---

---

2021-08-01

# Molecular Dynamics Of Nucleic Acids, Proteins, And Their Complexes: From Atomistic And Coarse-Grained Descriptions To New Methodologies

Parth Chaturvedi
*University of Texas at El Paso*

---

## Recommended Citation

Chaturvedi, Parth, "Molecular Dynamics Of Nucleic Acids, Proteins, And Their Complexes: From Atomistic And Coarse-Grained Descriptions To New Methodologies" (2021). *Open Access Theses & Dissertations*. 3229.
https://scholarworks.utep.edu/open_etd/3229

# MOLECULAR DYNAMICS OF NUCLEIC ACIDS, PROTEINS, AND THEIR COMPLEXES: FROM ATOMISTIC AND COARSE-GRAINED DESCRIPTIONS TO NEW METHODOLOGIES

PARTH CHATURVEDI

Doctoral Program in Chemistry

APPROVED:

Lela Vuković, Ph.D., Chair

Ricardo A. Bernal, Ph.D.

Sreeprasad Sreenivasan, Ph.D.

Tunna Baruah, Ph.D.

Stephen L. Crites, Jr., Ph.D.
Dean of the Graduate School

# MOLECULAR DYNAMICS OF NUCLEIC ACIDS, PROTEINS, AND THEIR COMPLEXES: FROM ATOMISTIC AND COARSE-GRAINED DESCRIPTIONS TO NEW METHODOLOGIES

by

PARTH CHATURVEDI, MS

DISSERTATION

Presented to the Faculty of the Graduate School of
The University of Texas at El Paso
in Partial Fulfillment
of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

Department of Chemistry and Biochemistry
THE UNIVERSITY OF TEXAS AT EL PASO
August 2021

# Acknowledgements

I thank my advisor Dr Lela Vuković for supporting and motivating me throughout the projects. I also thank my lab mates Payam, Tara and Ali (former) for useful inputs and discussion. I acknowledge the Texas Advanced Computing Center (TACC) and the Blue Waters project at the National Center for Supercomputing Applications for computational resources. I also thank the Department of Chemistry and Biochemistry at the University of Texas at El Paso and acknowledge the financial support from UTEP (startup funding) and the National Institutes of Health (R03AI142553).

# Abstract

Computational modeling of biological macromolecules through molecular dynamics (MD) simulations is becoming increasingly useful to interpret the experimental results and guide the future experiments. Over the last decade, MD simulations have become much more powerful and accessible due to advancements in fields of structural biology, computer science and integrated chip design. In this thesis, I utilize different MD simulation techniques to model interaction of and antiviral mechanisms. Within the framework of nucleic acid modelling, we describe first dynamics of small circular single stranded nucleic acids and highlight key dynamical differences of backbone behavior in solution between DNA and RNA and their ability to capture ions for much longer timescales as compared with linear nucleic acids. Second, we model single stranded DNA corona on carbon nanotubes (CNT) and quantify the number of nucleotides that cover CNTs. Third, we modelled melting mechanisms of DNA duplexes on laser-excited gold nanoparticles, which can provide control over heating in biological systems through laser-based sources. Within the framework of modeling of antiviral mechanisms, first, I describe computational models of APOBEC3G protein, which is known to restrict HIV-1 virus, and examine its interactions with ssDNA. Next, we model broad-spectrum antivirals nanoparticles, mimicking heparan sulfate proteoglycans against human papillomavirus (HPV) in which we identify viral capsid regions that facilitate binding with such nanoparticles. Lastly, I propose peptide-based therapeutics for fast mutating SARS-CoV-2 virus and design an adaptive algorithm to generate peptide templates against different mutants of SARS-CoV-2.

# Table of Contents

# List of Tables

# List of Figures

# Chapter 1: Introduction

The understanding of the biomolecule structures provides key knowledge to our understanding of the living world. Biomolecule structures, determined by interactions of particles comprising the biomolecules, determine the biological function..[1] Since initial structure determinations of biomolecules in the 1950s, our knowledge of such macromolecular structures has been increasing. At present, protein data bank (PDB)[2] holds more than 155,000 entries, including more than 130,000 proteins, approximately 2,100 nucleic acids and 7200 complexes of protein-nucleic acids. Such structural knowledge advances our understanding of basic biological phenomena such as translation, transcription, transport across membranes, enzyme regulation etc. Despite their enormous impact, structures stored at the PDB do not capture complete dynamics of such flexible biomolecules, which play a key role in their functionality. Ideally, one would like to watch these biomolecules in action, to perturb them at the atomic level, and to see how they respond. As observation of the motions of individual atoms of biomolecules and perturbing them in a desired fashion is difficult to achieve in real-life experiments, an increasingly attractive alternative is to work with an atomic-level or coarse-grained molecular dynamics (MD) simulations of such biomolecules.[3]

MD simulations tries to predict how every atom in a molecular system will move over time, based on a general physics model governing interatomic interactions,[4] which is known as a molecular mechanics force field (FF) function.[5] These FF functions can be both all-atomistic (AA) such as Chemistry at Harvard Macromolecular Mechanics (CHARMM)[6] FF and Assisted Model Building with Energy Refinement (AMBER)[7] FF or coarse-grained (CG) such as MARTINI[8] FF. MD simulations utilize classical laws of motion to predict the spatial position of each atom as a function of time.

In last two decades, MD simulations have become substantially more accessible, and they have increasingly been used to interpret both experimental results and to guide experimental work. There are two major contributors to wider use of MD. First, technical improvements in field of structural biology have led to an explosion of new structures discovered in the last decade with significantly improved resolutions, spearheaded by advancements in cryo-electron microscopy (cryo-EM).[9] [10] Cryo-EM technique has enabled structure determination of different states of flexible proteins, but is currently limited to larger proteins with atomic masses greater than 50 kDa.[11] Nuclear magnetic resonance (NMR) spectroscopy is also used determine three-dimensional model of a protein in solution by obtaining ensemble of structures but NMR is restricted to smaller proteins with molecular weight of less than 30 kDa.[12] A combination of increasing structural data and limitations of such expensive experimental techniques have lead to interest in utilizing MD simulations to understand molecular systems.

Secondly, MD simulations have become much more powerful and accessible over the past decade due to improvements in force field parameters and utilization of powerful graphics processing units (GPUs)[13] to run MD simulations, which has allowed larger and longer simulations to be run locally at a modest cost.[14] [15] These advances have allowed atomistic MD simulations of large macromolecular systems with biological relevance such as entire proteins in solution with explicit solvent, nucleosomes[16] [17], virus capsids[18] [19], ribosomes[20] [21], among others. MD simulations can today perform millisecond long simulation of proteins[22] or describe models of entire cell organelle[23]. Computational limitations of all atomistic simulations are increasingly being addressed by improving coarse-grained forcefields, which reduces the degrees of freedom and greatly accelerates MD simulations. Coarse-grained models have been utilized to understand large scale systems such as membrane proteins, lipid bilayers, nanoparticle aggregates etc.[24]

In my present work, I have utilized different MD simulation techniques to model nucleic acids and antiviral mechanisms. In subsequent chapters, I describe: modelling of small single stranded circular nucleic acids where I highlight key dynamical differences of backbone behavior in solution between DNA and RNA and ability to capture ions for much longer timescales as compared with linear nucleic acids; modelling of single stranded DNA corona on carbon nanotubes (CNT) to quantify the number of nucleotides that cover CNTs, modelling melting mechanisms of DNA duplexes on laser-excited gold nanoparticles which have potential for drug delivery applications. The next part of my thesis describes projects related to different antiviral mechanisms. These projects include: building computational models of APOBEC3G protein, which is known to restrict HIV-1 virus and understanding its interactions with ssDNA; modeling of broad-spectrum antivirals employing nanoparticles, mimicking heparan sulfate proteoglycans against human papillomavirus (HPV), in which we identify viral capsid regions that facilitate binding with such nanoparticles; developing peptide-based therapeutics for fast mutating SARS-CoV-2 virus which showcases our innovative adaptive algorithm that generates peptides targeting against different mutants of SARS-CoV-2.

# Chapter 2: Methods

All the systems described in this proposal were investigated using atomistic and coarse-grained classical molecular dynamics (MD) simulations. This chapter provides overview of the principles of MD simulations, force fields involved, parameters used to define interactions between atoms, coarse-grained beads and the integration method used to solve the equations of motion for a chosen thermodynamic ensemble.

## 2.1 CLASSICAL MOLECULAR DYNAMICS

Classical molecular dynamics are based on classical equation of motion using Newton's equations. Solving Newton’s for a system of N particles gives us time evolving functions of positions and momenta of particles which is sufficient to completely describe the system. Newton's equations of motion for a system of N particles are defined as

$$m_i \frac{\partial^2 \vec{r_1}}{\partial t^2} = \vec{f_i}$$
$$\vec{f_i} = -\frac{\partial}{\partial \vec{r_1}} U(\vec{r_1}, \vec{r_2}, \dots, \vec{r_N})$$

where $m_i$ is mass of particles, $\vec{f_i}$ denotes force, $r_i$ denotes position vectors and $U$ is the potential of the system. To solve one needs to calculate forces $\vec{f_i}$ exerted on particle *i* by all the other particles present in the system. Assuming that the masses ($m_i$) are already known, one needs to define a potential energy function $U$, which is a function of 3N coordinates ($\vec{r_1}, \vec{r_2}, \dots, \vec{r_N}$) which accurately takes into account all the possible interactions within the system. Obtaining potential forms of $U(\vec{r_1}, \vec{r_2}, \dots, \vec{r_N})$ and solving the above equations numerically and efficiently are the focal areas of all MD engines, such as NAnoscale Molecular Dynamics (NAMD)[25] and GROningen MAchine for Chemical Simulations (GROMACS)[26]. Potential functions are defined as force fields which approximate interaction of particles through functional form of known simpler potential containing parameters that may be derived from experiments, quantum calculations, ab initio methods etc.

## 2.2 ATOMISTIC FORCE FIELD

For atomistic force fields, the particles are considered rigid atoms with fixed atomic mass (m). The parameterization of force fields, for example CHARMM is done through means of quantum mechanical calculations. The calibration of force field parameters is done by comparing experimental and simulated thermodynamics properties including heats of vaporization and solvation free energy, etc.[27] [28] Force field parameters are divided into two groups, including bonded (defining intramolecular interactions) and non-bonded (defining intermolecular interactions) ones. In classical sense, the potential of the system can be consisting of additive potentials can be categorized as bonded and non-bonded potentials. Consider a full system potential $V_N = U(\overrightarrow{r_1}, \overrightarrow{r_2}, \dots, \overrightarrow{r_N})$ as sum of bonded and non-bonded interaction energies:

$$V_N = \sum V_{N,bonded} + \sum V_{N,non-bonded}$$

Bonded potentials ($V_{N,bonded}$) describe intramolecular interactions that includes stretching, bending and torsions of chemical bonds. Chemical bonds are considered explicitly in molecular dynamics and are treated as springs which is a minimal description of chemical bonds hence computationally efficient. $V_{N,bonded}$ of CHARMM force field is composed of bond, angle, dihedral and improper dihedral potentials, as shown in Figure 1, which are explained as follows:

$$V_{bond} = \sum_{bond\ i} k_i^{bond}(r_i - r_{0i})^2$$

$$V_{angle} = \sum_{angle\ i} k_i^{angle}(\theta_i - \theta_{0i})^2$$

$$V_{dihedral} = \sum_{dihedral\ i} k_i^{dihedral}[1 + cos(n_i\phi_i - \gamma_i)] \quad n_i \neq 0$$

$$V_{improper} = \sum_{improper\ i} k_i^{improper} (\phi_i - \gamma_i)^2 \quad n_i = 0$$

Above, $V_{bond}$ , $V_{angle}$ and $V_{improper}$ are described by weak harmonic potential forms with spring constants ( $k_i^{bond}$, $k_i^{angle}$ , $k_i^{improper}$) and associated instantaneous internal coordinates ($r_i$), angles ($\theta_i$), dihedrals and improper dihedrals ($\phi_i$), which differ from their equilibrium values ($r_i$, $\theta_i$, $\gamma_i$) in case of external interaction. The equilibrium values r, theta, and gamma represent stable equilibrium state with minimum stretching, bending and torsional energies. Cosine function is used to define $V_{dihedral}$, which is depends force constant $k_i^{dihedral}$, periodicity $n_i$ and dihedrals ($\phi_i$) varying from $\gamma_i$.

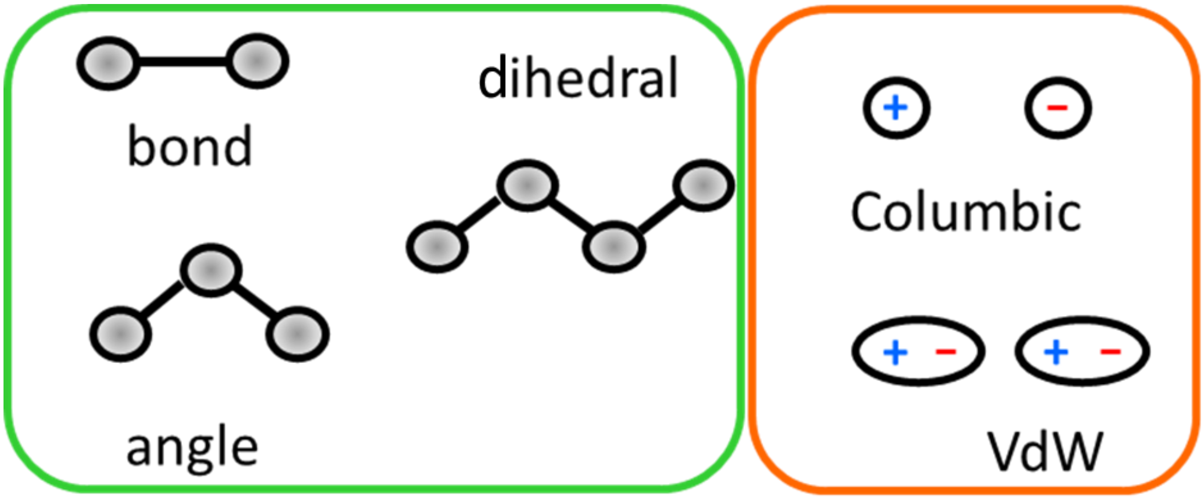


**Figure 1.** Pictorial representation of bonded and non-bonded interaction parameters.

Non-bonded potential energies include Coulomb and van der Waals (vdW) potentials, which are pair-wise interactions. Columbic interactions are calculated using classical coulombs law for point charges. Coulomb potential between atoms i and j, which carry $q_i$ and $q_j$ charges, respectively is defined as

$$V_{coul}(r_{ij}) = \sum_i \sum_{j>i} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}$$

Above, $\epsilon_0$ and $r_{ij}$ denote vacuum permittivity and distance between the center of atoms. The short-ranged Lennard-Jones (LJ) potential is used to describe van der Waals (vdW) interactions between atom pair $i - j$. The Lennard-Jones potential is composed of two terms - Pauli repulsion at short ranges due to overlapping electron orbitals and the attractive long-range term which describes attraction at long ranges (van der Waals force or dispersion force). LJ potential is defined as

$$V_{LJ}(r_{ij}) = \sum_{i}\sum_{j>i} \tau_{ij}\left[\left(\frac{R_{min,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{min,ij}}{r_{ij}}\right)^{6}\right]$$

where $r_{ij}$ is the distance between atoms i and j, and $R_{min,ij}$ and $\tau_{ij}$ are determined from the Lorentz-Berthelot:

$$R_{min,ij} = \frac{R_{min,i} + R_{min,j}}{2} \quad \tau_{ij} = \sqrt{\tau_i \tau_j}$$

where $R_{min,i}$ , $R_{min,j}$ and $\tau_i$, $\tau_j$ represent the radii and potential well depths of particles i and j, respectively.

### 2.3 COARSE GRAINED FORCE FIELD

The use of coarse grained (CG) models has proven to be a valuable tool to probe the time and length scales of systems beyond what is feasible with traditional all atom (AA) models.[29] We have used MARTINI coarse-grained force field for our research. This model is based on a four-to-one mapping, i.e., on average, four heavy atoms are represented by a single interaction center[30] and can be applied on water, lipds, protein, nucleic acids.[31] There are four main types of interaction sites: polar (P), nonpolar (N), apolar (C), and charged (Q). Each particle type has a number of subtypes, which allow for a more accurate representation of the chemical nature of the underlying atomic structure. Within a main type, subtypes are either distinguished by a letter denoting the hydrogen-bonding capabilities (d = donor, a = acceptor, da = both, 0 = none), or by a number indicating the degree of polarity (from 1, low polarity, to 5, high polarity).

Non bonded interactions include short range interactions described by Lennard-Jones potential and long-range coulombic interaction. These interactions are described by same mathematical formulation as CHARMM which is also used in many all atomistic force fields. Bonds and chain stiffness are described by a weak harmonic potential and the mathematical formulation is same as for all atomistic force field. Dihedral potential described for all atomistic models is only used to describe ring particles.[29]

Solvent water is described by a 4 to 1 mapping to coarse grained (CG) particles called P4 particles, and they lead to higher than normal freezing temperature of water. To rectify this problem, Anti-freeze water particles are introduced (denoted by BP4) for which the LJ parameter $\tau_{ij}$ for BP4−P4 interactions is scaled up. The BP4-BP4 interaction remains the same. Usually, 10% CG particles of solvent should be anti-freeze water particles.[29]

## 2.4 INTEGRATION METHOD FOR MD

The first integral of Newton's equations of motion provides momenta and second integral provides space coordinates of the particle. To compute the integrals numerically, velocity-Verlet integration algorithm is used. For a given particle's position ($R_n$), velocity ($v_n$) and force ($f_n$) at time step (n) the algorithm outputs same values for the next step ($R_{n+1}$, $v_{n+1}$, $f_{n+1}$). The difference between the two consecutive steps is small but finite. The velocity-Verlet algorithm is implemented according to the following steps:

"half-kick" $v_{n+1/2} = v_n + m^{-1} f_n . \Delta t/2$

"drift" $R_{n+1} = R_n + v_{n+1/2} . \Delta t$

"compute force" $f_{n+1} = f(R_{n+1})$

"half-kick" $v_{n+1} = v_{n+1/2} + m^{-1} f_{n+1} . \Delta t/2$

The advantages of the velocity-Verlet algorithm are time reversibility and symplecticity leading to conserved momentum.

## 2.5 PERIODIC BOUNDARY CONDITIONS (PBC)

MD simulations have limitations to simulate macroscale systems containing a large number of atoms (Avogadro's number) due to large number of degrees of freedom. A finite system will also have boundary effects that are undesirable in most situations. To resolve these issues, periodic boundary conditions (PBC) are defined in simulations in three dimensions or two dimensions (slab boundary condition). Through PBC condition, the original simulation unit cell interacts with its adjacent periodic images in each direction.

## 2.6 ENSEMBLES

According to the condition of the experiments, different thermodynamics ensembles can be applied to implement MD simulations. There are several simulation ensembles, such as NVE (the number of particles N, volume V and total energy E of the system are held constant), NVT (N, V and temperature T are held constant), and NPT (N, V, and pressure P are held constant). All our simulations used NPT ensemble. The coupling of pressure and temperature of the systems are maintained by using Langevin dynamics method. This method damps T and P fluctuations by adding a damping and fluctuating terms to the Newton's equation of motion:

$$m\frac{\partial^2\vec{r}}{\partial t^2} = m\dot{v} = F(r) - \gamma_{Lang}v + \sqrt{2\gamma_{Lang}k_BTm}\,G(t)$$

where r, t, m, v, and F denotes coordinate, time, mass, velocity and force respectively. $\gamma_{\mathrm{Lang}}$, $k_B$ and $T$ denote friction factor (depending on the systems and defined by user), Boltzmann constant and temperature respectively. G(t) denotes a single variable Gaussian process.

The second and third terms in the equation are damping and fluctuating terms, respectively. $\gamma_{\text{Lang}}$ controls the magnitudes of damping and fluctuating terms. For MARTINI simulations, Parrinello–Rahman system is used for pressure coupling, which is a modified version of Langevin dynamics.[32]

### 2.7 SIMULATION PARAMETERS

All the force fields used in the proposal have non-bonded interaction energies including both Coulombic (long range) and vdW (short range) potentials. The cut-off distance for vdW potential energies are explicitly defined to ignore the interaction energies beyond the cut-off distance. In CHARMM and MARTINI force field, both vdW cut-off distance is usually set to 1.2 nm. Smoothing this potential beyond the cut-off distance is necessary to avoid potential artifacts due to sharp truncation of vdW potentials. The smoothness occurs between the cut-off distance and the switching distance (usually 0.8 nm), which is defined by the end user. Coulomb potential is evaluated for only those atoms whose distances are more than the vdW cut-off distance. Particle-mesh Ewald (PME) method[33] is used to efficiently calculate the Coulombic electrostatic interactions in periodic systems. We used explicit solvent (water) molecules, which were modeled by the TIP3P water model[34] for all atomic simulations and MARTINI model for coarse grained simulations. The time step for integrator is set to 1-2 fs for CHARMM and 20-30 fs for MARTINI.

# Chapter 3 : Structural and dynamical properties of small single stranded circular nucleic acids

## 3.1 INTRODUCTION

Nucleic acid biology is presently in a very exciting period: more and more genomic information has been determined by advanced next-generation sequencing techniques[35], properties of deoxyribonucleic acid (DNA) on molecular and cellular levels are under intense investigation, and numerous novel ribonucleic acid (RNA) species, RNA functions and RNA targets are being discovered in living cells. RNA is been used to develop vaccine as a much faster pace.[36] Many new classes of plant and mammals infecting viruses have been found, which have novel genomes consisting of small single stranded nucleic acids.[37] [38] Besides DNA and RNA molecules existing in cells, these molecules have a great potential for use in nanotechnology and biomedical fields. Progress has been made to incorporate nucleic acids in molecular origami[39], nanoparticles[40] and carbon nanotubes[41]. Studies are slowly revealing how small RNAs play important role in regulation of proteins in eukaryotic cells.[42] [43] [44]

As an emerging class of nucleic acid molecules with potential biomedical applications, not much is known about the structural and dynamical properties of circular single-stranded nucleic acids and constraints that their geometry imposes on these properties. Our goal is to examine these properties for small single-stranded circular nucleic acids using classical atomistic molecular dynamics (MD) simulations. Reliable parameters have been determined and validated for atomistic descriptions of DNA and RNA molecules.[45] [46]

High stability in biological systems is a major criterion for why circular nucleic acids especially circular RNA (circRNAs) are becoming interesting for RNA-centered medical applications. Also, circularity is associated also with some other molecular properties that may be useful for therapeutic purposes.[47] Potential applications for circular nucleic acids are summarized as (a) MicroRNA sponging in competing-endogenous RNAs, including circRNAs, prevents homologous target mRNAs from degradation or allows their translation.[48] [49] (b) Due to stability of circular nucleic acids, circularization of classically employed linear RNAs is possible because compared to the rigid dsDNA helix, the flexibility of RNA molecules allows circularization of even very small RNAs.[50] [51] (c) In the absence of in-frame stop codons and termination signals, or after read-through, rolling-circle-translation can occur on circular RNAs which can lead to translation of multimeric repetitive protein motifs, which could be useful for cell engineering, but which can become toxic for cells if happening in an uncontrolled fashion.[52] [53] (d) CircRNAs can be used as tools to purposefully boost innate immune signaling to counteract disease-induced immunosuppression.[54] [55] (e) Circular nucleic acids (both circular ssDNA (circDNA) and circRNA) can be used for creating aptamers that bind with high specificity which are useful for selective drug delivery or protein activity control and circularizing them might make them more potent.[56] For example, circular DNA aptamer that were targeted against thrombin have half-lives that are extended beyond 10 h in serum and plasma, making such constructs viable for therapeutic and diagnostic applications.[57] The average half-life of endogenously produced 3′-5′-linked circRNA was found to amount to 19–24 h[58] and can be up to 48 h.[59] (f) Lastly, they can serve as a versatile material for designing and building novel nanodevices such as molecular switches, molecular motors and walkers.[60]

For our study, we have constructed several non-base-paired single stranded circular nucleic acids of varying length and sequence. The purpose of choosing non base-paired sequence is to understand the dynamics of the backbone of the nucleic acids without interference from paired bases as they will dominate the structure conformation owing to strong interaction. The sequences examined include ssDNAs: $(GT)_n$ (n=3-24), $(C)_{12}$, $(T)_{12}$,$(AC)_8$, and ssRNAs: $(GU)_n$ (n=3-24), $(C)_{12}$, $(U)_{12}$, $(AC)_8$. The set of non-base paired circular nucleic acid molecules was relaxed, solvated in water and simulated for 1 μs in 0.15 M NaCl.

## 3.2 METHODS

### 3.2.1 Building Circular Nucleic Acids

Linear single-stranded DNA and RNA molecules were prepared by 3DNA software[61] [62] and make-na web server based on B-form of duplex DNA which is the most abundant form. Circular potential density was obtained from dummy sodium atoms of adjusted diameters according to the length of the sequence. All linear molecules except $(GU)_{24}$ and $(GT)_{24}$ were then docked with these respective circular potentials using colores (Situs) software [63] with a resolution of 5 Å.

Molecular dynamics flexible fitting (MDFF) method[64] was used force fit these docked linear molecules into circular shape. In this method, an Electron Microscopy(EM) density map as a potential so that high density areas in the map correspond to energy minima, so that the atoms in the structure are subject to forces proportional to the gradient of the EM map. The potential is defined on a 3-D grid with 3D vector $R$ by

$$U_{EM}(R) = \sum_j \omega_j \, V_{EM}(r_j)$$

where

$$V_{EM}(r) = \xi \left[1 - \frac{\Phi(r) - \Phi_{thr}}{\Phi_{max} - \Phi_{thr}}\right] \quad \text{if} \quad \Phi(r) \geq \Phi_{thr}$$

$$V_{EM}(r) = \xi \qquad \text{if} \quad \Phi(r) < \Phi_{thr}$$

Here $\omega_j$ corresponds to a per-atom weigh, typically set to the atomic mass, $\xi$ is a force scaling, $\Phi(r)$ is the EM density at position $r$, $\Phi_{max}$ is the maximum value of the EM density map, and $\Phi_{thr}$ is a density threshold. The purpose of the density threshold is to remove from the EM data the solvent contribution.

An atom placed in this external potential feels a force of

$$\begin{aligned} f_i^{EM} &= -\frac{\partial}{\partial r_i} U_{EM}(R) \\ &= -\omega_i \frac{\partial}{\partial r_i} V_{EM}(r_i) \end{aligned}$$

MDFF introduces harmonic restraints to preserve the secondary structure of proteins and nucleic acids where restraints are imposed to seven dihedral angles and two interatomic distances between base pairs.

In MDFF simulations, our linear nucleic acid molecules were coupled to the density obtained from dummy atoms arranged in ring shapes. In some of the systems, soft harmonic potentials were used to fit the linear molecules into the potential. Then, after the linear molecules were fitted to ring shapes, a covalent bond between the O3′ atom of the last nucleotide and the P atom of the first nucleotide was created by a patch that we added to the force field, which mimics the usual nucleotide–nucleotide bond.

### 3.2.2 Molecular Dynamics simulations

The prepared circular molecules were solvated with TIP3P water and neutralized with 0.15 M NaCl with solvate and ionize VMD plugins, respectively.[65] Separately, systems of multiple $(GT)_6$ circDNAs and $(GU)_6$ circRNAs were simulated in 0.15 M NaCl aqueous solutions; these molecules were present at either 5.9- or 7.9-mM concentrations. The aqueous solution with monovalent ions (NaCl) was selected as solvent for simplicity. The systems were described with CHARMM36 force field parameters.[45] MD simulations were performed with the NAMD2.12 package.[27] All simulations were conducted with Langevin dynamics (Langevin constant $\gamma_{\mathrm{Lang}}$ = 1.0 $\mathrm{ps}^{-1}$) in the NpT ensemble, where temperature and pressure remained constant at 310 K and 1 bar, respectively. The particle-mesh Ewald (PME) method was used to calculate the Coulomb interaction energies, with periodic boundary conditions applied in all directions.[33] The time step was set to 2.0 fs. The evaluation of long-range van der Waals and Coulombic interactions was performed every 1- and 2-time steps, respectively. After 2000 steps of minimization, solvent molecules were equilibrated for 2 ns around the circular molecules, which were restrained by using harmonic forces with a spring constant of 1 kcal/(mol Å$^2$). Next, the systems were equilibrated in production MD runs, with center-of-mass (COM) restraints applied to the whole molecule COM with a force constant of 0.5 kcal/mol, so the molecule remains in the original unit cell.

### 3.2.3 Data Analysis

*Contact Area* - To analyze the contact area with the surrounding solvent, the solvent-exposed fraction of the nitrogenous bases of circular nucleic acids was calculated in time t and defined as

$$\frac{\mathrm{a}_{base}(\mathrm{t}) + \mathrm{a}_{water}(\mathrm{t}) - \mathrm{a}_{base \cup water}(t)}{2}$$

where $\mathrm{a}_{base}(\mathrm{t})$, $\mathrm{a}_{water}(\mathrm{t})$, and $\mathrm{a}_{base \cup water}(t)$ are the solvent accessible surface areas (SASA) of the bases, water, and bases and water together, respectively. The evaluation was done by the SASA built-in VMD plugin[65], where the van der Waals radius of 1.4 Å was assigned to atoms to identify the points on a sphere that are accessible to the solvent.

*Radial Distribution Function (RDF)* - g(r) was calculated for all phosphorus atoms in the circular nucleotides as a function of time by using g(r) GUI Plugin, v1.3, a built-in VMD plugin.[65]

*Radius of Gyration* - The radius of gyration of circular nucleotide chain was evaluated by measure, a built-in VMD plugin,[65] as a function of time. The radius of gyration $r_{gyr}^2$ is defined as

$$r_{gyr}^2 = \sum_{i=1}^{n} (r(i) - \bar{r})^2$$

where $r(i)$ is the position of the $i$th atom and $\bar{r}$ is the geometric center of the circular nucleic acid.

*Ion Residence Times* - The residence time, $\tau_r$, is evaluated using a standard time correlation function $c_i(t)$ for sodium ion binding to nucleic acid $i$:

$$C(t) = \sum_{i} c_i(t)$$

$$c_i(t) = \sum_{n} p_i(t_0)p_i(t_0 + n\Delta t)$$

where $p_i(t)$ is unity if a sodium ion is within 3 Å of backbone of nucleic acid i and zero otherwise, $n$ is an integer, and $\Delta t$ is 20 ps. The correlation function calculations were performed for the last 0.8 µs of trajectories. Residence times of hovering ions were calculated by evaluating $c_i(t)$ over 3 ns long segments of trajectories. Typically, the $C(t)$ function takes a form of a decaying exponential:

$$C(t) \sim \exp\left(-t/\tau_R\right)$$

where $\tau_r$ constant represents the residence time. When the ion–nucleic acid binding event occurs for longer than 3 ns, such ions are considered to be bridged/trapped ions. These binding events have much longer residence times and do not behave as a decaying exponential function, as these events are stochastic and rare for the time scale of the simulations performed.

## 3.3 RESULTS

### 3.3.1 Conformations of Circular Nucleic Acids in Aqueous Solution.

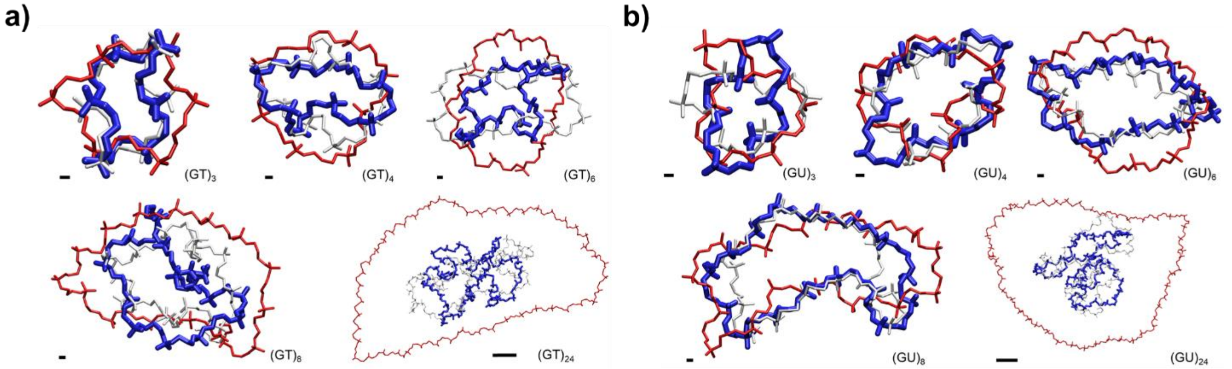


**Figure 2.** Conformations of (a) circDNA-(GT)n and (b) circRNA-(GU)n at 0 µs (red), 0.5 µs (white), and 1 µs (blue) as observed in MD simulations. The scale bars for n = 3, 4, 6, and 8 are 1 Å in length, and the scale bar for n = 24 is 10 Å in length.

