



Host-Directed Anti-Fusion Aptamers and Small Molecules as Respiratory Syncytial Virus (RSV) Inhibitors: An *in silico*-based study

Ssemuyiga Charles^{1,2,3*}, Mulumba Pius Edgar^{1,2}, Rajani Kanta Mahapatra^{3*}

Abstract

Despite significant medical advancements, RSV continues to strain healthcare systems and society. For over 60 years, developing a vaccine for RSV has been a top priority. In 2023, a milestone of two vaccines of Abrysvo and Arexvy were introduced¹. The physical interactions between Nucleolin (NCL) RBD1,2 and RSV-F protein during virus entry were identified by peptide arrays as two antiparallel strands of β -sheet in RBD1. We screened for novel ligands targeting NCL to inhibit this particular interaction. In a nutshell, small molecule inhibitors and nucleic acid aptamers were specifically docked to NCL RBD1, 2 targeting the previously identified strands. A total of 528 aptamers were docked to NCL RBD1,2 on the HDock website; the top 180 aptamers with docking scores equal to or better than -200 were then re-docked at the HADDOCK website, and the top 58 aptamers were chosen. Additionally, 976,450 small molecules were docked using Schrodinger Virtual Screening Workflow, and the top 50 ligands underwent post-docking validation using Prime MM_GBSA. The 16 best ligands were chosen from the top 30 using binding affinity from Molegro Virtual Docker. For molecular simulation dynamics studies, four protein-ligand complexes and six protein-aptamer complexes were randomly chosen and the results were analyzed. The results show that the selected NCL ligands are potential NCL binders and can inhibit NCL-RSV-F interactions. Results suggest that aptamers are better binders than small molecules. BDBM50308336, 1exd, and 5uza formed the most stable complexes. With many benefits as host-directed therapeutics, these ligands need in-vitro testing for potential inhibition of RSV entry besides some probable drawbacks.

Keywords: Respiratory Syncytial Virus; Nucleolin; Aptamers; Molecular Docking; Molecular Simulation Dynamics

Introduction

Human RSV is a highly contagious seasonal ubiquitous respiratory pathogen and a prevalent cause of severe lower respiratory tract infection including pneumonia and bronchiolitis in all age groups but the infants, young, and immunocompromised are the most vulnerable groups [1]. It belongs to the genus Orthopneumovirus within the family Pneumoviridae and order Mononegavirales with two major antigenic subtypes (A and B) primarily determined by antigenic drift and duplications in RSV-G sequences, but accompanied by genome-wide sequence divergence, including within RSV-F [2]. It is an enveloped virus with a linear non-segmented -ssRNA genome ~15.2kb, pleomorphic with a diameter of approximately 150 nm (range 120–300 nm) while spherical, filamentous species have been identified [3]. RSV

Affiliation:

¹PharmaQsar Bioinformatics Firm, Kampala, Uganda

²Department of Biological and Environmental Sciences, School of Natural Sciences, Kampala International University, Kampala, Uganda

³School of Biotechnology, KIIT Deemed to be University, Bhubaneswar, Odisha, India.

*Corresponding author:

Ssemuyiga Charles, PharmaQsar Bioinformatics Firm, Kampala, Uganda.

Citation: Ssemuyiga Charles, Mulumba Pius Edgar, Rajani Kanta Mahapatra. Host-Directed anti-fusion aptamers and small molecules as Respiratory Syncytial Virus (RSV) inhibitors: An *in silico*-based study. Journal of Biotechnology and Biomedicine. 7 (2024): 485-497.

Received: November 12, 2024

Accepted: November 21, 2024

Published: November 28, 2024

has 10 genes that encode for 11 proteins of non-structural protein 1 (NS1) and 2 (NS2), nucleocapsid protein (N), phosphoprotein (P), matrix protein (M), small hydrophobic protein (SH), attachment glycoprotein (G), Surface fusion protein (F), M2-1 and M2-2 protein (both translated from the M2 transcript that contains two partially overlapping reading frames), and polymerase protein (L) [4] in that respective order. F and G glycoproteins are major surface proteins that are crucial for infectivity and pathogenesis (control attachment and initial infection stages) and induce in vitro neutralizing antibodies [5]. RSV-F is highly conserved and is responsible for viral fusion with host cell membranes, as well as syncytium formation [6]. While viral attachment appears to involve both F and G proteins, F fusion occurs independently of G and exists in multiple conformational forms [7,8]. In the prefusion state (PreF), the protein exists in a trimeric form and contains the major antigenic site Ø which serves as a primary target of neutralizing antibodies. After binding to its target on the host cell surface, PreF undergoes a conformational change during which Ø is lost [7,8]. This change enables the protein to insert itself into the host cell membrane and leads to the fusion of the viral and host cell membranes. A final conformational shift results in a more stable and elongated form of the protein (post-fusion, PostF). Opposite of the RSV-G protein, the RSV F protein also binds to and activates toll-like receptor 4 (TLR4), initiating the innate immune response and signal transduction [6,8] and interacts with NCL during viral fusion. Nucleolin (NCL) (mainly nucleolus localized) is a multi-localized and multifactorial protein unusually secreted and involved in countless cellular functions and processes with no surviving Knockouts (KOs) while its expression and localization are cell cycle stage-dependent [9]. It is around 700 amino acids long (710 for humans and 707 for mice) with the central domain between N and C terminal regions containing two low-complexity, intrinsically disordered domains (IDDs). The 300 amino acid long N-terminal IDD is unique with four acidic tracts having phosphorylation sites and crucial in interacting with basic proteins like histones. Additionally, the N-terminal IDD contains eight repetitions of the motif [ST]PxK[KA] (TPxKK in 75% of instances) [10], which is recognized by proline-directed kinases and is important for NCL localization. The central domain is constituted by four RNA-recognition motifs (RRMs) [11], the most common RNA-binding domain (RBD) of higher vertebrates typically 8 amino acids, and arranged in four-stranded antiparallel beta-sheet, among two alpha helices. Each domain recognizes between two and eight nucleotides and the combination of multiple domains lengthens the stretch of recognized RNA, which may or may not be continuous [12]. The C-terminal IDD consists of a glycine/arginine-rich (GAR/RGG) domain involved in protein binding and nucleolar localization signal and also contributes to the interaction with RNA [9,13,14]. It

receipts and co-receptors various viruses, some bacteria, toxins, and many cellular factors on the cell surface including RSV, EVA71, HIV, IAV, HPIV-3, RHDV, CCHFV, CVB, Enterobacteriaceae, and Francisella tularensis, and various toxins including H.pylori Tpa, B.asper Mt-II, Cathelicidin-BF, LPS, Acharan sulfate, cytokines, growth factors, and matrix proteins [10,15]. It controls a variety of aspects of RNA and DNA metabolism, chromatin remodeler, and histone chaperone that takes part in DNA replication, repair, and recombination [16]. It participates in ribosomal RNA transcription, maturation, ribosome synthesis, maturation, assembly, and transport [17]. It also plays a role in numerous mRNA's transcription, splicing, stability, transport, and translation processes [18]. It aids in the cytosolic microtubule polymerization and anchoring of interphase cells' microtubules to centrosomes [19]. It controls MAPK signal transduction and Ras protein assembly and interactions on the plasma membrane (PM) [20]. It is involved in gene silencing, senescence, cytokinesis, cell proliferation, and growth. Its DNA-dependent ATPase and auto-degradative capabilities, as well as helicase activity, and interaction with replication protein A and telomerase, are all associated with nucleolin's putative roles in cell growth and replication. Because of these and many other functions, NCL is involved in numerous pathologies. Its role has been studied especially in cancer and viral diseases [21,22], but studies on its involvement in neurodegenerative diseases are increasing [23,24]. It is reported to be overexpressed with altered subcellular localization in cancer and a candidate molecular target for cancer therapy [25]. This multifunctionality is due to its multivalence, numerous domains, short linear motifs, and post-translational changes, which confer and modulate its ability to interact with proteins, nucleic acids, carbohydrates, and lipids and bridges the nucleus and cell surface, hence a passport for many pathogens [10]. Numerous post-translational modifications (PTMs) affect roles and localization, 179 modification events are reported in the iPTMnet database (2021) [26], including phosphorylations, acetylations, ubiquitinations, methylations, and sumoylations as well as complex N-linked and O-linked glycosylation patterns [27] with 2745 PTM-dependent Protein-Protein Interactions (2022) [28]. It is reportedly fragmented and has 12 potential sites for enzyme cleavage as well as 10 potential sites for isomerization by the prolyl isomerase Pin1 (S/T-P). The presence of cell surface NCL shows interaction with many proteins and is dependent on an intact actin cytoskeleton and can internalize extracellular material, while associated with a variety of proteins and is associated with flotillin, a multifunctional lipid raft protein [28,29].

