hydrogen bonds, and minimum distance. The structural properties and dynamics of the HLA-epitope complexes from 200 ns simulations are shown in Fig. 5. The HLA-A\*03:01-<sup>79</sup>SSPDDQIGY<sup>87</sup> complex deviated less from its reference

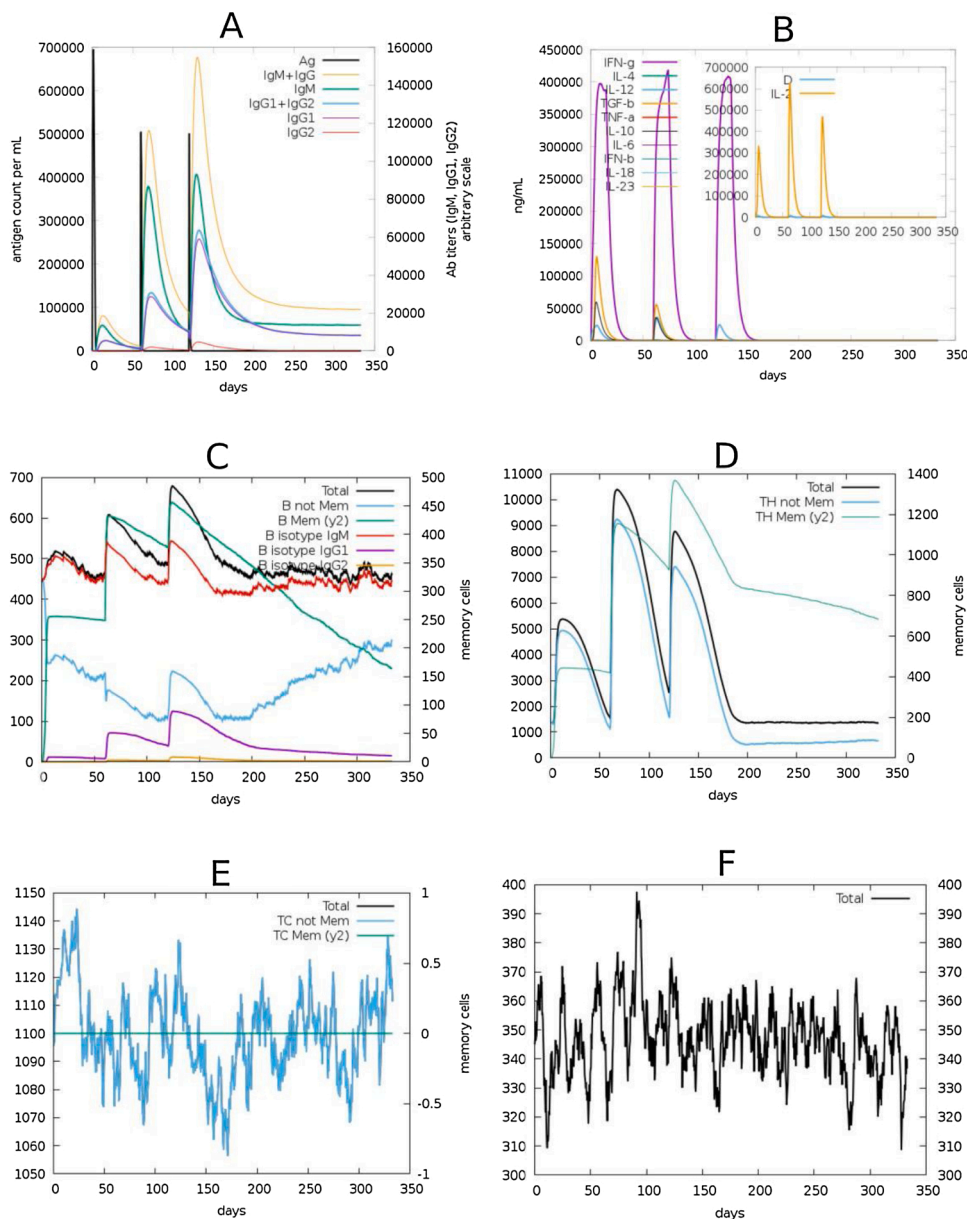


**Fig. 4.** Predicted binding conformations of MHC-restricted epitopes to HLA proteins. (A) HLA-DRB1\*09:01-AQFAPSASAFFGMSR complex (B) HLA-DRB1\*09:01-IAQFAPSASAFFGMS complex (C) HLA-DRB1\*09:01-PQIAQFAPSASAFFG complex (D) HLA-A\*03:01-SSPDDQIGY complex (E) HLA-B\*35:01-SSPDDQIGY complex. The 3D structures of the epitopes (red) and the surfaces of the HLA proteins (green) are shown.



**Fig. 5.** Molecular dynamics simulation analysis of MHC I-restricted epitope complexes and MHC II-epitope complexes. The MHC I-restricted epitope (<sup>78</sup>SSPDDQIGY<sup>86</sup>) is complexed to either HLA-A\*03:01 or HLA-B\*35:01 and the MHC II-restricted epitopes (<sup>305</sup>AQFAPSASAFFGMSR<sup>319</sup>, <sup>304</sup>IAQFAPSASAFFGMS<sup>318</sup> and <sup>302</sup>PQIAQFAPSASAFFG<sup>316</sup>) are complexed to HLA-DRB1\*09:01. (A, B) RMSD. (C, D) number of hydrogen bonds. (E, F) Evolution of the minimum distance between epitope and HLA.





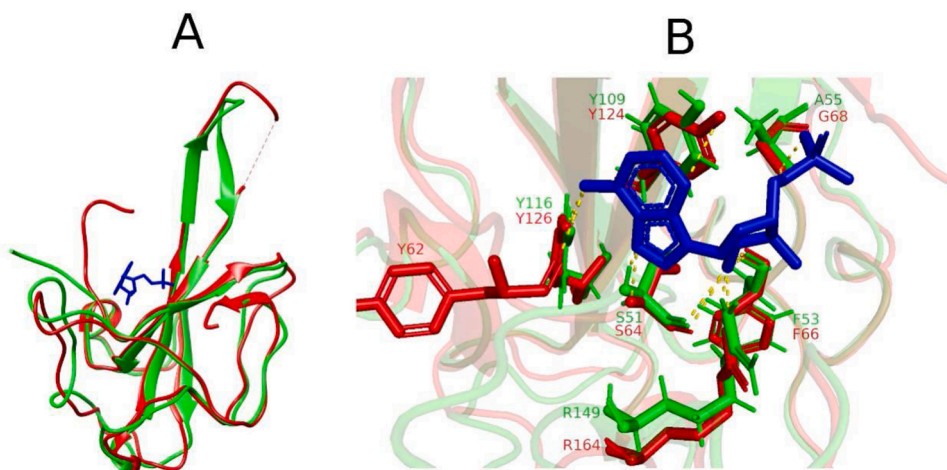
**Fig. 6.** Simulation of immune responses to SARS-CoV2 N-protein. **(A)** Virus-specific antibody responses along with the antigen counts are shown. **(B)** The various cytokines released against the antigen. Insert plot shows danger signals and IL-2 levels. **(C)** The different forms of B-cells including the memory B-cells are shown. **(D)** T-helper cell counts and kinetics of memory T-cells **(E)** T-cytotoxic cell levels **(F)** Evolution of natural killer cells.