In Figure 2, three representative structures of circDNA-(GT)$_n$ and circRNA-(GU)$_n$ molecules with n = 3, 4, 6, 8, and 24 are shown after 0, 0.5, and 1 µs of equilibration. Structures of the smallest circNAs, 6 and 8 nts in length, remained similar to the initial structures, as they experience the largest steric strain due to covalently bonded ends. For these smallest circNAs, limited refolding was observed, as the molecules sought the optimal placement of negatively charged backbones, while protecting the hydrophobic bases from contact with water. circNAs with a larger number of nucleotides, 12–48 nts, refolded significantly during the course of trajectories, as they sought more energetically favorable conformations. Equilibrated structures (blue) in Figure 2 indicate that backbones of circDNA-$(GT)_n$ are more likely to be bent than backbones of circRNA-$(GU)_n$. Also, backbones of circDNAs of intermediate lengths (12–16 nts) are more likely to have sharp kinks than the backbones of circRNAs.

Backbone flexibility of the molecules studied was analyzed by calculating the angles between phosphorus (P) atoms of three neighboring phosphate groups, averaged over all triplets of P atoms on neighboring nucleotides and over time (inset of Figure 3).

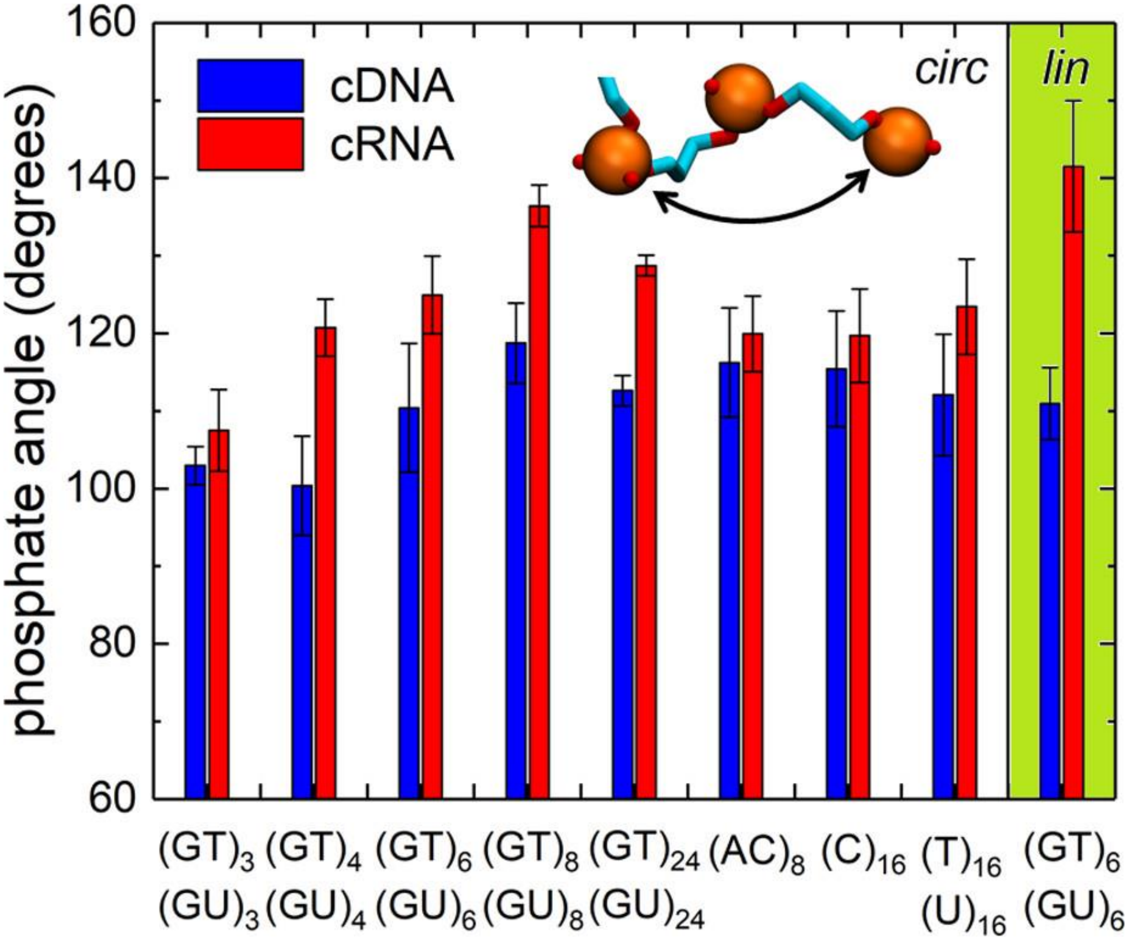


**Figure 3.** Phosphate angles in the nucleic acids examined, averaged over all P atom triplets and over the last 500 ns of trajectories.

The calculated average angles, shown in Figure 3, are consistently smaller in circDNA than in circRNA. The average angles of three consecutive P atoms in circDNAs range from 100° to 118°, while these angles in circRNAs range from 105° to 135°. For all circNAs, except for the shortest and the most sterically strained 6-nt circles, the distributions of the angle values (fluctuations) are also always wider (larger) for circDNAs than for circRNAs, as observed in Figure 3. The values of the measured angles and the comparative fluctuations of these angles indicate that circDNA backbones favor more bent structures (smaller angles) and are less rigid than circRNA backbones. The trends observed for (GT)n/(GU)n molecules are confirmed for other circDNAs ((AC)$_8$, (C)$_{16}$, (T)$_{16}$) and circRNAs ((AC)$_8$, (C)$_{16}$, (U)$_{16}$) sequences, as shown in Figure 3. The trends observed here for circular nucleic acids are consistent with previous experimental observations for linear single-stranded nucleic acids, where end-to-end FRET measurements showed that (T)$_{40}$ ssDNA is more flexible and has shorter persistence length than $(U)_{40}$ ssRNA.[64] However, recent atomistic and coarse-grained simulations demonstrate the opposite behavior of ssRNA and ssDNA for the selected simulation parameters and force fields.[66] Because elasticity and flexibility of single-stranded nucleic acids have great sensitivity to nucleotide sequence and whether base pairing is possible, salt species, and salt concentrations in the solution[67 68], more detailed studies are required to examine whether our observations, which agree with experimental observations as reported, are general.

### 3.3.2 Positioning of Phosphates, Sugars, and Bases in Circular Nucleic Acids

Binding of circNAs to other molecules can occur via different interactions, including Coulomb, hydrophobic, and hydrogen-bonding interactions. The phosphate groups of circNAs should play a key role in their Coulomb interactions. Because conformations of circDNAs and circRNAs are different (Figure 2), the positioning of their phosphate groups and negative charge densities should be different as well.

The arrangement of phosphate groups in circNA molecules is analyzed in the form of radial distribution functions (RDFs) of P atoms on phosphate backbones of circNAs, as shown in Figure 4.

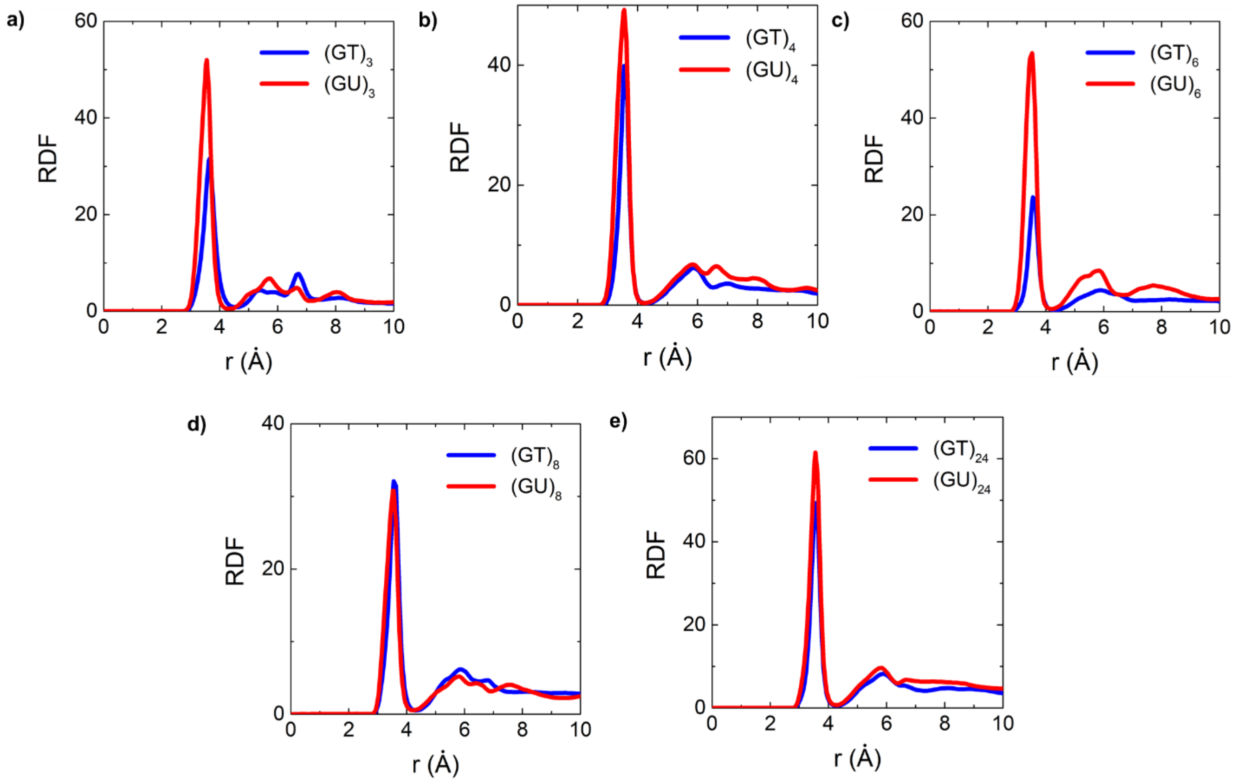


**Figure 4.** Radial distribution functions (RDFs) of phosphorous atoms of circNAs calculated over 1μs in MD simulations. RDFs are computed with a 10 ns time step. (GU)n circRNAs have P-atoms that are more likely to be closer to each other than P-atoms of (GT)n circDNAs, for all circNAs examined.

Consistently, RDFs have peaks at smaller P–P distances in circRNAs than in circDNAs. The probability to observe P–P distances less than 5 Å is significant for circRNAs but is noticeably smaller in circDNAs. $Na^+$ ions are often found in the solvation shell of phosphate ions but can also bridge the backbones of circNAs, bringing close together the phosphate groups (either neighboring or non-neighboring). The circRNA-(GU)$_n$ molecules with n < 24 have a larger average number of $Na^+$ counterions present at close distances than circDNA-(GT)$_n$ molecules, as shown in Figure 5a, likely because of the more hydrophilic nature of an additional 2′-OH group on the RNA phosphate backbone. circRNA-(GU)$_n$ molecules are observed to have 0.4–1.3 more $Na^+$ ions per molecule than circDNA-(GT)$_n$ molecules.

Next, the behavior of the hydrophobic bases of circNAs was examined. As in all nucleic acids, the bases in circNA molecules often exhibit stacking. Non-neighboring bases can also form hydrogen bonds but not base pair, due to the choice of non-base-pairing sequences in this study.

Besides stacking, bases in circNA molecules are also partly exposed to water due to the circular geometry and the flexibility of these molecules. Figure 5d show the areas of nucleotide bases exposed to the aqueous solution (water).

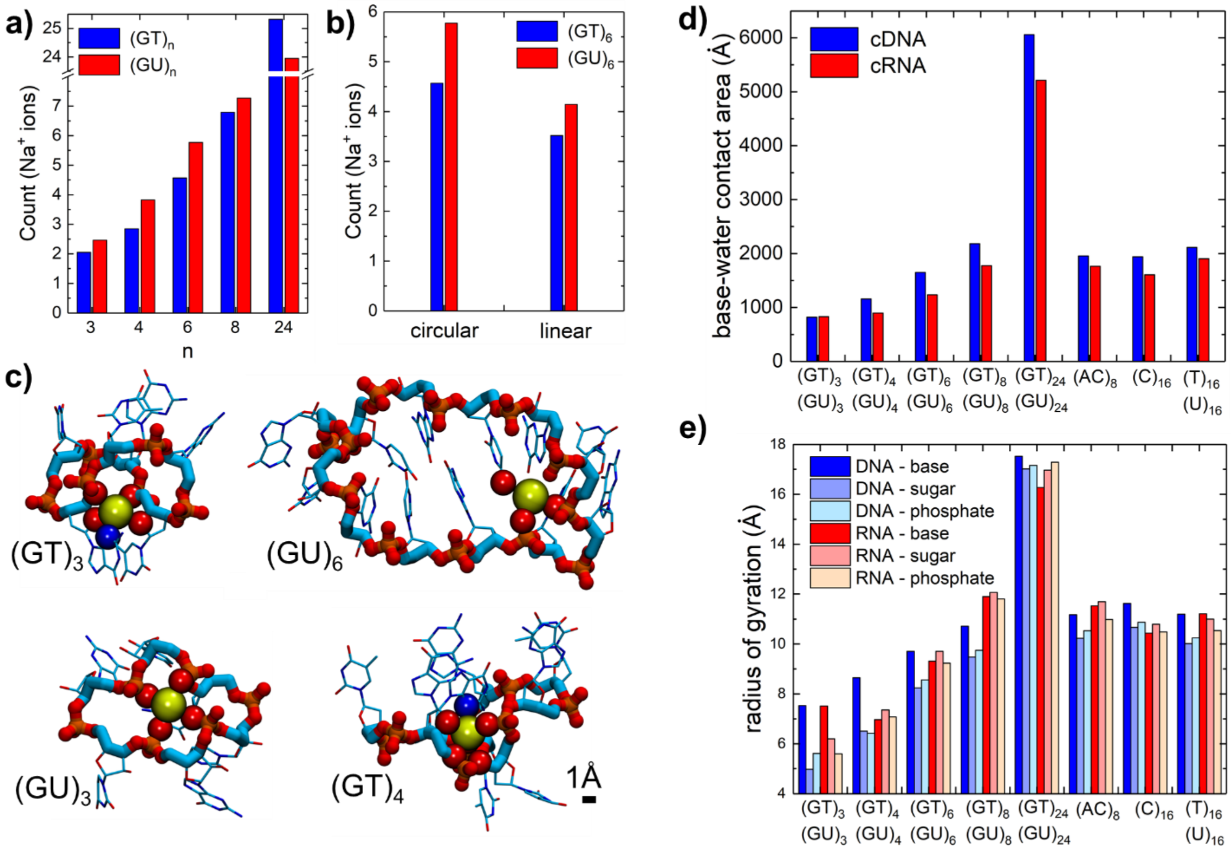


**Figure 5.** (a, b) Average number of sodium ions within 4 Å of nucleic acids examined, averaged over the last 500 ns of 1 μs MD trajectories. (c) Conformations of small circNAs with bridged/trapped sodium ions that have the longest residence times (Table 1). Polar atoms (P, O, and N) of nucleic acids within 3 Å of Na+ ions are highlighted as vdW spheres. (d) Contact areas between nucleotide bases and aqueous solvent for circNAs observed in MD simulations, averaged over the last 500 ns of 1 μs simulations. (e) Average radii of gyration of different parts of nucleotides (bases, sugars, and phosphate groups), averaged over the last 500 ns of 1 μs simulations.

circDNA bases are consistently (10–25%) more exposed to the aqueous solution than circRNA bases for molecules of different lengths and sequences. The exposure of bases to the solvent is affected by the folded arrangement of circNAs. Figure 5e shows the relative positioning of base, sugar, and phosphate groups in the circNAs examined in the form of the average radii of gyration of these groups, $R_{gyr}$, i.e., the average distances of geometric centers of these groups from geometric centers of circNA molecules. Consistently, in circDNAs, the bases have significantly larger $R_{gyr}$ than phosphate groups (by 1–2 Å). Therefore, in circDNAs, phosphate groups are likely to be positioned closer to COMs of the circular molecules, and the bases are more likely to be further away from COMs. In circRNAs, the situation is similar for the smallest circRNA, $(GU)_3$. However, for $(GU)_n$, with n = 4, 6, 8, and 24, the bases and phosphate groups either have similar $R_{gyr}$ or the bases have smaller $R_{gyr}$ than the phosphate groups. Therefore, bases are either closer or at similar distances to geometric centers of circRNA molecules, in comparison to the phosphate groups. The observed trends are also confirmed by the visual analysis of representative circDNA-$(GT)_6$ and circRNA-$(GU)_6$ molecules in Figure 8a, which clearly shows that the backbone of circDNA is accommodated within the folded structure, whereas the circRNA backbone is found more on the outside of the molecule.

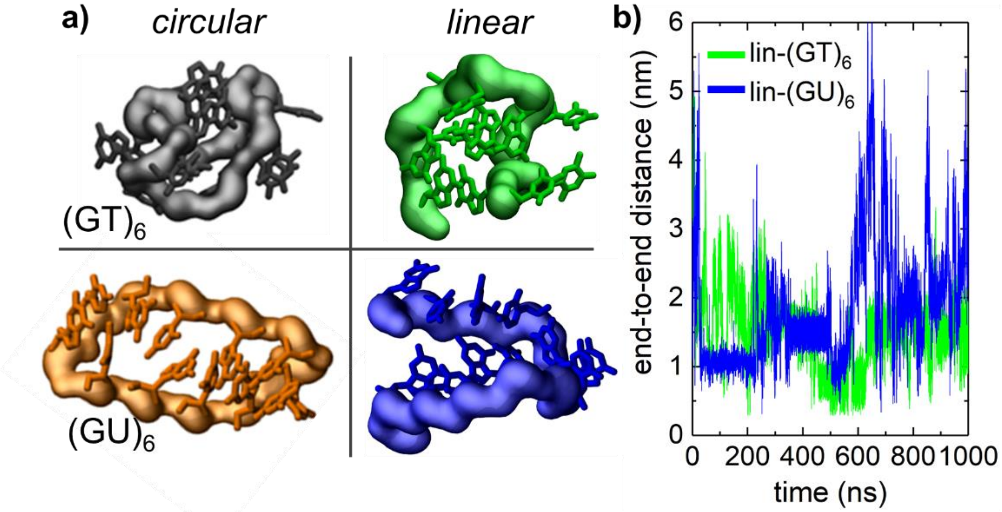


**Figure 6.** Comparison of circular and linear oligonucleotides. (a) Representative snapshots of circular and linear $(GT)_6$ and $(GU)_6$ DNA and RNA molecules. (b) End-to-end distance (O5T atom of the first nucleotide and O3′ atom of the last nucleotide) observed in 1 μs MD simulations of linear $(GT)_6$ DNA and $(GU)_6$ RNA molecules.

### 3.3.3 Comparison of Circular and Linear Oligonucleotides

The key difference between circular and linear molecules is the covalent bonding between 3′- and 5′-end nucleotides. In circular molecules, the 3′- and 5′-end nucleotides remain stably bound to each other. However, the distance between end nucleotides of linear molecules examined for $(GT)_6$ and $(GU)_6$, fluctuates between <0.5 and 6 nm (Figure 6b). Therefore, covalent bonding clearly presents a significant geometric constraint within flexible nucleic acid molecules without base pairing. Both linear and circular DNAs are found to be more bent and less rigid than the corresponding RNA molecules. However, circular RNA has a significantly more bent backbone than linear RNA, as its P–P–P angles assume on average 120° in comparison to 140° that the P–P–P angle of linear $(GU)_6$ assumes on average (Figure 3).

Negatively charged backbones of circNAs encircle the space within these molecules. Given the circular arrangement of the negative charge in circular molecules versus a more disordered arrangement of the negative charge in linear molecules, the number and coordination of counterions may be different for circular and linear molecules. In fact, Figure 5b shows that circular $(GT)_6$ and $(GU)_6$ molecules have on average a greater number (1–1.5) of sodium counterions within 4 Å of these molecules than their linear counterparts and are thus more effective at sequestering counterions.

**Table 1.** Three Longest Residence Times of $Na^+$ Ions Stably Binding in Pockets of Circular and Linear $(GT)_n$ and $(GU)_n$ Molecules, Reported in nanoseconds.

| Circular | | | | | | | | | | Linear | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $(GT)_3$ | $(GU)_3$ | $(GT)_4$ | $(GU)_4$ | $(GT)_6$ | $(GU)_6$ | $(GT)_8$ | $(GU)_8$ | $(GT)_{24}$ | $(GU)_{24}$ | $(GT)_6$ | $(GU)_6$ |
| 320 | 180 | 410 | 387 | 95 | 443 | 307 | 215 | 800 | 757 | 27 | 73 |
| 176 | 156 | 188 | 77 | 68 | 223 | 68 | 41 | 539 | 479 | 26 | 49 |
| 162 | 74 | 83 | 74 | 57 | 75 | 59 | 35 | 459 | 339 | 18 | 22 |

In both circular and linear molecules, counterions either can be hovering or can assume bridging/trapped conformations. Figure 5c shows that small circular nucleic acids can form stable pockets for counterions, which can remain for hundreds of nanoseconds, as shown in Table 1. The smallest circular nucleic acids seem to form especially interesting ion pockets, resembling crown ethers. While the trapped ions remain bound to circular $(GT)_6$ and $(GU)_6$ molecules for up to ≈95 and ≈443 ns, they remain bound to analogous linear molecules for significantly shorter times of up to ≈27 and ≈57 ns. We note that the measurements of single ion residence time are stochastic, where obtaining better statistics would require significantly better sampling of long binding events between ions and nucleic acids. While the geometry of the smallest circular nucleic acids can lead to ion sequestration via the existence of preformed pockets (Figure 5c), small linear single stranded molecules have also been observed to bend and fold around ion[69] [70], and ion binding is generally known to determine the tertiary structures and folding of ssRNAs.[71]

### 3.3.4 Circular Nucleic Acids at Millimolar Concentrations

Besides examining the interactions of circNA molecules with $Na^+$ counterions, we also explored the behavior of multiple circNA molecules in 0.15 M NaCl solution. Solutions containing three/four circDNA-$(GT)_6$ or three/four circRNA-$(GU)_6$ molecules were prepared at concentrations of 5.9 mM/7.9 mM and allowed to equilibrate for 1 μs. After 1 μs, three out of four circDNA-$(GT)_6$ molecules aggregated, while one molecule diffused in the solution. On the other hand, all four circRNA-$(GU)_6$ molecules aggregated. In the aggregated states, the molecules formed chains, as shown in Figure 7, rather than globular aggregates.

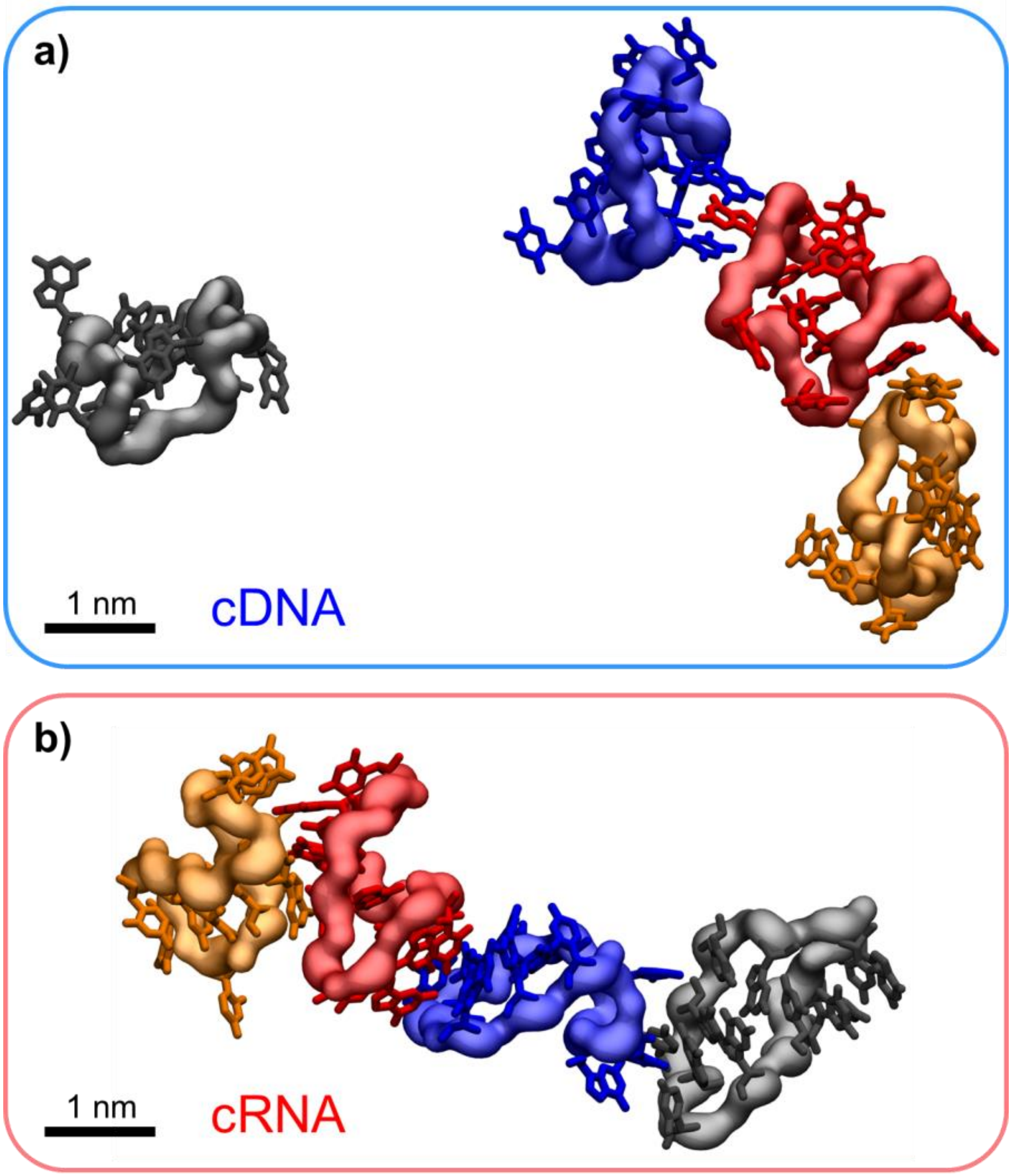


**Figure 7.** Circular nucleic acids at 7.9 mM. Four circDNA-(GT)$_6$ (a) and circRNA-(GU)$_6$ molecules (b) after 1 μs of equilibration.

Such chain formation, due to repulsion between negatively charged circDNA molecules, was observed for other negatively charged moieties, such as nanoparticles coated with negatively charged ligands[72]. Interestingly, the aggregation of circular molecules affects their phosphate angle distributions very likely because aggregated molecules adopt different conformations from single solvated molecules; the aggregated molecules have different solvation and can engage in intermolecular stacking. While circNAs are observed to aggregate at 7.9 mM concentrations, such aggregated states should be unfavorable or transient at concentrations at which small nucleic acid molecules, such as aptamers, are used in applications (nanomolar)[57]. The aggregated states are unlikely at low concentrations because translational entropy of separated molecules should overcome the favorable enthalpies of binding. In fact, even at 7.9 mM concentration, dissociation events are observed for circNAs.

## 3.4 CONCLUSION

In summary, we examined small covalently closed single-stranded circular DNAs and RNAs in microsecond long MD simulations. circDNAs were found to be more bent and more flexible than circRNAs. These molecules have distinctly different backbone and base arrangements in the observed conformations. Small circular nucleic acids are observed to be more effective in sequestering counterions than linear nucleic acids of same sequence and length, in conformations that resemble crown ethers for the smallest molecules examined. While the exposure of bases to aqueous solvent was more pronounced in circDNAs than in circRNAs, circRNAs were more prone to aggregation at millimolar concentrations. The dynamics, folding, three-dimensional structures, and finally biological/technological functions of circular nucleic acids are therefore expected to be dependent on molecular flexibility, ionic species, and ionic concentrations present, in analogy to linear single-stranded nucleic acids.[66] [73] [74] These circular molecules, expected to have slow degradation in cellular environments, are expected to be promising new constructs for therapeutic, diagnostic, and nanotechnology applications.

# Chapter 4: Understanding melting mechanisms of DNA on laser-excited gold nanoparticles

## 4.1 INTRODUCTION

Structure and function of biological materials is highly sensitive to local temperature. Modulation of the local temperature around biological materials can provide for precise control over both biological structure and activity.[75] One way is to utilize plasmonic nanoparticles (NPs) which have emerged as remote-controlled light-activated heat sources. Such NPs can be readily coupled to biological materials in order to control temperature.[76] Such photothermal heating systems can be designed to control the solution temperatures at nanometer length ranges allowing for applications such as photothermal therapy[77], molecular hyperthermia[78], and gene therapy[79]. Recently, it has been shown that femtosecond laser pulsed excitation of plasmonic NPs can generate local temperature increases that are capable of controlling the activity of biomolecules scaffolded on the NP surface.[80] [81] The increase in initial temperature around the NP can be tuned to the biologically relevant temperature range and the nanosecond duration of the temperature increase and steep temperature gradient extending from the NP surface create a highly dynamic temperature profile for biological materials.

Plasmonic NPs such as gold nanoparticles (AuNPs) can readily be conjugated with biological materials.[82] [83] AuNPs to act as an optically controlled heat source in biological applications such as photothermal therapy,[84] [85] membrane potential actuation,[86] [87] and molecular release/delivery.[88] [89] Previous theoretical work suggests design of photothermal systems using AuNPs that can be used to generate controlled temperature profiles relevant to biological activity.[90] Although continuous wave (CW) excitation leads to a steady-state temperature profile characterized by a 1/radius (r) temperature dependence, femtosecond (fs) laser pulse excitation can generate a temperature profile around a AuNP that follows a $1/r^3$ temperature dependence[91] leading to generation of temperature profiles that are temporally and spatially confined to the nanoscale.[90] [91] [92]

Ultrashort pulses of light can be used to release of nucleic acids from the surface of AuNPs which has shown to optically trigger nucleic acid release both in vitro and in vivo.[93] [94] Recently, Dr Igor Medintz and Dr Sebastian Diaz at Naval Research Laboratory (NRL), our collaborators on the present project, developed a quantitative local nanothermometer to understand the level of control afforded over denaturation of dsDNA displayed on gold nanoparticles (AuNP) under fs-laser pulsed excitation.[92] They used a modified dissociation equation to calculate a "sensed" temperature felted by dsDNA and compared to the theoretical average temperature.[95] Typically, the "sensed" temperature was found to be greater than theoretical average temperature, indicating that local peak or maximum temperatures play a significant role in the denaturing process.

In this work, we have utilized MD simulations to understand melting mechanisms of dsDNA after a single fs laser pulse excitation of dsDNA conjugated AuNP. We have simulated 19 and 13 base-pair long dsDNA attached to a 3T (Thymine) spacer ssDNA, which is in turn attached to an alkane chain linker. This alkane linker is covalently bonded to gold surface via gold-thiol bond. We carried out simulations at room temperatures, elevated temperatures and under a single laser pulse of energy fluences of 14.1 $J/m^2$ by mimicking heat profiles extracted from modified dissociation equation containing "sensed" DNA temperature. Our work shows that melting mechanism depends on length of dsDNA, orientation of relaxed dsDNA, intensity of the laser pulse, proximity to gold surface and bulk solution temperatures. In our simulations, we observe that dehybridization of dsDNA ($\geq$ 50% base-pair broken), which leads to subsequent release of single DNA strand in solution, is a rare event. We also observe that proximity to gold surface promotes melting of dsDNA. Thus, shorter dsDNA and increased proximity of relaxed dsDNA to the gold surface prior to laser pulse make them more susceptible for dehybridization.