NCL-RSV-F INTERACTIONS

Attachment and viral entry through RSV-F involves proteins such as IGFR1 and nucleolin. Because of both positively and negatively charged patches in site II of pre-

fusion F, polar and electrostatic interactions may be important in the binding of RSV site II to nucleolin RBD1,2 [30]. The development of compounds that precisely bind and block this viral interaction may be effective in RSV prophylaxis or therapy. Through the F-protein, RSV physically binds with NCL RBD1,2, and these interactions can be competitively blocked by the use of palivizumab or recombinant RBD1,2 [30]. Treatment with synthetic peptides derived from two 12-mer domains of RBD1,2 inhibited RSV infection in vitro and structural analysis suggests that these domains are potentially feasible for targeting in drug development hence providing essential groundwork toward RSV drug development [30]. Only a few groups in high-resource settings have access to the treatment and prevention alternatives (Palivizumab) [2]. The monoclonal antibody drug nirsevimab is also registered for clinical use [31]. A DNA aptamer AS1411 has been shown to successfully lower infection in cell culture and lung inflammation in two animal models according to several reports [32]. Recently, it has also been demonstrated that targeting a pathway that regulates the shuttling of NCL to the cell surface is a viable strategy to lower RSV infection [33]. The goal of this research was to identify potential compounds and aptamers that can bind the identified NCL domains that interact with RSV-F.

Methods

Protein-Nucleic acid docking

HDOCK

The pdb structures were downloaded from PDB (Receptor PDB ID: 2KRR, aptamer PDB IDs are shown in Table 1, the aptamers were downloaded by searching PDB with G-Quadruplex as the main search string) and cleaned with pymol. They were then uploaded to the HDOCK server³⁴ for specific docking. Only aptamers with docking scores better than -200 and confidence scores above 0.700 were retained for further steps. There was no small-angle X-ray scattering (SAXS) experimental data provided and no post-docking processes were performed [35].

Table 1: The results of Protein-Nucleic acid docking, the HDOCK confidence score is docking score dependent.

Aptamer PDB ID	HDOCK Docking score	Confidence score	Haddock score	z-score	Nature
6vvv	-253.38	0.887	-90.4	-1.8	RNA
7rwr	-213.24	0.780	-89.2	-2.2	RNA
484d	-206.27	0.755	-87.0	-2.3	RNA
7mk1	-254.13	0.889	-74.9	-1.5	DNA
7kvt	-235.18	0.846	-69.7	-2.2	RNA
7eoh	-245.51	0.871	-76.5	-1.7	RNA
7e9e	-247.91	0.876	-97.8	-1.8	RNA
7d82	-231.97	0.838	-48.8	-2.0	RNA
6zlc	-200.87	0.735	-78.9	-1.9	RNA

7d7v	-264.64	0.908	-45.0	-2.1	RNA
6xkn	-249.80	0.880	-108.2	-2.2	RNA
6wlj	-219.97	0.802	-104.3	-1.7	RNA
6wjrr	-255.61	0.892	-55.6	-2.2	RNA
5uza	-214.64	0.785	-97.5	-2.0	RNA
6sy6	-225.04	0.818	-83.0	-2.2	RNA
5swd	-227.11	0.824	-92.7	-1.5	RNA
6q57	-237.21	0.851	-51.6	-1.3	RNA
6pq7	-206.74	0.757	-51.6	-1.2	RNA
6j2w	-206.38	0.755	-90.6	-1.7	DNA
6dvk	-237.44	0.852	-66.6	-1.9	RNA
6db9	-221.15	0.806	-43.3	-1.9	RNA
6b14	-220.49	0.804	-77.1	-1.7	RNA
6b3k	-207.84	0.761	-87.5	-1.6	RNA
5voe	-210.45	0.770	-96.4	-2.0	RNA
5ob3	-220.77	0.805	-39.0	-2.6	RNA
5e54	-206.90	0.757	-87.1	-2.7	RNA
5kpy	-224.39	0.816	-74.0	-1.5	RNA
5kh8	-210.33	0.770	-69.6	-2.6	RNA
4lx5	-208.07	0.762	101.7	-1.5	RNA
4m4o	-257.59	0.896	-58.3	-1.9	RNA
2b57	-200.32	0.732	-89.5	-1.6	RNA
4m6d	-201.55	0.737	-53.2	-1.4	RNA
1exd	-234.85	0.845	-96.4	-1.5	RNA
3iqp	-211.56	0.774	-111.1	-1.5	RNA
1y26	-222.16	0.809	-116.2	-1.7	RNA
4fe5	-254.82	0.891	-30.9	-1.9	RNA
3sd3	-215.53	0.788	-92.7	-1.9	RNA
3zh2	-222.36	0.810	-94.0	-2.7	DNA
4i7y	-212.43	0.777	-141.7	-2.3	DNA
3iqr	-217.84	0.795	-75.6	-1.5	RNA
3iqn	-273.47	0.922	-68.7	-1.4	RNA
4kqy	-242.05	0.863	-77.0	-1.6	RNA
4lvv	-260.50	0.901	-40.7	-1.9	RNA
2lun	-260.57	0.901	-97.4	-2.6	RNA
5f5f	-206.88	0.757	-73.4	-1.6	RNA
3hxx	-220.79	0.805	-60.5	-1.3	DNA
2au4	-199.16	0.728	-61.4	-1.9	RNA
1ooa	-249.48	0.880	-73.3	-2.0	RNA
2qbz	-258.18	0.897	-113.9	-1.5	RNA
3ds7	-231.00	0.835	-28.5	-2.1	RNA
1nbk	-206.58	0.756	-59.5	-1.7	RNA
2fy1	-207.59	0.760	-59.9	-1.5	RNA
1fmn	-202.41	0.740	-88.5	-1.9	RNA
1ull	-207.79	0.761	-74.6	-1.6	RNA
1db6	-210.80	0.771	-84.9	-1.4	DNA
2gis	-218.43	0.797	-96.1	-2.2	RNA
1nta	-228.81	0.829	-81.8	-1.6	RNA
6gzk	-221.60	0.807	-69.8	-2.4	RNA