structure during the simulation than the HLA-B\*35:01-<sup>79</sup>SSPDDQIGY<sup>87</sup> complex (Fig. 5A) and a higher number of hydrogen bonds were formed between <sup>79</sup>SSPDDQIGY<sup>87</sup> and HLA-A\*03:01 (Fig. 5C), suggesting a better interaction and a more stable complex when compared to the HLA-B\*35:01-<sup>79</sup>SSPDDQIGY<sup>87</sup> complex. The MHCII-restricted epitopes <sup>305</sup>AQFAPSASAFFGMSR<sup>319</sup>, <sup>304</sup>IAQFAPSASAFFGMS<sup>318</sup> and <sup>302</sup>PQIAQFAPSASAFFG<sup>316</sup> were docked to HLA-DRB1\*09:01 and simulated to study the epitope-HLA interactions. Although structural deviation was not significantly different for the MHCII-epitope complexes (Fig. 5B), the HLA-DRB1\*09:01-<sup>305</sup>AQFAPSASAFFGMSR<sup>319</sup> complex seems to form the highest number of hydrogen bonds among the MHCII-restricted epitopes with  $\approx 11$  hydrogen bonds (Fig. 5D). The results also suggest that the epitopes remained in close proximity to the HLA alleles with minimum distances ranging from 0.15 to 0.20 nm (Fig. 5E, F).

### 3.1.8. Simulation of immune response

We used the C-ImmSim server to simulate the immune response to

the SARS-CoV2 N-protein in order to predict the immune response profile of the selected epitopes. The response elicited by antibodies, B-cells, T-cells and cytokines are shown in Fig. 6. We simulated the response for a period of 1 year after three series of antigen injections. The second dose was given 2 months after the first and the third given in the 6th month (0, 2, 6 months). High levels of IgG + IgM immunocomplex and IgM were observed. The levels of these antibodies were low after the first injection but increased significantly following subsequent doses and this was accompanied by a concomitant reduction of antigen levels (Fig. 6A). The major cytokines observed were IFN- $\gamma$  and IL-2 (Fig. 6B). Protection from B-cells and T-cells were notable along with development of immune memory. The immune memory developed by both B-cells and T-cells lasted for the period of the simulation (Fig. 6C, D). Other immune cells, including T-cytotoxic cells and natural killer cells, also elicited strong immune response to the antigen (Fig. 6E, F).



**Fig. 7.** Binding site analysis of AMP. (A) Structural superposition of SARS-CoV2 N-protein N-terminal domain model (green) to crystal structure of hCoV N-protein N-terminal domain (red) in complex to AMP (blue). (B) Superposition of active sites showing polar interactions (yellow dashes) between AMP and active site residues in SARS-CoV2 (green) and hCoV (red).

### 3.2. Structure-based drug design

#### 3.2.1. Ligand binding site analysis

The N-terminal domain (NTD) of the nucleocapsid protein of the coronaviruses has a binding pocket for ribonucleoside 5'-monophosphates and forms complexes with them during the packing of viral RNA. While, the crystal structure of SARS-CoV2 NTD in complex with ribonucleoside 5'-monophosphates remains to be elucidated, the NTD of hCoV-OC43 N-protein in complex with adenosine 5'-monophosphate (AMP) has been solved and available at the PDB (ID: 4LI4). Using superposition and structural alignment analysis, we predicted the binding pocket of AMP in the SARS-CoV2 N-protein NTD based on the structural configuration of the PDB entry 4LI4. The structure of the N-protein NTD deviated from that of hCoV-OC43 with RMSD of 2.2 nm. The binding conformation of the AMP in its binding pocket was then determined through docking analysis. The structural alignment, superposition and AMP interactions at the binding site have been provided in Fig. 7. AMP has polar interactions with residues S51, F53, A55, Y109, Y116 and R149 in the SARS-CoV2 N-protein NTD. The corresponding interacting residues in the hCoV N-protein NTD are also shown (see Fig. 7B).

#### 3.2.2. Discovery of potential inhibitors the N-terminal domain

A library of 21 antiviral drugs with varying targets was used in our virtual screening analysis. The nomenclature, molecular weight, chemical structure and autodock score (estimated free energy of binding) are provided in Table 2. Among the studied ligands, pleconaril (PLE), raltegravir (RAL) and zidovudine triphosphate (ZTP) recorded higher free energy of binding than AMP. The binding energy is referred to as 'autodock score' because the value, although efficient at discriminating suitable ligands from non-suitable ones, does not accurately represent the binding affinity.

#### 3.2.3. MM/PBSA binding energy analysis

The molecular mechanics with Poisson-Boltzmann and surface area solvation (MM/PBSA) method of binding energy calculation is designed to be more accurate than the docking and scoring approach. For g\_mmpbsa binding energy calculations, three energy terms are used namely molecular mechanics, polar solvation and non-polar solvation energies. The molecular mechanics potential energy includes energy of bonded and non-bonded interactions (Van-der Waals and electrostatic energies). The polar solvation term is calculated by solving the Poisson-Boltzmann equation [52]. The non-polar solvation term was calculated using the solvent accessible surface area (SASA) model and indicates repulsion and attraction between solvent and solute [52]. The total

binding energy of ZTP (-1312.837 kJ/mol) is significantly higher as compared to AMP (-108.234 kJ/mol) with major contributions from the electrostatic terms in both cases (Table 3).

#### 3.2.4. Simulation of protein-ligand complex

The 100 ns simulation revealed that the NTD-ZTP complex seems to be the most stable with RMSD of  $\approx 0.3$  nm while the NTD-AMP complex recorded RMSD of  $\approx 0.35$  nm (Fig. 8A). Evolution of the total hydrogen bonds between the ligands and the NTD are shown in (Fig. 8B) and ranged from 2 to 8 hydrogen bonds for NTD-AMP and NTD-ZTP complexes. The minimum distance between ligands and the protein (Fig. 8C) ranged from 0.15 to 0.25 nm, suggesting close proximity especially for ZTP ( $\approx 0.15$  nm). The compactness of the ZTP-NTD complex seems to be significantly low as compared to AMP-ZTP complex (see Fig. 8D). Root Mean Square Fluctuations (RMSF) values for residues 91–102 (residues of the  $\beta$ -hairpin) were significantly higher upon binding of AMP than ZTP (Fig. 9A). The flexibility of the  $\beta$ -hairpin residues was similar for the NTD and NTD-AMP complex. ZTP occupied the binding pocket of AMP and interacts with residues T49, S51, F53, R88, Y109, Y111 and R149 (Fig. 9B).