## 4.2 METHODS

### 4.2.1 Building dsDNA conjugated to gold surface.

Structure of gold atoms arranged in face centered cube (fcc) lattice was directly obtained from Crystallography Open Database (COD)[96] and was cut to design gold layer consisting of three layers of atoms along (1,1,1) direction using MERCURY CSD 2.0 software.[97] The bottom gold layer was heated to mimic laser pulse. The alkane linker was build using AMPAC 8.0 software[98] and was covalently bonded to gold surface with gold-thiol bond. The other end of the linker was covalently bonded to ssDNA spacer (3T) which is linked to dsDNA. All initial DNA structures were build using 3DNA software.[62] Two DNA sequences one longer 19 base-pair (3'-TCTCTACCTACTCACCTCA-5') and one shorter 13 base-pairs (3'-TCTCTACCTACTC-5') were tested in this work. An identical layer of gold surface was placed on top the simulation box after minimization such that there is a vacuum gap between gold layer of adjacent periodic cells along z-axis. The system was solvated with TIP3P water with desired ion concentrations between these two gold layers. The top gold layer ensures that during simulations, the water in the system remains in liquid phase (no vaporization) and no heat is propagated to periodic cells along z-axis due the presence of vacuum layer.

### 4.2.2 Mimicking Single Laser Pulse to heat Gold Surface

The 13-bp systems in aqueous solution in between two gold layers and equilibrated at 22°C as described above were used as the starting points for non-equilibrium MD simulations. The non-equilibrium simulations examine the effects of the nonhomogeneous heat pulses in the system. To mimic the localized heating induced in the gold nanoparticle by short laser pulses with 14.1 $J/m^2$ and 20 $J/m^2$ energy fluences, we applied oscillatory force, $F_z^h$, on all atoms of the bottom gold layer under the DNA construct along the z-axis:

$$F_z^h = Asin\omega t$$

where $A$ and $\omega$ are constant parameters that are optimized to mimic the local temperature profile obtained from the modified dissociation equation. Force $F_z^h$ was applied for 20 ps, leading to the local heating of the gold atoms, with $\gamma_{\text{Lang}} = 1.0$ ps$^{-1}$. The cooling temperature profiles in MD simulations are then generated by applying oscillatory decaying force $F_z^c$ along z-axis to all the gold atoms of the bottom gold layer surface for the next 2 ns:

$$F_z^c = A_i e^{-MDt} sin\omega t$$

where decay factor D depends on boundary conditions at t = 0 and t = 2 ns such that

$$D = \frac{1}{MT} ln \frac{A_i}{A_f}$$

where $A_i$ is the amplitude at initial time (t = 0) and $A_f$ is the amplitude at final time (t = 2 ns). The value of $A_i$ and $A_f$ are chosen such that the cooling profile of water mimic the profile obtained from the modified dissociation equation. $M$ is a constant number which modulates the decay function and $T$ is the total number of timesteps in a 2 ns long simulation. All three parameters $A_i$, $A_f$ and $M$ are constant numbers that are optimized to mimic the cooling profiles. The constant $\omega$ has the same constant value as in during heating. The parameters to mimic in MD simulations the effects of two laser pulses with 14.1 J/m$^2$ and 20 J/m$^2$ energy fluences are provided in Table 2. Force $F_z^c$ was applied for 2 ns, with $\gamma_{\text{Lang}} = 0.01$ ps$^{-1}$. After 2 ns, application of force to the gold layer was removed, and the systems were allowed to relax for 100 ns, using $\gamma_{\text{Lang}} = 0.0005$ ps$^{-1}$. In all cases and at all steps, atoms of the gold layer were restrained using harmonic potential with a spring constant of 1 kcal/(mol Å$^2$). cooling.

During the non-equilibrium simulations, the presence of the top gold layer ensures that the water in the system remains in the liquid phase (no vaporization) and that no heat is propagated along the z-axis between the periodic cells due to the presence of the vacuum layer.

**Table 2.** Optimized force parameter for heating and cooling under laser pulse

| Pulse intensity | Heating | | Cooling | | |
|---|---|---|---|---|---|
| | $A$ (kcal/mol Å$^2$) | $\omega$ (fs$^{-1}$) | $A_i$ (kcal/mol Å$^2$) | $A_f$(kcal/mol Å$^2$) | $M$ |
| 14.1 J/m$^2$ | 31.0 | 0.0225 | 8.0 | 2.1 | 1.1 |
| 20 J/m$^2$ | 40.0 | 0.0225 | 9.0 | 3.0 | 1.0 |

### 4.2.3 Molecular Dynamics simulations

The prepared systems were solvated with TIP3P water[34] and neutralized with 0.25 M NaCl and 0.05 M $MgCl_2$ to mimic ionic concentration of HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) buffer and $Mg^{2+}$ used in experimental setup with solvate and ionize VMD plugins, respectively.[65] All systems were described with CHARMM36 force field parameters.[45] MD simulations were performed with the NAMD2.13 package.[27] Initial systems were built without placing the top gold layer. These systems were minimized for 5000 steps and after that the solvent were equilibrated around solute progressively, first, for 0.005 ns with timestep set to 1.0 fs and cutoff electrostatics, followed by 0.01 ns long relaxation with full electrostatics with timestep set to 1.0 fs. The systems are further relaxed for 0.075 ns with timestep set to 1.5 fs and then finally for 2 ns with timestep of 2.0 fs.

After equilibrations, the top gold layer was placed, and after that all systems where further minimized for 5000 steps and 2 ns long equilibration of solvent around the solute. For all the systems. All the systems were then equilibrated for 100 ns at room temperature (295 K) to allow DNA relaxation with Langevin dynamics (Langevin constant $\gamma_{\text{Lang}}$ = 1.0 ps$^{-1}$) in the NPT ensemble, where temperature and pressure remained constant at 310 K and 1 bar, respectively. The particle-mesh Ewald (PME) method was used to calculate the Coulomb interaction energies, with periodic boundary conditions applied in all directions.[33]

The relaxed DNA was utilized for production MD runs which were 100 ns long including triplicates of MD runs at elevated temperatures (360 K), with and without gold layer. The relaxed DNA structures were utilized to simulate laser pulse involving 20 ps heating with force $F_z^h$ applied to bottom gold layer at $\gamma_{\mathrm{Lang}} = 1.0$ ps$^{-1}$ followed by cooling for 2 ns cooling with force $F_z^c$ applied to bottom gold layer at $\gamma_{\mathrm{Lang}} = 0.01$ ps$^{-1}$ and finally relaxed for 100 ns at $\gamma_{\mathrm{Lang}} = 0.0005$ ps$^{-1}$. In all cases and at every step, the gold layer was restrained using harmonic potential with a spring constant of 1 kcal/(mol Å$^2$).

## 4.3 RESULTS

### 4.3.1 Effect of presence of gold layer

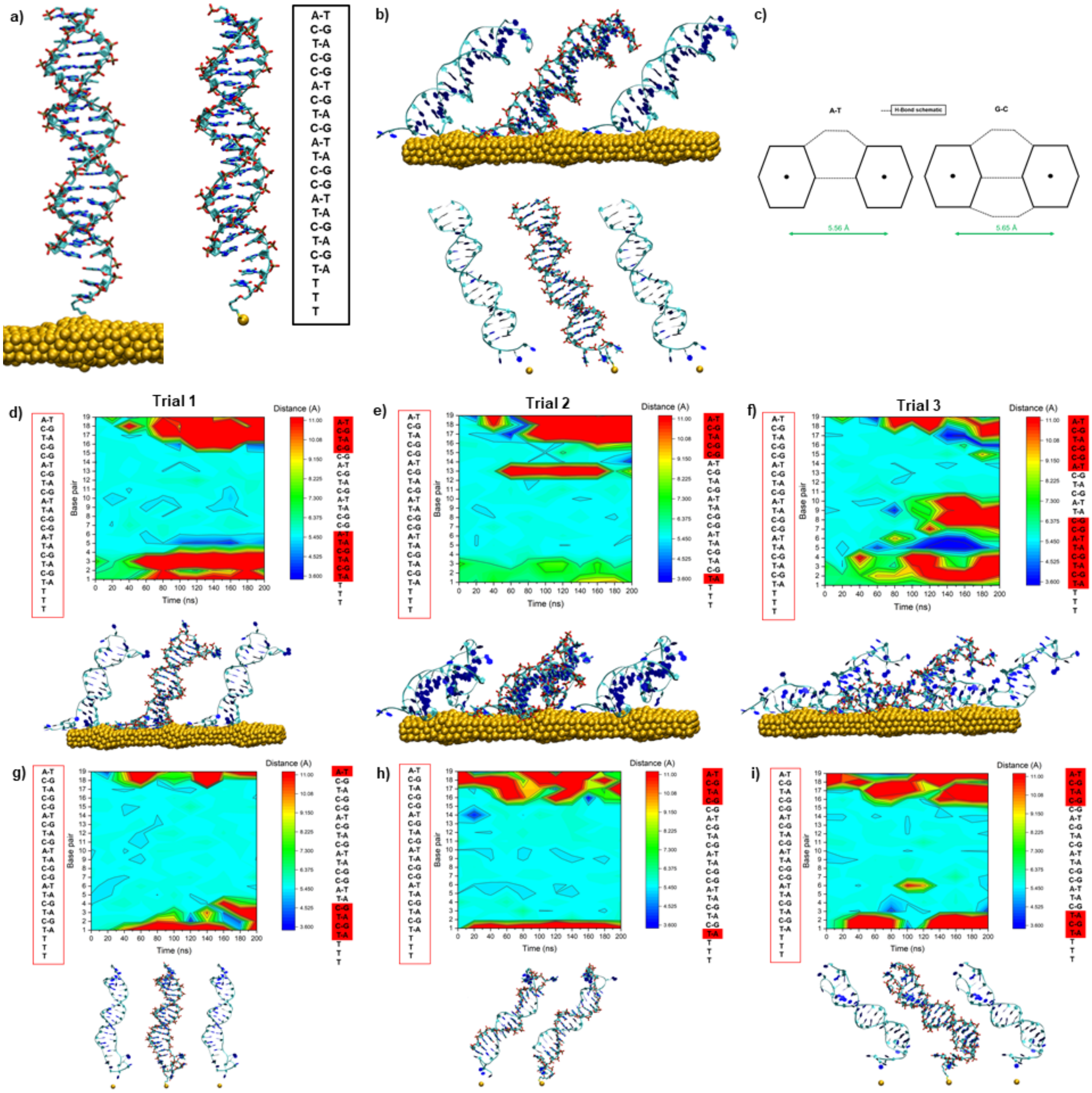


**Figure 8.** Dynamics of 19 nucleotide long dsDNA - (a) Initial structure of dsDNA with and without gold layer. Gold atoms are denoted by yellow beads. (b) dsDNA after 100 ns long simulation at 295 K. (c) Schematic diagram six membered ring of nitrogenous base which are involved in formation of base pair with bond lengths. (d-f, top) Heat maps of distance (D) between the base-pairs for three trials of 100 ns long simulations at 360 K with gold layer. Cyan colored regions indicate intact base pairs. Base-pairs broken at 100 ns are highlighted in red on right hand side of respective heat map, (d-f, bottom) Snapshots of DNA on gold surface at 100 ns. (g-i) Same as panel (d-f) except without gold layer.

DNA interacts strongly with and facilitates melting of dsDNA from the bottom. After 100 ns of relaxation at 295 K (room temperature), the 19 nucleotide long DNA prefers to tilt and interact with gold surface, which leads to some helix distortion and breakage of bottom most base pair (T-A) (Figure 8b, top), whereas in case of the system without gold layer, DNA retains its initial structure without any significant distortion of the DNA helix as shown in Figure 8b, bottom. No base-pairs are broken in absence of gold layer at 295 K. Interaction with gold layer is very stable over the course of trajectory as DNA does not prefers to move once its relaxed on the gold surface.

To measure time evolution of base-pair breaking, we analyzed heat map of distance (D), measured at fixed time interval over the course of the trajectory as shown in Figure 8(d-i). D is defined as distance between the geometric center of 6 membered ring of nitrogenous base which are involved in hydrogen bonds leading to formation of base pair (Figure 8c). Based on previous work by Guerra et al, the estimated value of D is 5.56 Å for A-T base pair and 5.65 Å G-C base pair.[99] Thus, regions in cyan color in the heatmap refer to the intact base-pairs of the duplex DNA (Figure 8(d-i)).

At 360 K and in presence of gold layer we see significant melting from both top and bottom as shown in three trials in figure 8(d-f). Trial 2 (figure 8e) show minimal melting from the bottom unlike trial 3 (figure 8f) which captures the rare event of DNA dehybridization (≥50 % of base-pairs are melted). This indicates that melting of dsDNA depends on its orientation with respect to the gold surface. In all three trials, we observe that DNA can “lay down” on gold surface which facilitate the melting process by disrupting the helical structure of dsDNA. Absence of gold layer, as shown in three trials in figure 8(g-i), show significantly less melting on average as compared to with gold layer trials. As no gold layer is present, the helical structure remains intact except for broken base-pair regions. We do not observe any DNA dehybridization in 100 ns long simulation in absence of gold layer.

### 4.3.2 Melting mechanism due single laser pulse.

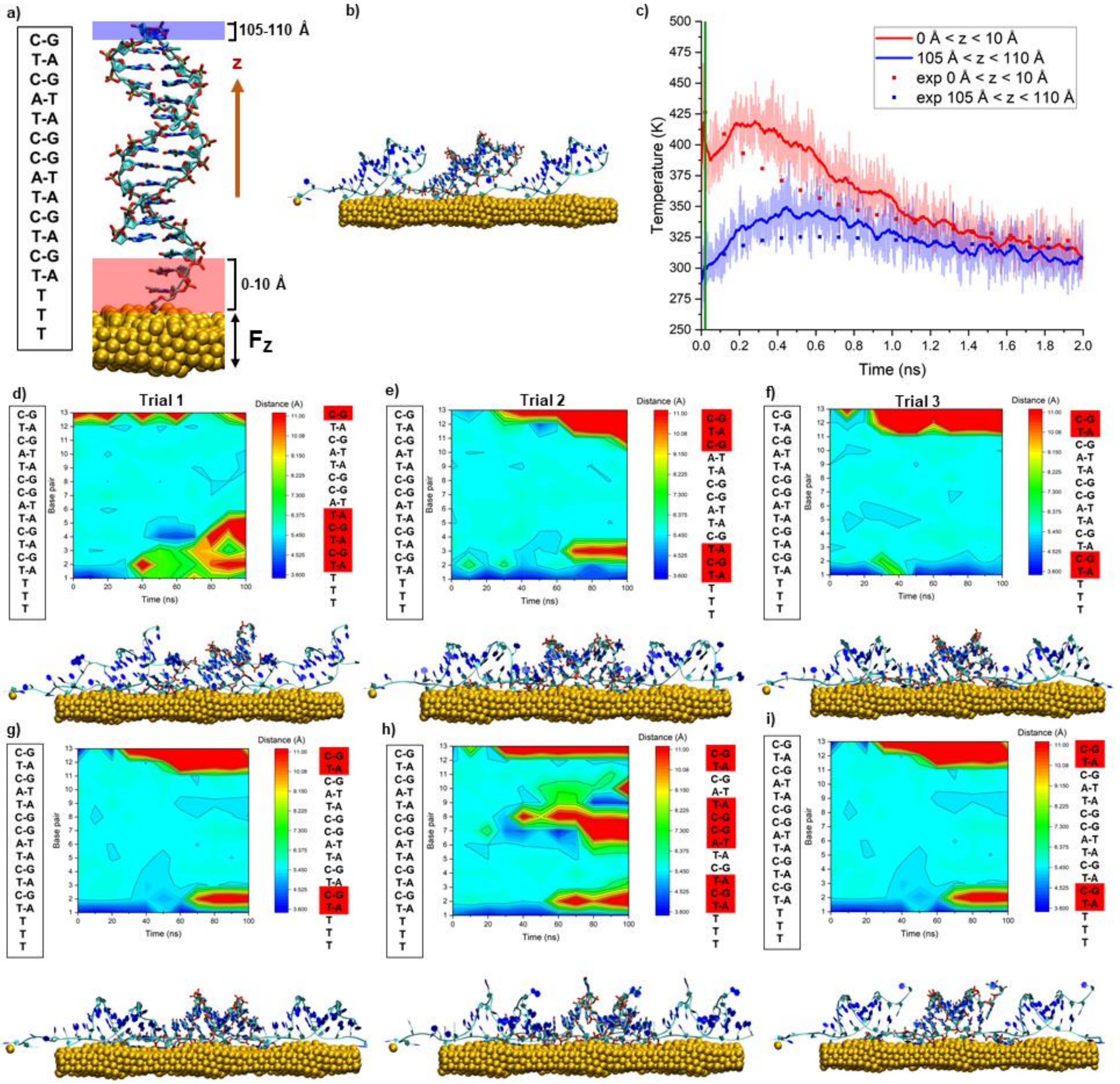


**Figure 9.** Dynamics of 13 nucleotide long dsDNA under a laser pulse - (a) Initial structure of dsDNA. Gold atoms are denoted by yellow beads. (b) dsDNA after 100 ns long simulation at 295 K. (c) Average solvent temperature in MD simulations (lines) compared with theoretical temperature profile (dots). z represents distance from gold layer. Vertical green line denotes the time when heating was stopped. (d-f, top) Heat maps of distance (D) between the base-pairs for three trials of 102 ns long simulations after heating by a laser pulse of intensity 14 J/m$^2$ . Cyan colored regions indicate intact base pairs. Base-pairs broken at 100 ns are highlighted in red on right hand side of respective heat map, (d-f, bottom) Snapshots of DNA on gold surface at 100 ns. (g-i) Same as panel (d-f) except simulating at 322 K instead of heating the gold layer with a laser pulse.

We utilized shorter 13 nucleotide long base pair to study melting mechanism under a single pulse of laser of intensity 14J/m$^2$. Figure 9b shows at after relaxation of 100 ns at 295 K, the DNA “lays down” on the gold surface with bottom most base pairs broken at 100 ns. The shorter DNA interacts significantly more with the gold layer allowing almost entire duplex to lay horizontal on the gold surface. This orientation of relaxed DNA is also observed albeit to less degree in longer DNA (figure 8b, top). Such orientation makes DNA more susceptible to melting under laser pulse as it increases the proximity of base-pairs to the gold layer allowing access to solvent regions that will have higher local temperatures during heating and subsequent cooling under laser pulse.

As shown in figure 9c, we measured average temperature of solvent close to gold surface ($0 < z < 10$ Å) and far away from gold surface ($105 < z < 110$ Å) in MD simulations. These measurements were consistent with theoretical local temperature profiles showing that water close to the gold surface is quickly rises to 425 K due to laser pulse and falls rapidly due to heat dissipation in bulk solution. Solvent far away from the solution do not see such step rise in temperature which peaks at around 350 K as compared with water close to the gold surface (figure 9c). The difference in temperature profiles of solvents in both regions confirm the highly localized nature heating by fs pulsed laser. These temperature profiles indicate that dsDNA is exposed to very high temperatures for short period of time which can induce melting and eventual dehybridization of dsDNA and proximity to gold surface during laser pulse should facilitate DNA melting.

Figure 9(d-f) shows results for three trials of 100 ns relaxation after laser pulse of 14 J/m$^2$ intensity. In all three trials, we observe that base-pair breaking is not a spontaneous event in response to the heating due to laser pulse. In trial 1, we observe minimal melting from the top as compared with other two trials due to fact that the upper portion of the dsDNA is farther away from gold surface than in other two trials indicating the positive role of gold layer on DNA melting (figure 9d). Dehybridization of DNA, which is a rare event, is only observed in trial 1. In such events, DNA start smelting significantly from one end at around 50 ns and progressively increases to cause dehybridization of duplex DNA (figure 9d). Additional simulations mimicking the localized heating of the gold nanoparticle by a laser pulse with 20 $J/m^2$ energy fluency demonstrate immediate effects on dsDNA duplex stability, where breaking of base pair nucleotides at the duplex end closer to the Au-S bond is observed within the first few nanoseconds after the pulsed heating.

Our collaborators at NRL lab, based on their theoretical model, estimated that the melting effect of dsDNA under single laser pulse of 14 $J/m^2$ will be similar if, instead of laser pulse, the average temperature of the bulk solution is kept at 322 K. Figure 9(g-i) shows melting of DNA in three trials of 100 ns long simulations at 322 K. The melting of dsDNA at 322 K is very similar to melting under laser pulse of intensity of 14 $J/m^2$ validating the theoretical model. Such models can utilize to calculate local "average sensed" temperature parameter which can provide better understanding of actual biological responses in such systems and allow researchers to design precision nanoscale photothermal heating sources.

## 4.4 CONCLUSION

MD simulations can be utilized to simulate melting mechanisms of dsDNA bound to AuNPs due to femtosecond pulsed laser, with excellent agreement with both theory and experiments. Our innovative approach can be applied to similar systems involving heating by pulsed lasers. Our work demonstrates that dsDNA melting under fs pulsed laser is an uncommon event which does not happen spontaneously after heating by laser pulse. The melting mechanism does not have any preferred ends of the dsDNA but proximity to gold surface seems to facilitate DNA melting. MD simulations validated the local "averaged sensed" temperature felt by the DNA calculated using theoretical model involving dissociation equation for DNA. Accurate knowledge of such parameters can be used to design better heat sources for such plasmonic NP systems.

# Chapter 5: Realistic modeling of ssDNA corona on Carbon Nanotubes (CNTs)

## 5.1 INTRODUCTION

Functional nanomaterials are often hybrid materials consisting of a solid core material and an organic interface that guarantees colloidal stability and functionality. Therefore, the properties of such hybrid materials are more complex than those of bulk materials. Single-walled carbon nanotubes (SWCNTs) are a highly interesting building block for functional nanomaterials such as biosensors, drug delivery agents, and optoelectronic devices. [100] [101] Semiconducting SWCNTs fluoresce in the near-infrared (nIR), which is a useful property for optical applications such as imaging, biosensing, or optoelectronic applications.[102] [103] However, SWCNTs are hydrophobic, and therefore, different concepts have been developed to ensure water solubility and colloidal stability.[104] [105] One of the most common noncovalent functionalization approaches uses single stranded DNA (ssDNA). ssDNA functionalization has been extremely successful in separation of different chiralities of SWCNTs.[106] [107] Therefore, the interface between ssDNA and SWCNTs is of special interest. One of the basic metrics of ssDNA/SWCNT conjugates is the number of ssDNA molecules per SWCNT, or in other words, the SWCNT segment length one ssDNA molecule occupies. Reported numbers vary, for example, for $(GT)_{16}$, from ∼6 nm (adsorbed DNA train length) to 16−20 nm.[108] [109] Additionally, the length of one ssDNA pitch at the SWCNT surface has been reported to differ from ∼2 to ∼14 nm up to 20 nm.[110] [111]

SWCNTs are versatile building blocks for optical sensors.[101] Molecular recognition is essential for sensing, and therefore, various (noncovalent) functionalization concepts have been used with different macromolecules. These conjugates have shown to be very successful for sensors including neurotransmitters such as dopamine, proteins, and small molecules.[112] [113] Again, ssDNA functionalization has served as a key concept. The recognition and photophysical mechanism has been attributed on the one side to conformation changes that varies exciton decay pathways.[114] On the other side, the influence of redox chemistry has been discussed.[115] [116] Interestingly, it was shown that for sensors of the important neurotransmitter dopamine, sequence and length affect sensitivity and also selectivity.[117] In this context, the ssDNA wrapping and the design of an organic corona have been discussed as a guide to create general recognition motifs. The exact coverage, conformation, and dynamics of ssDNA on SWCNTs should play an important role, especially in light of recent simulations that show the importance of kinetics (rate constants) for sensing and imaging with such sensors.[118]

Even though ssDNA functionalization of SWCNTs is widely used, many fundamental aspects of the ssDNA phase around SWCNTs are still unclear. Most importantly, quantitative data regarding the adsorption of ssDNA are not fully available. Therefore, the simple question how many ssDNA molecules are bound to a single SWCNT has not been fully addressed.

Here, we determine conformations of ssDNA molecules present on CNT surface in experimentally determined number densities. The number densities were determined by our collaborator, S. Krauss, as reported in Ref 101.

## 5.2 METHODS

### 5.2.1 Molecular Dynamics simulations

We used classical atomistic MD simulations to examine the structure of $(GT)_{15}$ ssDNA/(6,5)-SWCNT conjugates. We determined the likely structures of these conjugates and calculated the contact areas between ssDNA and SWCNTs. The simulated systems contained (6,5)-SWCNTs wrapped with $(GT)_{15}$. The initial configurations of all simulations were generated by using Material Studio (Accelrys Software Inc. Materials Studio, Release 4.5; Accelrys Software Inc.: San Diego, CA, 2007) and VMD software.[65] First, a helical structure of $(GT)_{15}$ ssDNA was constructed by using Material Studio. Then, a (6,5)-SWCNT was wrapped by the helical ssDNA by using VMD plugins. The systems were solvated and ionized by solvate and ionize VMD plugins, respectively;[65] the salt concentration of 150 mM NaCl was selected to mimic physiological conditions. The final systems contained approximately 29,725 atoms.

The CHARMM36 force-field parameters[119] and TIP3P water model were used to define interactions for all the simulated molecules.[6] MD simulations were performed with NAMD2.12.[25] Long-range electrostatics were evaluated by the particle-mesh Ewald method.[33] The evaluation of van der Waals and long-range Coulomb interactions was performed every 1 and 2 time steps, respectively; the time step was set to 2 fs. All the simulations were conducted in the NpT ensemble and with periodic boundary conditions applied. Temperature and pressure remained constant at 310 K and 1 bar, respectively; the Langevin constant was set to $\gamma$Lang = 1.0 ps$^{-1}$. The first prepared system, containing the (6,5)-SWCNT wrapped by a helically arranged $(GT)_{15}$, was initially minimized for 10,000 steps. After minimization and warming the system, water and ions were equilibrated for 2 ns around the hybrid systems, which were restrained using harmonic forces with a spring constant of 1 kcal/(mol Å$^2$). Then, the system was equilibrated for 100 ns, while only SWCNT atoms were held restrained. The second system was simulated to examine the structure of a $(GT)_{15}$ molecule wrapping (6,5)-SWCNTs, while occupying an SWCNT fragment 2.24 nm in length, as determined by experiments. In the second set of simulations, the above system was simulated for additional 327 ns. In these simulations, all the heavy atoms of (GT)15 were restrained to occupy a 2.24 nm long fragment of SWCNTs, by means of a square potential

$$V(z) = 0; z > \ -z_0 \cap z < z_0$$

$$V(z) = V_0; z < -z_0 \cap z > z_0$$

where $V_0$ was gradually increased from 0.01 kcal/mol per atom to 0.1 kcal/mol per atom. Overall, the second system was progressively simulated for 100 ns ($V_0$ = 0.01 kcal/mol, $z_0$ = 1.1 nm), 100 ns ($V_0$ = 0.02 kcal/mol, $z_0$ = 1.1 nm), 100 ns ($V_0$ = 0.03 kcal/mol, $z_0$ = 1.1 nm), 100 ns ($V_0$ = 0.04 kcal/mol, $z_0$ = 1.1 nm), and 200 ns ($V_0$ = 0.1 kcal/mol, $z_0$ = 1.12 nm).

### 5.2.2 Data analysis

To analyze the stacking of $(GT)_{15}$ on SWCNT surface in two simulations above, the contact area between ssDNA and SWCNTs was calculated in time t, defined as

$$\frac{\mathrm{a}_{DNA}(\mathrm{t}) + \mathrm{a}_{SWCNT}(\mathrm{t}) - \mathrm{a}_{DNA \cup SWCNT}(t)}{2}$$

where $\mathrm{a}_{DNA}(\mathrm{t})$, $\mathrm{a}_{SWCNT}(\mathrm{t})$, and $\mathrm{a}_{DNA \cup SWCNT}(t)$ are the solvent-accessible surface areas of the ssDNA, SWCNTs, and the ssDNA/SWCNT conjugate, respectively, at time t. The analyses were performed during the last 50 ns of production runs. The contact area evaluation was performed by the SASA built-in VMD plugin,[25] where the van der Waals radius of 1.4 Å was assigned to atoms to identify the points on a sphere that are accessible to the solvent.

## 5.3 RESULTS

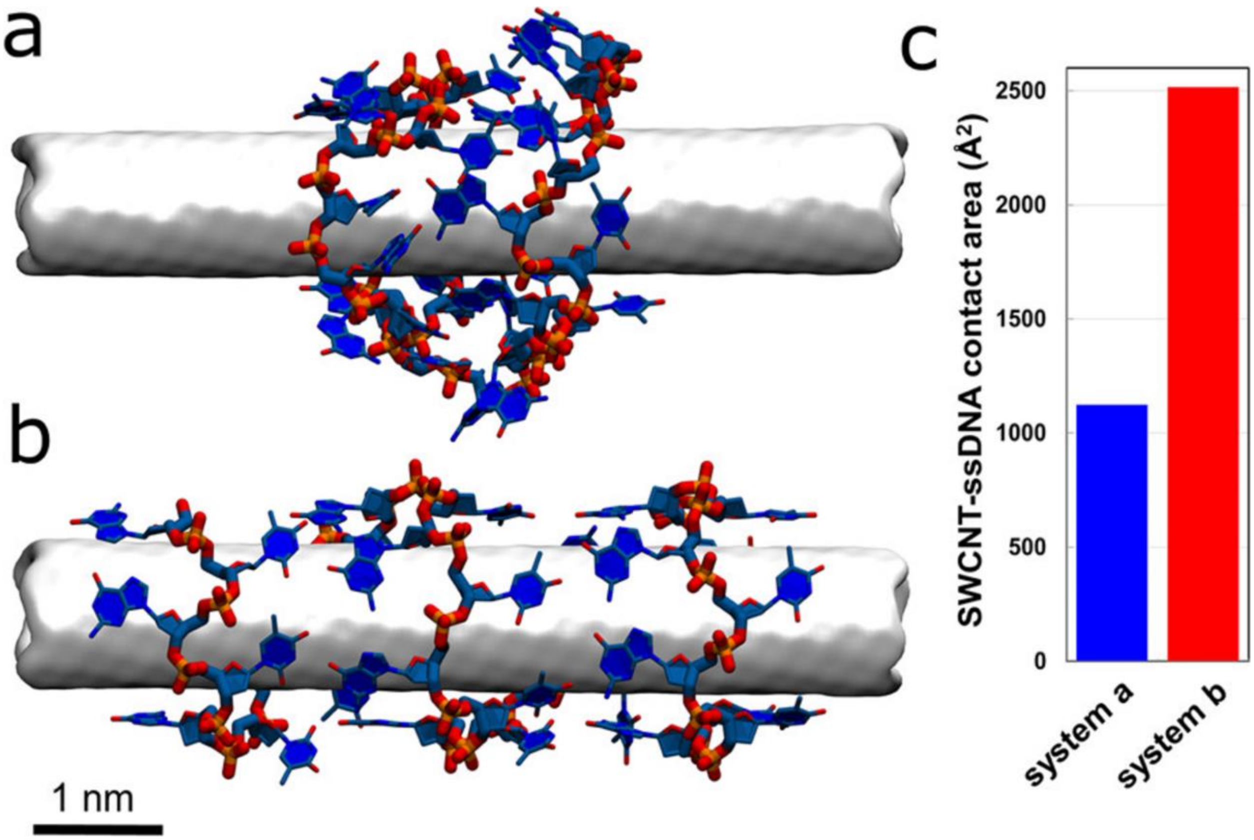


**Figure 10.** Molecular dynamics simulations of $(GT)_{15}$ ssDNA on a (6,5)-SWCNT with experimentally determined adsorption numbers as parameters. (a) Representative view of a $(GT)_{15}$–ssDNA molecule adsorbed on a (6,5)-SWCNT. ssDNA is restrained to stack on an SWCNT segment of the length determined in experiments. When $(GT)_{15}$ occupies the SWCNT segment as determined in experiments, only a fraction of bases stack on the SWCNT, while the rest of the bases stack on top of each other (less than half). (b) Representative view of a $(GT)_{15}$ molecule that fully adsorbs on a (6,5)-SWCNT. (c) Contact area between $(GT)_{15}$ and (6,5)-SWCNTs for systems shown in panels (a,b). In panels (a,b), the SWCNT is shown as a white surface, and ssDNA atoms are shown in subdued blue (C), red (O), orange (P), and blue (N) colors. Hydrogen atoms are not shown for clarity.