HADDOCK

The thoroughly cleaned pdb structures were uploaded for specific docking, and the molecules chosen for docking were protein-nucleic acid without coarse-graining the molecules. The buried active/passive residues were set to be excluded from the selection, and the Minimum percentage of relative solvent accessibility (RSA) to consider a residue as accessible was set to 15.0. Automatic selection within 6.5Å° of the passive residues with a minimum relative solvent accessibility of 40.0 was made. To preserve the nucleic acid structure, no energy minimization (EM) was set. Non-crystallographic symmetry was used and no restraints were subjected to dihedral, Hydrogen Bond, relaxing anisotropy, pseudo-contact shifts, radius of gyration, center of mass, and surface contact. Non-polar hydrogens were eliminated, and a fraction of the ambiguous constraints (AIRs) were set to be excluded randomly without random patches. For rigid body docking, number of structures for semi-flexible refinement, final refinement, analysis, and trials for rigid body minimization was set to 1000, 200, 200, 200, and 5, respectively. It was also set to sample 180° rotated solutions during rigid body Energy Minimization and semi-flexible Simulated Annealing while final refinements were carried out without the use of molecular dynamics. The clustering method used was Fraction of Common Contacts (FCC) and RMSD cut-off for clustering set to 0.6Å° with a minimum cluster size of 4 and without residual dipolar coupling (RDC). The Epsilon constants for the electrostatic energy term in it0 and it1 were both optimized to 78.0 from defaulted values (10.0), while other energy and interaction, scoring, and advanced sampling parameters were defaulted. Analysis of the results was limited to clusters and the proton-acceptor cut-off distance to define H-bonds was set to 2.5Å° while the cut-off Carbon-Carbon distance to define hydrophobic contacts was set to 3.9Å° and without advanced water analysis [36]. To further verify the obtained receptor-aptamer complexes before molecular simulation Dynamics, the receptor and ligands were subjected to PIPER³⁷ for protein– Nucleic acid docking in Maestro Bioluminate 4.3 [38], the number of ligand rotations to probe was set to 7000, no restraints and set to return 30 maximum poses.

Small molecule docking

Protein preparation and Grid generation

The protein preparation wizard's default parameters were used to pre-process the protein, optimize hydrogen bonding, and minimize energy. OPLS4 (optimized potentials for liquid simulations) force field³⁹ was used for geometry optimization. The centroid of the active site residues was used to build the receptor grid, with a van-der-Waals scaling factor of 1.0 and a partial charge cut-off of 0.25 [40-42].

Small molecule preparation

OPLS4 forcefield was used to expand protonation and tautomeric states after the ligands were prepared by LigPrep [43] with Epik [44] at 7 ± 2.0 pH units. For each ligand, low-energy stereoisomers were created with a maximum conformation set to 32, and those that held low-energy 3D structures with the proper chiralities were kept. The best poses were retained and taken for further studies. The prepared protein and ligands were subjected to Glide docking with standard precision (SP) followed by Extra precision (XP) with flexible ligand sampling and without Bias sampling of torsions and docking scores subjected to Epik penalties without constraints [41].

Prime MM_GBSA

The prime MM_GBSA (Molecular Mechanics Generalized Born Surface Area) was employed as a post-docking tool to determine the energy of different complexes. The binding free energies of receptor-ligand complexes were calculated using the Prime [45] module of the Schrodinger suite with the OPLS4 force field and VSGB solvation model. The binding free energy of a ligand (L) to a protein (P) to form the complex (PL) is obtained as the difference.

$$\Delta G_{\text{binding}} = \Delta G (\text{complex}) - \Delta G (\text{Protein}) - \Delta G (\text{Ligand})$$

Where $\Delta G_{\text{Binding}}$ is the binding free energy; whereas $\Delta G_{\text{complex}}$, $\Delta G_{\text{protein}}$, and ΔG_{ligand} represent the free energy of complex, protein, and ligand respectively.

Binding affinity determination

Briefly, to determine the binding affinities, the protein was imported into the Molegro Virtual Docker/MVD (2013 6.0.1) [46] and prepared by repairing amino acid residues. The ligands in MOL format were imported and auto-prepared. With a user-defined Ligand origin, radius of 13Å°, and Ligand evaluation including internal ES, Internal HBond, and sp2-sp2 Torsions, the docking function was set to MolDock Score and to enforce Hbond directionality. The algorithm utilized was MolDock Optimizer with 20 runs and was configured to minimize energy after docking while returning a maximum of 50 poses per run with docking being carried out in separate processes. The Ligand pose with the highest re-rank for each Ligand was reported rather than that with the highest MolDock Score/Binding Affinity [47], however, the ligand with the best/lowest binding affinity was still considered superior.

Molecular Simulation Dynamics

Protein-small molecule complexes

All simulations were done in GROMACS 2021.4 [48]. The topology for protein was generated using AMBER99SB [49] force field from *pdb2gmx* module while ligands were parametrized by using Generalized Amber Force Field

(GAFF2) [50] at antechamber website. The complexes were placed in the hexahedral box using *gmx editconf*, solvation was done using *gmx solvate*, TIP3P water molecules were added to the systems, and Na^+ and Cl^- ions were added using *genion* to a concentration of 0.15M. Equilibration was done with steps set to 20,000 in 10ns at 298K and 1atm with position restraints Force Constant set to 700KJmol^{-1} . This allowed water molecules and ions to move freely. Production MD simulations lasted for 100ns and they were monitored by checking system energies during the simulations. Pymol, vmd, and Gromacs binaries [48] were used for the analysis of the results.

Protein-aptamer complexes

This was done in Google Collaboratory [51], briefly, the protein-aptamer complexes were uploaded to Colab, the macromolecule topologies were generated using ff19SB forcefield [52], and solvation was done with TIP3P water model in 12\AA box. Using amber tleap, NaCl was added to the systems to 0.15M. To equilibrate temperature and pressure and to allow the solvent to accommodate around the protein for proper solvation layers, the systems were subjected to energy minimization with maximum steps set to 50,000 steps followed by equilibration for 10ns at 298K and 1atm with a force constant of 500KJmol^{-1} . The prepared systems were then subjected to production MD for 50ns. A summary of the followed procedures is shown in Figure 1.

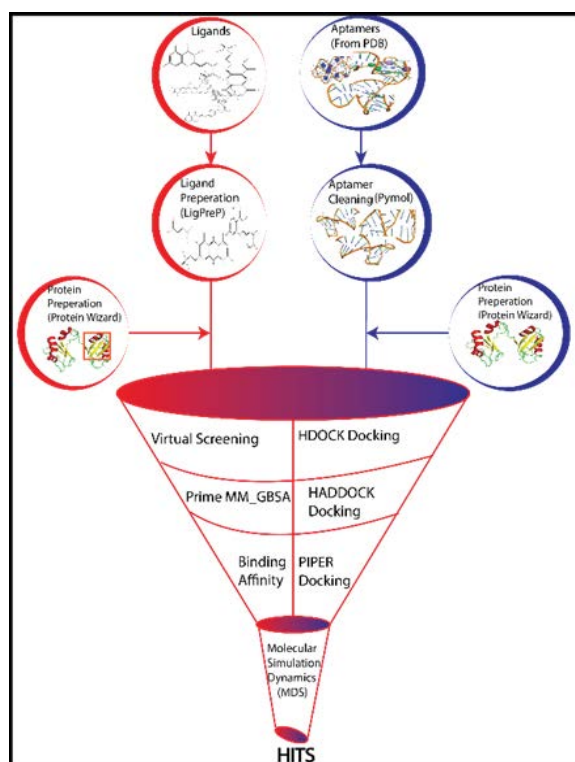


Figure 1: A summary of the steps and methods followed in this study

Results and Discussion

Protein-Nucleic acid Docking

A total of 528 aptamers were docked against NCL and a total of 58 aptamers were presented in Table 1 with HDock scores better than -200.00 , $\sim 10\%$ were DNA (1DB6, 7MK1, 3HXQ, 4I7Y, 3ZH2, and 6J2W), a statistic not surprising as NCL is a re-known RNA Binding Protein. The HADDOCK results are in agreement with HDock results supporting the possibility that these aptamers can bind the suggested receptor protein. Riboswitches had better docking scores than non-regulatory RNAs, which may be attributed to their ability to bind other molecules [53].