## 4. Discussion

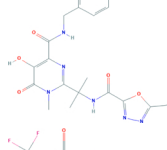
A number of B-cell and T-cell epitopes have been identified using immuno-informatics which were subjected to stringent assessment and evaluation analysis based on desirable immunogenic properties such as allergenicity, toxicity, IFN- $\gamma$  induction and structural stability. The N-protein epitopes <sup>404</sup>SKQLQSMSSADS<sup>416</sup> and <sup>92</sup>RRIRGGDGKMKDL<sup>104</sup> were selected as potential epitope-based vaccine candidates capable of inducing B-cell immunity against SARS-CoV2 infection. Based on evaluation performance, MHCII-restricted epitope <sup>79</sup>SSPDDQIGY<sup>87</sup> and MHCII-restricted epitopes <sup>305</sup>AQFAPSASAFFGMSR<sup>319</sup>, <sup>304</sup>IAQFAPSASAFFGMS<sup>318</sup> and <sup>302</sup>PQIAQFAPSASAFFG<sup>316</sup> were selected for further evaluations on the basis of cross reactivity across related coronaviruses and population coverage of their response.

A possibility of memory T-cell cross-reactivity specific to the S-protein has been suggested by Braun et al. [53] and a scenario of cross-reactivity in relations to SARS-CoV1 and endemic CoV has also been reported [54]. Based on sequence alignment analysis, our results suggest a possible MHCII-restricted epitope-induced T-cell cross-reactivity to SARS-CoV1, MERS-CoV, hCoV-OC43 and hCoV-NL63. Moreover, the HLA-I restricted epitope is predicted to induce T-cell immunity in SARS-CoV1.

Presentation of antigenic peptides by MHC for recognition by

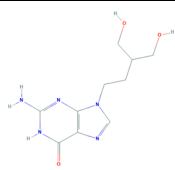
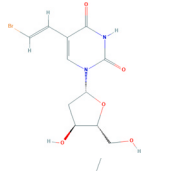
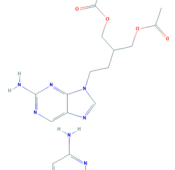
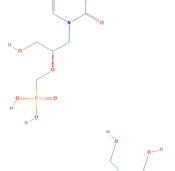
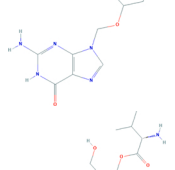
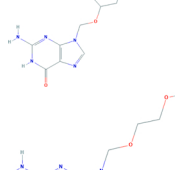
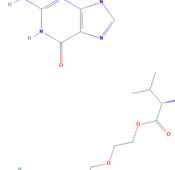
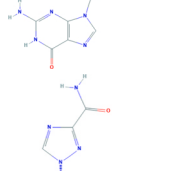

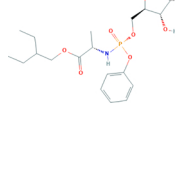
**Table 2**

List of ligands used for the virtual screening. Their binding energies based on autodock analysis are also shown.

Ligand	IUPAC nomenclature	M.W(g/mol)	Chemical representation	Binding energy (kcal/mol)
Adenosine 5'-monophosphate	[(2R,3S,4R,5R)-5-(6-aminopurin-9-yl)-3,4-dihydroxyoxolan-2-yl]methyl dihydrogen phosphate	347.22		-4.94
Amprenavir	[(3S)-oxolan-3-yl] N-[(2S,3R)-4-[(4-aminophenyl)sulfonyl-(2-methylpropyl)amino]-3-hydroxy-1-phenylbutan-2-yl]carbamate	505.6		-3.24
Entecavir	2-amino-9-[(1S,3R,4S)-4-hydroxy-3-(hydroxymethyl)-2-methylidencyclopentyl]-1H-purin-6-one	277.28		-3.28
Indinavir	(2S)-1-[(2S,4R)-4-benzyl-2-hydroxy-5-[[[(1S,2R)-2-hydroxy-2,3-dihydro-1H-inden-1-yl]amino]-5-oxopentyl]-N-tert-butyl-4-(pyridin-3-ylmethyl)piperazine-2-carboxamide	613.8		-2.85
Pleconaril	3-[3,5-dimethyl-4-[3-(3-methyl-1,2-oxazol-5-yl)propoxy]phenyl]-5-(trifluoromethyl)-1,2,4-oxadiazole	381.3		-5.44
Raltegravir	N-[2-[4-[(4-fluorophenyl)methylcarbamoyl]-5-hydroxy-1-methyl-6-oxopyrimidin-2-yl]propan-2-yl]-5-methyl-1,3,4-oxadiazole-2-carboxamide	444.4		-5.5
Trifluridine	1-[(2R,4S,5R)-4-hydroxy-5-(hydroxymethyl)oxolan-2-yl]-5-(trifluoromethyl)pyrimidine-2,4-dione	296.2		-4.26
Umifenovir	ethyl 6-bromo-4-[(dimethylamino)methyl]-5-hydroxy-1-methyl-2-(phenylsulfanylmethyl)indole-3-carboxylate	477.4		-3.62
Vicriviroc	(4,6-dimethylpyrimidin-5-yl)-[4-[(3S)-4-[(1R)-2-methoxy-1-[4-(trifluoromethyl)phenyl]ethyl]-3-methylpiperazin-1-yl]-4-methylpiperidin-1-yl]methanone	533.6		-4.3
Zalcitabine	4-amino-1-[(2R,5S)-5-(hydroxymethyl)oxolan-2-yl]pyrimidin-2-one	211.22		-3.48
Penciclovir	2-amino-9-[4-hydroxy-3-(hydroxymethyl)butyl]-1H-purin-6-one	253.26		-2.87

(continued on next page)

Table 2 (continued)

Ligand	IUPAC nomenclature	M.W(g/mol)	Chemical representation	Binding energy (kcal/mol)
Brivudine	5-[(E)-2-bromoethenyl]-1-[(2R,4S,5R)-4-hydroxy-5-(hydroxymethyl)oxolan-2-yl]pyrimidine-2,4-dione	333.13		-4.87
Famciclovir	[2-(acetyloxymethyl)-4-(2-aminopurin-9-yl)butyl] acetate	321.33		-3.61
Cidofovir	[(2S)-1-(4-amino-2-oxopyrimidin-1-yl)-3-hydroxypropan-2-yl]oxymethylphosphonic acid	279.19		-2.83
Ganciclovir	2-amino-9-(1,3-dihydroxypropan-2-yloxymethyl)-1H-purin-6-one	255.23		-3.58
Valganciclovir	[2-[(2-amino-6-oxo-1H-purin-9-yl)methoxy]-3-hydroxypropyl] (2S)-2-amino-3-methylbutanoate	354.36		-2.18
Acyclovir	2-amino-9-(2-hydroxyethoxymethyl)-1H-purin-6-one	225.2		-3.17
Valacyclovir	2-[(2-amino-6-oxo-1H-purin-9-yl)methoxy]ethyl (2S)-2-amino-3-methylbutanoate	324.34		-2.26
Ribavirin	1-[(2R,3R,4S,5R)-3,4-dihydroxy-5-(hydroxymethyl)oxolan-2-yl]-1,2,4-triazole-3-carboxamide	244.2		-3.02
Remdesivir	2-ethylbutyl (2S)-2-[[[(2R,3S,4R,5R)-5-(4-aminopyrrolo[2,1-f][1,2,4]triazin-7-yl)-5-cyano-3,4-dihydrooxolan-2-yl]methoxy-phenoxyphosphoryl]amino]propanoate	602.6		-3.02
Zidovudine triphosphate	[[[(2S,3S,5R)-3-azido-5-(5-methyl-2,4-dioxypyrimidin-1-yl)oxolan-2-yl]methoxyhydroxyphosphoryl] phosphono hydrogen phosphate	469.2		-7.83

