Nanomechanical Analysis of Coronavirus Spike proteins and Correlation with Infectivity and Lethality

by
Yiwen Hu

B.S. Theoretical and Applied Mechanics
Peking University, 2019

Submitted to the Department of Mechanical Engineering in Partial Fulfillment of the Requirements for the Degree of

Master of Science in Mechanical Engineering
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2021

Signature of Author:_____
Yiwen Hu
Department of Mechanical Engineering
May 14, 2021

Certified by:_____
Markus J. Buehler
Jerry McAfee Professor in Engineering
Thesis Supervisor

Certified by:_____
Nicolas Hadjiconstantinou
Professor of Mechanical Engineering
Thesis Reader

Accepted by:_____
Nicolas Hadjiconstantinou
Professor of Mechanical Engineering
Chairman, Department Committee on Graduate Theses

Nanomechanical Analysis of Coronavirus Spike proteins and Correlation with Infectivity and Lethality

by

Yiwen Hu

Submitted to the Department of Mechanical Engineering on May 14, 2021
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Mechanical Engineering

ABSTRACT

The novel coronavirus disease, COVID-19, has spread rapidly around the world. Its causative virus, SARS-CoV-2, enters human cells through the physical interaction between the receptor-binding domain (RBD) of its spike protein and the human cell receptor ACE2. As an increasing number of variants of SARS-CoV-2 circulates globally, estimates of infectiousness and lethality of newly emerging strains are important. Here, we provide a novel way to develop a deeper understanding of coronavirus spike proteins, connecting their nanomechanical features – specifically the vibrational spectrum and quantitative measures of mobility – with virus lethality and infection rate. The key result of our work is that both, the overall flexibility of upward RBD and the mobility ratio of RBDs in different conformations, represent two significant factors that show a positive scaling with virus lethality and an inverse correlation with the infection rate. A quantitative model is presented to make predictions on the infectivity and lethality of SARS-CoV-2 variants based on molecular motions and vibrational patterns of the virus spike protein. Our analysis shows that epidemiological virus properties can be linked directly to pure nanomechanical, vibrational aspects, offering an alternative way of screening new viruses and mutations against high threat levels, and potentially exploring novel ways to prevent infections from occurring by interfering with the nanoscale motions.

Thesis Supervisor: Markus J. Buehler
Title: Jerry McAfee Professor in Engineering

Acknowledgements

I would like to thank my advisor, Prof. Markus J. Buehler, for his helpful guidance and support. It's a great pleasure to work in the Laboratory for Atomistic and Molecular Mechanics (LAMM). I also want to thank all my labmates for being so kind and supportive all the time. Last but not least, I really appreciate my project sponsors MIT-IBM AI lab, ONR, AFOSR, NIH, as well as ARO.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Background

The novel coronavirus disease 2019 (COVID-19), has spread rapidly around the world [1–5] . As an increasing number of variants of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), the causative virus of COVID-19, circulates globally, how to understand and predict the virus infectiousness and lethality becomes an even more important question. (Figure 1.1)

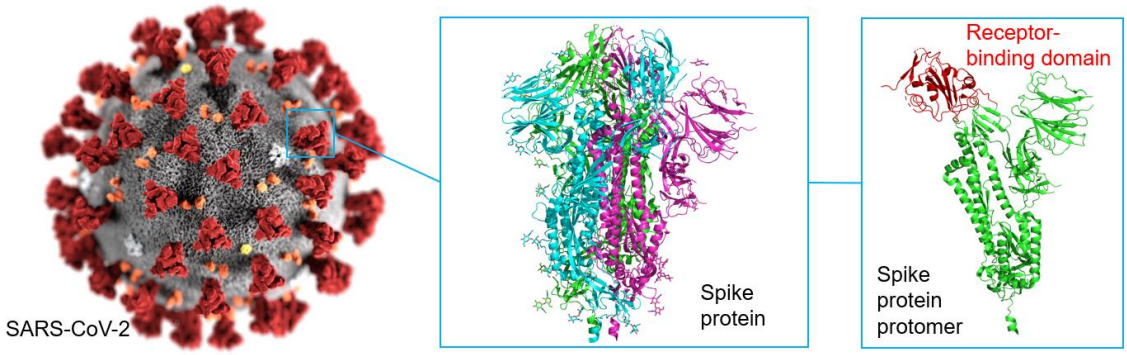

Figure 1. 1: Overview of COVID-19 pathogen, SARS-CoV-2. Detailed structure of SARS-CoV-2 spike protein (PDB: 6VXX) is shown.