Small molecule docking














The results of molecular docking are shown in Table II, only small molecules with docking scores better than -9.000KCal/mol and Binding Affinity better than -150.0KCal/mol . Since the affinity more likely scales with the surface area of the molecule that can interact with the protein rather than the volume, which is what the number of heavy atoms is essentially measuring, the ligand efficiency sa (surface area) metric approximates the effect of surface area. In contrast, the \ln metric has no real physical justification but appears to better fit the experimental data for maximal affinity of ligands [54]. The ligand with the best docking score was DB15919 from DrugBank with $-10.557\text{KCalMol}^{-1}$ while the ligand with the best binding affinity was 6603812 from PubChem with $-218.826\text{KCalMol}^{-1}$.

Molecular simulation dynamics

Small molecule complexes

The outcomes of molecular simulation dynamics concur with the docking outcomes and demonstrate that the compounds examined in this work are potential RSV inhibitors. The difference between the backbones of a protein from its initial structural conformation to its ultimate position is measured using the root mean square deviation (RMSD) [42]. The variations generated during the simulation of the protein can be used to gauge its stability with respect to its conformation. The RMSDs of the complexes all stabilized during simulation (Figure 2A), demonstrating the stability of the generated complexes. The ligand RMSDs (Figure S1F) additionally demonstrate that the ligands gained stability through simulation at higher RMSDs, a justification for their small sizes. The RMSD distribution of the complexes is also shown in Figure S3A and results show that the complex of J1.089.466B is less stable while BDBM50308336 complex was the most stable among all. The RMSD of the apoprotein over simulation stabilized at higher magnitudes than those of the complexes (Figure S1D) showing the binding of the Ligands stabilized the receptor.

Table 2: Results of small molecule docking with NCL; ligand efficiency sa is (docking score) / (number of heavy atoms)^{2/3}, and ligand efficiency ln is (docking score) / (1 + ln (number of heavy atoms)).

Ligand structure and Database ID	Glide XP Docking score (KCal/mol ⁻¹)	ligand efficiency sa	ligand efficiency ln	Glide gscore	Prime MMGBSA dG Bind	Binding Affinity MolDock Score (KCal/mol ⁻¹)	Rerank Score	HBond Energy (KCal/mol ⁻¹)	Database
 DB15919	-10.557	-0.788	-2.158	-10.561	-55.02	-176.367	-105.605	-9.221	Drug Bank
 CHEMBL4087771	-10.404	-0.890	-2.219	-10.573	-36.01	-176.561	-82.401	-14.688	CHEMBL
 BDBM50308336	-10.405	-0.777	-2.127	-10.405	-50.13	-204.404	-109.469	-10.848	Binding Database
 DB03495	-10.064	-0.808	-2.104	-10.571	-33.18	-155.559	-89.801	-5.305	Drug Bank
 DB04158	-10.045	-0.761	-2.062	-10.078	-28.82	-208.69	-133.163	-14.665	Drug Bank
 DB01753	-10.027	-0.749	-2.050	-11.419	-24.92	-189.425	-109.374	-7.478	Drug Bank
 118753825	-9.891	-0.739	-2.022	-9.951	-54.50	-191.742	-138.468	-15.883	PubChem
 DB02844	-9.744	-1.032	-2.231	-9.754	-45.09	-151.691	-107.938	-9.629	Drug Bank
 DB02547	-9.470	-0.838	-2.042	-9.470	-49.52	-170.373	-125.792	-12.355	Drug Bank
 4532	-9.461	-0.809	-2.018	-9.614	-53.85	-176.824	-76.104	-7.305	Brenda
 ZINC00000819	-9.383	-0.665	-1.888	-9.384	-41.93	-216.941	-126.304	-10.445	ZINC20
 6603812	-9.279	-0.658	-1.867	-9.280	-48.46	-218.826	-95.790	-8.656	PubChem
 ZINC00016935	-9.253	-0.656	-1.862	-9.254	-52.08	-216.941	-126.304	-10.645	ZINC20

J1.089.466B 	-9.231	-0.847	-2.014	-9.290	-35.42	-189.573	-85.697	-12.548	ZINC20
DB04983 	-9.076	-0.687	-1.863	-10.245	-54.79	-170.547	-100.568	-8.435	Drug Bank
HMDB0001211 (Diadenosine tetraphosphate) 	-9.053	-0.642	-1.821	-10.090	-58.06	-215.036	-139.818	-9.548	Human Metabolome Database