(continued on next page)

Table 2 (continued)

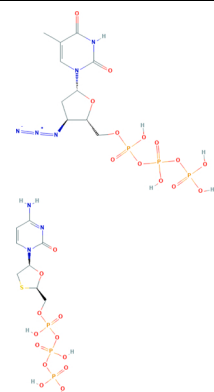
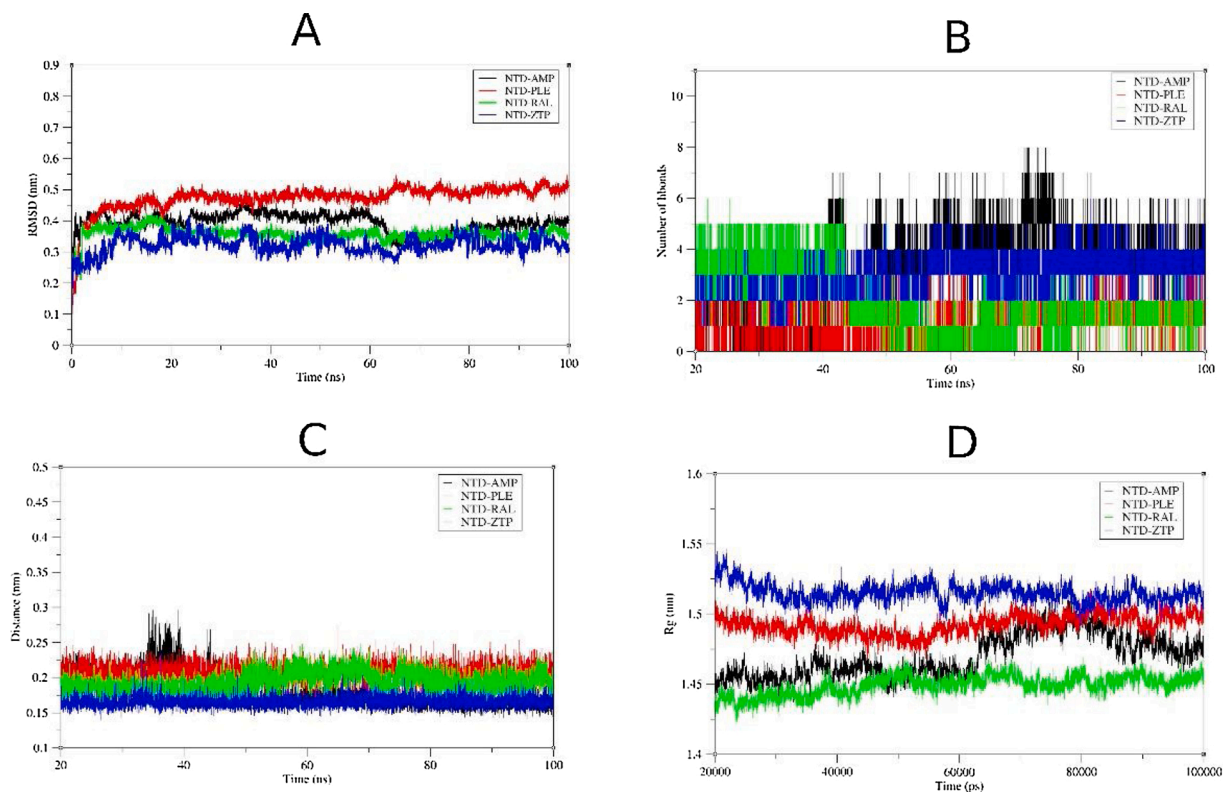
Ligand	IUPAC nomenclature	M.W(g/mol)	Chemical representation	Binding energy (kcal/mol)
Lamivudine triphosphate	[[[(2R,5S)-5-(4-amino-2-oxopyrimidin-1-yl)-1,3-oxathiolan-2-yl]methoxy-hydroxyphosphoryl] phosphono hydrogen phosphate	469.2		-3.09

Table 3

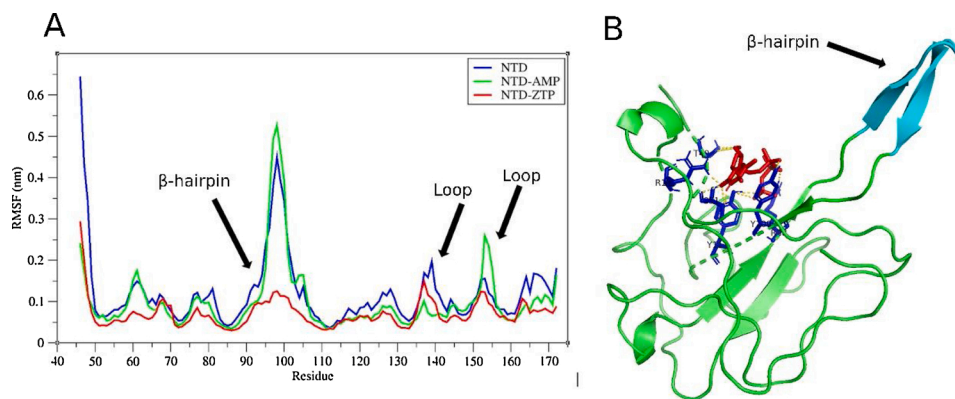
MM/PBSA calculations of energy terms and binding energy governing the protein-ligand interactions.

Ligand	Energy terms (kJ/mol)				Binding energy (kJ/mol)
	Van der Waal	Electrostatics	Polar solvation	Non-polar solvation (SASA)	
AMP	-0.001	-106.428	-1.855	0.051	-108.234
RAL	-0.002	0.175	-8.177	0.068	-7.936
ZTP	-15.851	-2012.207	721.787	-6.565	-1312.837
PLE	-43.228	-3.938	23.653	-6.002	-29.514

immune cells during COVID-19 infection is not clearly defined, however, extrapolating knowledge from previous coronaviruses suggests presentation by both MHCI and MHCII alleles [55]. The relevance of demonstrating the HLA-epitope complex formation and interactions cannot be disregarded. The epitopes were docked to specific HLA-alleles and the best docked conformations were further characterized by simulation analysis. The findings indicate that the HLA-A\*03:01-<sup>79</sup>SSPDDQIGY<sup>87</sup> complex is a more stable structure than HLA-B\*35:01-<sup>79</sup>SSPDDQIGY<sup>87</sup> based on RMSD analysis, hydrogen bonds formed and proximity of epitope to HLA protein. Similarly, the HLA-DRB1\*09:01-<sup>305</sup>AQFAPSASAFFGMSR<sup>319</sup> complex was the most stable among the MHCII-restricted epitope candidates. Collectively, our results show that the epitopes (<sup>79</sup>SSPDDQIGY<sup>87</sup> and <sup>305</sup>AQFAPSASAFFGMSR<sup>319</sup>) can bind efficiently and form stable structures with their respective HLA proteins.