The experimentalist determined that on average one $(GT)_{15}$ molecule occupies 2.2 nm long segment of (6,5) SWCNT. To determine the most likely conformations of ssDNA molecules on SWCNTs at such number densities, we performed above-described atomistic MD simulations of $(GT)_{15}$ molecules adsorbing on (6,5)-SWCNTs. $(GT)_{15}$ molecules were adsorbed and simulated on the SWCNT surface, either with restraints imposed, so that $(GT)_{15}$ molecule spans the experimentally determined segment length of SWCNTs, or without any restraints present. In both simulations, ssDNA molecules initially wrapped SWCNTs in helical conformations.[120] [121] In a 100 ns simulation of the $(GT)_{15}$/SWCNT conjugate without any restraints present, ssDNA remained helically wrapped and the majority of the bases stacked on the SWCNT, as shown in Figure 7b. However, in a ~500 ns simulation of the $(GT)_{15}$/SWCNT conjugate, where $(GT)_{15}$ was restrained to occupy a 2.24 nm long segment of SWCNTs, only a fraction of bases can stack on the SWCNT, while the rest of the bases stack on top of each other (Figure 7a). The analysis of the $(GT)_{15}$/SWCNT contact area, shown in Figure 7c, indicates that less than half of ssDNA bases stack directly on the SWCNT surface in the system as shown in Figure 7a. Furthermore, the thickness of the ssDNA corona around the SWCNT increased indicating that CNT surface is less accessible at experimentally determined number densities than the $(GT)_{15}$ bases stacked directly on CNT. This shows that the experimentally determined number of bound ssDNA molecules is important to understand the organic phase around SWCNTs and that this parameter affects MD simulations significantly.

## 5.4 CONCLUSION

Our work shows that the mean available space for ssDNA on SWCNT should be considered as an important factor of the molecular understanding of the events on the SWCNT surface. Our results suggest that interactions between the ssDNA molecule itself play a more important role than previously thought. In the future, MD simulations could use our experimentally determined numbers as starting parameters. We have quantified and reported the numbers for ssDNA bases forming a corona on SWCNT. Such parameters will be helpful to assess further surface modification and can serve as parameters for MD simulations.

# Chapter 6: Computational modeling of broad-spectrum non-toxic antiviral nanoparticles with a virucidal inhibition mechanism

## 6.1 INTRODUCTION

Among all the various infectious disease threats humans face from microorganisms, viral infections are arguably the biggest pandemic threat in the modern. Viruses have high replication rates, high rate of mutations and transmissibility.[122] Even in absence of a pandemic, viral infections kill millions of people every year. Available antiviral drugs are highly targeted against a specific virus or, in some cases, members of a viral family. Current antiviral drugs include small molecules, such as nucleoside analogues and peptidomimetics, monoclonal antibodies, proteins that are able to stimulate the immune response, such as interferon, and oligonucleotides.[123] These drugs act intracellularly to allow for selectivity for viral enzymes but as viruses mostly depend for their replication on infected host cells, the specificity of antiviral drugs for viral proteins is not ideal, often causing general intrinsic toxicity upon administration.[124] [125] Furthermore, due to high mutation rates, most viruses develop drug resistance.[126] Due to the use of virus-specific proteins as targets of antiviral drugs, it is difficult to develop broad-spectrum therapeutics capable of acting against a large number of viruses that are phylogenetically unrelated and structurally different. Thus, there is a critical need to develop a broad-spectrum and non-toxic antiviral therapeutics.

One approach is to utilize substances known to interact with viruses outside of host cells, where they interfere with the first phases of the viral replication cycle but many potential substances, which can be broad-spectrum and non-toxic, exhibit only virustatic properties.[127] Virustatic molecules, such as heparin and polyanions, can interact with a broad spectrum of viruses because they target virus-cell interactions that are common to many viruses.[128] [129] [130] [131] [132] Activity of virustatic materials depends on a reversible binding event, which makes them therapeutically ineffective: upon dilution, these materials detach from intact viral particles allowing the viruses to infect again.

Virucidal materials are more promising for use as drugs against viral infections. Virucidal molecules cause irreversible viral deactivation, since their effect is retained even if their dilution occurs after the initial interaction with the virus.[133] Virucidal materials include simple detergents, strong acids, polymers[134], and nanoparticles[135] [136] [137] [138]. All these materials have intrinsic cellular toxicity since they attempt to chemically damage the virus and they cannot selectively damage the virus without affecting the host in which the virus replicates.

One way to achieve a broad-spectrum efficacy can be development of novel therapeutic materials that can target specific virus-cell interactions that are common to many viruses. One of these interactions is between a viral attachment ligand and an associated cell receptor called heparan sulfate proteoglycans (HSPG) which are responsible for the initial steps of the virus replication cycle.[127] Many viruses, including human papillomavirus type 16 (HPV-16), human immunodeficiency virus 1 (HIV-1), herpex simplex virus 1 and 2 (HSV-1 and HSV-2), attach to HSPGs, which are expressed on the surfaces of almost all eukaryotic cell types. It has been shown that nanoparticles (NPs) with long and flexible ligands mimicking HSPG effectively associate with such viruses and eventually lead to irreversible viral deformation. These non-toxic nanoparticles (NPs) show in vitro irreversible virucidal activity against HSV-2, HPV-16, RSV, Dengue and lentivirus. They are active ex vivo in human cervicovaginal histocultures infected by HSV-2 and in vivo in mice infected with RSV as well.[127]

Here, we studied molecular interactions between gold nanoparticle with core diameter of 2.4 nm decorated with ligands mimicking HSPGs with capsid segments of two, three and four HPV16 major late (L1) pentamer proteins using nanosecond scale atomistic molecular dynamics (MD) simulations. Although over the simulation timescales, internal pentamer stability is not compromised in the presence of NP, we identify the interface of two pentamers as preferred site of interaction with NPs. Interaction at this site is facilitated by two solvent exposed loop regions - one consisting of mostly polar or charged amino acids allowing long rang electrostatic interaction with sulphate groups present on NPs and other consisting of mostly non-polar amino acids interaction with alkyl groups present on NP surface. Such interactions cause observable perturbation in capsid segments and can lead to irreversible viral deformation.

## 6.2 METHODS

### 6.2.1 Constructing HPV capsid segments

Three capsid segments were constructed, having either two, three or four pentamers. The arrangement of pentamers in these capsid segments was based on the Human Papillomavirus Type 16 capsid structure resolved by cryo-electron microscopy (PDB:3J6R)[139]. Disulfide bonds were built between residue 161 and residue 324 for all segments, based on visual inspection.

For each pentamer in all the three systems, crystal structure of HPV16 L1 pentamer (PDB:5W1O)[140] was fitted in place of the relaxed pentamers using molecular dynamics flexible fitting (MDFF) method.[64] In MDFF simulations, L1 pentamer was coupled to the density obtained from relaxed pentamers. For all the three systems, L1 pentamer was first docked with the relaxed pentamer potential using colores (Situs) software with a resolution of 5 Å [63]. Next, MDFF simulations were performed in vacuum with scaling factor for potential ($g_{scale}$ ($\xi$)) set to 0.3. The secondary structure along with cispeptide bonds and chirality of L1 pentamer were restrained during the simulations.

The crystal structure of L1 pentamer was missing one loop region as well as some residues at both C and N terminal of all segments of the pentamer. The missing residues were built in VMD and covalently added at ends manually to the final obtained pentamer structure from the MDFF simulations.

The prepared system was solvated with TIP3P water[34] and neutralized with 0.15 M NaCl with solvate and ionize VMD plugins, respectively.[65]

### 6.2.2 Ligand Docking on HPV Pentamer

Sulphonated ligands ($CH_3CH_2SO_3^-$) were docked on the HPV pentamer using the Autodock Vina software.[141] Docking was performed using the crystal structure of the L1 pentamer, using the structural models of the ligand constructed in AMPAC GUI. This part of the model building was performed by Payam Kelich. In Autodock Vina, the grid box was centered at various positions on a grid, scanning the pentamer surface with five fixed Z-coordinates and changing the X and Y coordinates from -40 Å to 40 Å for both with an interval of 1 nm, which the Linux shell scripting automized the procedure. The coordinates with the highest affinity were taken as the initial conformation for the MD simulation. Each grid box dimension was 1 $nm^3$ × 1 $nm^3$ × 1 $nm^3$ with a default spacing and exhaustiveness of 0.0375 nm and 8, respectively.

### 6.2.3 Nanoparticle Model

The model of the nanoparticle was prepared by Tara A. Nitka. Atomistic models of spherical nanoparticles were prepared by cutting a face centered cubic lattice of gold (Au) atoms (Au-Au bond length 2.88 Å) into spheres with diameter 2.4 nm and ligated with a random spherical array of 1-octanethiol (OT) and 11-mercapto-1-undecanesulfonate ligands (MUS) in 1:1 ratio forming MUS:OT-NP. All ligands were built in the AMPAC GUI and arranged into a random spherical array with 0.3949 per Au atom surface density using our own code. These MUS:OT NPs were then solvated with TIP3P water and ionized with sodium ($Na^+$) and chloride ($Cl^-$) ions at a 0.15 M NaCl concentration using the VMD solvate and ionize plugins.[65] After solvation and ionization, MUS:OT NPs were minimized for 2,000 steps using NAMD2.13 software[25] and equilibrated for 20 ns in the NPT ensemble with Langevin dynamics (Langevin constant $\gamma_{Lang}$ = 1.0 $ps^{-1}$), where temperature and pressure remained constant at 310 K and 1 bar, respectively. The particle-mesh Ewald (PME) method[33] was used to calculate the Coulomb interaction energies, with periodic boundary conditions applied in all directions. Long range van der Waals and Coulombic interactions was performed every 1- and 2-time steps, respectively. A 2 fs time step was used for equilibration of MUS:OT NPs.

### 6.2.4 Molecular Dynamics Simulations

Atomistic simulations were conducted to investigate interaction of nanoparticle and HPV capsid segments. All the systems were described with CHARMM36[6] force field parameters. MD simulations were performed with the NAMD2.13 package. All simulations were conducted with Langevin dynamics (Langevin constant $\gamma_{\mathrm{Lang}} = 1.0\ \mathrm{ps}^{-1}$) in the NPT ensemble, where temperature and pressure remained constant at 310 K and 1 bar, respectively. The particle-mesh Ewald (PME) method was used to calculate the Coulomb interaction energies, with periodic boundary conditions applied in all directions. The evaluation of long-range van der Waals and Coulombic interactions was performed every 1- and 2-time steps, respectively.

After 10,000 steps of minimization with timestep set to 2.0 fs, the 6 pentamer system was progressively relaxed, first, for 0.61 ns with the protein backbone restrained by using harmonic forces with a spring constant of 1 kcal/(mol Å$^2$) and timestep set to 1.0 fs. Next, the system was further relaxed for 3.075 ns with the protein backbone restrained by using harmonic forces with a smaller spring constant of 0.1 kcal/(mol Å$^2$) and timestep was increased to 1.5 fs. Next, the system was further relaxed for 0.975 ns with the backbone of the secondary structure of the protein restrained by using harmonic forces with a smaller spring constant of 0.1 kcal/(mol Å$^2$) and timestep was increased to 1.5 fs. Lastly, the system was relaxed in solvent without any restrains applied for 3.9 ns with timestep set to 1.5 fs.

In total, six systems were prepared for final MD simulation runs. They consisted of two, three and four pentamers with and without single MUS:OT NP. The systems with MUS:OT NP were prepared by placing a single MUS:OT NP 10 Å above the interface of three pentamers (two in case of two pentamer system) based on docking results. For all systems, 10,000 steps of minimization were performed and after that solvent molecules were equilibrated for 2 ns around the pentamers, which were restrained by using harmonic forces with a spring constant of 1 kcal/(mol Å$^2$). Next, the systems were equilibrated in production MD runs, with center-of-mass (COM) restraints applied to α-carbon of a buried residue (chosen to be residue 331) for all segments of the pentamers with a force constant of 2.0 kcal/mol, so the molecule remains in the original unit cell. The atoms within 5 Å of the XY plane containing the origin of the solvent box were constrained along z-axis with a square walled potential centered at the origin with width of 10 Å and wall potential of 100.0 kcal/mol, so that the pentamer system does not tilt on whole leading to interaction with neighboring unit cell. The length of production MD runs were 200 ns for two and three pentamer systems and 115 ns for 4 pentamer system.

## 6.3 RESULTS

### 6.3.1 Effect of nanoparticle on dynamics of capsid segments

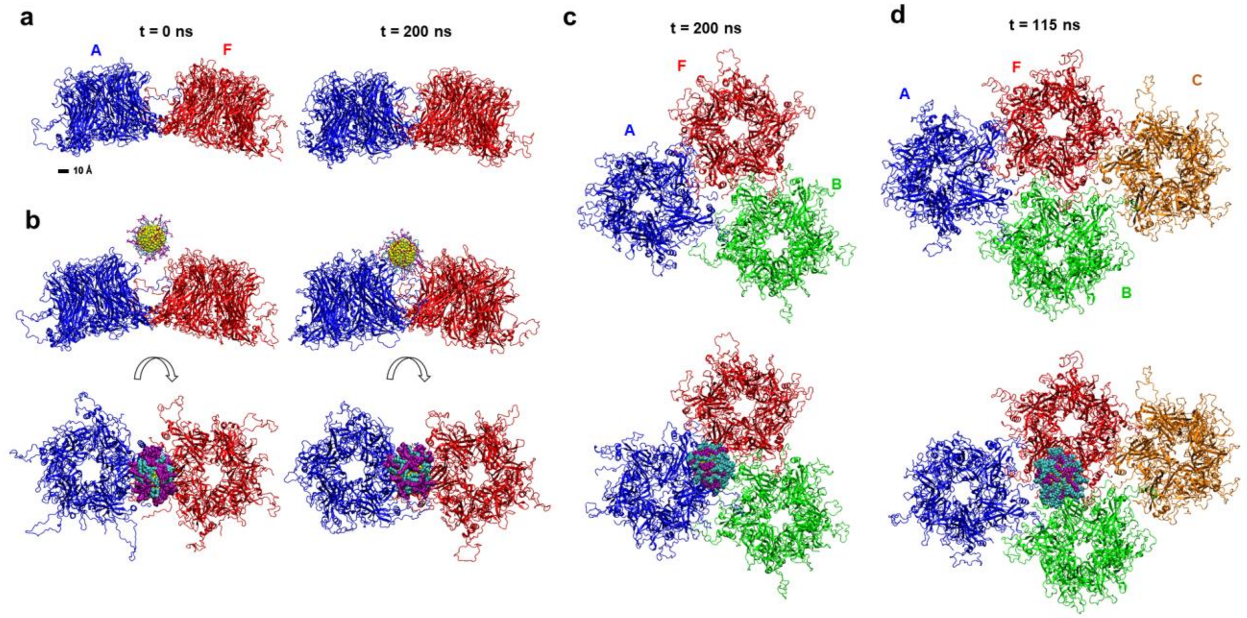


**Figure 11.** Conformations of different capsid segments drawn as ribbon, in presence and absence of MUS:OT nanoparticle drawn as surf at initial and final time. a) Two pentamer system consisting of pentamer A (blue) and pentamer F (red) without MUS:OT NP at initial time (left) and final time (right). b) Side view (top) and top view (bottom) of two pentamer system with MUS:OT NP. MUS ligand is colored in purple while OT ligand is teal colored. c) Final conformations (at 200 ns) of three pentamer system consisting of pentamer A, F and B (green) with (bottom) and without (top) MUS:OT NP. d) Final conformations (at 115 ns) of four pentamer system consisting of pentamers A, F, B and C (yellow ochre) with (bottom) and without (top) MUS:OT NP.

As shown in Figure 11, the MUS:OT NP interacts strongly and form stable interactions with HPV 16 L1 capsid protein pentamers. However, the MUS:OT NP is unable to distort individual LI pentamers structurally within the simulated time. Based on MD simulations and docking results, the MUS:OT NP prefers to interact with only at the junction of two L1 pentamers as shown in Figure 11. In case of three and four pentamer systems, we initially placed MUS:OT NP was above junction of three L1 pentamers - A, B and F. Over the course of simulation, we observe that in both cases, the MUS:OT NP shifts from three pentamer junction to two pentamer junction. Figure 11c shows that in case of three pentamer system the MUS:OT NP shifts and preferably bind to A-

F pentamer junction whereas Figure 11d shows that MUS:OT NP shifts to B-F pentamer junction in four pentamer system. The shift of MUS:OT NP to two pentamer junction happens at around 30 ns for three pentamer system and around 70 ns for four pentamer system indicating the strong preference MUS:OT NP with core diameter of 2.4 nm, to interact with two pentamer junction. Such preference indicates that a single L1 pentamer can effectively interact with only five MUS:OT NPs at a due to the geometry of the capsid segment. MUS type NPs with larger size may be able to interact with three pentamer junctions effectively and can lead to increased distortion of capsid segment.

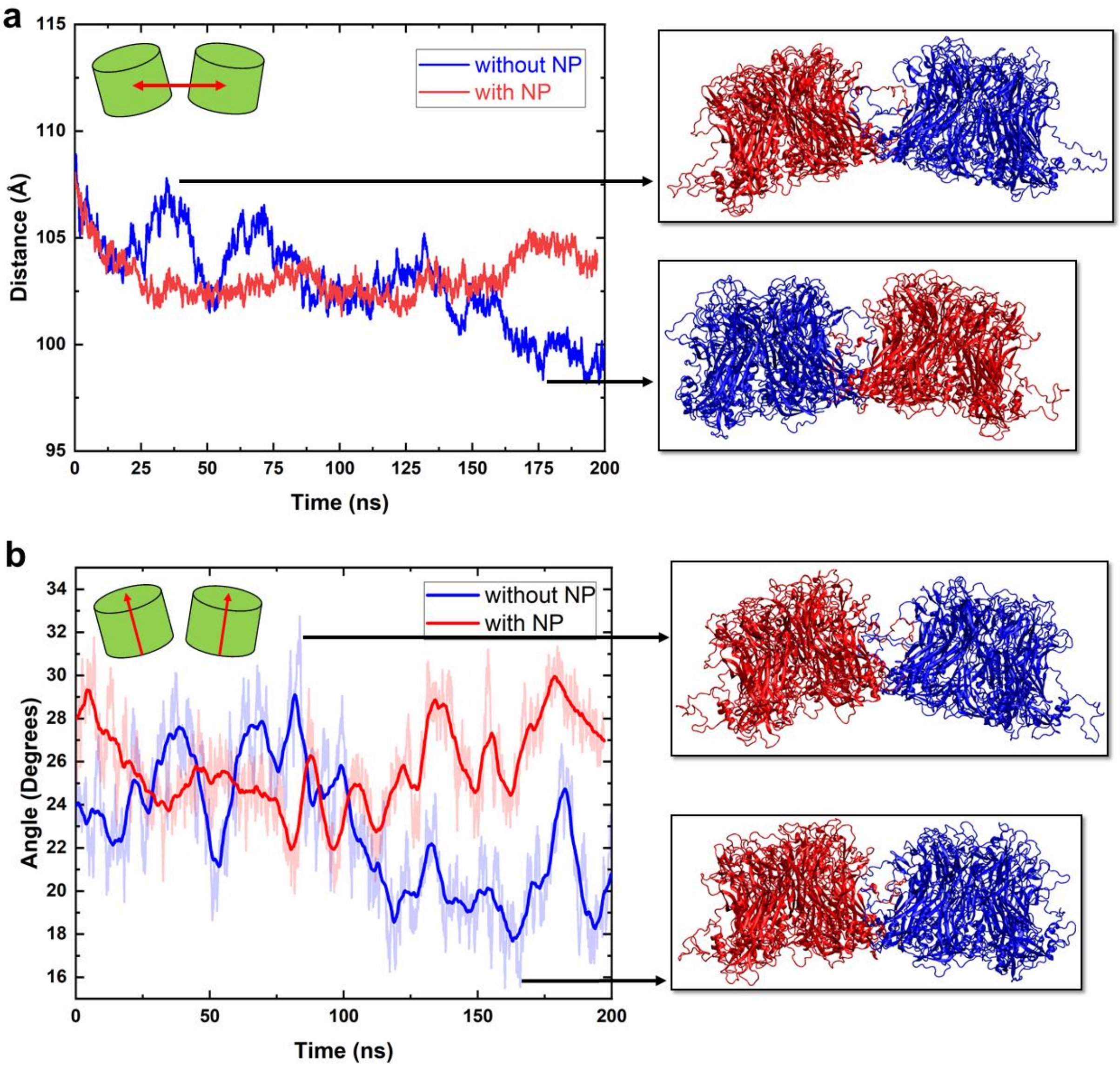


**Figure 12.** a) Distance between the center of mass of two pentamers calculated over 200 ns long simulation for two pentamer system. b) Angle between the planes containing respective pentamers for 200 ns long simulation of two pentamer system.

In order to measure capsid segment distortion in presence of MUS:OT NP, we calculated distance between center of mass (COM) of two adjacent pentamers and angle between the planes containing L1 opening for two adjacent pentamers. Figure 12 shows that in two pentamer system, the MUS:OT NP is able to "wedge in" to a certain extend between the two pentamers distorting the capsid segment as shown in Figure 11a, right. Both the distance between the interacting pentamers and angle between the planes of pentamers increases in presence of NP indicating that MUS:OT

NP can change orientation of pentamers with respect to each other and perturb the capsid segment. At two pentamer junction, MUS:OT NP can force pentamers to be further apart and more tilted with respect to each other as compared with the control.

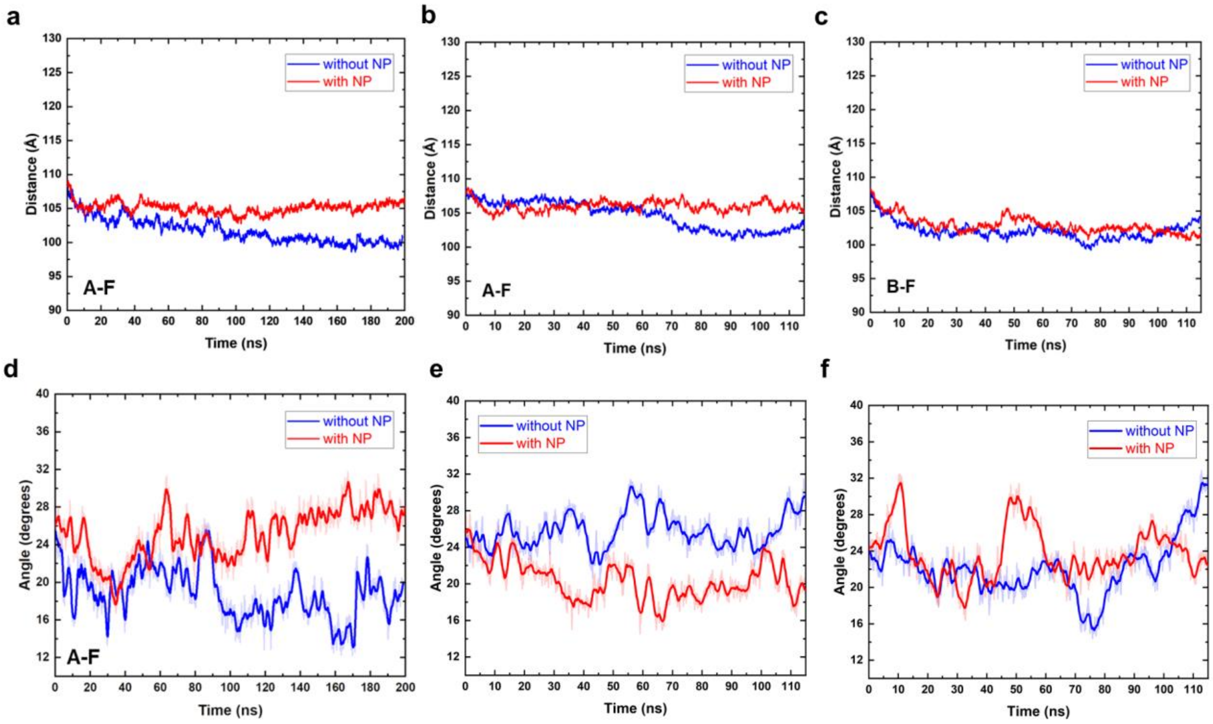


**Figure 13.** a) Distance between the center of mass of two pentamers A and F that participate in interaction with MUS:OT NP calculated over 200 ns long simulation for three pentamer system. b-c) Distance between the center of mass of two pentamers pairs (A-F and B-F) that participate in interaction with MUS:OT NP calculated over 115 ns long simulation for four pentamer system. d) Angle between the planes containing respective pentamers mentioned in (a) calculated over 200 ns long simulation of three pentamer system. e-f) Angle between the planes containing respective pentamers mentioned in (b-c) calculated over 115 ns long simulation of three pentamer system.

Similar observations are made for three pentamer and four pentamer systems as well. Figure 13a,d shows that both the distance and angle between pentamers A and F increases in presence of MUS:OT NP in three pentamer system. Note that pentamer C does not interact with MUS:OT NP in this system, and hence is not included in analysis. This “wedge in” effect starts after 100 ns in two pentamer system and after 60 ns in three pentamer system (Figure 12, 13a,d). In case of four pentamer system, the pentamer moves from A-F-B pentamer junction to A-B pentamer junction (Figure 11d). Analysis of distance and angle between pentamers involved in A-F and A-B junction which are affected in presence of MUS:OT NP; show much less pronounced effect on capsid segment as compared with smaller systems, as shown in Figure 13b,c,e,f. This indicates that stable interaction with two pentamer junction is essential for capsid segment distortion as seen in two and three pentamer systems whereas in case of four pentamer system, the MUS:OT NP seems to scan for preferred binding site (B-F pentamer junction) within the simulation time which does not allow it to effectively interact with any two pentamer junction to cause disruption of capsid segment within 115 ns. We believe that with longer simulation time MD runs, should be able to effectively capture capsid segment disruption caused by MUS type NPs in larger systems consisting of many L1 pentamers. In all cases we observe that in presence of MUS:OT NP, the distance between center of mass (COM) of two adjacent pentamers can vary from 1 Å to 5 Å as compared with control and the angle between two pentamers vary from 2 to 8 degrees as compared to the control.

### 6.3.1 Interaction of nanoparticle with capsid protein

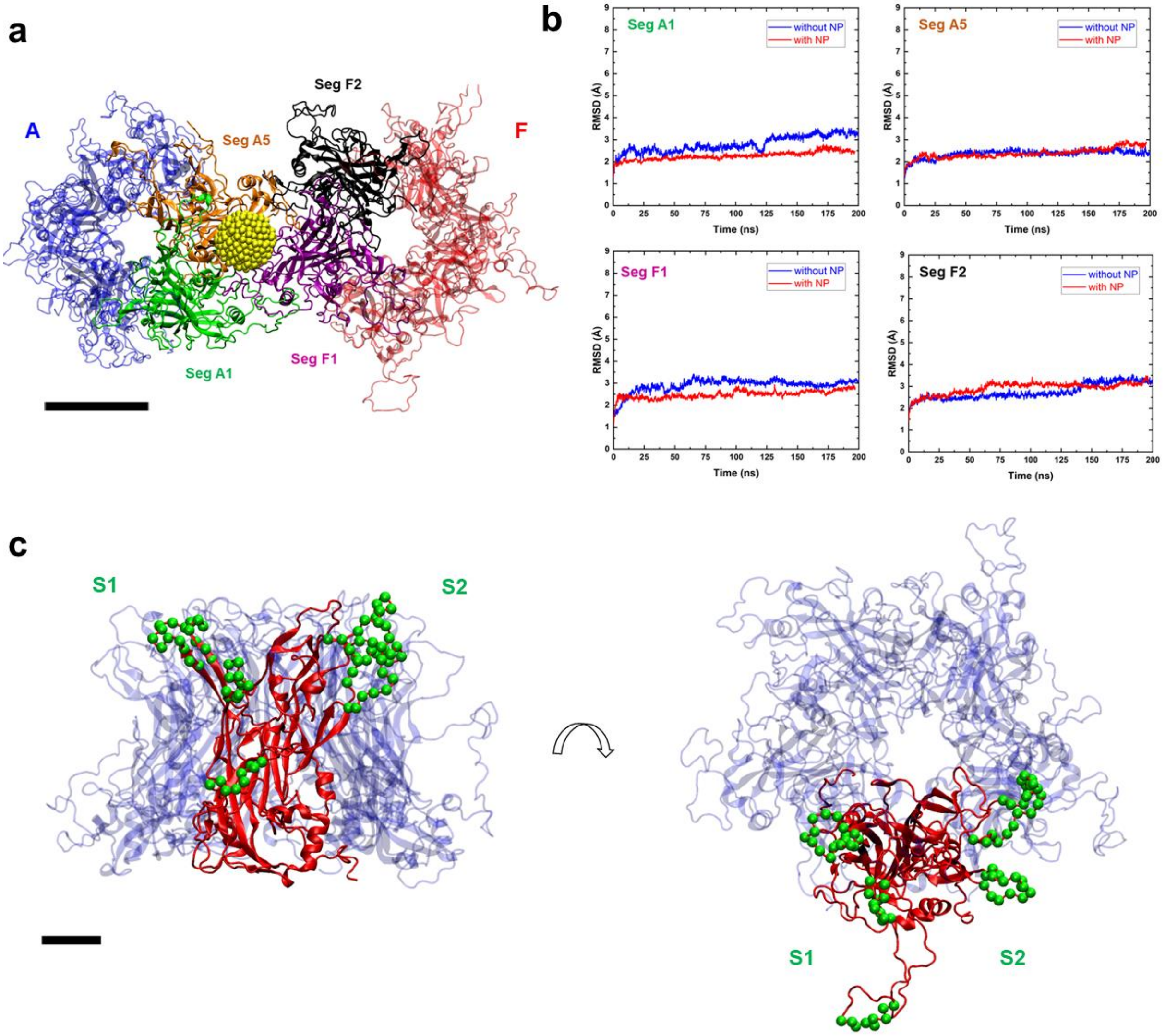


**Figure 14.** a) Typical interaction of MUS:OT NP with L1 pentamer protein (blue and red). The MUS:OT NP "wedges in" between two pentamers and interact with upto four exposed surfaces of four different protein segments of the pentamers denoted by seg A5 (yellow ochre), seg F2 (black), seg F1 (pink) and seg A1 (green). b) RMSD values of secondary structure calculated for these protein segments. c) Structure of a typical segment shown in red. Five such segments form one L1 pentamer capsid (blue). The green spheres denote the α-carbon of residues that interact with MUS:OT NP. The segment has two solvent exposed surface planes S1 and S2. The scale bar is 10 Å (black).

MUS:OT NP interacts with mostly flexible loop regions that are protruding from the capsid surface into the solvent at two pentamer junction as shown in Figure 14a,c These loop regions are also known to interact with heparin oligosaccharides[140] indicating that MUS:OT type NPs can mimic HSPG cell receptors allowing for effective viral association.[126] Each pentamer consists of five structurally similar segments. A two pentamer junction consist of interaction of four such segments from two different pentamers (two segments from each) as shown in Figure 14a. In case of two pentamer system, the MUS:OT NP is able to “wedge in” between the center of these four segments (segments A1, A5, F1 and F2) as shown in Figure 14a. Similar interaction dynamics were observed in case on three pentamer system (MUS:OT NP interacts with A-F junction consisting of segments A1, A5, F1 and F2) and four pentamer system (MUS:OT NP interacts with B-F junction consisting of segments B1, B2, F1 and F2) as well. The secondary structure of individual segment is preserved during the interaction with MUS:OT NP as shown in Figure 14b. Overall, the MUS:OT NP interaction with capsid proteins does not affect secondary structure of individual segments and pentamers within the simulation time but does affect the capsid segment.