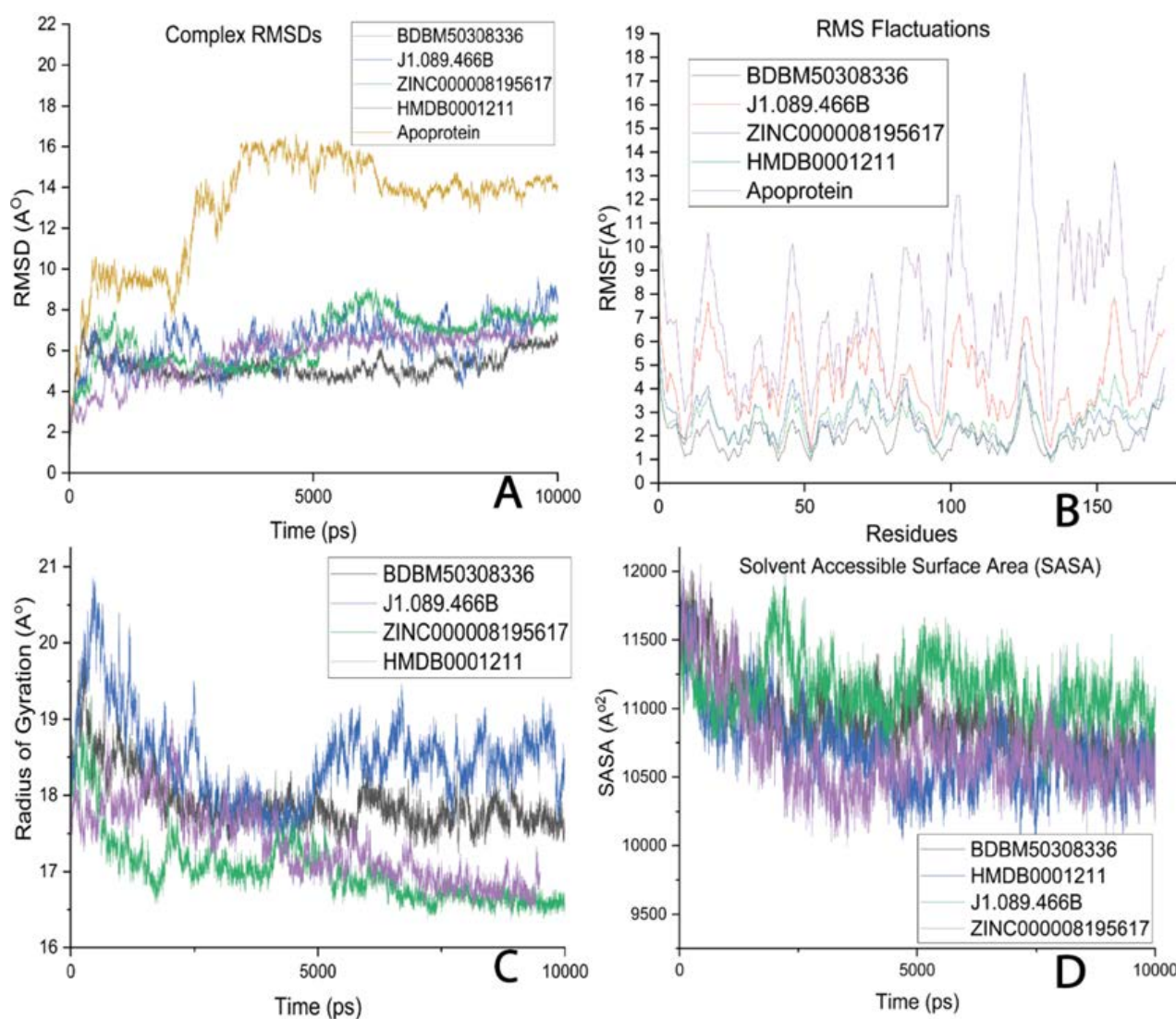


Figure 2: (A) Shows the variation of RMSDs of all the small molecule complexes over simulation. (B) Shows RMSFs of the receptor in respective complexes. (C) Shows the variation of Rg for small molecule complexes over simulation time. (D) Shows the variation of SASA for small molecule complexes over simulation.

With 1D RMSD, it is easy to think that two structures with the same RMSD from a reference frame are also similar, but in fact, they can be very different. Instead, calculating the RMSD of each frame in the trajectory to all other frames in the other trajectory can contain much more information [55]. This is pairwise/2D RMSD. Pairwise RMSDs of each trajectory were calculated to itself and the results are shown in Figure S2, the diagonal represents 0 (RMSD of a structure to its self), while the occupation of a given state shows up as a block of similar RMSD along the diagonal, BDBM50308336 complex has more structures occupying and revisiting a nearly similar state (RMSD close to but not necessarily 0) shown by many blocks of low RMSDs below and above the diagonal compared to apoprotein and other complexes, suggesting more stability. 2D RMSD further confirms the more stability of BDBM50308336 (0-8.5Å°) complex and lower stability of J1.089.466B complex (0-14.3Å°) compared to all complexes. All complexes were more stable than the apoprotein (0-16.5Å°, Figure S1A). Complexes of BDBM50308336 and J1.089.466B seem to have single main states without any major state transitions. HMDB0001211 and ZINC000008195617 complexes have two last major transitions occurring at ~60ns and ~52ns respectively after minor state transitions at ~18ns and ~17ns respectively. However, there are many other small transitions reflected among these complexes, these results of 2D RMSD agree with those of 1D RMSD. The average deviation of a particle over time from a reference position is measured by the root-mean-square fluctuation (RMSF). As a result, RMSF examines the structural elements that deviate the most from their mean structure (or least). Less fluctuation means more stability [41,42]. The interacting residues are observed to have low RMSF values hence more stable and interacting with the ligands, The RMSFs of the apoprotein are higher than those of the complexes showing that the formed complexes are stable over simulation. The results show that the complex of BDBM50308336 is more stable while that of J1.089.466B is less stable than the rest of the complexes (Figure 2B). There were no significant differences in the average RMSFs of Domains 1 and 2 (3.0 and 2.9Å° respectively) and this is further supported by the presence of interacting residues in both domains (Figure 4) lowering their RMSFs compared to the apoprotein. The radius of gyration (Rg) is the distribution of atoms of a protein around its axis. It is frequently used to evaluate a protein's compactness. A lower value indicates a highly compact structure, but a value that stays relatively constant across the simulation demonstrates a stable system. Its value can also change when a ligand binds a receptor. The Rg of all systems had an average value of ~19.034Å° and the values relatively reduced over the 100ns simulation period showing a relative increase in stability of the systems over simulation (Figure 2C). The Rg distribution over simulation is also shown in Figure S3B. All these results show that the ZINC0000081956 complex is more compact while that of J1.089.466B was less compact than the rest of the systems.

Theoretically, changes in the solvent accessibility of proteins can be detected via the calculation of solvent-accessible surface area (SASA). To understand how residues interact with the surrounding solvent, the polar and non-polar molecular surface areas are calculated using the SASA method. A lower value indicates more compacted protein hence more stable and a low fluctuation is expected over the simulation time [42]. Any small chemical could alter SASA upon binding, and occasionally this could have a significant impact on the protein structure. The SASA of complexes relatively reduced over the simulation time showing an increase in compactness and therefore stability of all the systems. The SASA distribution over simulation is shown in Figure S3C. The similar trends of SASA and Rg predict the correctness of molecular dynamics simulation results. To quantify the strength of the interaction between the ligand and the protein, non-bonded Molecular Mechanics (MM) interaction energy between the Ligands and their receptor was computed [51]. All the energy terms of electrostatic, van der Waals, and total energy were negative over 100ns simulation except for a complex of HMDB0001211 which had positive values of total and electrostatic energy for about 30ns (50-80ns) and relative positive values of van der Waals energy, but with better average Electrostatic and total MM energy than other systems over the simulation period. These results show that the simulated complexes were stable over simulation (Figure 3 and S3D). It is however important to note that this quantity is NOT a free energy or a binding energy [51]. The details of molecular interactions in the middle of the simulation are shown in Figure 4. To evaluate the conformational subspace of the complexes and identify the various regions of the energy landscape sampled during the MD simulation, Principal Component Analysis (PCA) was done on the energy and structural data of the trajectories on only alpha carbons [41,56,57]. This was done to better understand the dynamic behavior of the complexes and apo-protein. The compounds encompassed relatively small subspaces and characterized conformational clusters. The type of motion and displacement of atomic fluctuation between the complexes were identified by graphing the two eigenvectors 1 and 2. The first few eigenvalues had bigger values and accounted for the majority of the energy; however, subsequent eigenvalues were in decreasing order. The complexes also occupied a small subspace compared to the apo-protein, suggesting more stabilized complexes and leading to well-defined internal motion vital for complex stabilization (Figures S1C and S4). All PC2 values for complexes were lower than PC1 values. There is a displacement to the positive PC1 and PC2 (top right) of the major eigenvalues for the BDBM50308336 complex. ZINC0000095617 complex forms a small unique cluster in the positive of PC1. HMDB0001211 complex had a unique cluster for main eigenvalues displaced to the negative side of the PCs.