**Fig. 8.** Molecular dynamics analysis of selected antiviral drugs complexed with the N-terminal domain (NTD) of SARS-CoV2 N-protein. (A) RMSD of complexes after least square fitting to reference structure. (B) Number of hydrogen bonds formed between the ligands and protein. (C) The minimum distance between the ligands and the protein are shown (D) Radius of gyration (Rg) of the protein-ligand complexes.



**Fig. 9.** The interaction between ZTP and the N-terminal domain (NTD). **(A)** RMSF analysis of the NTD. RMSF of the N-protein is shown alongside the RMSF of the protein in complex with ZTP or AMP. Residual fluctuations for the residues of the  $\beta$ -hairpin of the SARS-CoV2 N-protein NTD and loop regions are demonstrated. **(B)** The active site of ZTP (red) and the polar interaction (yellow dashes) with active site residues (blue) are shown. The structure of the  $\beta$ -hairpin is also demonstrated (cyan).

Although we have an incomplete understanding of the immunological responses during the SARS-CoV2 infection, findings from research on previous coronavirus outbreaks (SARS-CoV1 and MERS-CoV) and the emerged findings on immunologic characteristics of COVID-19 provide substantial answers. Available information suggests the involvement of innate immunity, T- and B-cell immunity as well as specific antiviral neutralizing antibodies in the immunological responses during SARS-CoV2 infection [56–58]. Simulation of the immune responses to the N-protein of SARS-CoV2 yielded results that correlate well with the emerging data on immunological profile of COVID-19. Upon repeated injection of the antigen (N-protein), aspects of innate immunity (Natural killer cells), B- and T-cell immunity and antibody-mediated neutralization were observed accompanied by development of immunological memory that lasted for several months. Both T-helper and T-cytotoxic cell-mediated immunity were observed upon exposure to the N-protein, however, only T-helper cells developed memory for the antigen. Protein specific T-cell responses with induction of IFN- $\gamma$ , TNF- $\alpha$ , and IL-2 have been detected in COVID-19 patients [59]. Recognition of S and N protein epitopes by virus specific T-helper and T-cytotoxic cells during the COVID-19 infection have also been reported [60]. Immunological characteristics of SARS-CoV1 infection include the development of CD4 and CD8 memory T-cells which persisted for over 4 years [61]. The extent of memory T-cell immunity against COVID-19 is yet to be elucidated, however, our results indicate a different scenario as observed for the SARS-CoV1 infection. The B-cells also elicited a robust response to the SARS-CoV2 N-protein just that the memory B-cell faded significantly over time as compared to the memory T-helper cells. Long term protection courtesy of memory B-cells have been confirmed in SARS-CoV1 infections [62,63] and emerging evidence are supporting a similar situation for COVID-19 [64–66] but the exact life-span is yet to be known.

The B-cell immunity during SARS-CoV2 infection is further confirmed by the detection of virus-specific immunoglobulin titers with particular affinity for S- and N- protein epitopes after about a week of infection [22,24,64]. Current reports suggest IgM, IgG and IgA responses as the main antibody neutralization during the SARS-CoV2 infection with the detection of IgM response at the early stages (7 days of infection) followed by IgG and IgA upon progression to an acute stage [56, 58]. Our computational results agree well with these findings. According to the immune simulation analysis, immunocomplex IgG + IgM and IgM are the antibodies responsible for antigen neutralization during the early stages of exposure. IgA was, however not detected but the source of this discrepancy may be attributed to the antigenic protein used as target.

Exposure to the N-protein also elicited certain aspects of both adaptive and innate immune response. According to our computational results, IFN- $\gamma$  and IL-2 elicited robust responses to the antigen with significant levels of natural killer (NK) cells also being detected. Possible NK-mediated response to the SARS-CoV2 infection has been reported by Pinto et al. [67] and evidence of reduced levels of NK cells in the peripheral blood of infected individuals have also emerged [68,69].

Taking into consideration the novelty of the SARS-CoV2 infection and our limited understanding of its dynamics, other therapeutic options are mandatory for complete containment of the pandemic. If the effects of the vaccine fades over time provided the virus develops means to evade immunity, drugs can be used to eliminate symptoms faster. Moreover, populations or individuals that do not respond well to the vaccine would rely on drugs for managing the pandemic. In view of this, we targeted the N-terminal domain (NTD) of SARS-CoV2 N-protein for potential drug candidates. The NTD, otherwise known as the RNA-binding domain, plays important roles during package of viral genome into a ribonucleoprotein complex [70,71]. Kang et al. [72] solved the structure of the NTD of SARS-CoV2 N-protein (PDB ID: 6M3M) and identified potential pockets for drug targeting. Considering the urgent need to contain the SARS-CoV2 infection, drug repurposing seems to be the best tool in the search for therapeutic options. Existing drugs for different targets such as chloroquine, remdesivir, lopinavir, ribavirin and ritonavir have been re-oriented to target the SARS-CoV2 and several *in-vitro* and *in-vivo* studies are providing positive outcomes [73–75]. Based on MM/PBSA binding energy calculations and molecular dynamics simulation analysis, zidovudine triphosphate (ZTP) has higher affinity and interacts better with the NTD of N-protein than AMP. ZTP is an anti-HIV agent that targets the viral reverse transcriptase, thus interfering with the reverse transcription process [76]. Residual fluctuation analysis from the simulation studies revealed that upon binding of ZTP, the flexibility of the NTD  $\beta$ -hairpin (basic palm) residues significantly reduce. The impact could be detrimental for the NTD considering the roles of the flexible, positively charged  $\beta$ -hairpin. The  $\beta$ -hairpin has been proposed to be essential for RNA packaging by interacting with phosphate groups of the viral RNA [77–79]. This observation together with the high binding affinity depicts ZTP as a promising N-protein inhibitor and should be further validated by experimental studies.

## 5. Conclusion

We have explored the SARS-CoV2 N-protein for potential epitope-based vaccine candidates using immuno-informatics and identified an antiviral drug that binds to the ribonucleoside 5'-monophosphate binding pocket in the N-terminal domain (NTD) of SARS-CoV2 N-protein with much greater affinity and stability. In view of this, we have identified B-cell epitopes (<sup>404</sup>SKQLQSQSMSSADS<sup>416</sup> and <sup>92</sup>RRIRGGDGMKDL<sup>104</sup>), MHC I-restricted epitope (<sup>76</sup>SSPDDQIGY<sup>86</sup>) and MHC II-restricted epitope (<sup>305</sup>AQFAPSASAFFGMSR<sup>319</sup>) based on performance after stringent evaluation and profiling analysis on the basis of certain desirable vaccine characteristics. We also identified zidovudine triphosphate as a promising inhibitor of the NTD of SARS-CoV2 N-protein. The limitations of our study cannot be disregarded as experimental data are needed to validate our findings. Nonetheless, the integrative bioinformatics methods we used will significantly save time

and cost for research communities using wet-laboratory approaches and clinical studies. As the need for COVID-19 therapeutics grow by the day, our findings will assist research scientists to fast-track the discovery of vaccines and drugs to contain the SARS-CoV2 infection.