We can consider L1 pentamer shaped as a cylindrical object with 5 arms protruding in space belonging to five structurally similar segments. Due the presence of this arm, each segment can be considered as regular tetrahedron shaped with two surfaces exposed to solvent, one buried and protruding arm at one of the vertex as shown in Figure 14c,d. The MUS:OT NP can only interact with maximum four such surfaces at a time belonging to four different segments (two of each belonging to different pentamers) of two different segments of two pentamers and as the pentamers begin to tilt with respect to each other the MUS:OT NP interacts with any one pentamer more strongly than other due to size limitation. We believe, as observed in previous subsection, that larger sized NP will be able to disrupt capsid segment more effectively.

Figure 14c show the position of residues that interact with MUS:OT NP. As each segment is structurally similar, the MUS:OT NP interact with the two unique surfaces of the segment named S1 and S2 as shown in Figure 14c. For the S1 surface, the MUS:OT NP interacts with three loop regions (residue id 52-62, 345-362 and 424-432) denoted by S1:52-62, S1:345-362 and S1:424-432 respectively, and for surface S2, the MUS:OT NP interact with two loop regions (residue id 172-186 and 265-286) denoted by S2:172-186 and S2:265-286 respectively. These loop regions are a mix of charged and non-polar amino acids which correlates well with previous MD study of single L1 pentamer with MUS:OT NP[127] which demonstrated interaction between the negative sulfonate groups of MUS:OT-NP and the positive HSPG-binding lysine residues of loop regions and interaction between non-polar alkyl chains of NP ligands and L1 proteins.

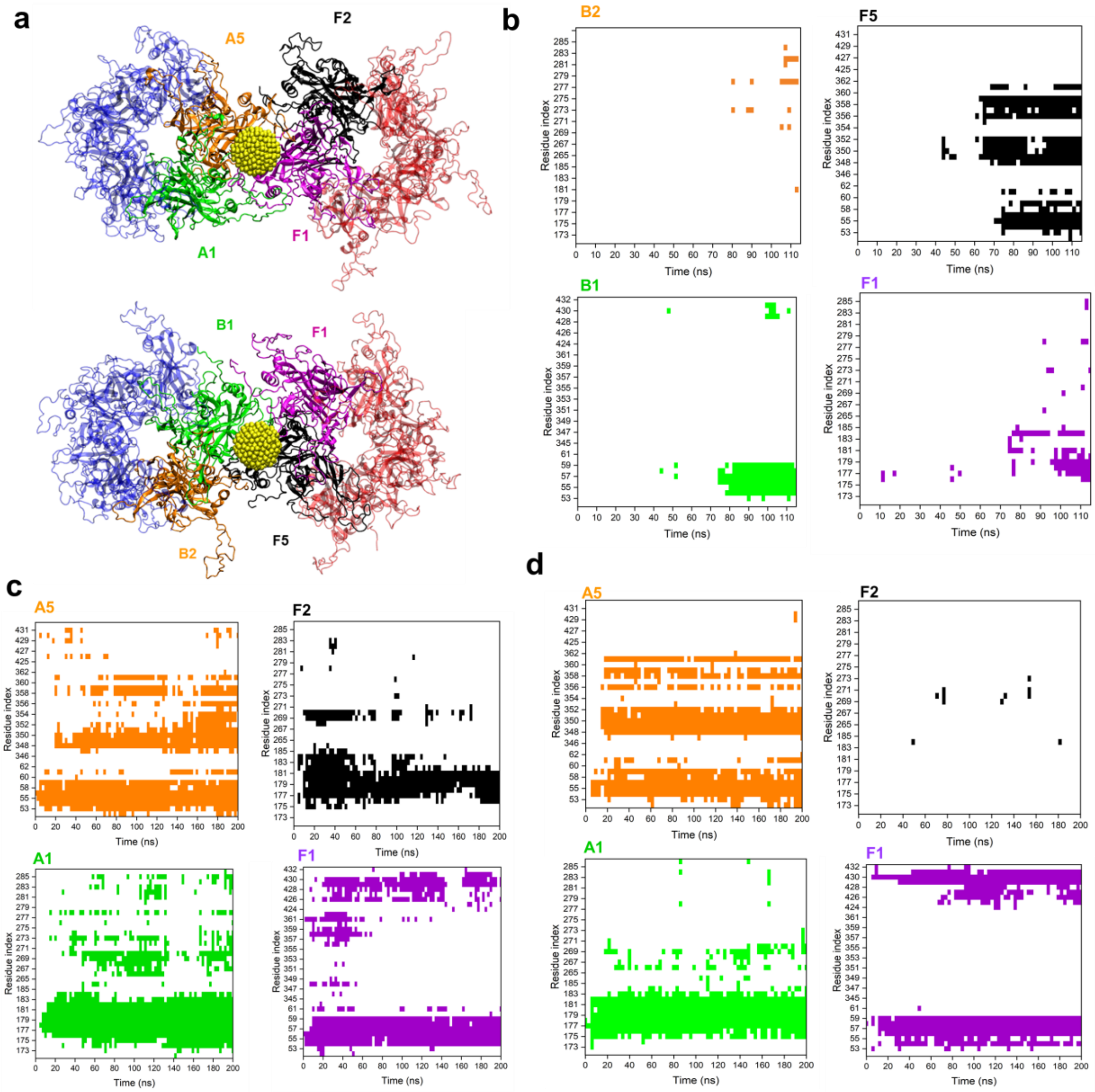


**Figure 15.** a) Typical interaction of NP with L1 pentamer junction for two and three pentamer system (top) and for four pentamer system (bottom). Interacting segments are highlighted in different colors. Contact map of amino acid residues present in interacting loop regions of different protein segments of L1 pentamer capsid protein with MUS:OT NP calculated over 115 ns long simulations for b) four pentamer system and for 200 ns long simulation of c) two pentamer system and d) three pentamer system. A contact is defined as any part of the residue within 5 Å of MUS:OT NP.

Figure 15 shows that interaction of MUS:OT NP with two pentamer junction is biased towards one pentamer than the other - pentamer A in both two and three pentamer systems and pentamer F in four pentamer system. We observe that MUS:OT NP interact with loop regions of all four surfaces (S1 and S2 of both pentamers) in two and three pentamer system but interact with only three surface regions in larger system. This is due the fact that interplay of dynamics between pentamers force these surface regions to be further apart from each other as compared to two pentamer system which only allows MUS:OT NP to interact with upto three surface regions.(Figure 15).

S1 surface interaction with MUS:OT NP is significantly larger as compared to with S2 surface. This is due to the significant interaction with S1:52-62 loop region (Figure 15) which contain positively charged Lysine (residue id: 53,54 and 59) and polar Asparagine (residue id: 56-58) which interact strongly with the negative sulfonate groups of MUS:OT-NP. Interaction with S1:52-62 is significant in all three systems indicating the importance of this region in establishing initial binding of L1 pentamer with MUS:OT NP through long range electrostatic interactions.

In case of S2 surface, the loop region S2:172-186 interacts more strongly as compared to other regions as shown in Figure 15 due to interaction between non-polar residues and alkyl chains of NP ligands. Figure 15c shows MUS:OT NP starts binding to B-F pentamer at around 70 ns with biased binding to F pentamer. S1:52-62 loop regions on segments B1 and F2 play significant role in binding followed by S2:172-186 loop region of segment F1. These two loop regions which face each other at two pentamer junction seem to play key role in binding of MUS type NPs to capsid proteins allowing "wedge in" effect which should disrupt the capsid segment.

### 6.1 CONCLUSION

In this work, we have modelled the interactions between nanomaterials coated with ligands mimicking HSPGs and HPV-16 proteins that recognize host cell HSPGs using classical atomistic molecular dynamics (MD) simulations. We determined initial molecular mechanisms responsible for the virucidal activity of these nanomaterials and identify the key regions on viral capsid surface that facilitate such interactions. Such knowledge can be utilized for designing new generations of related optimized virucidal materials and contribute to the development of new broad-spectrum and non-toxic virucidal materials.

# Chapter 7: Modeling structure of APOBEC3G protein and its interactions with ssDNA molecules

## 7.1 INTRODUCTION

APOBEC ("apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like") proteins belong to a family of evolutionarily conserved cytidine deaminases. They have diverse and important functions in human health and disease.[142] Out of 11 members of the family, APOBEC3G (A3G) has been shown to have the highest anti-viral activity against HIV-1 virus. A3G restricts the replication of HIV-1, hepatitis B virus, retrotransposons, and other DNA-based parasites.[143] In infected cells, A3G inhibits viral replication through the specific deamination of Cytosines (C) in viral minus-strand DNA resulting in huge G to A hypermutation of the viral DNA genome during reverse transcription.[144] [145]

A3G has a conserved zinc-dependent deaminase sequence motif (ZDD) within a α-β-α super secondary structural element that forms the catalytic site of a C-terminal domain (CTD) domain as shown in Figure 16.[146] [147] [142] A3G deaminates two consecutive cytidines (bold) in a strict 3′ to 5′ order in 5′-**CC**CA-3′ motif in ssDNA.[148] Of the two homologous domains, the catalytically inactive N-terminal domain (NTD) required for Vif, DNA and RNA binding and is positively charged, and the catalytically active C-terminal domain (CTD) required for catalysis and base sequence specificity.[149]

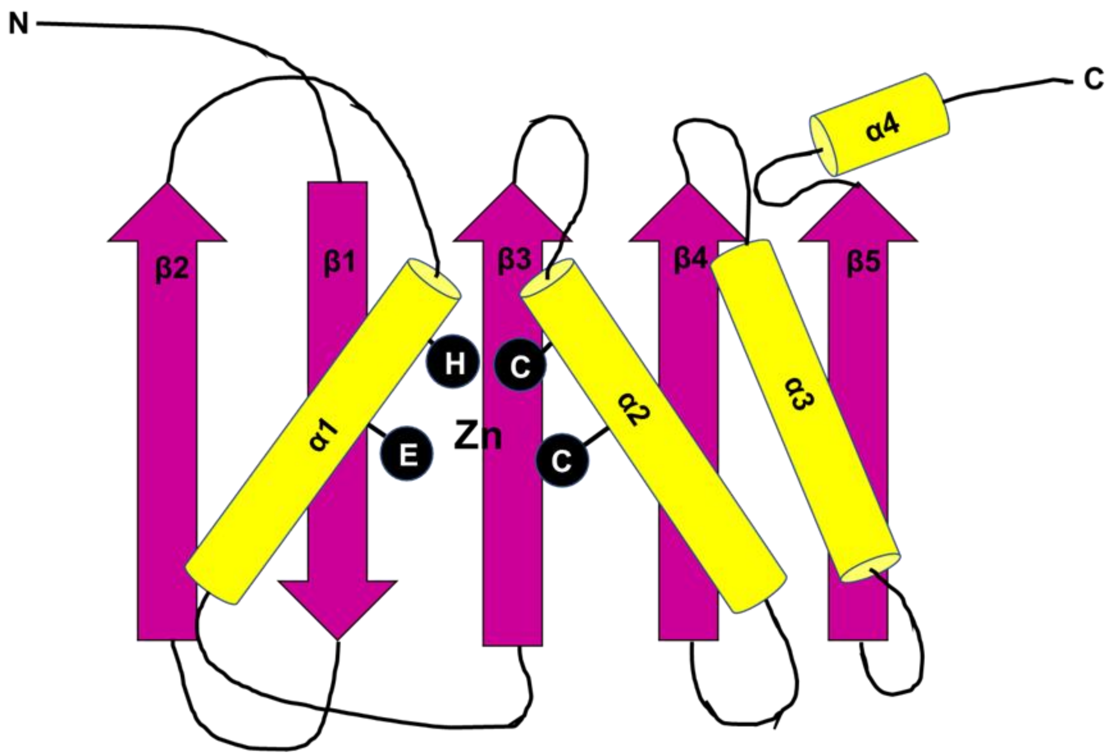


**Figure 16.** Schematic representation of ZDD domain with beta sheets represented by arrows while cylinder denotes the alpha helix. H, C and E denote amino acids histidine, cysteine and glutamic acid respectively.

High-resolution atomic structure of a full-length functional human A3G remains undetermined, due to oligomerization and precipitation of A3G at concentrations required for crystallization.[150] Expression and purification in bulk has also been very difficult using bacterial expression systems as these proteins are genotoxic and are frequently insoluble as full-length proteins.[151] [152] However, structures of individual mutated CTD and NTD have been determined by NMR spectroscopy and X-ray crystallography[153] [154], and the shape of the full-length A3G protein was revealed by small-angle X-ray experiments and advanced envelope restoration methods.[155] [156]

Previous work done by our lab in collaboration with Prof. Yuri Lyubchenko (U. Nebraska) resulted in the atomic scale model of the full-size A3G via computational modeling and docking of A3G C-terminal and N-terminal domains based on available structures of CTD and NTD of A3G, followed by microsecond-long molecular dynamics (MD) simulations.[157] These simulations revealed a highly dynamic feature of A3G monomer that can lead to extended conformations in which two domains are separated by distances as large as 4.5 nm.

According to recent AFM studies[158], such dynamical conformation change should allow A3G to assume dumbbell conformation of A3G may allow CTD domain to efficiently search for the deamination spot on ssDNA substrate, while NTD domain remains strongly attached to the ssDNA substrate. A3G is known to scan for preferred motif (5′-CCC-3′) for deamination by different mechanisms that are categized as sliding, jumping and intersegmental transfer.[159] [160] The dumbbell conformation should allow enzyme to be more flexible and positively charged NTD domain to effectively interact with negatively charge backbone of ssDNA without the interference of overall negatively charge CTD domain.

Current computational model, however, is not ideal to study such dynamics of protein:ssDNA complex as initial ensemble of A3G structures has predominantly globular conformations because the docking procedure favors structures where CTD and NTD interact with each other, rather than being fully immersed in the solvent. Both A3G and A3G: ssDNA complex behavior in solution during MD simulations does not correlate with experimentally determined properties: A3G remains in globular conformation on microsecond timescale. Computational cost of running atomistic MD simulations on longer time scales is very high, hence impractical. Thus, our goal is to develop a better computational model of A3G which is more dynamic and correlates with experimental data.

In this chapter, I describe how we improve upon previous computational model using structural information from recently solved crystal structure of Rhesus Macaque full length A3G[161], and utilized coarse grained MD simulations to access longer timescales and understand interaction of A3G with ssDNA.

## 7.2 METHODS

### 7.2.1 All atomistic (AA) models

Due to lack of structural data of full length human, we chose computational model of full length A3G based on previous work involving MD simulations of docked ensembles CTD and

NTD as our starting model.[157] This docked model consisted of C-terminal domain with the linker (PDB:2KBO, residues 196-384)[162] and N-terminal domain (PDB:5K83, residues 1-195).[163] The CTD structure had a wild-type sequence and NTD crystal structure contained 57 mutations from the wild-type sequence (UniProtKB code: Q9HC16)[164] which were mutated back to their corresponding wild-type residues within VMD.[65] A missing loop (residues 139–142) on NTD was added with VMD, by alignment to another solved NTD structure (PDB:2MZZ)[154]. The NTD was relaxed and simulated for 100 ns prior to docking. $Zn^{2+}$ ions were placed into the active site pockets after docking.

In starting structure was further refined by structural alignment by full-length structure of A3G (rA3G) found in rhesus macaque (PDB: 6P3X).[161] For this, the starting structure was divided into two parts, one part containing residues 1 to 207 and other part containing residues 208 to 384. These two parts were structurally aligned into their corresponding domains present in rA3G and then combined to build the initial A3G structure.

The initial structure was solvated with TIP3P water[34] and neutralized with 0.15 M NaCl. After 2,000 steps of minimization, the solvent molecules were equilibrated for 1 ns around the initial structure, which were restrained by using harmonic forces with a spring constant of 1 kcal/(mol $Å^2$). After that the linker region (residues 196-218) was relaxed with the rest of the protein restrained by using harmonic forces with a spring constant of 1 kcal/(mol $Å^2$). The relaxed structure was then simulated for 100 ns in three different trials with different seed value, with center-of-mass (COM) restraints applied to three random buried residues of CTD with a force constant of 1.0 kcal/mol, so the molecule remains in the original unit cell. NTD was unrestrained.

The final structures obtained from the three trials were used to build initial structures of A3G-ssDNA complexes by placing ssDNA adjacent to A3G model guided by the alignment of available crystal structures of individual domains in complexes with ssDNA. These structures include CTD with 9 nucleotide long ssDNA (PDB: 6BUX)[165] and NTD with 9 nucleotide long ssDNA (PDB: 5K83)[163]. In total, six systems were prepared ; three systems with ssDNA bound to CTD and three systems with ssDNA bound to NTD. All the six prepared systems were solvated with TIP3P water and neutralized with 0.15 M NaCl with solvate and ionize VMD plugins, respectively.[65]

### 7.2.2 AA Molecular Dynamics simulations

Atomistic simulations were performed to investigate dynamics of A3G protein complexes with ssDNA. All the systems were described with CHARMM36 force field parameters.[6] MD simulations were performed with the NAMD v2.13 package.[25] All simulations were conducted with Langevin dynamics (Langevin constant $\gamma_{\mathrm{Lang}} = 1.0\ \mathrm{ps}^{-1}$) in the NPT ensemble, where temperature and pressure remained constant at 310 K and 1 bar, respectively. The particle-mesh Ewald (PME) method[33] was used to calculate the Coulomb interaction energies, with periodic boundary conditions applied in all directions. The evaluation of long-range van der Waals and Coulombic interactions was performed every 1- and 2-time steps, respectively.

All six systems were relaxed for 10,000 steps after which the solvent molecules were equilibrated for 1 ns around the initial structure, which were restrained by using harmonic forces with a spring constant of 1 kcal/(mol Å$^2$). After that the ssDNA was relaxed with the protein restrained by using harmonic forces with a spring constant of 1 kcal/(mol Å$^2$). The relaxed structure was then simulated for 100 ns, with center-of-mass (COM) restraints applied to three random buried residues of CTD with a force constant of 1.0 kcal/mol, so the molecule remains in the original unit cell. NTD was unrestrained.

### 7.2.1 Coarse grained (CG) models and MD simulations

CG models were built for atomistic structures described above, obtained after 100 ns AA simulations. These structures include three trials each of A3G, A3G with ssDNA bound to CTD and A3G with ssDNA bound to NTD. CG models were built using python scripts available on MARTINI website.[166] [31] All nine systems were solvated with standard MARTINI water model having 10 % anti-freeze water and neutralized with 0.15 M NaCl using program package

GROMACS (version 2019.x).[167] Zinc ions were removed from the final structure and secondary structure of A3G was distance restrained by harmonic potentials between atoms with a distance cutoff of 1.5 nm. All simulations were performed with GROMACS (version 2019.x) using the Martini force field v2.1 in NPT ensemble with temperature and pressure set to 310 K and 1 bar, respectively. The time step is set to 30 fs for protein systems and 20 fs for protein:ssDNA complex. All simulations were performed using velocity rescale thermostat[168], Berendsen (for minimization) and Parrinello–Rahman barostat.[169] After 50,000 steps of minimization, the solvent was equilibrated for 5 ns with time step of 1 fs. All systems were then equilibrated in production MD runs of 10 μs with distance restraints applied as aforementioned.

## 7.3 RESULTS

### 7.2.2 Dynamics of A3G monomer

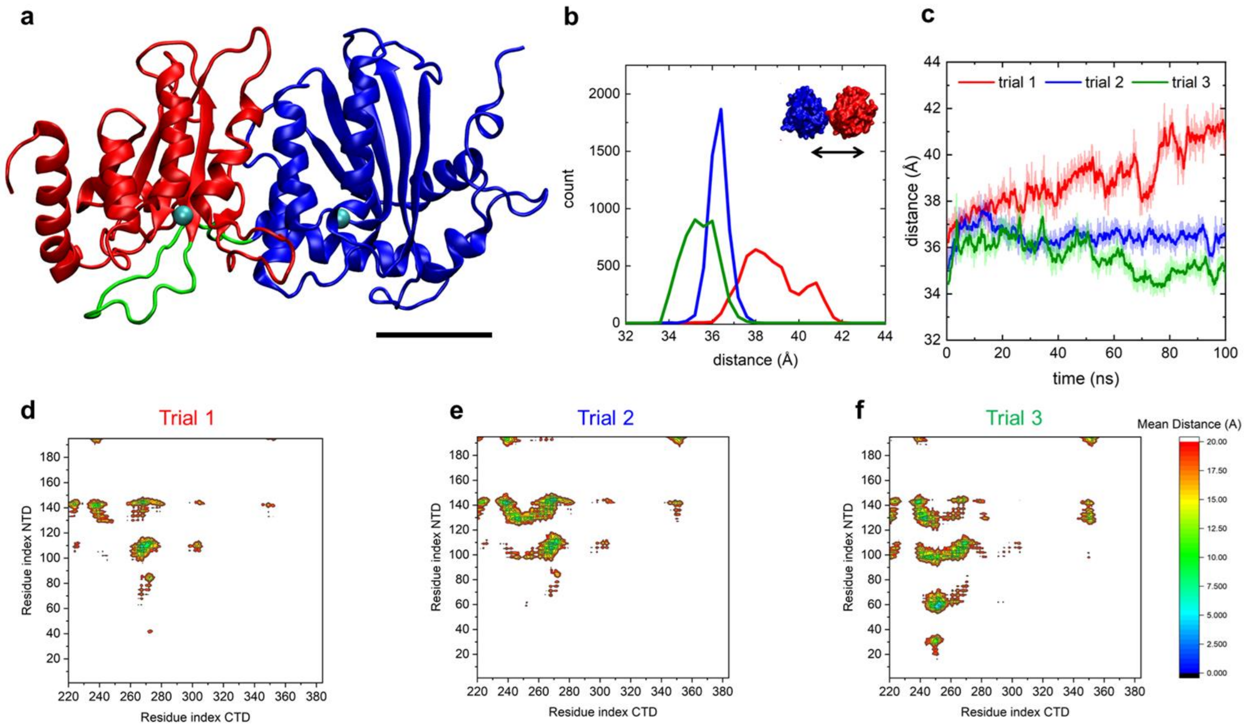


**Figure 17.** a) Initial structure of the A3G consisting of two domains CTD (red), and NTD (blue), linked by a flexible loop region (green). $Zn^{2+}$ ions are shown as cyan spheres b) Histograms of distance d between center of mass of CTD and NTD for three 100 ns long AA simulations. c) Time evolution of same. d-f) Contact maps between CTD and NTD at 100 ns for different trials.

The conformational space of A3G is likely highly rugged and depend on domain-domain interactions. Figure 17b-c shows that three 100 ns AA simulations lead to different final conformations of A3G domains based on distances between the center of mass(COM) of CTD and NTD. Trial 1 shows that CTD and NTD domains are able to break domain-domain interactions (Figure 17d) and attain a dumbbell conformation with distance between COM of two domains reaching around 42 Å at the end of the simulations. Trial 3 shows that new domain-domain interactions can be formed (Figure 17f) and the distance between the COM of two domains is reduced to 35 Å leading to a highly globular conformation. Trial 3, on the other hand, is stuck in a globular conformation (Figure 17b) for the entire length of the simulation with unchanging domain-domain interactions (Figure 17e).

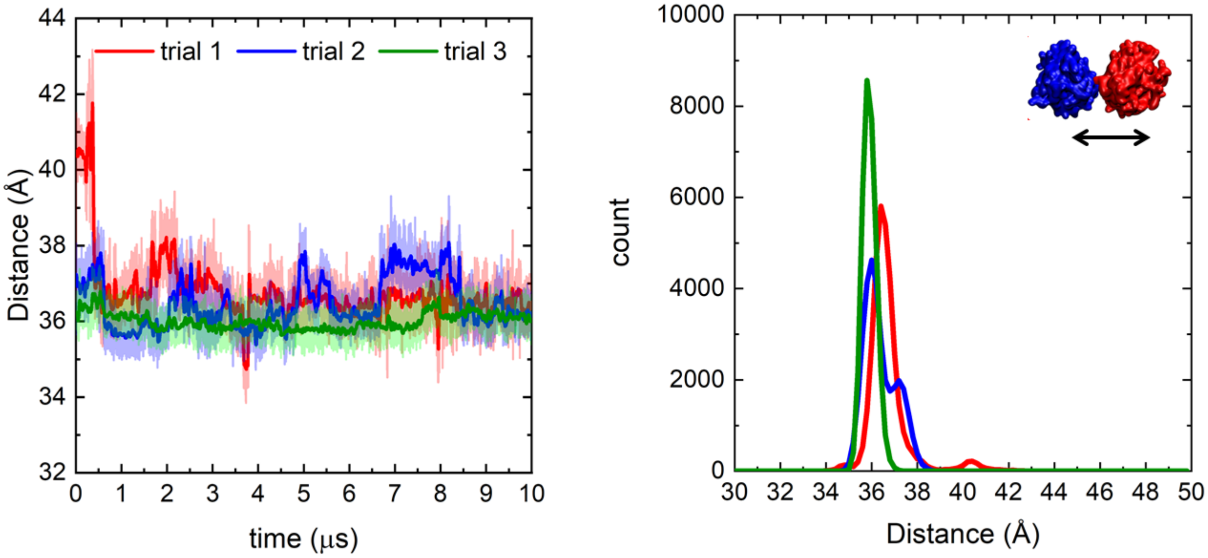


**Figure 18.** Distance between the center of mass of CTD and NTD calculated over 10 μs long CG simulations (left) and histogram of the same (right).

In 10 μs CG simulations of A3G starting from final conformations from AA simulations, all the conformations return to globular shapes with mean distance between domain varying from 36 to 38 Å (Figure 18). The conformational space of globular A3G is highly dynamical, with A3G domains readjusting with respect to each other.(Figure 18a). The globular shape seems to be energetically favorable and indicates the presence of high energy barrier between globular and dumbbell conformations.

### 7.2.2 Interaction of ssDNA with A3G monomer

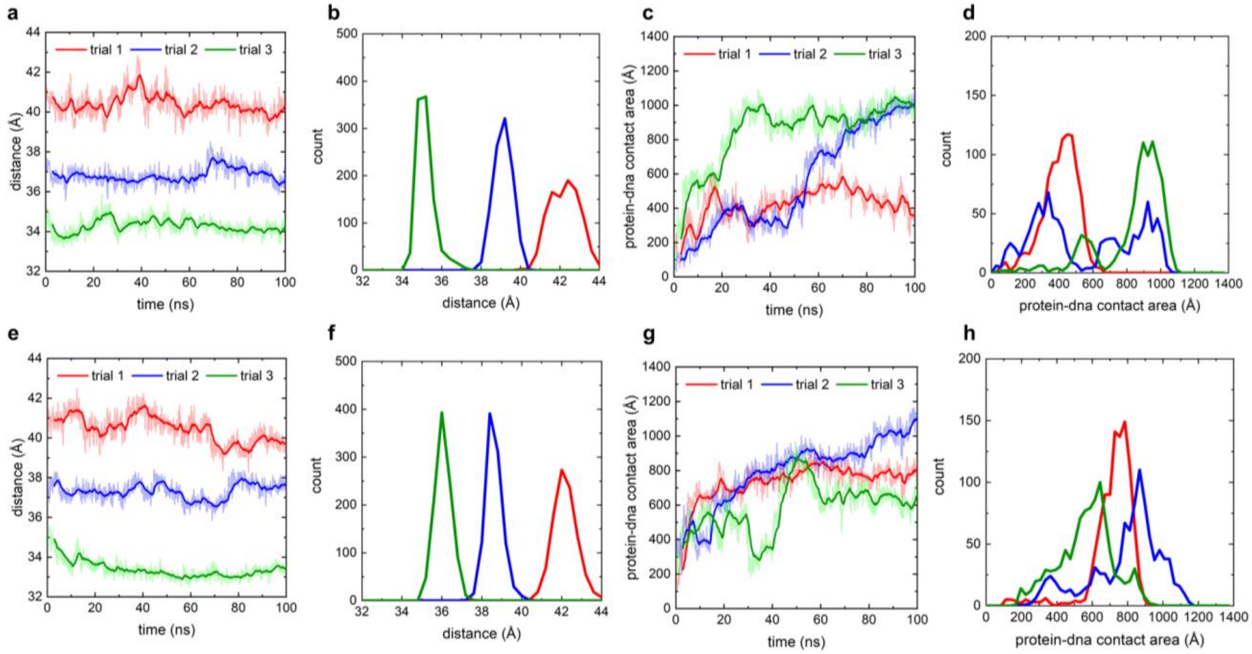


**Figure 19.** a) Time evolution of distance between center of mass of CTD and NTD for three trials of 100 ns long AA simulations with ssDNA in complex with CTD. b) Histogram of the same. c) Protein-DNA contact area for three trials of 100 ns long AA simulations with ssDNA in complex with CTD. d) Histogram of the same. e) Same as (a) except the DNA is in complex with NTD instead of CTD. f) Histogram of the same. g) Same as (c) except the DNA is in complex with NTD instead of CTD. h) Histogram of the same.

In presence of DNA, there is minimal conformational change in A3G, as observed in distances between the domains in 100 ns long AA simulations, shown in Figure 19a,e. Interaction of ssDNA with A3G is highly dynamic. In case of ssDNA in complex with CTD (Figure 19c,d), the ssDNA interacts with A3G more in trials 2 and 3, which are in globular conformations as compared to trial 1, where A3G is in dumbbell conformation. This indicates that ssDNA has more interaction with globular conformations which are thought to be essential for catalytic activity as opposed to dumbbell conformations which allow CTD domain to reduce interaction with ssDNA and scan for preferred motif in ssDNA at different site as proposed by recent AFM studies.[158] Figure 19g,h shows that the interaction with NTD bound DNA does not seem to depend on the whole A3G conformation, which is correlated with previous experimental studies showing that NTD is a non catalytic domain and binds non-specifically to ssDNA.[163]

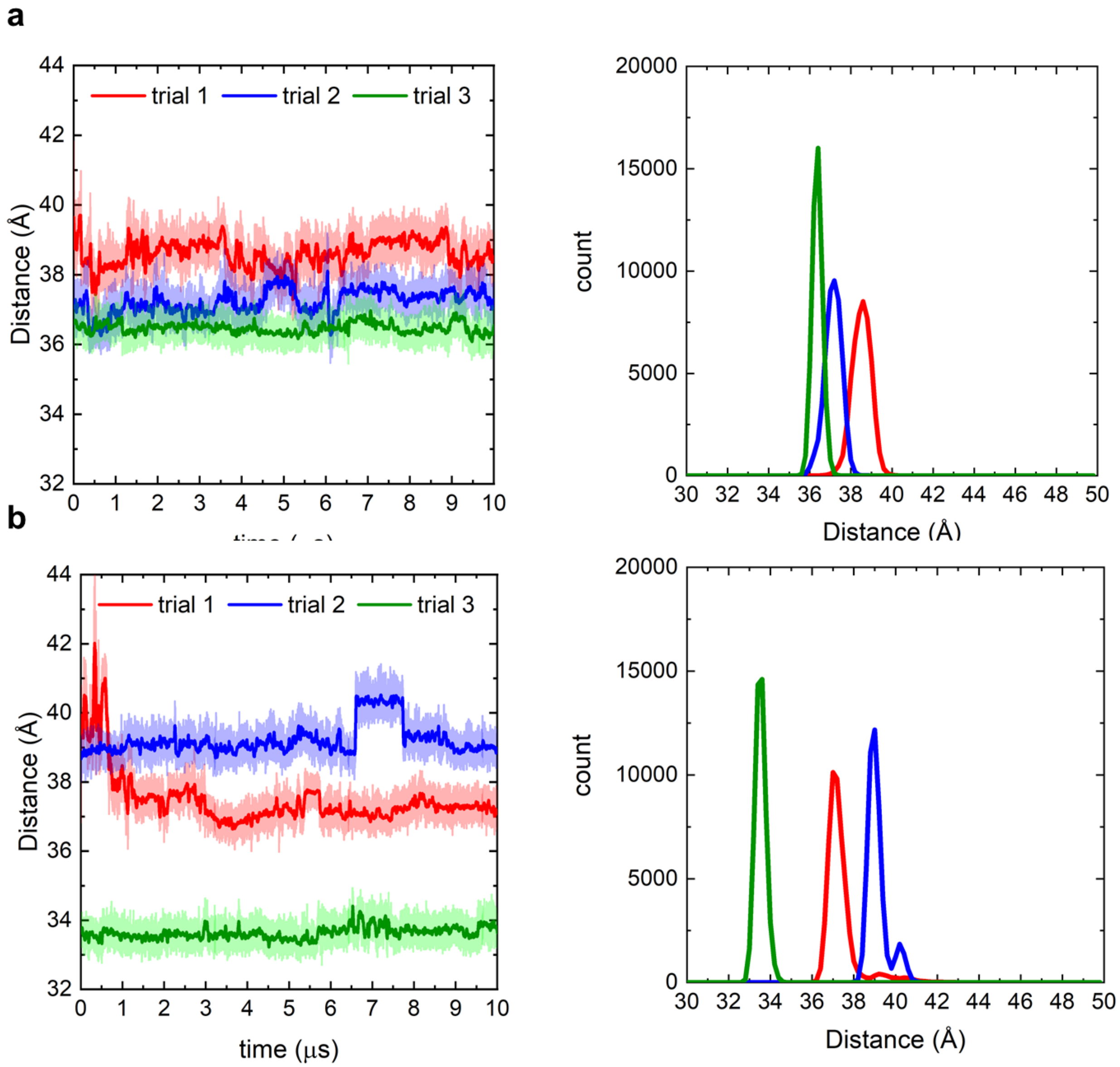


**Figure 20.** a) Distance between the center of mass of CTD and NTD, calculated over 10 µs long CG simulations with ssDNA in complex with CTD (left) and histogram of the same (right). b) same as a except ssDNA in complex with NTD instead of CTD.