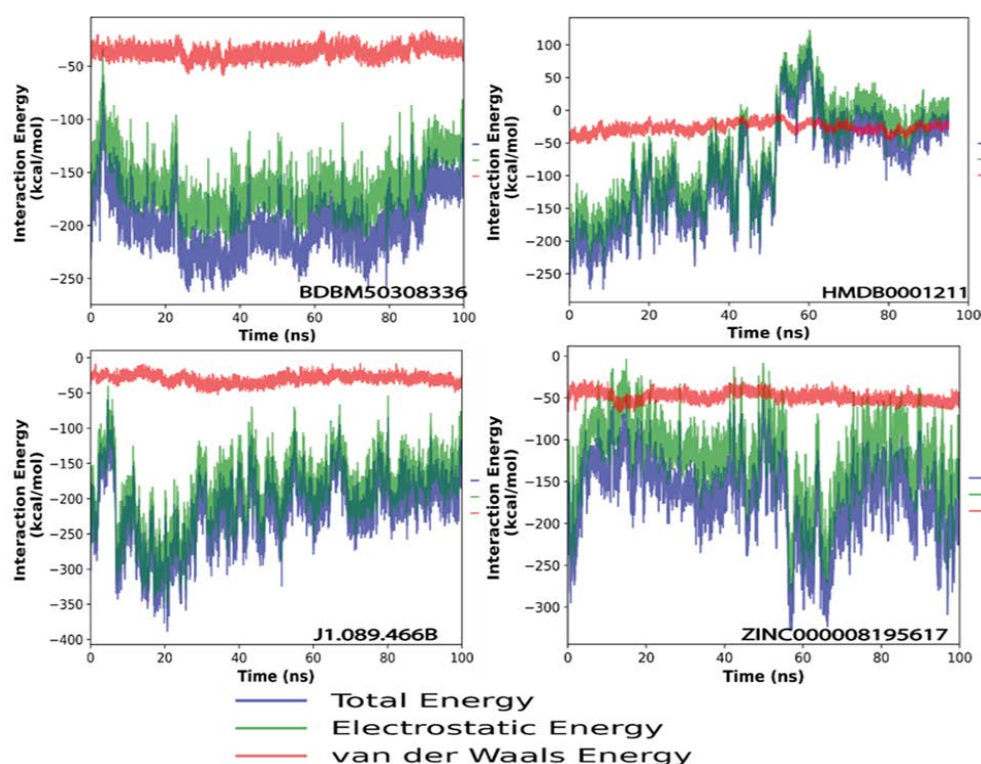


Figure 3: The variation of Molecular Mechanics energy for all simulated systems over simulation period.

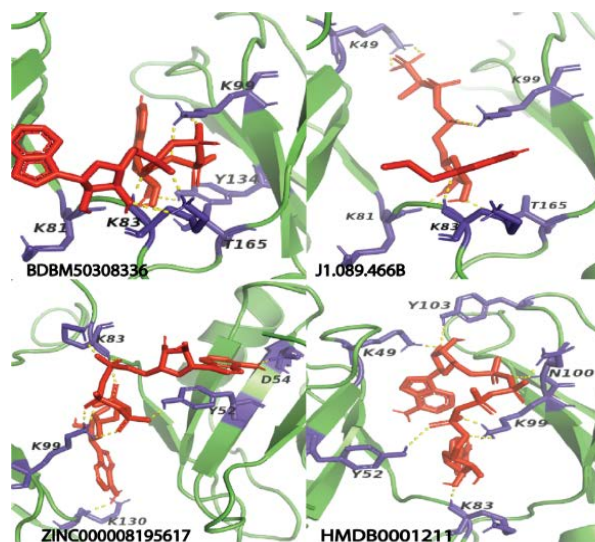


Figure 4: The interaction snapshots taken at 50ns to view the main interactions between NCL and ligands under study. The interacting interface receptor residues are presented in Licorice colored purple while ligands are presented in licorice and colored red. The H-bonds are presented as yellow dashed lines.

To further capture protein dynamic information, Global motion analysis was carried out [56], the computation of the correlation matrix is a frequently used method to represent the dynamic information of proteins in two dimensions. To observe the correlation in the dynamics of NCL complexes, the cross-correlation was computed for each trajectory and

plotted (Figure S5). When the CCs of Apo-Proteins are compared to the complexes, it is evident that interacting residues have increased positive correlation, validating their interactions with the suggested targets. There was a relatively stronger correlation between residues in each domain but a weaker correlation between residues in domains 1 and 2

for complexes of HMDB0001211 and J1.089.466B and this may be due to the nature of their interactions (Figures S5 and 4). There was no significant difference in residue correlation displayed by complexes of ZINC0000095617 and BDBM50308336 complexes and this may be due to the unsymmetrical nature of their interactions across the two domains (Figures S5 and 4).

Aptamer complexes

The RMSDs of the simulated complexes all show that all simulated complexes attained reasonable stability at RMSDs lower than that of the apoprotein hence more stable compared to the latter. The RMSD distribution of the complexes is also shown in Figure S7A with 1exd and 5uza complexes being the most stable among all complexes (Figure 5). 2D RMSD results show that 1exd (0-5.5Å) and 5uza (0-6.3Å) had the lower RMSDs between the structures over simulation (Figure S6). However, 4m4o and 484d had more similar structures that are occupying and revisiting a nearly similar state shown by many blocks of low RMSDs below and above the diagonal compared to apoprotein and other complexes, suggesting more stability (Figure S6). All complexes were more stable than the apoprotein (0-16.5Å, Figure S3A). Most complexes lack sounding state transitions with many minor transitions except 5uza with a major state transition at ~30ns (Figure S6). These results of 2D RMSD partly agree with those of 1D RMSD. The computed radius of gyration of complexes ranged from 16.5-23.0Å with an average of 18.42Å and showed that all attained stability with 1exd complex having the smallest radius over simulation, however, the sizes of these complexes differ and depend on the sizes of the aptamers which differs significantly, hence stability and compactness of these systems may not necessarily be influenced by their Rg values. The distribution of complex Rg is also shown in Figure S7B. The observed low RMSF values of the interacting residues and the fact that the apoprotein's RMSF is higher than that of the complexes indicate that the produced complexes are stable over simulation. The computed RMSF shows a relatively same pattern of domain stability showing the same interacting residues in all complexes, the results further show that 1exd is more stable while the 484d complex is less stable than the rest of the complexes (Figure 5). The presence of interacting residues in both domains (Figure 6) which lower their RMSFs relative to the apoprotein further supports the lack of any significant variations in the average RMSFs of Domains 1 and 2 (2.3 and 2.5Å, respectively). These RMSFs are however lower than those of small molecule complexes hence more stable and their interaction patterns agree with these results as aptamers interact with many residues compared to small molecules (Figure 6). Throughout the simulation, the SASA of complexes decreased comparatively, indicating an increase in folding and, consequently, stability of all the systems. Figure S7C displays its distribution throughout the simulation. The similarities between SASA and Rg's trends

indicate that the results of the molecular dynamics simulation are accurate (Figure 5). PCA Results from aptamer complexes reveal that all complexes occupy comparably smaller conformational subspaces than the apoprotein, indicating that they are more stabilized (Figures S1C and S8). Except for 1exd complex whose PC1 main PCs are shifted to the right (positive) of PC1, other complexes have their main PCs displaced to the negative of PC1 with uniformly distributed PC2s between -40 to 70nm². Complexes of 1exd, 5swd, and 5uza aptamers increased positive correlation among residues of domains 1 and 2 but reduced correlation between residues in domains 1 and 2 showing that these sections are more rigid. The correlations of the rest of the complexes are not significantly different from those of the apoprotein (Figure S8).