### Authors contribution

AK and EA: conception and design of study, AK, EA, SOP, SAS: Acquisition of data: AK, EA, SOP, SAS: Analysis and interpretation of data, AK, EA, SOP, SAS: Drafting and final revision of manuscript.

### Data availability

Data is available upon request

### Declaration of Competing Interest

The authors declare they have no competing interest.

### Acknowledgement

AK is a grant recipient on EDCTP and African Research Network for Neglected Tropical and Diseases SGP-III. We are also indebted to the staff of the Kwarteng Lab (Kumasi Centre for Collaborative Research in Tropical Medicine) for their unflinching support in the preparation of this manuscript.

### References

- L.E. Gralinski, V.D. Menachery, Return of the coronavirus: 2019-nCoV, *Viruses* 12 (2) (2020) 135, <https://doi.org/10.3390/v12020135>.
- A. Wu, Y. Peng, D. Huang, X. Ding, X. Wang, P. Niu, et al., Genome composition and divergence of the Novel Coronavirus (2019-nCoV) originating in China, *Cell Host Microbe* 27 (2020) 325–328, <https://doi.org/10.1016/j.chom.2020.02.001>.
- A.A.T. Naqvi, K. Fatima, T. Mohammad, U. Fatima, I.K. Singh, A. Singh, M. I. Hassan, Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: structural genomics approach, *Biochimica et Biophysica Acta (BBA) – Mol. Basis of Dis.* 165878 (2020), <https://doi.org/10.1016/j.bbadis.2020.165878>.
- D. Wang, B. Hu, C. Hu, F. Zhu, X. Liu, J. Zhang, B. Wang, H. Xiang, Z. Cheng, Y. Xiong, Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China, *JAMA* 323 (2020) 1061–1069, <https://doi.org/10.1001/jama.2020.1585>.
- C. Huang, et al., Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China, *Lancet* 395 (10223) (2020) 497–506, <https://doi.org/10.1016/j.chom.2020.02.001>.
- E. Ortiz-Prado, K. Simbaña-Rivera, L. Gómez-Barreno, M. Rubio-Neira, L. P. Guaman, N.C. Kyriakidis, A. López-Cortés, Clinical, molecular and epidemiological characterization of the SARS-CoV2 virus and the Coronavirus disease 2019 (COVID-19), a comprehensive literature review, *Diagn. Microbiol. Infect. Dis.* 115094 (2020), <https://doi.org/10.1016/j.diagmicrobio.2020.115094>.
- P.S. Masters, L.S. The molecular biology of coronaviruses, *Adv. Virus Res.* 64 (2006) 193–292, [https://doi.org/10.1016/S0065-3527\(06\)66005-3](https://doi.org/10.1016/S0065-3527(06)66005-3).
- F.K. Yoshimoto, The Proteins of Severe Acute Respiratory Syndrome Coronavirus-2 (SARS CoV-2 or n-COV19), the Cause of COVID-19, *Protein J.* 39 (3) (2020) 198–216, <https://doi.org/10.1007/s10930-020-09901-4>.
- C.A. De Haan, P.J. Rottier, Molecular interactions in the assembly of coronaviruses, *Adv. Virus Res.* 64 (2005) 165–230, [https://doi.org/10.1016/S0065-3527\(05\)64006-7](https://doi.org/10.1016/S0065-3527(05)64006-7).
- R. McBride, M. Van Zyl, B. Fielding, the coronavirus nucleocapsid is a multifunctional protein, *Viruses* 6 (8) (2014) 2991–3018, <https://doi.org/10.3390/v6082991>.
- C.A. De Haan, L. Kuo, P.S. Masters, H. Vennema, P.J. Rottier, Coronavirus particle assembly: primary structure requirements of the membrane protein, *J. Virol.* 72 (1998) 6838–6850, <https://doi.org/10.1128/jvi.72.8.6838-6850.1998>.
- M. Surjit, R. Kumar, R.N. Mishra, M.K. Reddy, V.T. Chow, S.K. Lal, The severe acute respiratory syndrome coronavirus nucleocapsid protein is phosphorylated and localizes in the cytoplasm by 14-3-3-mediated translocation, *J. Virol.* 79 (17) (2005) 11476–11486, <https://doi.org/10.1128/JVI.79.17.11476-11486.2005>.
- D. Cavanagh, Coronaviruses and toroviruses, in: A.J. Zuckerman, J.E. Banatvala, P. D. Griffiths, J.R. Pattison, B.D. Schoub (Eds.), *Principles and Practice of Clinical Virology*, John Wiley & Sons Ltd., Chichester, 2004, pp. 379–397.
- Y. Van der Meer, E.J. Snijder, J.C. Dobbe, S. Schleich, M.R. Denison, W.J. Spaan, J. K. Locker, Localization of mouse hepatitis virus nonstructural proteins and RNA synthesis indicates a role for late endosomes in viral replication, *J. Virol.* 73 (1999) 7641–7657, <https://doi.org/10.1128/JVI.73.9.7641-7657.1999>.
- M.B. Serafin, A. Bottega, V.S. Foletto, T.F. da Rosa, A. Hörner, R. Hörner, Drug repositioning an alternative for the treatment of coronavirus COVID-19, *Int. J. Antimicrob. Agents* 105969 (2020), <https://doi.org/10.1016/j.ijantimicag.2020.105969>.
- World Health Organization, Draft of the Landscape of COVID-19 Candidate Vaccines, 2020, <https://www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines>.
- A. Vartak, S.J. Sucheck, Recent advances in subunit vaccine carriers, *Vaccines* 4 (2) (2016) 12, <https://doi.org/10.3390/vaccines4020012>.
- W. Shang, Y. Yang, Y. Rao, et al., The outbreak of SARS-CoV-2 pneumonia calls for viral vaccines, *Npj Vaccines* 5 (2020) 18, <https://doi.org/10.1038/s41541-020-0170-0>.