As shown in Figure 20a, A3G with ssDNA bound to CTD remains in globular conformation at microsecond scales. Trial 1, which was initially in dumbbell shape transitions to globular conformation as well. In Figure 18b, A3G conformations that have greater distance between the domains (trial 1 and 2) are more dynamic and seem to switch between conformations which are globular and “dumbbell-like”. Overall, the average distance between the domains for all three trials is the largest in the case of A3G in which ssDNA bound to NTD, followed by A3G in which ssDNA is bound to CTD, and the smallest in A3G alone (Figure 19, 20). This supports the experimental data which indicates that A3G population shift towards dumbbell conformation in presence of ssDNA and even more if ssDNA is bound to NTD which allows CTD to scan for preferred motifs in ssDNA for catalytic activity.

## 7.4 CONCLUSION

We demonstrated that CG simulations are faster and computationally efficient to understand dynamics of A3G and its interaction with ssDNA. CG simulations allow access to microsecond time scales for such systems to better sample the conformational energy landscape, which is highly rugged in such systems. Our studies show that CG computational model of A3G is able to capture some dynamics observed in experiments, but further studies are required and are ongoing to understand mechanism of interaction of ssDNA with A3G.

# Chapter 8: Adaptive Evolution of Peptide Inhibitors for Mutating SARS-CoV-2

## 8.1 INTRODUCTION

The very fast spreading of a severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in the human population has already led to hundreds of thousands of fatalities and dire socioeconomic effects worldwide. Therefore, many concepts and strategies were initiated to identify drug targets[170], develop effective therapeutics and vaccines against SARS-CoV-2 [171] [172], including the optional use of therapeutics developed for other purposes.[173] [174] Some of the strategies have focused on direct targeting of the spike (S) protein[175] [176], located on the outer surface of the SARS-CoV-2 virion, which initiates the cell entry process. This process starts by binding of the receptor binding domain (RBD) on the S protein to a human host receptor angiotensin converting enzyme 2 (ACE2), followed by a proteolytic cleavage and release of its S1 subunit.[177] Then, the remaining S2 subunit undergoes a conformational change, which eventually leads to virus–host membrane fusion.[178] The S protein RBD of SARS-CoV-2 binds more strongly to ACE2 than SARS-CoV.[179] This fact, together with other innovative characteristics of the S protein, including a furin preactivation of the cleavage site, and a hidden RBD, likely contributes to the large infectivity of SARS-CoV-2.[179]

SARS-CoV-2 is mutating and further adjusting to the human environment, like other novel viral pathogens. Many strains of SARS-CoV-2 have already been detected.[180] Some mutations, such as D614G on the S protein, lead to rapid and enhanced viral transmission[181], causing this strain to locally dominate. The mutating SARS-CoV-2 coronavirus could also adapt to new hosts, such as domestic animals.[182] Therefore, to mitigate the large spreading and effects of SARS-CoV-2, it is important to identify classes of therapeutics that could be rapidly developed to act against multiple coronavirus strains.

In this chapter, I describe how we address this problem by introducing libraries of adaptive peptide therapeutics that could block different S protein strains from interacting with ACE2. With the use of advanced computational methods, we show that adaptive evolution of suitable peptide templates can provide multiple inhibitors for competitive binding to different S protein variants. Using libraries of peptides adapted to different S protein variants could prevent their mutational escape, in analogy to using cocktails of S protein antibodies.[175] Our algorithm can evolve peptide inhibitors that competitively bind to (block) a set of desired targets, such as different S protein variants, S proteins with glycan shielded sites[183] [184], and other related systems.[185]

## 8.2 METHODS

### 8.2.1 Molecular Dynamics Simulations

The simulated peptides template-1 (amino acids 21–43 of ACE2) and template-2 (amino acids 19–83 of ACE2) were separately bound to the S protein RBD. Template-1 size is typical of peptides that can be chemically synthesized easily and quickly for experimental testing. Previous simulations have demonstrated that the extracted ACE2-based helix of template-1 loses the curvature and shape complementarity to the RBD binding site, a larger template-2 was also selected for sequence optimization.[176] Template-2 contains two helices that can preserve helix curvature and shape complementarity to the RBD binding site. All structures were directly based on the crystal structure of the human ACE2 protein bound to the SARS-CoV-2 Spike protein RBD (pdbID: 6LZG).[177] The mutations in peptides and RBD were introduced using the psfgen plugin in VMD.[65]

The systems were simulated using NAMD2.13[25] and the CHARMM36 protein force field.[6] The simulations were conducted with the Langevin dynamics ($\gamma_{\text{Lang}} = 1\ \text{ps}^{-1}$) in the NpT ensemble, at a temperature of $T = 310$ K and a pressure of $p = 1$ bar. The particle-mesh Ewald (PME) method was used to evaluate Coulomb coupling, with periodic boundary conditions applied.[33] The time step was set to 2 fs. The long-range van der Waals and Coulombic coupling were evaluated every 1- and 2-time steps, respectively. After 2000 steps of minimization, the solvent molecules were equilibrated for 1 ns, while the complexes were restrained using harmonic forces with a spring constant of 1 kcal (mol Å)$^{-1}$. Next, the systems were equilibrated in 100 ns production MD runs with no restraints.

### 8.2.2 MMGB-SA Calculations

The molecular mechanics generalized Born–surface area (MMGB-SA)method[186] [187] was used to estimate the relative binding free energies between peptides (or ACE2) and RBDs. The free energies were estimated from separate MMGB-SA calculations for three systems (peptide/ACE2, RBD, and the whole complex) in configurations extracted from the MD trajectories of the whole complex in the explicit solvent. The MMGB-SA free energies of the extracted configurations of the three systems were calculated as

$$G_{tot} = E_{MM} + G_{solv-p} + G_{solv-np} - T\Delta S_{conf}$$

where $E_{MM}$, $G_{solv-p}$, $G_{solv-np}$, and $\Delta S_{conf}$ are the sum of bonded and Lennard-Jones energy terms, the polar contribution to the solvation energy, the nonpolar contribution, and the conformational entropy, respectively. The $E_{MM}$, $G_{solv-p}$, and $G_{solv-np}$ terms were calculated using the NAMD 2.13 package[25] generalized Born implicit solvent model[188], with a dielectric constant of the solvent of 78.5. The $G_{solv-np}$ term for each system configuration was calculated in NAMD as a linear function of the solvent-accessible surface area (SASA), determined using a probe radius of 1.4 Å, as $G_{solv-np} = \gamma$SASA, where $\gamma$ = 0.00542 kcal mol$^{-1}$ Å$^{-2}$ is the surface tension. The $\Delta S_{conf}$ term was neglected, since the entropic contribution differences nearly cancel when considering protein–protein binding of single residue mutants.[189] [190] Moreover, the entropy term, which is often calculated with a large computational cost and low prediction accuracy, is likely to be similar for the studied systems, which differ in single or several mutations. Since the $G_{tot}$ values are obtained for peptide configurations extracted from the trajectories of complexes, $G_{tot}$ does not include the free energies of peptide reorganization; the correct free energies of binding should consider configurations of separately relaxed systems. The approximate binding free energies of the studied complexes were calculated as $\langle \Delta G_{MMGB-SA} \rangle = \langle G_{tot}(P/ACE2 - RBD) -$

$G_{tot}(P/ACE2) - G_{tot}(RBD)\rangle$ where P/ACE2 represents peptides or complete ACE2, and the $\langle$ averaging $\rangle$ is performed over configurations within the second half of the calculated trajectories.

### 8.2.3 Adaptive Evolution Algorithm

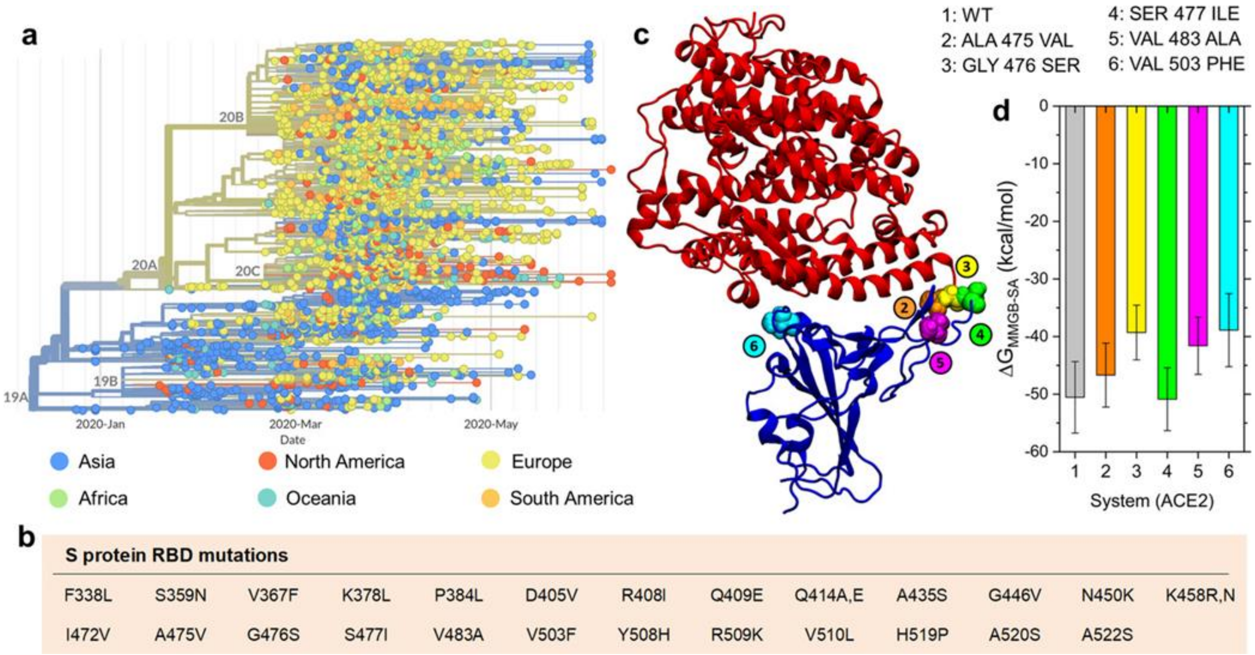
A mutation/selection algorithm was developed and used to iteratively increase the affinity of binding between peptide templates and the S protein RBD. The algorithm involved sequences of steps combining molecular dynamics simulations and MC decisions using the Metropolis criterion[191 192], which resulted in MC sampling of the peptide sequence space. Initially, the selected peptide template (directly extracted from ACE2) complexed with RBD was equilibrated for 100 ns. Then, the free energy of binding of the peptide:RBD complex, $\Delta G_{MMGB-SA}^{after}$ was evaluated. Next, a random mutation was introduced at a random position in the peptide, followed by a short 1–10 ns equilibration in MD simulations of the complex, and evaluation of the $\Delta G_{MMGB-SA}^{after}$ free energy of binding of the complex. Finally, the mutation was accepted if $\Delta G_{MMGB-SA}^{after} < \Delta G_{MMGB-SA}^{after}$ or if the Metropolis criterion was satisfied,

$$exp\left[-\left(\Delta G_{MMGB-SA}^{after} - \Delta G_{MMGB-SA}^{before}\right)/kBT\right] > r$$

where $r$ is a random number in the (0,1) interval. If the mutation is accepted, then the new peptide becomes the new reference peptide and its $\Delta G_{MMGB-SA}$ becomes the reference value $\Delta G_{MMGB-SA}^{before}$ for the next attempted mutation. In each run of the algorithm, 100–200 mutations were attempted on peptide templates, as stated in the results.

## 8.3 RESULTS

### 8.3.1 S protein variants



**S protein RBD mutations**

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| F338L | S359N | V367F | K378L | P384L | D405V | R408I | Q409E | Q414A,E | A435S | G446V | N450K | K458R,N |
| I472V | A475V | G476S | S477I | V483A | V503F | Y508H | R509K | V510L | H519P | A520S | A522S | |

**Figure 21.** Mutations of SARS-CoV-2. a) Time-dependent mutation tree of SARS-CoV-2 colored according to geographical regions of origin (through June 2020).[21] b) Twenty-five single mutations identified in RBD of the S protein. c) Five amino acid mutations on RBD in contact with the ACE2 receptor. d) Binding free energies are evaluated with the MMGB-SA method for the ACE2-RBD complexes, including the originally reported RBD (wild type, labeled as WT) and the five mutant RBDs listed in panel (c).

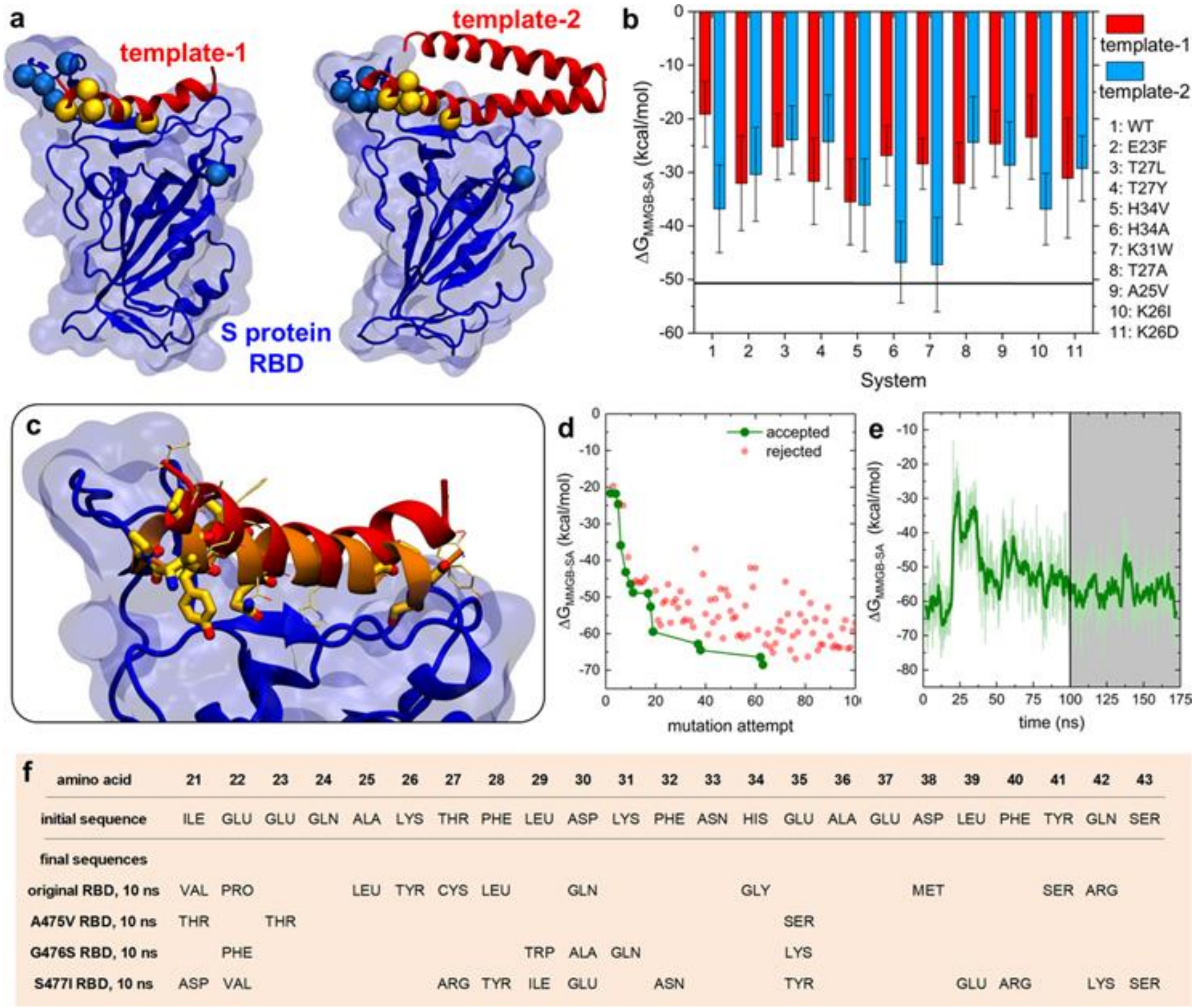
Over the time and geographical regions, SARS-CoV-2 virus has so far evolved into more than 104 mutated strains, shown in the mutation tree of publicly available unique genome sequences (through June 2020) in Figure 21a.[193] Out of these mutations, so far (June 2020) 25 have been recognized in the S protein RBD, as summarized in Figure 21b. Five of these mutations include amino acids that form a part of the ACE2-binding surface (A475V, G476S, S477I, V483A, and V503F), as highlighted in Figure 1c. The information above was identified by a collaborator Yanxiao Han.

In Figure 21d, we present the RBD-ACE2 host-receptor binding free energies, $\Delta G_{MMGB-SA}$, obtained in the presence of these five mutations. The 5 RBD:ACE2 complexes are simulated for 30 ns, and their average binding energies are obtained in the last 15 ns. The originally reported RBD and the S477I RBD have the strongest binding to the human ACE2, $\Delta G_{MMGB-SA}$, $\approx -50$ kcal mol$^{-1}$, while the other systems bind with $\Delta G_{MMGB-SA}$, ≈ −(40–50) kcal mol$^{-1}$. In order to block all these RBD variants, the peptide inhibitors should have comparable or lower ΔGMMGB-SA values.

### 8.3.2 Adaptation of Peptides by Single Mutations

Two ACE2-based peptide structures, shown in Figure 22a, are selected as templates for the first generation peptide inhibitors of the S protein.[176] The smaller template-1 includes single truncated $\alpha 1$ helix of ACE2 (amino acids 21–43), and the larger template-2 includes two $\alpha 1 \alpha 2$ helices of ACE2 (amino acids 19–83). In the adaptive evolution search for optimized therapeutics, the selected ACE2-extracted peptide templates will be gradually mutated to increase their binding strength to RBD.

In recent mutagenesis experiments, the whole ACE2 with single mutations in regions directly contacting RBD were examined for their binding to the original S protein.[194] To perform preliminary testing of our adaptive evolution search of peptide therapeutics, we first selected the most fit mutants from these experiments and implemented their mutations in our templates- 1,2.We simulated 22 peptides, that is, the original templates and their 10 single mutants, complexed with the original S protein RBD. Their free energies of binding, $\Delta G_{MMGB-SA}$, were evaluated in 100 ns simulations and presented in Figure 22b. Template-1 binds to RBD with $\Delta G_{MMGB-SA} \approx -19$ kcal mol$^{-1}$, while its mutants have higher affinities giving $\Delta G_{MMGB-SA} \approx -(24–35)$ kcal mol$^{-1}$. In all cases, template-1 significantly changes its conformation in the bound configuration, as the helix loses the curvature observed when within ACE2, and the hydrogen bonding between Glu35 (template-1) and Gln493 (RBD), enabled by the helix curvature, is broken. In contrast, template-2 has more direct contacts with RBD than the shorter template-1 variants, so it binds to it more strongly, $\Delta G_{MMGB-SA} \approx -36$ kcal mol$^{-1}$. However, only two template-2 mutants (H34A and K31W) have higher affinities to RBD than the original template-2, having $\Delta G_{MMGB-SA} \approx -45$ kcal mol$^{-1}$. These simulations also revealed that peptides with a stronger binding covered larger RBD sections and reduced the RBD exposure to other potential binding partners. These results show that the experimental results obtained formulated ACE2[194] can provide a good guidance in the mutation of template-1, but the same mutations are less effective in the larger template-2.



| amino acid | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| initial sequence | ILE | GLU | GLU | GLN | ALA | LYS | THR | PHE | LEU | ASP | LYS | PHE | ASN | HIS | GLU | ALA | GLU | ASP | LEU | PHE | TYR | GLN | SER |
| final sequences | | | | | | | | | | | | | | | | | | | | | | | |
| original RBD, 10 ns | VAL | PRO | | | LEU | TYR | CYS | LEU | | GLN | | | | GLY | | | | MET | | | SER | ARG | |
| A475V RBD, 10 ns | THR | | THR | | | | | | | | | | | | SER | | | | | | | | |
| G476S RBD, 10 ns | | PHE | | | | | | | TRP | ALA | GLN | | | | LYS | | | | | | | | |
| S477I RBD, 10 ns | ASP | VAL | | | | | ARG | TYR | ILE | GLU | | ASN | | | TYR | | | | GLU | ARG | | LYS | SER |

**Figure 22.** Modeling of peptide-RBD complexes. a) Complexes of S protein RBD (blue) and two peptide templates (red). Locations of five S protein mutations examined in the present work are marked by blue spheres. Amino acid residues changed in singly mutated peptides are marked by yellow spheres. b) Free energies of binding, ΔGMMGB-SA, between the originally reported S protein RBD and the wild type or singly mutated ACE2-based peptides. Locations of mutated peptide amino acids are highlighted in panel a. c) Snapshots of initial and optimized template-1 peptides binding to the original RBD. The sequence of the optimized peptide was obtained after 100 mutation attempts, with 10 ns long MD simulation after each mutation. The final peptide with the optimized sequence was further relaxed in a 175 ns MD simulation. The initial peptide is shown as a red helix, with amino acids that are subsequently mutated shown in thin faded yellow licorice.

The optimized peptide is shown as an orange helix, with mutated amino acids shown in thick yellow licorice. d) Adaptive evolution of template-1. The plot shows the binding free energies, ΔGMMGB-SA, between the peptide and the original RBD, presented as a function of the performed mutation, where the peptide:RBD complexes are relaxed for 10 ns after each mutation attempt. e) The time evolution of ΔGMMGB-SA between the final optimized peptide and the original RBD. The average value, obtained from the last 75 ns of the trajectory (gray), is ΔGMMGB-SA = −57 kcal mol−1. The faded green line shows the data points calculated every 0.1 ns, and the dark green line shows the running average. f) Initial and optimized sequences of template-1 peptides. The final peptides were optimized for binding to the original and mutant RBDs, with peptide-RBD complexes relaxed in 10 ns MD simulations after each attempted mutation.

### 8.3.3 Adaptive Evolution of Peptide Inhibitors

The above results have clearly demonstrated that suitable peptide templates with appropriate mutations can acquire a strong binding to specific targets. To optimize such peptides against specific viral strains, we have developed combined mutation/selection (evolution) computational algorithms which can guide a multi-step adaptive evolution of the peptides: 1) Random mutations are introduced into random positions of the peptide templates. 2) Short MD simulation trajectory of the mutated-peptide:RBD complex are run and followed by a selection or rejection of the mutation via Monte Carlo (MC) sampling using a Metropolis criterion applied to the change of the free energy of peptide-RBD binding, $\Delta G_{MMGB-SA}$. 3) The mutation/selection process is iteratively repeated until the binding affinity of peptides to the target S protein RBD is satisfactory. 4) Additional evolution of the molecules might be considered after the MC decisions to allow for a better internal relaxation of the molecules. Due to a partly stochastic nature of MD simulations, the randomness in mutations, and the MC selection, different peptides can be obtained in separate trajectories that correspond to separate local minima of the free energy surface. These peptides form a pool (ensemble) of potential therapeutics evolved for a selected viral strain, which can be further enriched by considering multiple viral strains.

In Figures 22c–e, coupling of template-1 to RBD was optimized in the adaptive evolution, where 10 ns MD simulation trajectories were generated after each trial mutation of template-1. Of the 100 mutations attempted, 13 mutations were accepted, and 11 amino acids were changed (individual residues can be mutated more than once). Figure 22d reveals the progression of binding free energies with the mutations of template-1, starting from $\Delta G_{MMGB-SA}= -19$ kcal mol$^{-1}$. As detailed in Figure 22c, during the adaptive evolution, the mutating helical peptide lost its bending (this change is independent of mutations) and multiple initial contacts with RBD. At the same time, it shifted with respect to its initial position and formed many new contacts. Peptide residue E37 formed a salt bridge with the original RBD, while residues Q24, Y26 (mutated), Q30 (mutated), S41 (mutated), and R42 (mutated) formed hydrogen bonds of varying stability with the original RBD. The resulting peptide bound to RBD with $\Delta G_{MMGB-SA}= -70$ kcal mol$^{-1}$ at the end of thirteen 10 ns long trajectories (associated with individual accepted mutations).

As shown in Figure 22e, additional 175 ns relaxation of this peptide resulted in a slightly less favorable $\Delta G_{MMGB-SA}\approx -57$ kcal mol$^{-1}$. Therefore, adaptive evolution requires sufficiently long relaxation times for a good stabilization of the whole system. Short relaxation times may result in incomplete peptide adjustments and free energies that can be misleadingly favorable. Moreover, a faster MC convergence could be achieved by considering the whole free energy changes rather than the peptide-RBD binding free energies. However, internal reorganizations of molecules inevitably take part in long trajectories, so the difference in binding energies alone might be sufficient for the MC decision, as long as additional relaxation is allowed between individual MC steps.

Next, we adaptively evolved template-1 coupled to three separate singly-mutated RBDs, chosen from Figure 21d. For simplicity, 100 mutations were attempted, followed by 10 ns simulations after each attempt. The adaptive evolution gave peptides with $\Delta G_{MMGB-SA}\approx -(45$ to $70)$ kcal mol$^{-1}$, as summarized in Figure 23a–c and Figure 22f. Peptides targeting A475V and

G476S RBDs each had five accepted mutations, respectively, while peptides targeting S477I RBD had 19 accepted mutations. In the A475V RBD case, one of the early accepted mutations lead to breaking of the helix secondary structure, and thus to a different peptide-RBD binding mode. This shows that individual alpha helices without additional stabilization, such as by side branching, might be too simplistic therapeutics.

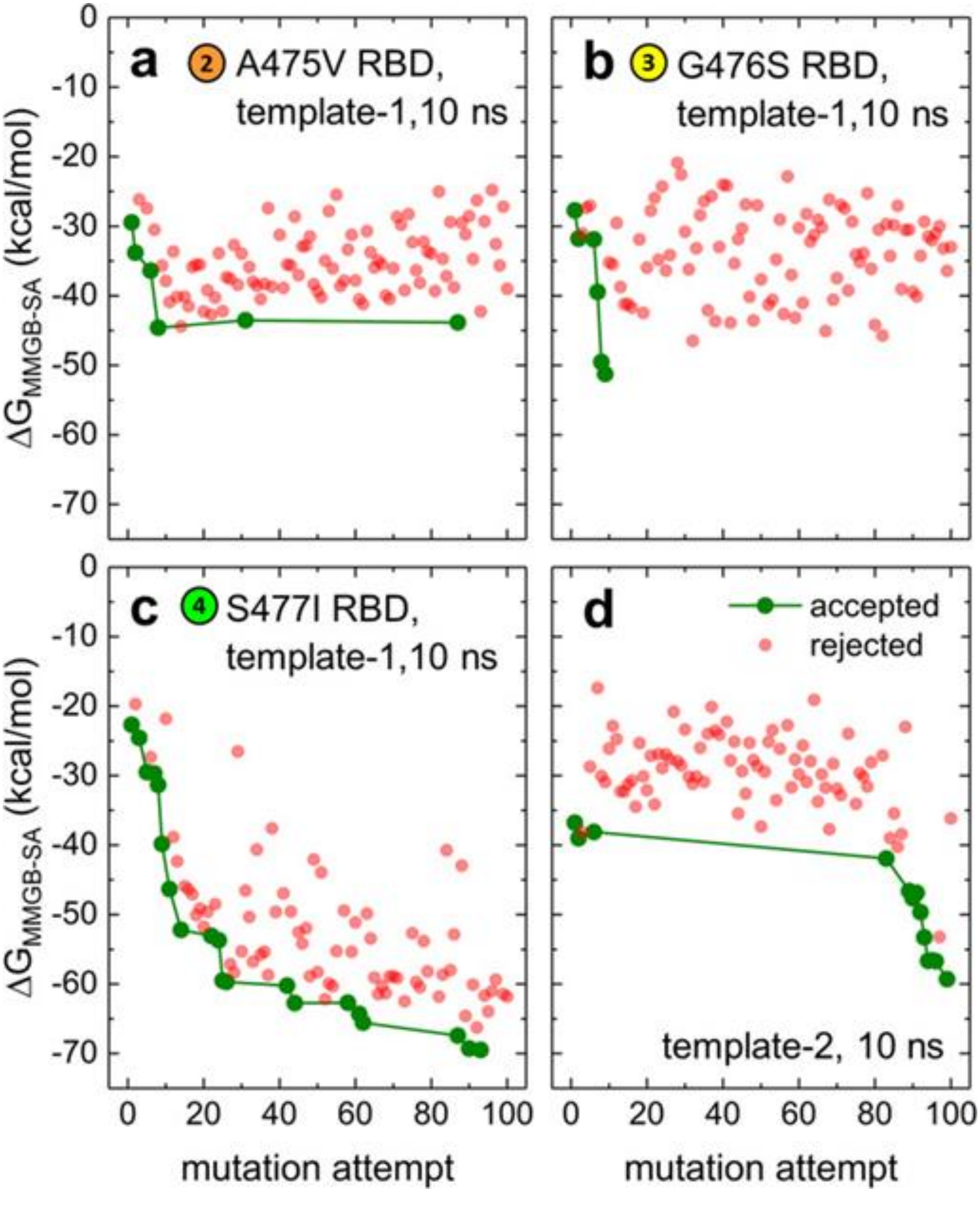


**Figure 23.** a–c) Adaptive evolution of template-1 coupled with singly mutated RBDs. d) Adaptive evolution of template-2 coupled with the original RBD. After attempted mutations, peptide:RBD complexes were relaxed in 10 ns simulation steps.

In Figure 23d, the adaptive evolution was performed with a more stable template-2 ($\alpha$1 $\alpha$2), but random mutations were only introduced into the $\alpha$1 helix, which was in direct contact with the original RBD. After 12 accepted mutations and 10 changed amino acids, the binding strength increased from $\Delta G_{MMGB-SA} = -36$ kcal mol$^{-1}$ to $-60$ kcal mol$^{-1}$. Therefore, the adaptively evolved template-2 can compete with the whole ACE2, having $\Delta G_{MMGB-SA} \approx -50$ kcal mol$^{-1}$ (Figure 1c). Both the initial and optimized template-2 preserve the curvature of the $\alpha$1 helix, despite the mutation of E35 (to Y35), which is observed to interact with Q493 (RBD). This feature preserves the binding pattern observed in ACE2:S protein RBD complex.[170] However, the salt bridge between D30 (initial peptide) and K417 (RBD) is lost in the peptide optimized with 10 ns simulations after mutations.

The work described was performed together with Yanxiao Han, where we shared the work associated with running and analyzing the computational work, in order to accelerate the completion of the project and its publication during the COVID-19 pandemic.

## 7.3 CONCLUSION

In this work, we have demonstrated that ACE2-based peptide templates can be adaptively evolved by computation using mutation/selection processes to form optimized inhibitors for a strong and competitive S protein RBD binding. The developed approach can be used to design peptide inhibitors based on templates extracted from different ACE2 polymorphs, including those from other species, and other proteins binding to viral pathogens. The optimized inhibitors obtained in different evolution runs can be collected to form libraries of suitable therapeutics for different RBD variants. Cocktails (ensembles) of peptide therapeutics could be delivered by different means to provide a broad-spectrum protection against different SARS-CoV-2 strains.