Conclusions

There is an urgent need for alternative therapeutic strategies to control the unprecedented spread of viral infections due to their persistent recurrence, challenging variants due to resistance, and rapid rate of viral evolution. Broad-spectrum targets for the creation of new antiviral medications may be found in the targeting of host pathways and cellular proteins that are commandeered by viruses since they are not prone to viral resistance. We have presented possible compounds that can bind NCL, developing such compounds may also be useful in a wider range of viral infections because nucleolin is involved in the attachment and entrance of several other

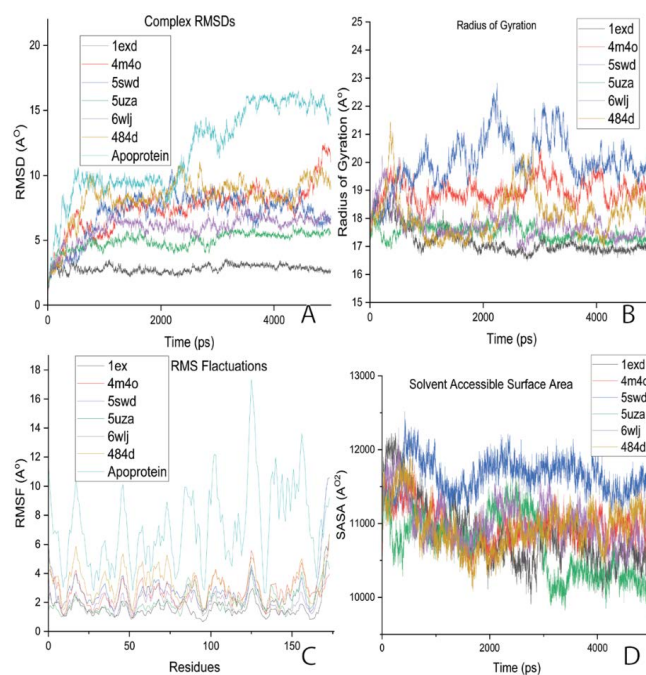


Figure 5: (A) Shows the variation of RMDs of aptamer complexes over simulation time. (B) Shows the variation Rg over simulation. (C) Shows the RMSF for the receptor in different respective complexes. (D) Shows a variation of SASA over simulation.

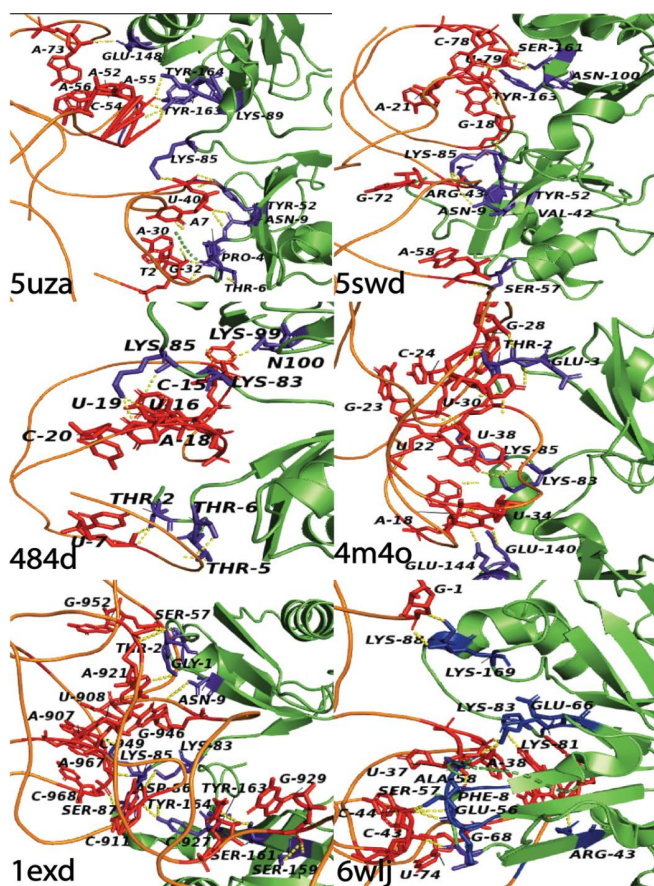


Figure 6: The molecular interaction details of the simulated aptamers with NCL. The interacting residues of the receptor and ligands are presented in licorice in purple and red respectively. The H-Bonds are presented as yellow dashed lines. Some domains might have been deflected from their inherent positions for visual acuity.

viral pathogens [10]. Targeting the NCL secretory pathway needs a kin look to unlock its therapeutic potential in various diseases. These host-directed treatments (HDTs) have many benefits, but they also come with some drawbacks and are likely to present a greater barrier to the development and selection of drug-resistant mutations. Targeting host factors could lead to more severe side effects and/or cytotoxicity due to their relevance in the host's numerous life processes. It requires personalized treatment, which necessitates a thorough health profile evaluation of the patients. It cannot be utilized as a full-fledged solo therapy, but it may typically be employed as an addition to the pathogen's present treatment plan. The development of HDTs also suffers from the lack of suitable model systems for trials. More efforts to screen for universal host targets are necessary and making use of Artificial Intelligence for host-target screening and de-novo drug design would be a necessity due to very few de-novo drugs approved in past years.

Supplementary Information

All supplemental information is accessible and available.

Funding

This research was funded by PharmaQsar Bioinformatics Firm, Kampala, Uganda. The financial support provided by the firm significantly contributed to the execution of this study.

Data availability

Relevant data associated with this study is contained within the article or in the supplementary material.

Conflict of interest

The authors declare that they have no conflicts of interest.

References

1. Wang Y, Fekadu G, You JHS. Comparative Cost-Effectiveness Analysis of Respiratory Syncytial Virus Vaccines for Older Adults in Hong Kong. *Vaccines* 11 (2023).
2. Respiratory Syncytial Virus (RSV) disease. (n.d.) (2023).
3. Griffiths C, Drews SJ, Marchant DJ. Respiratory Syncytial Virus: Infection, Detection, and New Options for Prevention and Treatment. *Clinical Microbiology Reviews* 30 (2017): 277.
4. Collins PL, Fearn R, Graham BS. Respiratory syncytial virus: Virology, reverse genetics, and pathogenesis of disease. *Current Topics in Microbiology and Immunology* 372 (2013): 3–38.
5. Azzari C, Baraldi E, Bonanni P, et al. Epidemiology and prevention of respiratory syncytial virus infections in children in Italy. *Italian Journal of Pediatrics* 47 (2021): 1–12.
6. Collins PL, Fearn R, Graham BS. Respiratory Syncytial Virus: Virology, Reverse Genetics, and Pathogenesis of Disease. *Current Topics in Microbiology and Immunology* 372 (2013): 3.
7. Coultas JA, Smyth R, Openshaw PJ. Respiratory syncytial virus (RSV): a scourge from infancy to old age State of the art review 74 (2019): 986-993.
8. Griffiths C, Drews SJ, Marchant DJ. Respiratory Syncytial Virus: Infection, Detection, and New Options for Prevention and Treatment. *Clinical Microbiology Reviews* 30 (2017): 277.
9. Ginisty H, Sicard H, Roger B, et al. Structure and functions of nucleolin. *Journal of Cell Science* 112 (1999): 761–772.
10. Tonello F, Massimino ML, Peggion C. Nucleolin: a cell portal for viruses, bacteria, and toxins. In *Cellular and Molecular Life Sciences* 79 (2022).