- A. Grifoni, J. Sidney, Y. Zhang, R.H. Scheuermann, B. Peters, A. Sette, A sequence homology and bioinformatic approach can predict candidate targets for immune responses to SARS-CoV-2, *Cell Host Microbe* 27 (4) (2020) 671–680, <https://doi.org/10.1016/j.chom.2020.03.002>, e2.
- C.K.F. Li, H. Wu, H. Yan, S. Ma, L. Wang, M. Zhang, X. Tang, N.J. Temperton, R. A. Weiss, J.M. Brencley, et al., T cell responses to whole SARS coronavirus in humans, *J. Immunol.* 181 (2008) 5490–5500.
- M.Z. Tay, C.M. Poh, L. Renia, P.A. MacAry, L.F.P. Ng, The trinity of COVID-19: immunity, inflammation and intervention, *Nat. Rev. Immunol.* (2020).
- F. Amanat, D. Stadlbauer, S. Strohmaier, et al., A serological assay to detect SARS-CoV-2 seroconversion in humans, *Nat. Med.* 26 (2020) 1033–1036, <https://doi.org/10.1038/s41591-020-0913-5>.
- B. Ju, Q. Zhang, J. Ge, et al., Human neutralizing antibodies elicited by SARS-CoV-2 infection, *Nature* 584 (2020) 115–119, <https://doi.org/10.1038/s41586-020-2380-z>.
- K.K. To, O.T. Tsang, W.S. Leung, et al., Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study, *Lancet Infect. Dis.* 20 (5) (2020) 565–574, [https://doi.org/10.1016/S1473-3099\(20\)30196-1](https://doi.org/10.1016/S1473-3099(20)30196-1).
- M.C. Jespersen, B. Peters, M. Nielsen, P. Marcantili, BepiPre-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes, *Nucleic Acids Res.* 45 (W1) (2017) W24–W29, <https://doi.org/10.1093/nar/gkx346>.
- J. Ponomarenko, H.H. Bui, W. Li, N. Füsseder, P.E. Bourne, A. Sette, B. Peters, ElliPro: a new structure-based tool for the prediction of antibody epitopes, *BMC Bioinformatics* 9 (1) (2008) 514, <https://doi.org/10.1186/1471-2105-9-514>.
- S. Gupta, P. Kapoor, K. Chaudhary, A. Gautam, R. Kumar, et al., In silico approach for predicting toxicity of peptides and proteins, *PLoS One* 8 (9) (2013) e73957, <https://doi.org/10.1371/journal.pone.0073957>.
- I. Dimitrov, I. Bangov, D.R. Flower, I. Doytchinova, AllerTOP v.2—a server for in silico prediction of allergens, *J. Mol. Model.* 20 (6) (2014), <https://doi.org/10.1007/s00894-014-2278-5>.
- S.K. Dhandu, P. Vir, G.P. Raghava, Designing of interferon-gamma inducing MHC class-II binders, *Biol. Direct* 8 (1) (2013), <https://doi.org/10.1186/1745-6150-8-30>.
- A. Lamiable, P. Thévenet, J. Rey, M. Vavrusa, P. Derreumaux, P. Tufféry, PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex, *Nucleic Acids Res.* 44 (W1) (2016) W449–W454, <https://doi.org/10.1093/nar/gkw329>.
- D. Kozakov, D.R. Hall, B. Xia, K.A. Porter, D. Padhorny, C. Yueh, S. Vajda, The ClusPro web server for protein–protein docking, *Nat. Protoc.* 12 (2) (2017) 255–278, <https://doi.org/10.1038/nprot.2016.169>.
- R. Brenke, D.R. Hall, G.Y. Chuang, S.R. Comeau, T. Bohnuud, D. Beglov, D. Kozakov, Application of asymmetric statistical potentials to antibody–protein docking, *Bioinformatics* 28 (20) (2012) 2608–2614, <https://doi.org/10.1093/bioinformatics/bts493>.
- D. Kozakov, R. Brenke, S.R. Comeau, S. Vajda, PIPER: an FFT-based protein docking program with pairwise potentials, *Proteins* 65 (2006) 392–406.
- E. Katchalski-Katzir, et al., Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques, *Proc. Natl. Acad. Sci. U.S.A.* 89 (1992) 2195–2199.
- B.R. Brooks, et al., Charmm - a program for macromolecular energy, minimization, and dynamics calculations, *J. Comput. Chem.* 4 (1983) 187–217.
- M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindahl, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, *SoftwareX* 1-2 (2015) 19–25, <https://doi.org/10.1016/j.softx.2015.06.001>.
- F. Celada, P.E. Seiden, A computer model of cellular interactions in the immune system, *Immunol. Today* 13 (1992) 56–62.
- P.E. Seiden, F. Celada, A model for simulating cognate recognition and response in the immune system, *J. Theor. Biol.* 158 (1992) 329–357.
- F. Celada, P.E. Seiden, Affinity maturation and hypermutation in a simulation of the humoral immune response, *Eur. J. Immunol.* 26 (1996) 1350–1358.
- K. Haskins, R. Kubo, J. White, M. Pigeon, J. Kappler, et al., The major histocompatibility complex-restricted antigen receptor on T cells, *J. Exp. Med.* 157 (1983) 1149–1162.
- C. Bere, M. Ziegner, The maturation of the immune response, *Immunol. Today* 14 (1993) 400–404.
- F. Celada, The cellular basis of immunological memory, *Prog. Allergy* 15 (1971) 223–267.
- J. Sprent, D. Lo, F.K. Gao, Y. Ron, T cell selection in the thymus, *Immunol. Rev.* 101 (1988) 173–190.
- P.M. Allen, Antigen processing at the molecular level, *Immunol. Today* 8 (1987) 270–273.
- J.A. Berzofski, The nature and role of antigen processing, in: J.M. Cruse, R.E. Lewis (Eds.), *T Cell Activation, Year in Immunology*, 1985, pp. 18–24.