# References

(1) Hospital, A.; Goñi, J. R.; Orozco, M.; Gelpí, J. L. Molecular Dynamics Simulations: Advances and Applications. *Adv. Appl. Bioinforma. Chem. AABC* **2015**, *8*, 37–47. https://doi.org/10.2147/AABC.S70333.

(2) Berman, H. M.; Battistuz, T.; Bhat, T. N.; Bluhm, W. F.; Bourne, P. E.; Burkhardt, K.; Feng, Z.; Gilliland, G. L.; Iype, L.; Jain, S.; Fagan, P.; Marvin, J.; Padilla, D.; Ravichandran, V.; Schneider, B.; Thanki, N.; Weissig, H.; Westbrook, J.; Zardecki, C. The Protein Data Bank. *Acta Crystallogr. D Biol. Crystallogr.* **2002**, *58* (6), 899–907. https://doi.org/10.1107/S0907444902003451.

(3) Hollingsworth, S. A.; Dror, R. O. Molecular Dynamics Simulation for All. *Neuron* **2018**, *99* (6), 1129–1143. https://doi.org/10.1016/j.neuron.2018.08.011.

(4) Karplus, M.; McCammon, J. A. Molecular Dynamics Simulations of Biomolecules. *Nat. Struct. Biol.* **2002**, *9* (9), 646–652. https://doi.org/10.1038/nsb0902-646.

(5) Lopes, P. E. M.; Guvench, O.; MacKerell, A. D. Current Status of Protein Force Fields for Molecular Dynamics. *Methods Mol. Biol. Clifton NJ* **2015**, *1215*, 47–71. https://doi.org/10.1007/978-1-4939-1465-4_3.

(6) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. CHARMM General Force Field: A Force Field for Drug-like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2010**, *31* (4), 671–690. https://doi.org/10.1002/jcc.21367.

(7) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25* (9), 1157–1174. https://doi.org/10.1002/jcc.20035.

(8) Souza, P. C. T.; Alessandri, R.; Barnoud, J.; Thallmair, S.; Faustino, I.; Grünewald, F.; Patmanidis, I.; Abdizadeh, H.; Bruininks, B. M. H.; Wassenaar, T. A.; Kroon, P. C.; Melcr, J.; Nieto, V.; Corradi, V.; Khan, H. M.; Domański, J.; Javanainen, M.; Martinez-Seara, H.; Reuter, N.; Best, R. B.; Vattulainen, I.; Monticelli, L.; Periole, X.; Tieleman, D. P.; de Vries, A. H.; Marrink, S. J. Martini 3: A General Purpose Force Field for Coarse-Grained Molecular Dynamics. *Nat. Methods* **2021**, *18* (4), 382–388. https://doi.org/10.1038/s41592-021-01098-3.

(9) Callaway, E. 'It Opens up a Whole New Universe': Revolutionary Microscopy Technique Sees Individual Atoms for First Time. *Nature* **2020**, *582* (7811), 156–157. https://doi.org/10.1038/d41586-020-01658-1.

(10) Callaway, E. Revolutionary Cryo-EM Is Taking over Structural Biology. *Nature* **2020**, *578* (7794), 201–201. https://doi.org/10.1038/d41586-020-00341-9.

(11) Liu, Y.; Huynh, D. T.; Yeates, T. O. A 3.8 Å Resolution Cryo-EM Structure of a Small Protein Bound to an Imaging Scaffold. *Nat. Commun.* **2019**, *10* (1), 1864. https://doi.org/10.1038/s41467-019-09836-0.

(12) Howard, M. J. Protein NMR Spectroscopy. *Curr. Biol.* **1998**, *8* (10), R331–R333. https://doi.org/10.1016/S0960-9822(98)70214-3.

(13) Friedrichs, M. S.; Eastman, P.; Vaidyanathan, V.; Houston, M.; Legrand, S.; Beberg, A. L.; Ensign, D. L.; Bruns, C. M.; Pande, V. S. Accelerating Molecular Dynamic Simulation on Graphics Processing Units. *J. Comput. Chem.* **2009**, *30* (6), 864–872. https://doi.org/10.1002/jcc.21209.

(14) Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald | Journal of Chemical Theory and Computation https://pubs.acs.org/doi/10.1021/ct400314y (accessed 2021 -04 -09).

(15) Stone, J. E.; Hallock, M. J.; Phillips, J. C.; Peterson, J. R.; Luthey-Schulten, Z.; Schulten, K. Evaluation of Emerging Energy-Efficient Heterogeneous Computing Platforms for Biomolecular and Cellular Simulation Workloads. *IEEE Int. Symp. Parallel Distrib. Process. Workshop Phd Forum Proc. IEEE Int. Symp. Parallel Distrib. Process. Workshop Phd Forum* **2016**, *2016*, 89–100. https://doi.org/10.1109/IPDPSW.2016.130.

(16) Structural flexibility of the nucleosome core particle at atomic resolution studied by molecular dynamics simulation - Roccatano - 2007 - Biopolymers - Wiley Online Library https://onlinelibrary.wiley.com/doi/full/10.1002/bip.20690 (accessed 2021 -04 -09).

(17) Multiscale Modeling of Nucleosome Dynamics - ScienceDirect https://www.sciencedirect.com/science/article/pii/S0006349507709566 (accessed 2021 -04 -09).

(18) Zhao, G.; Perilla, J. R.; Yufenyuy, E. L.; Meng, X.; Chen, B.; Ning, J.; Ahn, J.; Gronenborn, A. M.; Schulten, K.; Aiken, C.; Zhang, P. Mature HIV-1 Capsid Structure by Cryo-Electron Microscopy and All-Atom Molecular Dynamics. *Nature* **2013**, *497* (7451), 643–646. https://doi.org/10.1038/nature12162.

(19) Perilla, J. R.; Goh, B. C.; Cassidy, C. K.; Liu, B.; Bernardi, R. C.; Rudack, T.; Yu, H.; Wu, Z.; Schulten, K. Molecular Dynamics Simulations of Large Macromolecular Complexes. *Curr. Opin. Struct. Biol.* **2015**, *31*, 64–74. https://doi.org/10.1016/j.sbi.2015.03.007.

(20) Tinoco, I.; Wen, J.-D. Simulation and Analysis of Single-Ribosome Translation. *Phys. Biol.* **2009**, *6* (2), 025006. https://doi.org/10.1088/1478-3975/6/2/025006.

(21) Brandman, R.; Brandman, Y.; Pande, V. S. A-Site Residues Move Independently from P-Site Residues in All-Atom Molecular Dynamics Simulations of the 70S Bacterial Ribosome. *PloS One* **2012**, *7* (1), e29377. https://doi.org/10.1371/journal.pone.0029377.

(22) Shaw, D. E. Millisecond-Long Molecular Dynamics Simulations of Proteins on a Special-Purpose Machine. *Biophys. J.* **2013**, *104* (2, Supplement 1), 45a. https://doi.org/10.1016/j.bpj.2012.11.289.

(23) Singharoy, A.; Maffeo, C.; Delgado-Magnero, K. H.; Swainsbury, D. J. K.; Sener, M.; Kleinekathöfer, U.; Vant, J. W.; Nguyen, J.; Hitchcock, A.; Isralewitz, B.; Teo, I.; Chandler, D. E.; Stone, J. E.; Phillips, J. C.; Pogorelov, T. V.; Mallus, M. I.; Chipot, C.; Luthey-Schulten, Z.; Tieleman, D. P.; Hunter, C. N.; Tajkhorshid, E.; Aksimentiev, A.; Schulten, K. Atoms to Phenotypes: Molecular Design Principles of Cellular Energy Metabolism. *Cell* **2019**, *179* (5), 1098-1111.e23. https://doi.org/10.1016/j.cell.2019.10.021.

(24) Joshi, S. Y.; Deshmukh, S. A. A Review of Advancements in Coarse-Grained Molecular Dynamics Simulations. *Mol. Simul.* **2020**, *0* (0), 1–18. https://doi.org/10.1080/08927022.2020.1828583.

(25) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26* (16), 1781–1802. https://doi.org/10.1002/jcc.20289.

(26) Spoel, D. V. D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26* (16), 1701–1718. https://doi.org/10.1002/jcc.20291.

(27) Huang, J.; MacKerell, A. D. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *J. Comput. Chem.* **2013**, *34* (25), 2135–2145. https://doi.org/10.1002/jcc.23354.

(28) An Improved Empirical Potential Energy Function for Molecular Simulations of Phospholipids | The Journal of Physical Chemistry B https://pubs.acs.org/doi/abs/10.1021/jp0007843 (accessed 2021 -04 -09).

(29) May, A.; Pool, R.; van Dijk, E.; Bijlard, J.; Abeln, S.; Heringa, J.; Feenstra, K. A. Coarse-Grained versus Atomistic Simulations: Realistic Interaction Free Energies for Real Proteins. *Bioinforma. Oxf. Engl.* **2014**, *30* (3), 326–334. https://doi.org/10.1093/bioinformatics/btt675.

(30) The MARTINI Force Field:  Coarse Grained Model for Biomolecular Simulations | The Journal of Physical Chemistry B https://pubs.acs.org/doi/10.1021/jp071097f (accessed 2021 -04 -09).

(31) Martini Coarse-Grained Force Field: Extension to DNA | Journal of Chemical Theory and Computation https://pubs.acs.org/doi/abs/10.1021/acs.jctc.5b00286 (accessed 2021 -04 -09).

(32) Quigley, D.; Probert, M. I. J. Langevin Dynamics in Constant Pressure Extended Systems. *J. Chem. Phys.* **2004**, *120* (24), 11432–11441. https://doi.org/10.1063/1.1755657.

(33) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089–10092. https://doi.org/10.1063/1.464397.

(34) Price, D. J.; Brooks, C. L. A Modified TIP3P Water Potential for Simulation with Ewald Summation. *J. Chem. Phys.* **2004**, *121* (20), 10096–10103. https://doi.org/10.1063/1.1808117.

(35) Levy, S. E.; Myers, R. M. Advancements in Next-Generation Sequencing. *Annu. Rev. Genomics Hum. Genet.* **2016**, *17* (1), 95–115. https://doi.org/10.1146/annurev-genom-083115-022413.

(36) Full article: Developing mRNA-vaccine technologies https://www.tandfonline.com/doi/full/10.4161/rna.22269 (accessed 2021 -04 -09).

(37) Rosario, K.; Duffy, S.; Breitbart, M. A Field Guide to Eukaryotic Circular Single-Stranded DNA Viruses: Insights Gained from Metagenomics. *Arch. Virol.* **2012**, *157* (10), 1851–1871. https://doi.org/10.1007/s00705-012-1391-y.

(38) Sanger, H. L.; Klotz, G.; Riesner, D.; Gross, H. J.; Kleinschmidt, A. K. Viroids Are Single-Stranded Covalently Closed Circular RNA Molecules Existing as Highly Base-Paired Rod-like Structures. *Proc. Natl. Acad. Sci. U. S. A.* **1976**, *73* (11), 3852–3856. https://doi.org/10.1073/pnas.73.11.3852.

(39) Han, D.; Qi, X.; Myhrvold, C.; Wang, B.; Dai, M.; Jiang, S.; Bates, M.; Liu, Y.; An, B.; Zhang, F.; Yan, H.; Yin, P. Single-Stranded DNA and RNA Origami. *Science* **2017**, *358* (6369). https://doi.org/10.1126/science.aao2648.

(40) Chen, Z.; He, Y.; Luo, S.; Lin, H.; Chen, Y.; Sheng, P.; Li, J.; Chen, B.; Liu, C.; Cai, Q. Label-Free Colorimetric Assay for Biological Thiols Based on SsDNA/Silver Nanoparticle System by Salt Amplification. *Analyst* **2010**, *135* (5), 1066–1069. https://doi.org/10.1039/B925683K.

(41) Beyene, A. G.; Alizadehmojarad, A. A.; Dorlhiac, G.; Goh, N.; Streets, A. M.; Král, P.; Vuković, L.; Landry, M. P. Ultralarge Modulation of Fluorescence by Neuromodulators in

Carbon Nanotubes Functionalized with Self-Assembled Oligonucleotide Rings. *Nano Lett.* **2018**, *18* (11), 6995–7003. https://doi.org/10.1021/acs.nanolett.8b02937.

(42) Memczak, S.; Jens, M.; Elefsinioti, A.; Torti, F.; Krueger, J.; Rybak, A.; Maier, L.; Mackowiak, S. D.; Gregersen, L. H.; Munschauer, M.; Loewer, A.; Ziebold, U.; Landthaler, M.; Kocks, C.; le Noble, F.; Rajewsky, N. Circular RNAs Are a Large Class of Animal RNAs with Regulatory Potency. *Nature* **2013**, *495* (7441), 333–338. https://doi.org/10.1038/nature11928.

(43) Lasda, E.; Parker, R. Circular RNAs: Diversity of Form and Function. *RNA N. Y. N* **2014**, *20* (12), 1829–1842. https://doi.org/10.1261/rna.047126.114.

(44) Soma, A. Circularly Permuted TRNA Genes: Their Expression and Implications for Their Physiological Relevance and Development. *Front. Genet.* **2014**, *5*. https://doi.org/10.3389/fgene.2014.00063.

(45) Denning, E. J.; Priyakumar, U. D.; Nilsson, L.; Mackerell, A. D. Impact of 2'-Hydroxyl Sampling on the Conformational Properties of RNA: Update of the CHARMM All-Atom Additive Force Field for RNA. *J. Comput. Chem.* **2011**, *32* (9), 1929–1943. https://doi.org/10.1002/jcc.21777.

(46) Ponder, J. W.; Case, D. A. Force Fields for Protein Simulations. *Adv. Protein Chem.* **2003**, *66*, 27–85. https://doi.org/10.1016/s0065-3233(03)66002-x.

(47) Holdt, L. M.; Kohlmaier, A.; Teupser, D. Circular RNAs as Therapeutic Agents and Targets. *Front. Physiol.* **2018**, *9*. https://doi.org/10.3389/fphys.2018.01262.

(48) Salmena, L.; Poliseno, L.; Tay, Y.; Kats, L.; Pandolfi, P. P. A CeRNA Hypothesis: The Rosetta Stone of a Hidden RNA Language? *Cell* **2011**, *146* (3), 353–358. https://doi.org/10.1016/j.cell.2011.07.014.

(49) Tay, Y.; Kats, L.; Salmena, L.; Weiss, D.; Tan, S. M.; Ala, U.; Karreth, F.; Poliseno, L.; Provero, P.; Di Cunto, F.; Lieberman, J.; Rigoutsos, I.; Pandolfi, P. P. Coding-Independent Regulation of the Tumor Suppressor PTEN by Competing Endogenous MRNAs. *Cell* **2011**, *147* (2), 344–357. https://doi.org/10.1016/j.cell.2011.09.029.

(50) Synthesis, Structure, and Biological Activity of Dumbbell-Shaped Nanocircular RNAs for RNA Interference | Bioconjugate Chemistry https://pubs.acs.org/doi/abs/10.1021/bc2003154 (accessed 2021 -04 -09).

(51) Egli, M.; Gessner, R. V.; Williams, L. D.; Quigley, G. J.; van der Marel, G. A.; van Boom, J. H.; Rich, A.; Frederick, C. A. Atomic-Resolution Structure of the Cellulose Synthase Regulator Cyclic Diguanylic Acid. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87* (8), 3235–3239.

(52) Jeck, W. R.; Sharpless, N. E. Detecting and Characterizing Circular RNAs. *Nat. Biotechnol.* **2014**, *32* (5), 453–461. https://doi.org/10.1038/nbt.2890.

(53) Abe, N.; Matsumoto, K.; Nishihara, M.; Nakano, Y.; Shibata, A.; Maruyama, H.; Shuto, S.; Matsuda, A.; Yoshida, M.; Ito, Y.; Abe, H. Rolling Circle Translation of Circular RNA in Living Human Cells. *Sci. Rep.* **2015**, *5* (1), 16435. https://doi.org/10.1038/srep16435.

(54) Le Bon, A.; Etchart, N.; Rossmann, C.; Ashton, M.; Hou, S.; Gewert, D.; Borrow, P.; Tough, D. F. Cross-Priming of CD8+ T Cells Stimulated by Virus-Induced Type I Interferon. *Nat. Immunol.* **2003**, *4* (10), 1009–1015. https://doi.org/10.1038/ni978.

(55) Hochheiser, K.; Klein, M.; Gottschalk, C.; Hoss, F.; Scheu, S.; Coch, C.; Hartmann, G.; Kurts, C. Cutting Edge: The RIG-I Ligand 3pRNA Potently Improves CTL Cross-Priming and Facilitates Antiviral Vaccination. *J. Immunol.* **2016**, *196* (6), 2439–2443. https://doi.org/10.4049/jimmunol.1501958.

(56) Hentze, M. W.; Preiss, T. Circular RNAs: Splicing's Enigma Variations. *EMBO J.* **2013**, *32* (7), 923–925. https://doi.org/10.1038/emboj.2013.53.

(57) Da, D. G.; Sm, K.; Y, L.; Gd, T.; Mt, A.; Gc, K. Multitasking by Multivalent Circular DNA Aptamers. *Chembiochem Eur. J. Chem. Biol.* **2006**, *7* (3), 535–544. https://doi.org/10.1002/cbic.200500316.

(58) Enuka, Y.; Lauriola, M.; Feldman, M. E.; Sas-Chen, A.; Ulitsky, I.; Yarden, Y. Circular RNAs Are Long-Lived and Display Only Minimal Early Alterations in Response to a Growth Factor. *Nucleic Acids Res.* **2016**, *44* (3), 1370–1383. https://doi.org/10.1093/nar/gkv1367.

(59) Jeck, W. R.; Sorrentino, J. A.; Wang, K.; Slevin, M. K.; Burd, C. E.; Liu, J.; Marzluff, W. F.; Sharpless, N. E. Circular RNAs Are Abundant, Conserved, and Associated with ALU Repeats. *RNA* **2013**, *19* (2), 141–157. https://doi.org/10.1261/rna.035667.112.

(60) Krishnan, Y.; Simmel, F. C. Nucleic Acid Based Molecular Devices. *Angew. Chem. Int. Ed Engl.* **2011**, *50* (14), 3124–3156. https://doi.org/10.1002/anie.200907223.

(61) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures | Nucleic Acids Research | Oxford Academic https://academic.oup.com/nar/article/31/17/5108/1192167 (accessed 2021 -04 -09).

(62) Lu, X.-J.; Olson, W. K. 3DNA: A Versatile, Integrated Software System for the Analysis, Rebuilding and Visualization of Three-Dimensional Nucleic-Acid Structures. *Nat. Protoc.* **2008**, *3* (7), 1213–1227. https://doi.org/10.1038/nprot.2008.104.

(63) Wriggers, W. Conventions and Workflows for Using Situs. *Acta Crystallogr. D Biol. Crystallogr.* **2012**, *68* (Pt 4), 344–351. https://doi.org/10.1107/S0907444911049791.

(64) Trabuco, L. G.; Villa, E.; Mitra, K.; Frank, J.; Schulten, K. Flexible Fitting of Atomic Structures into Electron Microscopy Maps Using Molecular Dynamics. *Struct. Lond. Engl. 1993* **2008**, *16* (5), 673–683. https://doi.org/10.1016/j.str.2008.03.005.

(65) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14* (1), 33–38. https://doi.org/10.1016/0263-7855(96)00018-5.

(66) Chen, H.; Meisburger, S. P.; Pabit, S. A.; Sutton, J. L.; Webb, W. W.; Pollack, L. Ionic Strength-Dependent Persistence Lengths of Single-Stranded RNA and DNA. *Proc. Natl. Acad. Sci.* **2012**, *109* (3), 799–804. https://doi.org/10.1073/pnas.1119057109.

(67) Salt Species-Dependent Electrostatic Effects on ssDNA Elasticity | Macromolecules https://pubs.acs.org/doi/abs/10.1021/ma1028196 (accessed 2021 -04 -09).

(68) Bao, L.; Zhang, X.; Jin, L.; Tan, Z.-J. Flexibility of Nucleic Acids: From DNA to RNA. *Chin. Phys. B* **2016**, *25* (1), 018703. https://doi.org/10.1088/1674-1056/25/1/018703.

(69) Influence of Sodium Ions on the Dynamics and Structure of Single-Stranded DNA Oligomers:  A Molecular Dynamics Study | Journal of the American Chemical Society https://pubs.acs.org/doi/10.1021/ja0108786 (accessed 2021 -04 -09).

(70) Chakraborty, K.; Khatua, P.; Bandyopadhyay, S. Exploring Ion Induced Folding of a Single-Stranded DNA Oligomer from Molecular Simulation Studies. *Phys. Chem. Chem. Phys.* **2016**, *18* (23), 15899–15910. https://doi.org/10.1039/C6CP00663A.

(71) Misra, V. K.; Shiman, R.; Draper, D. E. A Thermodynamic Framework for the Magnesium-Dependent Folding of RNA. *Biopolymers* **2003**, *69* (1), 118–136. https://doi.org/10.1002/bip.10353.

(72) Controlling the Growth of Charged-Nanoparticle Chains through Interparticle Electrostatic Repulsion - Zhang - 2008 - Angewandte Chemie International Edition - Wiley Online

Library https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.200705537 (accessed 2021 -04 -09).
(73) Woodson, S. A. Metal Ions and RNA Folding: A Highly Charged Topic with a Dynamic Future. *Curr. Opin. Chem. Biol.* **2005**, *9* (2), 104–109. https://doi.org/10.1016/j.cbpa.2005.02.004.
(74) Monovalent ions modulate the flux through multiple folding pathways of an RNA pseudoknot | PNAS https://www.pnas.org/content/115/31/E7313 (accessed 2021 -04 -09).
(75) Ren, J.; Wang, Y.; Yao, Y.; Wang, Y.; Fei, X.; Qi, P.; Lin, S.; Kaplan, D. L.; Buehler, M. J.; Ling, S. Biological Material Interfaces as Inspiration for Mechanical and Optical Material Designs. *Chem. Rev.* **2019**, *119* (24), 12279–12336. https://doi.org/10.1021/acs.chemrev.9b00416.
(76) Thang, D. C.; Wang, Z.; Lu, X.; Xing, B. Precise Cell Behaviors Manipulation through Light-Responsive Nano-Regulators: Recent Advance and Perspective. *Theranostics* **2019**, *9* (11), 3308–3340. https://doi.org/10.7150/thno.33888.
(77) Gold nanoparticle-mediated photothermal therapy: applications and opportunities for multimodal cancer treatment - Riley - 2017 - WIREs Nanomedicine and Nanobiotechnology - Wiley Online Library https://onlinelibrary.wiley.com/doi/full/10.1002/wnan.1449 (accessed 2021 -04 -10).
(78) A review on hyperthermia via nanoparticle-mediated therapy - ScienceDirect https://www.sciencedirect.com/science/article/abs/pii/S0007455117300577 (accessed 2021 -04 -10).
(79) Recent Advances of Plasmonic Gold Nanoparticles in Optical Sensin...: Ingenta Connect https://www.ingentaconnect.com/content/ben/cpd/2019/00000025/00000046/art00004 (accessed 2021 -04 -10).
(80) Femtosecond Laser Pulse Excitation of DNA-Labeled Gold Nanoparticles: Establishing a Quantitative Local Nanothermometer for Biological Applications | ACS Nano https://pubs.acs.org/doi/abs/10.1021/acsnano.0c02899 (accessed 2021 -04 -10).
(81) Plasmonic Heating of Nanostructures | Chemical Reviews https://pubs.acs.org/doi/abs/10.1021/acs.chemrev.8b00738 (accessed 2021 -04 -10).
(82) Duncan, B.; Kim, C.; Rotello, V. M. Gold Nanoparticle Platforms as Drug and Biomacromolecule Delivery Systems. *J. Control. Release Off. J. Control. Release Soc.* **2010**, *148* (1), 122–127. https://doi.org/10.1016/j.jconrel.2010.06.004.
(83) Fratila, R. M.; Mitchell, S. G.; del Pino, P.; Grazu, V.; de la Fuente, J. M. Strategies for the Biofunctionalization of Gold and Iron Oxide Nanoparticles. *Langmuir* **2014**, *30* (50), 15057–15071. https://doi.org/10.1021/la5015658.
(84) El-Sayed, I. H.; Huang, X.; El-Sayed, M. A. Selective Laser Photo-Thermal Therapy of Epithelial Carcinoma Using Anti-EGFR Antibody Conjugated Gold Nanoparticles. *Cancer Lett.* **2006**, *239* (1), 129–135. https://doi.org/10.1016/j.canlet.2005.07.035.
(85) Gold nanoparticles: Optical properties and implementations in cancer diagnosis and photothermal therapy - ScienceDirect https://www.sciencedirect.com/science/article/pii/S2090123210000056 (accessed 2021 -04 -11).
(86) Paviolo, C.; Haycock, J. W.; Cadusch, P. J.; McArthur, S. L.; Stoddart, P. R. Laser Exposure of Gold Nanorods Can Induce Intracellular Calcium Transients. *J. Biophotonics* **2014**, *7* (10), 761–765. https://doi.org/10.1002/jbio.201300043.

(87) Carvalho-de-Souza, J. L.; Treger, J. S.; Dang, B.; Kent, S. B. H.; Pepperberg, D. R.; Bezanilla, F. Photosensitivity of Neurons Enabled by Cell-Targeted Gold Nanoparticles. *Neuron* **2015**, *86* (1), 207–217. https://doi.org/10.1016/j.neuron.2015.02.033.

(88) Quantification of Loading and Laser-Assisted Release of RNA from Single Gold Nanoparticles | Langmuir https://pubs.acs.org/doi/10.1021/acs.langmuir.8b01831 (accessed 2021 -04 -11).

(89) Goodman, A. M.; Hogan, N. J.; Gottheim, S.; Li, C.; Clare, S. E.; Halas, N. J. Understanding Resonant Light-Triggered DNA Release from Plasmonic Nanoparticles. *ACS Nano* **2017**, *11* (1), 171–179. https://doi.org/10.1021/acsnano.6b06510.

(90) Baffou, G.; Cichos, F.; Quidant, R. Applications and Challenges of Thermoplasmonics. *Nat. Mater.* **2020**, *19* (9), 946–958. https://doi.org/10.1038/s41563-020-0740-6.

(91) Baffou, G.; Rigneault, H. Femtosecond-Pulsed Optical Heating of Gold Nanoparticles. *Phys. Rev. B* **2011**, *84* (3), 035415. https://doi.org/10.1103/PhysRevB.84.035415.

(92) Hastman, D. A.; Melinger, J. S.; Aragonés, G. L.; Cunningham, P. D.; Chiriboga, M.; Salvato, Z. J.; Salvato, T. M.; Brown, C. W.; Mathur, D.; Medintz, I. L.; Oh, E.; Díaz, S. A. Femtosecond Laser Pulse Excitation of DNA-Labeled Gold Nanoparticles: Establishing a Quantitative Local Nanothermometer for Biological Applications. *ACS Nano* **2020**, *14* (7), 8570–8583. https://doi.org/10.1021/acsnano.0c02899.

(93) Understanding Resonant Light-Triggered DNA Release from Plasmonic Nanoparticles | ACS Nano https://pubs.acs.org/doi/abs/10.1021/acsnano.6b06510 (accessed 2021 -04 -11).

(94) Riley, R. S.; Dang, M. N.; Billingsley, M. M.; Abraham, B.; Gundlach, L.; Day, E. S. Evaluating the Mechanisms of Light-Triggered SiRNA Release from Nanoshells for Temporal Control Over Gene Regulation. *Nano Lett.* **2018**, *18* (6), 3565–3570. https://doi.org/10.1021/acs.nanolett.8b00681.

(95) Chen, C.; Wang, W.; Ge, J.; Zhao, X. S. Kinetics and Thermodynamics of DNA Hybridization on Gold Nanoparticles. *Nucleic Acids Res.* **2009**, *37* (11), 3756–3765. https://doi.org/10.1093/nar/gkp230.

(96) Ruano, M.; Díaz, M.; Martínez, L.; Navarro, E.; Román, E.; García-Hernandez, M.; Espinosa, A.; Ballesteros, C.; Fermento, R.; Huttel, Y. Matrix and Interaction Effects on the Magnetic Properties of Co Nanoparticles Embedded in Gold and Vanadium. *Phys. Chem. Chem. Phys.* **2012**, *15* (1), 316–329. https://doi.org/10.1039/C2CP42769A.

(97) Macrae, C. F.; Bruno, I. J.; Chisholm, J. A.; Edgington, P. R.; McCabe, P.; Pidcock, E.; Rodriguez-Monge, L.; Taylor, R.; Streek, J. van de; Wood, P. A. Mercury CSD 2.0 – New Features for the Visualization and Investigation of Crystal Structures. *J. Appl. Crystallogr.* **2008**, *41* (2), 466–470. https://doi.org/10.1107/S0021889807067908.

(98) Semichem, Inc. - The AMPAC People http://www.semichem.com/contact.php (accessed 2021 -04 -14).

(99) Fonseca Guerra, C.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. Hydrogen Bonding in DNA Base Pairs: Reconciliation of Theory and Experiment. *J. Am. Chem. Soc.* **2000**, *122* (17), 4117–4128. https://doi.org/10.1021/ja993262d.

(100) He, X.; Htoon, H.; Doorn, S. K.; Pernice, W. H. P.; Pyatkov, F.; Krupke, R.; Jeantet, A.; Chassagneux, Y.; Voisin, C. Carbon Nanotubes as Emerging Quantum-Light Sources. *Nat. Mater.* **2018**, *17* (8), 663–670. https://doi.org/10.1038/s41563-018-0109-2.

(101) Kruss, S.; Hilmer, A. J.; Zhang, J.; Reuel, N. F.; Mu, B.; Strano, M. S. Carbon Nanotubes as Optical Biomedical Sensors. *Adv. Drug Deliv. Rev.* **2013**, *65* (15), 1933–1950. https://doi.org/10.1016/j.addr.2013.07.015.

(102) Graf, A.; Held, M.; Zakharko, Y.; Tropf, L.; Gather, M. C.; Zaumseil, J. Electrical Pumping and Tuning of Exciton-Polaritons in Carbon Nanotube Microcavities. *Nat. Mater.* **2017**, *16* (9), 911–917. https://doi.org/10.1038/nmat4940.

(103) Hong, G.; Diao, S.; Antaris, A. L.; Dai, H. Carbon Nanomaterials for Biological Imaging and Nanomedicinal Therapy. *Chem. Rev.* **2015**, *115* (19), 10816–10906. https://doi.org/10.1021/acs.chemrev.5b00008.

(104) Singh, P.; Campidelli, S.; Giordani, S.; Bonifazi, D.; Bianco, A.; Prato, M. Organic Functionalisation and Characterisation of Single-Walled Carbon Nanotubes. *Chem. Soc. Rev.* **2009**, *38* (8), 2214–2230. https://doi.org/10.1039/B518111A.

(105) Vázquez, E.; Giacalone, F.; Prato, M. Non-Conventional Methods and Media for the Activation and Manipulation of Carbon Nanoforms. *Chem. Soc. Rev.* **2014**, *43* (1), 58–69. https://doi.org/10.1039/C3CS60164A.

(106) Deng, J.; Ghosh, K.; Wong, H.-P. Modeling Carbon Nanotube Sensors. *IEEE Sens. J.* **2007**, *7* (9), 1356–1357. https://doi.org/10.1109/JSEN.2007.905039.

(107) Zheng, M.; Jagota, A.; Strano, M. S.; Santos, A. P.; Barone, P.; Chou, S. G.; Diner, B. A.; Dresselhaus, M. S.; Mclean, R. S.; Onoa, G. B.; Samsonidze, G. G.; Semke, E. D.; Usrey, M.; Walls, D. J. Structure-Based Carbon Nanotube Sorting by Sequence-Dependent DNA Assembly. *Science* **2003**, *302* (5650), 1545–1548. https://doi.org/10.1126/science.1091911.

(108) Brunecker, F. K.; Schöppler, F.; Hertel, T. Interaction of Polymers with Single-Wall Carbon Nanotubes. *J. Phys. Chem. C* **2016**, *120* (18), 10094–10103. https://doi.org/10.1021/acs.jpcc.6b02198.