11. Arumugam S, Clarke Miller M, Maliekal J, et al. Solution structure of the RBD1,2 domains from human nucleolin. *Journal of Biomolecular NMR* 47 (2010): 79–83.
12. Afroz T, Cienikova Z, Cléry A, et al. One, Two, Three, Four! How Multiple RRM s Read the Genome Sequence. *Methods in Enzymology* 558 (2015): 235–278.
13. Jia W, Yao Z, Zhao J, et al. New perspectives of physiological and pathological functions of nucleolin (NCL). *Life Sciences* 186 (2017): 1–10.
14. Chong PA, Vernon RM, Forman-Kay JD. RGG/RG Motif Regions in RNA Binding and Phase Separation. *Journal of Molecular Biology* 430 (2018): 4650–4665.
15. Fujiki H, Watanabe T, Suganuma M. Cell-surface nucleolin acts as a central mediator for carcinogenic, anti-carcinogenic, and disease-related ligands. *Journal of Cancer Research and Clinical Oncology* 140 (2014): 689–699.
16. Mongelard F, Bouvet P. Nucleolin: a multiFACeTed protein. *Trends in Cell Biology* 17 (2007): 80–86.
17. Cong R, Das S, Ugrinova I, et al. Interaction of nucleolin with ribosomal RNA genes and its role in RNA polymerase I transcription. *Nucleic Acids Research* 40 (2012): 9441–9454.
18. Abdelmohsen K, Tominaga K, Lee EK, et al. Enhanced translation by Nucleolin via G-rich elements in coding and non-coding regions of target mRNAs. *Nucleic Acids Research* 39 (2011): 8513–8530.
19. Gaume X, Tassin AM, Ugrinova I, et al. Centrosomal nucleolin is required for microtubule network organization. *Cell Cycle*, 14 (2015): 902–919.
20. Inder KL, Lau C, Loo D, et al. Nucleophosmin and nucleolin regulate K-ras plasma membrane interactions and MAPK signal transduction. *Journal of Biological Chemistry* 284 (2009):28410–28419.
21. Berger CM, Gaume X, Bouvet P. The roles of nucleolin subcellular localization in cancer. *Biochimie* 113 (2015): 78–85.
22. Abdelmohsen K, Gorospe M. RNA-binding protein nucleolin in disease. 9 (2012): 799–808.
23. Peggion C, Massimino ML, Stella R, et al. Nucleolin Rescues TDP-43 Toxicity in Yeast and Human Cell Models. *Frontiers in Cellular Neuroscience* 15 (2021): 625665.
24. Caudle WM, Kitsou E, Li J, et al. A role for a novel protein, nucleolin, in Parkinson's disease. *Neuroscience Letters* 459 (2009): 11–15.
25. Bates PJ, Reyes-Reyes EM, Malik MT, et al. G-quadruplex oligonucleotide AS1411 as a cancer-targeting agent: Uses and mechanisms. *Biochimica et Biophysica Acta. General Subjects* 1861 (2017): 1414–1428.
26. Huang H, Arighi CN, Ross KE, et al. iPTMnet: an integrated resource for protein post-translational modification network discovery. *Nucleic Acids Research* 46 (2018): D542–D550.
27. Carpentier M, Morelle W, Coddeville B, et al. Nucleolin undergoes partial N- and O-glycosylations in the extranuclear cell compartment. *Biochemistry* 44 (2005): 5804–5815.
28. Chen X, Shank S, Davis PB, et al. Nucleolin-mediated cellular trafficking of DNA nanoparticle is lipid raft and microtubule dependent and can be modulated by glucocorticoid. *Molecular Therapy : The Journal of the American Society of Gene Therapy* 19 (2011): 93–102.
29. Hovanessian AG, Puvion-Dutilleul F, Nisole S, et al. The cell-surface-expressed nucleolin is associated with the actin cytoskeleton. *Experimental Cell Research* 261 (2000): 312–328.
30. Mastrangelo P, Chin AA, Tan S, et al. Identification of rsv fusion protein interaction domains on the virus receptor, nucleolin. *Viruses* 13 (2021).
31. Drysdale SB, Cathie K, Flamein F, et al. Nirsevimab for Prevention of Hospitalizations Due to RSV in Infants. *The New England Journal of Medicine* 389 (2023): 2425–2435.
32. Mastrangelo P, Norris MJ, Duan W, et al. Targeting Host Cell Surface Nucleolin for RSV Therapy: Challenges and Opportunities. *Vaccines* 5 (2017): 27.
33. Griffiths CD, Bilawchuk LM, McDonough JE, et al. IGF1R is an entry receptor for respiratory syncytial virus. *Nature* 583 (2020): 615–619.
34. Yan Y, Zhang D, Zhou P, et al. HDock: a web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Research* 45 (2017): W365–W373.
35. Yan Y, Tao H, He J, et al. The HDock server for integrated protein–protein docking. *Nature Protocols* 15 (2020): 1829–1852.
36. Van Zundert GCP, Rodrigues JPGLM, Trellet M, et al. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *Journal of Molecular Biology* 428 (2016): 720–725.
37. Kozakov D, Brenke R, Comeau SR, et al. PIPER: an FFT-based protein docking program with pairwise potentials. *Proteins* 65 (2006): 392–406.

38. Beard H, Cholleti A, Pearlman D, et al. Applying Physics-Based Scoring to Calculate Free Energies of Binding for Single Amino Acid Mutations in Protein-Protein Complexes. *PLOS ONE* 8 (2013): e82849.
39. Lu C, Wu C, Ghoreishi D, et al. OPLS4: Improving force field accuracy on challenging regimes of chemical space. *Journal of Chemical Theory and Computation* 17 (2021): 4291–4300.
40. Charles S, Paul S, Pius Edgar M, et al. (n.d.). The Hunt for antipox compounds against Monkeypox Virus Thymidylate Kinase and scaffolding protein leveraging Pharmacophore modeling, molecular docking, and molecular dynamics simulation studies. Abstract.
41. Charles S, Mahapatra RK. Artificial intelligence based de-novo design for novel Plasmodium falciparum plasmepsin (PM) X inhibitors. *Journal of Biomolecular Structure and Dynamics* (2023): 1–16.
42. Charles S, Edgar MP, Mahapatra RK. Artificial intelligence based virtual screening study for competitive and allosteric inhibitors of the SARS-CoV-2 main protease. *Journal of Biomolecular Structure and Dynamics* (2023).
43. Schrodinger. Schrödinger Press Glide User Manual Glide 6.7 User Manual Glide User Manual. (2015).
44. Shelley JC, Cholleti A, Frye LL, et al. Epik: A software program for pKa prediction and protonation state generation for drug-like molecules. *Journal of Computer-Aided Molecular Design* 21 (2007): 681–691.
45. Schrodinger. Schrödinger Press Prime User Manual Prime 4.0 User Manual Prime User Manual. (2015).
46. Thomsen R, Christensen MH. MolDock: A new technique for high-accuracy molecular docking. *Journal of Medicinal Chemistry* 49 (2006): 3315–3321.
47. Molegro virtual docker user manual. (2012).
48. Bauer P, Hess B, Lindahl E. GROMACS 2022.1 Manual (2022).
49. Lindorff-Larsen K, Piana S, Palmo K, et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* 78 (2010): 1950–1958.
50. Wang J, Wolf RM, Caldwell JW, et al. Development and testing of a general amber force field. *Journal of Computational Chemistry* 25 (2004): 1157–1174.
51. Arantes PR, Polêto MD, Pedebos C, Ligabue-Braun R. (n.d.). Making it rain: cloud-based molecular simulations for everyone Dynamics of bioactive compounds and impacts on molecular recognition View project Design and development of thrombin inhibitors View project.
52. Tian C, Kasavajhala K, Belfon KAA, et al. Ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *Journal of Chemical Theory and Computation* 16 (2020): 528–552.
53. Riboswitches, RNA Regulation | Learn Science at Scitable. (n.d.) (2023).
54. Kuntz ID, Chen K, Sharp KA, et al. The maximal affinity of ligands. *Proceedings of the National Academy of Sciences* 96 (1999): 9997–10002.
55. Grossfield A, Patrone PN, Roe DR, et al. Best Practices for Quantification of Uncertainty and Sampling Quality in Molecular Simulations [Article v1.0]. *Living Journal of Computational Molecular Science* 1 (2019).
56. Brown DK, Penkler DL, Amamuddy OS, et al. MD-TASK: a software suite for analyzing molecular dynamics trajectories. *Bioinformatics* 33 (2017): 2768.
57. Abdi H, Williams LJ. Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics* 2 (2010): 433–459.