- [46] E.R. Unanue, Antigen-presenting function of the macrophage, *Ann. Rev. Immunol.* 2 (1985) 395–428.
- [47] N. Rapin, O. Lund, M. Bernaschi, F. Castiglione, Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system, *PLoS One* 5 (4) (2010) e9862, <https://doi.org/10.1371/journal.pone.0009862>.
- [48] S. Yuan, H.C.S. Chan, Z. Hu, Using PyMOL as a platform for computational drug design, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 7 (2) (2017) e1298, <https://doi.org/10.1002/wcms.1298>.
- [49] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera?A visualization system for exploratory research and analysis, *J. Comput. Chem.* 25 (13) (2004) 1605–1612, <https://doi.org/10.1002/jcc.20084>.
- [50] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A. J. Olson, AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility, *J. Comput. Chem.* 30 (16) (2009) 2785–2791, <https://doi.org/10.1002/jcc.21256>.
- [51] P.A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D.A. Case, T.E. Cheatham 3rd, calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models, *Acc. Chem. Res.* 33 (2000) 889–897.
- [52] R. Kumari, R. Kumar, A. Lynn, g\_mmpbsa—a GROMACS tool for high-throughput MM-PBSA calculations, *J. Chem. Inf. Model.* 54 (7) (2014) 1951–1962, <https://doi.org/10.1021/ci500020m>.
- [53] J. Braun, L. Loyal, M. Frentsch, D. Wendisch, P. Georg, F. Kurth, S. Hippenstiel, M. Dingeldey, B. Kruse, F. Fauchere, et al., Presence of SARS-CoV-2 reactive T cells in COVID-19 patients and healthy donors, *MedRxiv* (2020), <https://doi.org/10.1101/2020.04.17.20061440>.
- [54] C. Gioia, D. Horejsh, C. Agrati, F. Martini, M.R. Capobianchi, G. Ippolito, F. Poccia, T-Cell response profiling to biological threat agents including the SARS coronavirus, *Int. J. Immunopathol. Pharmacol.* 18 (2005) 525–530.
- [55] P. Liu, F. Wu, et al., Novel immunodominant peptide presentation strategy: a featured HLA-A\*2402-restricted cytotoxic T-lymphocyte epitope stabilized by intrachain hydrogen bonds from severe acute respiratory syndrome coronavirus nucleocapsid protein, *J. Virol.* 84 (2010), <https://doi.org/10.1128/JVI.01464-10.11849e11857>.
- [56] A.Z. Azkur, M. Akdis, D. Azkur, M. Sokolowska, W. van de Veen, M. Brüggen, C. A. Akdis, Immune response to SARS-CoV-2 and mechanisms of immunopathological changes in COVID-19, *Allergy* (2020), <https://doi.org/10.1111/all.14364>.
- [57] X. Li, M. Geng, Y. Peng, L. Meng, S. Lu, Molecular immune pathogenesis and diagnosis of COVID-19, *J. Pharm. Anal.* 10 (2) (2020) 102–108, <https://doi.org/10.1016/j.jpha.2020.03.001>.
- [58] N. Vabret, G.J. Britton, C. Gruber, S. Hegde, J. Kim, R. Kuksin, R.M. Samstein, Immunology of COVID-19: Current State of the Science, *Immunity*, 2020, <https://doi.org/10.1016/j.immuni.2020.05.002>.
- [59] L. Ni, F. Ye, M.L. Chen, Y. Feng, Y.Q. Deng, H. Zhao, P. Wei, J. Ge, M. Gou, X. Li, et al., , Detection of SARS-CoV-2-Specific Humoral and Cellular Immunity in COVID-19 Convalescent Individuals, *Immunity*, 2020, <https://doi.org/10.1016/j.immuni.2020.04.023>.
- [60] L. Xu, Y. Shi, et al., Pathological findings of COVID-19 associated with acute respiratory distress syndrome, *Lancet Resp. Med.* (2020), [https://doi.org/10.1016/S2213-2600\(20\)30076-X](https://doi.org/10.1016/S2213-2600(20)30076-X).
- [61] Y.Y. Fan, Z.T. Huang, L. Li, et al., Characterization of SARS-CoV-specific memory T cells from recovered individuals 4 years after infection, *Arch. Virol.* 154 (2009), <https://doi.org/10.1007/s00705-009-0409-6>, 1093e1099.
- [62] Z. Cao, L. Liu, L. Du, C. Zhang, S. Jiang, T. Li, Y. He, Potent and persistent antibody responses against the receptor-binding domain of SARS-CoV spike protein in recovered patients, *Virol. J.* 7 (2010) 299.
- [63] J. Liu, S. Li, J. Liu, B. Liang, X. Wang, H. Wang, W. Li, Q. Tong, J. Yi, L. Zhao, et al., Longitudinal characteristics of lymphocyte responses and cytokine profiles in the peripheral blood of SARS-CoV-2 infected patients, *EBioMedicine* 55 (2020) 102763.
- [64] B. Ju, Q. Zhang, X. Ge, R. Wang, J. Sun, X. Ge, J. Yu, S. Shan, B. Zhou, S. Song, Human neutralizing antibodies elicited by SARS-CoV-2 infection, *Nature* (2020), <https://doi.org/10.1038/s41586-020-2380-z>.
- [65] L. Bao, W. Deng, H. Gao, C. Xiao, J. Liu, J. Xue, Q. Lv, J. Liu, P. Yu, Y. Xu, et al., Reinfection could not occur in SARS-CoV-2 infected rhesus macaques, *BioRxiv* (2020), <https://doi.org/10.1101/2020.03.13.990226>.
- [66] E.R. Adams, R. Anand, M.I. Andersson, K. Auckland, J.K. Baillie, E. Barnes, J. Bell, T. Berry, S. Bibi, M. Carroll, et al., Evaluation of antibody testing for SARS-CoV-2 using ELISA and lateral flow immunoassays, *MedRxiv* (2020), <https://doi.org/10.1101/2020.04.15.20066407>.
- [67] D. Pinto, Y.J. Park, M. Beltramello, A.C. Walls, M.A. Tortorici, S. Bianchi, S. Jaconi, K. Culap, F. Zatta, A. De Marco, Structural and functional analysis of a potent sarbecovirus neutralizing antibody, *bioRxiv* (2020), <https://doi.org/10.1101/2020.04.07.023903>.
- [68] C.Y. Song, J. Xu, J.Q. He, Y.Q. Lu, COVID-19 early warning score: a multi-parameter screening tool to identify highly suspected patient, *medRxiv* (2020), <https://doi.org/10.1101/2020.03.05.20031906>.
- [69] M. Zheng, Y. Gao, G. Wang, et al., Functional exhaustion of antiviral lymphocytes in COVID-19 patients, *Cell. Mol. Immunol.* 17 (2020) 533–535, <https://doi.org/10.1038/s41423-020-0402-2>.
- [70] K.S. Saikatend, J.S. Joseph, V. Subramanian, B.W. Neuman, M.J. Buchmeier, R. C. Steven, et al., Ribonucleocapsid formation of severe acute respiratory syndrome coronavirus through molecular action of the N-terminal domain of N protein, *J. Virol.* 81 (2007) 3913e21.
- [71] Y.W. Tan, S. Fang, H. Fan, J. Lescar, D.X. Liu, Amino acid residues critical for RNA-binding in the N-terminal domain of the nucleocapsid protein are essential determinants for the infectivity of coronavirus in cultured cells, *Nucleic Acids Res.* 34 (2006) 4816e25.
- [72] S. Kang, M. Yang, Z. Hong, L. Zhang, Z. Huang, X. Chen, et al., Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites, *Acta. Pharm. Sin. B* (2020), <https://doi.org/10.1016/j.apsb.2020.04.009>.
- [73] N. Yamamoto, S. Matsuyama, T. Hoshino, N. Yamamoto, Nelfi- navir inhibits replication of severe acute respiratory syndrome coronavirus 2 in vitro, *bioRxiv* (2020), <https://doi.org/10.1101/2020.04.06.026476>.
- [74] E. de Wit, F. Feldmann, J. Cronin, R. Jordan, A. Okumura, T. Thomas, D. Scott, T. Cihlar, H. Feldmann, Prophylactic and therapeutic remdesivir (GS-5734) treatment in the rhesus macaque model of MERS-CoV infection, *Proc. Natl. Acad. Sci. U.S.A.* 117 (2020) 6771–6776.
- [75] M.G.S. Borba, F. de Almeida Val, V.S. Sampaio, M.A.A. Alexandre, G.C. Melo, M. Brito, M.P.G. Moura o, J.D. Brito Sousa, D.C. Baía-da-Silva, M.V.F. Guerra, et al., Effect of high vs low doses of chloroquine diphosphate as adjunctive therapy for patients hospitalized with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection, *JAMA Netw. Open* 3 (2020) e208857.
- [76] E. De Clercq, G. Li, Approved antiviral drugs over the past 50 years, *Clin. Microbiol. Rev.* 29 (2016) 695–747, <https://doi.org/10.1128/CMR.00102-15>.
- [77] 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 (2004) 6059–6063.
- [78] H. Fan, A. Ooi, Y.W. Tan, S. Wang, S. Fang, D.S. Liu, J. Lescar, The nucleocapsid protein of coronavirus infectious bronchitis virus: crystal structure of its N-terminal domain and multimerization properties, *Structure* 13 (2005) 1859–1868.
- [79] Y.W. Tan, S. Fang, H. Fan, J. Lescar, D.X. Liu, Amino acid residues critical for RNA-binding in the N-terminal domain of the nucleocapsid protein are essential determinants for the infectivity of coronavirus in cultured cells, *Nucleic Acids Res.* 34 (2006) 4816–4825.