(109) Schöppler, F.; Mann, C.; Hain, T. C.; Neubauer, F. M.; Privitera, G.; Bonaccorso, F.; Chu, D.; Ferrari, A. C.; Hertel, T. Molar Extinction Coefficient of Single-Wall Carbon Nanotubes. *J. Phys. Chem. C* **2011**, *115* (30), 14682–14686. https://doi.org/10.1021/jp205289h.

(110) Jin, H.; Jeng, E. S.; Heller, D. A.; Jena, P. V.; Kirmse, R.; Langowski, J.; Strano, M. S. Divalent Ion and Thermally Induced DNA Conformational Polymorphism on Single-Walled Carbon Nanotubes. *Macromolecules* **2007**, *40* (18), 6731–6739. https://doi.org/10.1021/ma070608t.

(111) Bisker, G.; Ahn, J.; Kruss, S.; Ulissi, Z. W.; Salem, D. P.; Strano, M. S. A Mathematical Formulation and Solution of the CoPhMoRe Inverse Problem for Helically Wrapping Polymer Corona Phases on Cylindrical Substrates. *J. Phys. Chem. C* **2015**, *119* (24), 13876–13886. https://doi.org/10.1021/acs.jpcc.5b01705.

(112) Kruss, S.; Landry, M. P.; Vander Ende, E.; Lima, B. M. A.; Reuel, N. F.; Zhang, J.; Nelson, J.; Mu, B.; Hilmer, A.; Strano, M. Neurotransmitter Detection Using Corona Phase Molecular Recognition on Fluorescent Single-Walled Carbon Nanotube Sensors. *J. Am. Chem. Soc.* **2014**, *136* (2), 713–724. https://doi.org/10.1021/ja410433b.

(113) Bisker, G.; Dong, J.; Park, H. D.; Iverson, N. M.; Ahn, J.; Nelson, J. T.; Landry, M. P.; Kruss, S.; Strano, M. S. Protein-Targeted Corona Phase Molecular Recognition. *Nat. Commun.* **2016**, *7* (1), 10241. https://doi.org/10.1038/ncomms10241.

(114) Kruss, S.; Salem, D. P.; Vuković, L.; Lima, B.; Ende, E. V.; Boyden, E. S.; Strano, M. S. High-Resolution Imaging of Cellular Dopamine Efflux Using a Fluorescent Nanosensor Array. *Proc. Natl. Acad. Sci.* **2017**, *114* (8), 1789–1794. https://doi.org/10.1073/pnas.1613541114.

(115) Cognet, L.; Tsyboulski, D. A.; Rocha, J.-D. R.; Doyle, C. D.; Tour, J. M.; Weisman, R. B. Stepwise Quenching of Exciton Fluorescence in Carbon Nanotubes by Single-Molecule Reactions. *Science* **2007**, *316* (5830), 1465–1468. https://doi.org/10.1126/science.1141316.

(116) Kurnosov, N. V.; Leontiev, V. S.; Karachevtsev, V. A. Probing the Influence of Amino Acids on Photoluminescence from Carbon Nanotubes Suspended with DNA. *J. Fluoresc.* **2016**, *26* (6), 1951–1958. https://doi.org/10.1007/s10895-016-1888-y.

(117) Mann, F. A.; Herrmann, N.; Meyer, D.; Kruss, S. Tuning Selectivity of Fluorescent Carbon Nanotube-Based Neurotransmitter Sensors. *Sensors* **2017**, *17* (7), 1521. https://doi.org/10.3390/s17071521.

(118) Kinetic Requirements for Spatiotemporal Chemical Imaging with Fluorescent Nanosensors | ACS Nano https://pubs.acs.org/doi/abs/10.1021/acsnano.7b00569 (accessed 2021 -04 -12).

(119) Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain χ1 and χ2 Dihedral Angles | Journal of Chemical Theory and Computation https://pubs.acs.org/doi/abs/10.1021/ct300400x (accessed 2021 -04 -12).

(120) Gigliotti, B.; Sakizzie, B.; Bethune, D. S.; Shelby, R. M.; Cha, J. N. Sequence-Independent Helical Wrapping of Single-Walled Carbon Nanotubes by Long Genomic DNA. *Nano Lett.* **2006**, *6* (2), 159–164. https://doi.org/10.1021/nl0518775.

(121) Johnson, R. R.; Johnson, A. T. C.; Klein, M. L. Probing the Structure of DNA−Carbon Nanotube Hybrids with Molecular Dynamics. *Nano Lett.* **2008**, *8* (1), 69–75. https://doi.org/10.1021/nl071909j.

(122) Adalja, A.; Inglesby, T. Broad-Spectrum Antiviral Agents: A Crucial Pandemic Tool. *Expert Rev. Anti Infect. Ther.* **2019**, *17* (7), 467–470. https://doi.org/10.1080/14787210.2019.1635009.

(123) Clercq, E. D.; Li, G. Approved Antiviral Drugs over the Past 50 Years. *Clin. Microbiol. Rev.* **2016**, *29* (3), 695–747. https://doi.org/10.1128/CMR.00102-15.

(124) Clercq, E. D.; Herdewijn, P. Strategies in the Design of Antiviral Drugs. In *Pharmaceutical Sciences Encyclopedia*; American Cancer Society, 2010; pp 1–56. https://doi.org/10.1002/9780470571224.pse026.

(125) Fridland, A.; Connelly, M. C.; Robbins, B. L. Cellular Factors for Resistance against Antiretroviral Agents. *Antivir. Ther.* **2000**, *5* (3), 181–185.

(126) Spillmann, D. Heparan Sulfate: Anchor for Viral Intruders? *Biochimie* **2001**, *83* (8), 811–817. https://doi.org/10.1016/S0300-9084(01)01290-1.

(127) Cagno, V.; Andreozzi, P.; D'Alicarnasso, M.; Jacob Silva, P.; Mueller, M.; Galloux, M.; Le Goffic, R.; Jones, S. T.; Vallino, M.; Hodek, J.; Weber, J.; Sen, S.; Janeček, E.-R.; Bekdemir, A.; Sanavio, B.; Martinelli, C.; Donalisio, M.; Rameix Welti, M.-A.; Eleouet, J.-F.; Han, Y.; Kaiser, L.; Vukovic, L.; Tapparel, C.; Král, P.; Krol, S.; Lembo, D.; Stellacci, F. Broad-Spectrum Non-Toxic Antiviral Nanoparticles with a Virucidal Inhibition Mechanism. *Nat. Mater.* **2018**, *17* (2), 195–203. https://doi.org/10.1038/nmat5053.

(128) Rusnati, M.; Vicenzi, E.; Donalisio, M.; Oreste, P.; Landolfo, S.; Lembo, D. Sulfated K5 Escherichia Coli Polysaccharide Derivatives: A Novel Class of Candidate Antiviral Microbicides. *Pharmacol. Ther.* **2009**, *123* (3), 310–322. https://doi.org/10.1016/j.pharmthera.2009.05.001.

(129) Riblett, A. M.; Blomen, V. A.; Jae, L. T.; Altamura, L. A.; Doms, R. W.; Brummelkamp, T. R.; Wojcechowskyj, J. A. A Haploid Genetic Screen Identifies Heparan Sulfate

Proteoglycans Supporting Rift Valley Fever Virus Infection. *J. Virol.* **2016**, *90* (3), 1414–1423. https://doi.org/10.1128/JVI.02055-15.

(130) Klimyte, E. M.; Smith, S. E.; Oreste, P.; Lembo, D.; Dutch, R. E. Inhibition of Human Metapneumovirus Binding to Heparan Sulfate Blocks Infection in Human Lung Cells and Airway Tissues. *J. Virol.* **2016**, *90* (20), 9237–9250. https://doi.org/10.1128/JVI.01362-16.

(131) Cagno, V.; Donalisio, M.; Bugatti, A.; Civra, A.; Cavalli, R.; Ranucci, E.; Ferruti, P.; Rusnati, M.; Lembo, D. The Agmatine-Containing Poly(Amidoamine) Polymer AGMA1 Binds Cell Surface Heparan Sulfates and Prevents Attachment of Mucosal Human Papillomaviruses. *Antimicrob. Agents Chemother.* **2015**, *59* (9), 5250–5259. https://doi.org/10.1128/AAC.00443-15.

(132) Donalisio, M.; Quaranta, P.; Chiuppesi, F.; Pistello, M.; Cagno, V.; Cavalli, R.; Volante, M.; Bugatti, A.; Rusnati, M.; Ranucci, E.; Ferruti, P.; Lembo, D. The AGMA1 Poly(Amidoamine) Inhibits the Infectivity of Herpes Simplex Virus in Cell Lines, in Human Cervicovaginal Histocultures, and in Vaginally Infected Mice. *Biomaterials* **2016**, *85*, 40–53. https://doi.org/10.1016/j.biomaterials.2016.01.055.

(133) Shogan, B.; Kruse, L.; Mulamba, G. B.; Hu, A.; Coen, D. M. Virucidal Activity of a GT-Rich Oligonucleotide against Herpes Simplex Virus Mediated by Glycoprotein B. *J. Virol.* **2006**, *80* (10), 4740–4747. https://doi.org/10.1128/JVI.80.10.4740-4747.2006.

(134) Bastian, A. R.; Kantharaju; McFadden, K.; Duffy, C.; Rajagopal, S.; Contarino, M. R.; Papazoglou, E.; Chaiken, I. Cell-Free HIV-1 Virucidal Action by Modified Peptide Triazole Inhibitors of Env Gp120. *ChemMedChem* **2011**, *6* (8), 1335–1339. https://doi.org/10.1002/cmdc.201100177.

(135) Lara, H. H.; Garza-Treviño, E. N.; Ixtepan-Turrent, L.; Singh, D. K. Silver Nanoparticles Are Broad-Spectrum Bactericidal and Virucidal Compounds. *J. Nanobiotechnology* **2011**, *9* (1), 30. https://doi.org/10.1186/1477-3155-9-30.

(136) Broglie, J. J.; Alston, B.; Yang, C.; Ma, L.; Adcock, A. F.; Chen, W.; Yang, L. Antiviral Activity of Gold/Copper Sulfide Core/Shell Nanoparticles against Human Norovirus Virus-Like Particles. *PLOS ONE* **2015**, *10* (10), e0141050. https://doi.org/10.1371/journal.pone.0141050.

(137) Bromberg, L.; Bromberg, D. J.; Hatton, T. A.; Bandín, I.; Concheiro, A.; Alvarez-Lorenzo, C. Antiviral Properties of Polymeric Aziridine- and Biguanide-Modified Core–Shell Magnetic Nanoparticles. *Langmuir* **2012**, *28* (9), 4548–4558. https://doi.org/10.1021/la205127x.

(138) de Souza e Silva, J. M.; Hanchuk, T. D. M.; Santos, M. I.; Kobarg, J.; Bajgelman, M. C.; Cardoso, M. B. Viral Inhibition Mechanism Mediated by Surface-Modified Silica Nanoparticles. *ACS Appl. Mater. Interfaces* **2016**, *8* (26), 16564–16572. https://doi.org/10.1021/acsami.6b03342.

(139) Cardone, G.; Moyer, A. L.; Cheng, N.; Thompson, C. D.; Dvoretzky, I.; Lowy, D. R.; Schiller, J. T.; Steven, A. C.; Buck, C. B.; Trus, B. L. Maturation of the Human Papillomavirus 16 Capsid. *mBio* **2014**, *5* (4). https://doi.org/10.1128/mBio.01104-14.

(140) Dasgupta, J.; Bienkowska-Haba, M.; Ortega, M. E.; Patel, H. D.; Bodevin, S.; Spillmann, D.; Bishop, B.; Sapp, M.; Chen, X. S. Structural Basis of Oligosaccharide Receptor Recognition by Human Papillomavirus*. *J. Biol. Chem.* **2011**, *286* (4), 2617–2624. https://doi.org/10.1074/jbc.M110.160184.

(141) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization and Multithreading. *J. Comput. Chem.* **2010**, *31* (2), 455–461. https://doi.org/10.1002/jcc.21334.

(142) Salter, J. D.; Bennett, R. P.; Smith, H. C. The APOBEC Protein Family: United by Structure, Divergent in Function. *Trends Biochem. Sci.* **2016**, *41* (7), 578–594. https://doi.org/10.1016/j.tibs.2016.05.001.

(143) Holden, L. G.; Prochnow, C.; Chang, Y. P.; Bransteitter, R.; Chelico, L.; Sen, U.; Stevens, R. C.; Goodman, M. F.; Chen, X. S. Crystal Structure of the Anti-Viral APOBEC3G Catalytic Domain and Functional Implications. *Nature* **2008**, *456* (7218), 121–124. https://doi.org/10.1038/nature07357.

(144) Lecossier, D. Hypermutation of HIV-1 DNA in the Absence of the Vif Protein. *Science* **2003**, *300* (5622), 1112–1112. https://doi.org/10.1126/science.1083338.

(145) Zhang, H.; Yang, B.; Pomerantz, R. J.; Zhang, C.; Arunachalam, S. C.; Gao, L. The Cytidine Deaminase CEM15 Induces Hypermutation in Newly Synthesized HIV-1 DNA. *Nature* **2003**, *424* (6944), 94–98. https://doi.org/10.1038/nature01707.

(146) Bishop, K. N.; Verma, M.; Kim, E.-Y.; Wolinsky, S. M.; Malim, M. H. APOBEC3G Inhibits Elongation of HIV-1 Reverse Transcripts. *PLOS Pathog.* **2008**, *4* (12), e1000231. https://doi.org/10.1371/journal.ppat.1000231.

(147) Deaminase-independent inhibition of HIV-1 reverse transcription by APOBEC3G | Nucleic Acids Research | Oxford Academic https://academic.oup.com/nar/article/35/21/7096/2375914?login=true (accessed 2021 -04 -13).

(148) Lu, X.; Zhang, T.; Xu, Z.; Liu, S.; Zhao, B.; Lan, W.; Wang, C.; Ding, J.; Cao, C. Crystal Structure of DNA Cytidine Deaminase ABOBEC3G Catalytic Deamination Domain Suggests a Binding Mode of Full-Length Enzyme to Single-Stranded DNA*. *J. Biol. Chem.* **2015**, *290* (7), 4010–4021. https://doi.org/10.1074/jbc.M114.624262.

(149) Newman, E. N. C.; Holmes, R. K.; Craig, H. M.; Klein, K. C.; Lingappa, J. R.; Malim, M. H.; Sheehy, A. M. Antiviral Function of APOBEC3G Can Be Dissociated from Cytidine Deaminase Activity. *Curr. Biol.* **2005**, *15* (2), 166–170. https://doi.org/10.1016/j.cub.2004.12.068.

(150) Jarmuz, A.; Chester, A.; Bayliss, J.; Gisbourne, J.; Dunham, I.; Scott, J.; Navaratnam, N. An Anthropoid-Specific Locus of Orphan C to U RNA-Editing Enzymes on Chromosome 22. *Genomics* **2002**, *79* (3), 285–296. https://doi.org/10.1006/geno.2002.6718.

(151) Polevoda, B.; McDougall, W. M.; Bennett, R. P.; Salter, J. D.; Smith, H. C. Structural and Functional Assessment of APOBEC3G Macromolecular Complexes. *Methods* **2016**, *107*, 10–22. https://doi.org/10.1016/j.ymeth.2016.03.006.

(152) Harris, R. S.; Petersen-Mahrt, S. K.; Neuberger, M. S. RNA Editing Enzyme APOBEC1 and Some of Its Homologs Can Act as DNA Mutators. *Mol. Cell* **2002**, *10* (5), 1247–1253. https://doi.org/10.1016/S1097-2765(02)00742-6.

(153) Chen, K.-M.; Harjes, E.; Gross, P. J.; Fahmy, A.; Lu, Y.; Shindo, K.; Harris, R. S.; Matsuo, H. Structure of the DNA Deaminase Domain of the HIV-1 Restriction Factor APOBEC3G. *Nature* **2008**, *452* (7183), 116–119. https://doi.org/10.1038/nature06638.

(154) Kouno, T.; Luengas, E. M.; Shigematsu, M.; Shandilya, S. M. D.; Zhang, J.; Chen, L.; Hara, M.; Schiffer, C. A.; Harris, R. S.; Matsuo, H. Structure of the Vif-Binding Domain of the Antiviral Enzyme APOBEC3G. *Nat. Struct. Mol. Biol.* **2015**, *22* (6), 485–491. https://doi.org/10.1038/nsmb.3033.

(155) Salter, J. D.; Krucinska, J.; Raina, J.; Smith, H. C.; Wedekind, J. E. A Hydrodynamic Analysis of APOBEC3G Reveals a Monomer−Dimer−Tetramer Self-Association That Has Implications for Anti-HIV Function. *Biochemistry* **2009**, *48* (45), 10685–10687. https://doi.org/10.1021/bi901642c.

(156) Wedekind, J. E.; Gillilan, R.; Janda, A.; Krucinska, J.; Salter, J. D.; Bennett, R. P.; Raina, J.; Smith, H. C. Nanostructures of APOBEC3G Support a Hierarchical Assembly Model of High Molecular Mass Ribonucleoprotein Particles from Dimeric Subunits*,. *J. Biol. Chem.* **2006**, *281* (50), 38122–38126. https://doi.org/10.1074/jbc.C600253200.

(157) Gorle, S.; Pan, Y.; Sun, Z.; Shlyakhtenko, L. S.; Harris, R. S.; Lyubchenko, Y. L.; Vuković, L. Computational Model and Dynamics of Monomeric Full-Length APOBEC3G. *ACS Cent. Sci.* **2017**, *3* (11), 1180–1188. https://doi.org/10.1021/acscentsci.7b00346.

(158) Pan, Y.; S. Shlyakhtenko, L.; L. Lyubchenko, Y. Insight into the Dynamics of APOBEC3G Protein in Complexes with DNA Assessed by High Speed AFM. *Nanoscale Adv.* **2019**, *1* (10), 4016–4024. https://doi.org/10.1039/C9NA00457B.

(159) Chelico, L.; Pham, P.; Goodman, M. F. Mechanisms of APOBEC3G-Catalyzed Processive Deamination of Deoxycytidine on Single-Stranded DNA. *Nat. Struct. Mol. Biol.* **2009**, *16* (5), 454–455. https://doi.org/10.1038/nsmb0509-454.

(160) Ara, A.; Love, R. P.; Chelico, L. Different Mutagenic Potential of HIV-1 Restriction Factors APOBEC3G and APOBEC3F Is Determined by Distinct Single-Stranded DNA Scanning Mechanisms. *PLOS Pathog.* **2014**, *10* (3), e1004024. https://doi.org/10.1371/journal.ppat.1004024.

(161) Yang, H.; Ito, F.; Wolfe, A. D.; Li, S.; Mohammadzadeh, N.; Love, R. P.; Yan, M.; Zirkle, B.; Gaba, A.; Chelico, L.; Chen, X. S. Understanding the Structural Basis of HIV-1 Restriction by the Full Length Double-Domain APOBEC3G. *Nat. Commun.* **2020**, *11* (1), 632. https://doi.org/10.1038/s41467-020-14377-y.

(162) Furukawa, A.; Nagata, T.; Matsugami, A.; Habu, Y.; Sugiyama, R.; Hayashi, F.; Kobayashi, N.; Yokoyama, S.; Takaku, H.; Katahira, M. Structure, Interaction and Real-Time Monitoring of the Enzymatic Reaction of Wild-Type APOBEC3G. *EMBO J.* **2009**, *28* (4), 440–451. https://doi.org/10.1038/emboj.2008.290.

(163) Xiao, X.; Li, S.-X.; Yang, H.; Chen, X. S. Crystal Structures of APOBEC3G N-Domain Alone and Its Complex with DNA. *Nat. Commun.* **2016**, *7* (1), 12193. https://doi.org/10.1038/ncomms12193.

(164) The UniProt Consortium. UniProt: The Universal Protein Knowledgebase in 2021. *Nucleic Acids Res.* **2021**, *49* (D1), D480–D489. https://doi.org/10.1093/nar/gkaa1100.

(165) Maiti, A.; Myint, W.; Kanai, T.; Delviks-Frankenberry, K.; Sierra Rodriguez, C.; Pathak, V. K.; Schiffer, C. A.; Matsuo, H. Crystal Structure of the Catalytic Domain of HIV-1 Restriction Factor APOBEC3G in Complex with SsDNA. *Nat. Commun.* **2018**, *9* (1), 2460. https://doi.org/10.1038/s41467-018-04872-8.

(166) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J. Chem. Theory Comput.* **2008**, *4* (5), 819–834. https://doi.org/10.1021/ct700324x.

(167) GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation | Journal of Chemical Theory and Computation https://pubs.acs.org/doi/abs/10.1021/ct700301q (accessed 2021 -04 -14).

(168) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism

from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001.

(169) Coarse-graining polymers with the MARTINI force-field: polystyrene as a benchmark case - Soft Matter (RSC Publishing) DOI:10.1039/C0SM00481B https://pubs.rsc.org/en/content/articlehtml/2011/sm/c0sm00481b (accessed 2021 -04 -14).

(170) Jimenez, M.; Campillo, N. E.; Canelles, M. COVID-19 Vaccine Race: Analysis of Age-Dependent Immune Responses against SARS-CoV-2 Indicates that more than Just One Strategy May Be Needed https://www.ingentaconnect.com/content/ben/cmc/pre-prints/content-33109026 (accessed 2021 -04 -13). https://doi.org/10.2174/0929867327666201027153123.

(171) Research and Development on Therapeutic Agents and Vaccines for COVID-19 and Related Human Coronavirus Diseases | ACS Central Science https://pubs.acs.org/doi/abs/10.1021/acscentsci.0c00272 (accessed 2021 -04 -13).

(172) Le, T.; Andreadakis, Z.; Kumar, A.; Román, R.; Tollefsen, S.; Saville, M.; Mayhew, S. The COVID-19 Vaccine Development Landscape. *Nat. Rev. Drug Discov.* **2020**. https://doi.org/10.1038/d41573-020-00073-5.

(173) Morse, J. S.; Lalonde, T.; Xu, S.; Liu, W. R. Learning from the Past: Possible Urgent Prevention and Treatment Options for Severe Acute Respiratory Infections Caused by 2019-NCoV. *ChemBioChem* **2020**, *21* (5), 730–738. https://doi.org/10.1002/cbic.202000047.

(174) Wang, Y.; Zhang, D.; Du, G.; Du, R.; Zhao, J.; Jin, Y.; Fu, S.; Gao, L.; Cheng, Z.; Lu, Q.; Hu, Y.; Luo, G.; Wang, K.; Lu, Y.; Li, H.; Wang, S.; Ruan, S.; Yang, C.; Mei, C.; Wang, Y.; Ding, D.; Wu, F.; Tang, X.; Ye, X.; Ye, Y.; Liu, B.; Yang, J.; Yin, W.; Wang, A.; Fan, G.; Zhou, F.; Liu, Z.; Gu, X.; Xu, J.; Shang, L.; Zhang, Y.; Cao, L.; Guo, T.; Wan, Y.; Qin, H.; Jiang, Y.; Jaki, T.; Hayden, F. G.; Horby, P. W.; Cao, B.; Wang, C. Remdesivir in Adults with Severe COVID-19: A Randomised, Double-Blind, Placebo-Controlled, Multicentre Trial. *The Lancet* **2020**, *395* (10236), 1569–1578. https://doi.org/10.1016/S0140-6736(20)31022-9.

(175) Baum, A.; Fulton, B. O.; Wloga, E.; Copin, R.; Pascal, K. E.; Russo, V.; Giordano, S.; Lanza, K.; Negron, N.; Ni, M.; Wei, Y.; Atwal, G. S.; Murphy, A. J.; Stahl, N.; Yancopoulos, G. D.; Kyratsous, C. A. Antibody Cocktail to SARS-CoV-2 Spike Protein Prevents Rapid Mutational Escape Seen with Individual Antibodies. *Science* **2020**, *369* (6506), 1014–1018. https://doi.org/10.1126/science.abd0831.

(176) Han, Y.; Král, P. Computational Design of ACE2-Based Peptide Inhibitors of SARS-CoV-2. *ACS Nano* **2020**, *14* (4), 5143–5147. https://doi.org/10.1021/acsnano.0c02857.

(177) Wang, Q.; Zhang, Y.; Wu, L.; Niu, S.; Song, C.; Zhang, Z.; Lu, G.; Qiao, C.; Hu, Y.; Yuen, K.-Y.; Wang, Q.; Zhou, H.; Yan, J.; Qi, J. Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. *Cell* **2020**, *181* (4), 894-904.e9. https://doi.org/10.1016/j.cell.2020.03.045.

(178) Hoffmann, M.; Kleine-Weber, H.; Schroeder, S.; Krüger, N.; Herrler, T.; Erichsen, S.; Schiergens, T. S.; Herrler, G.; Wu, N.-H.; Nitsche, A.; Müller, M. A.; Drosten, C.; Pöhlmann, S. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **2020**, *181* (2), 271-280.e8. https://doi.org/10.1016/j.cell.2020.02.052.

(179) Shang, J.; Wan, Y.; Luo, C.; Ye, G.; Geng, Q.; Auerbach, A.; Li, F. Cell Entry Mechanisms of SARS-CoV-2. *Proc. Natl. Acad. Sci.* **2020**, *117* (21), 11727–11734. https://doi.org/10.1073/pnas.2003138117.

(180) Tung, H. Y. L. COVID-19 and SARS-COV-2 Infection and Virulence, Hypothesis II: Elevated Temperature, Humidity and Far Infrared Irradiation Will Significantly Lower the Viability, Infectivity and Virulence of SARS-COV-2. *Sci. Prepr.* **2020**. https://doi.org/10.14293/S2199-1006.1.SOR-.PP4VZ1S.v1.

(181) Korber, B.; Fischer, W. M.; Gnanakaran, S.; Yoon, H.; Theiler, J.; Abfalterer, W.; Hengartner, N.; Giorgi, E. E.; Bhattacharya, T.; Foley, B.; Hastie, K. M.; Parker, M. D.; Partridge, D. G.; Evans, C. M.; Freeman, T. M.; de Silva, T. I.; Angyal, A.; Brown, R. L.; Carrilero, L.; Green, L. R.; Groves, D. C.; Johnson, K. J.; Keeley, A. J.; Lindsey, B. B.; Parsons, P. J.; Raza, M.; Rowland-Jones, S.; Smith, N.; Tucker, R. M.; Wang, D.; Wyles, M. D.; McDanal, C.; Perez, L. G.; Tang, H.; Moon-Walker, A.; Whelan, S. P.; LaBranche, C. C.; Saphire, E. O.; Montefiori, D. C. Tracking Changes in SARS-CoV-2 Spike: Evidence That D614G Increases Infectivity of the COVID-19 Virus. *Cell* **2020**, *182* (4), 812-827.e19. https://doi.org/10.1016/j.cell.2020.06.043.

(182) Liu, Z.; Xiao, X.; Wei, X.; Li, J.; Yang, J.; Tan, H.; Zhu, J.; Zhang, Q.; Wu, J.; Liu, L. Composition and Divergence of Coronavirus Spike Proteins and Host ACE2 Receptors Predict Potential Intermediate Hosts of SARS-CoV-2. *J. Med. Virol.* **2020**, *92* (6), 595–601. https://doi.org/10.1002/jmv.25726.

(183) Watanabe, Y.; Mendonça, L.; Allen, E. R.; Howe, A.; Lee, M.; Allen, J. D.; Chawla, H.; Pulido, D.; Donnellan, F.; Davies, H.; Ulaszewska, M.; Belij-Rammerstorfer, S.; Morris, S.; Krebs, A.-S.; Dejnirattisai, W.; Mongkolsapaya, J.; Supasa, P.; Screaton, G. R.; Green, C. M.; Lambe, T.; Zhang, P.; Gilbert, S. C.; Crispin, M. Native-like SARS-CoV-2 Spike Glycoprotein Expressed by ChAdOx1 NCoV-19/AZD1222 Vaccine. *ACS Cent. Sci.* **2021**. https://doi.org/10.1021/acscentsci.1c00080.

(184) Grant, O. C.; Montgomery, D.; Ito, K.; Woods, R. J. Analysis of the SARS-CoV-2 Spike Protein Glycan Shield Reveals Implications for Immune Recognition. *Sci. Rep.* **2020**, *10* (1), 14991. https://doi.org/10.1038/s41598-020-71748-7.

(185) Enhanced Binding of SARS-CoV-2 Spike Protein to Receptor by Distal Polybasic Cleavage Sites | ACS Nano https://pubs.acs.org/doi/abs/10.1021/acsnano.0c04798 (accessed 2021 -04 -13).

(186) Vergara-Jaque, A.; Comer, J.; Monsalve, L.; González-Nilo, F. D.; Sandoval, C. Computationally Efficient Methodology for Atomic-Level Characterization of Dendrimer–Drug Complexes: A Comparison of Amine- and Acetyl-Terminated PAMAM. *J. Phys. Chem. B* **2013**, *117* (22), 6801–6813. https://doi.org/10.1021/jp4000363.

(187) Homeyer, N.; Gohlke, H. Free Energy Calculations by the Molecular Mechanics Poisson−Boltzmann Surface Area Method. *Mol. Inform.* **2012**, *31* (2), 114–122. https://doi.org/10.1002/minf.201100135.

(188) Tanner, D. E.; Chan, K.-Y.; Phillips, J. C.; Schulten, K. Parallel Generalized Born Implicit Solvent Calculations with NAMD. *J. Chem. Theory Comput.* **2011**, *7* (11), 3635–3642. https://doi.org/10.1021/ct200563j.

(189) Massova, I.; Kollman, P. A. Computational Alanine Scanning To Probe Protein−Protein Interactions:  A Novel Approach To Evaluate Binding Free Energies. *J. Am. Chem. Soc.* **1999**, *121* (36), 8133–8143. https://doi.org/10.1021/ja990935j.

(190) Abroshan, H.; Akbarzadeh, H.; Parsafar, G. A. Molecular dynamics simulation and MM–PBSA calculations of sickle cell hemoglobin in dimer form with Val, Trp, or Phe at the lateral contact. *J. Phys. Org. Chem.* **2010**, *23* (9), 866–877. https://doi.org/10.1002/poc.1679.

(191) Russo, A.; Scognamiglio, P. L.; Enriquez, R. P. H.; Santambrogio, C.; Grandori, R.; Marasco, D.; Giordano, A.; Scoles, G.; Fortuna, S. In Silico Generation of Peptides by Replica Exchange Monte Carlo: Docking-Based Optimization of Maltose-Binding-Protein Ligands. *PLOS ONE* **2015**, *10* (8), e0133571. https://doi.org/10.1371/journal.pone.0133571.

(192) Hong Enriquez, R. P.; Pavan, S.; Benedetti, F.; Tossi, A.; Savoini, A.; Berti, F.; Laio, A. Designing Short Peptides with High Affinity for Organic Molecules: A Combined Docking, Molecular Dynamics, And Monte Carlo Approach. *J. Chem. Theory Comput.* **2012**, *8* (3), 1121–1128. https://doi.org/10.1021/ct200873y.

(193) Hadfield, J.; Megill, C.; Bell, S. M.; Huddleston, J.; Potter, B.; Callender, C.; Sagulenko, P.; Bedford, T.; Neher, R. A. Nextstrain: Real-Time Tracking of Pathogen Evolution. *Bioinformatics* **2018**, *34* (23), 4121–4123. https://doi.org/10.1093/bioinformatics/bty407.

(194) Engineering human ACE2 to optimize binding to the spike protein of SARS coronavirus 2 | Science https://science.sciencemag.org/content/369/6508/1261.abstract (accessed 2021 -04 -13).

# Vita

Parth Chaturvedi obtained dual Bachelors/Masters in Physics from Indian Institute of Science education and Research, Bhopal, India in 2015 and joined UTEP in 2016 as doctoral student in chemistry. He joined research group of Lela Vuković in Department of chemistry and biochemistry working on computational modeling nucleic acids and anti-viral mechanisms in biology using molecular dynamics simulations. His research at UTEP have resulted in three published papers so far: Adaptive Evolution of Peptide Inhibitors for Mutating SARS-CoV-2 (https://doi.org/10.1002/adts.202000156), Structural properties of small single-stranded circular nucleic acids (https://doi.org/10.1021/acs.jpcb.9b06831) and Quantification of the number of adsorbed DNA molecules on single-walled carbon nanotubes (https://doi.org/10.1021/acs.jpcc.8b11058).

Contact Information : parth9022@gmail.com