To provide a comparative analysis – specifically focused on how nano-level features relate with macroscopic epidemiolocal observables – we focus on different coronavirus types within the same family of pathogens. Over the past decades, several types of coronaviruses have emerged. The virus types HCoV-NL63 and HCoV-HKU1 are often reported to cause lower respiratory tract infections, while HCoV-OC43 and HCoV-229E are usually associated with comparatively mild symptoms similar to the common cold [5–7]. The ones that threaten public health more seriously are three highly pathogenic human coronaviruses — namely, SARS-CoV, Middle East Respiratory Syndrome Coronavirus (MERS-CoV) and SARS-CoV-2. SARS-CoV was first reported in China in November 2002, then quickly spread globally, resulting in over 8,000 infections with about 800 deaths [8]. MERS-CoV was first identified in Saudi Arabia in June 2012, featuring limited transmission with case

fatality rate as high as 35% [9, 10]. SARS-CoV-2 was first reported in China in December 2019 [3, 11, 12]. It can easily transmit from human to human, resulting in more than 120 million global cases as of April 2021 [13, 14]. A general comparison of epidemiological properties of SARS-CoV, MERS-CoV and SARS-CoV-2 is depicted in Figure 1.2.

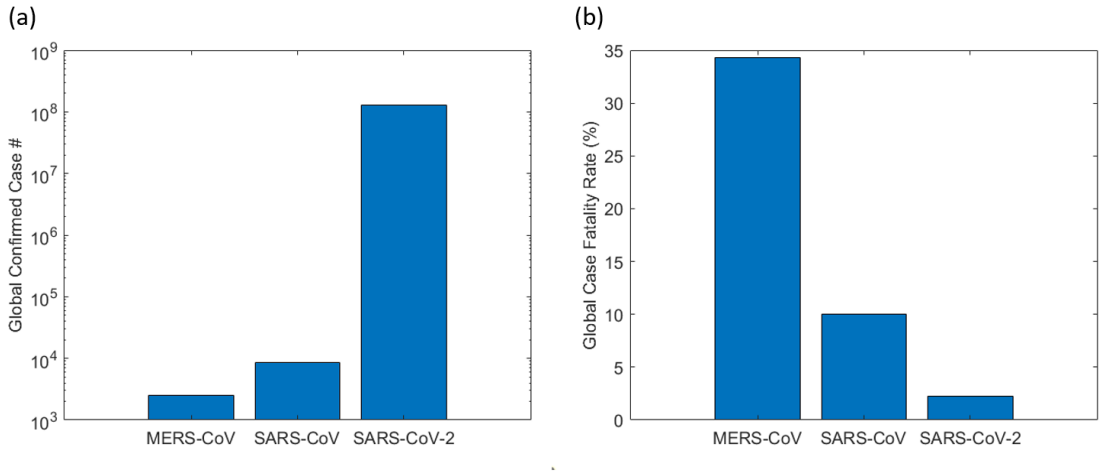

Figure 1. 2: Viral infectivity and lethality comparison between MERS-CoV, SARS-CoV and SARS-CoV-2. Data updated as of Apr. 1, 2021 [13].

Aside from the significant difference in infectiousness and lethality of various types of coronaviruses, even variants of the same coronavirus could behave differently in epidemiological aspect. Recent experimental research reveals that SARS-CoV-2 D614G variant, which emerged in late January or early February 2020 and has quickly spread around the world since then, shows enhanced infectivity and transmissibility compared to the wild-type virus[15, 16]. The lineage B.1.1.7, first reported in the United Kingdom in December 2020, is predicted to be 43 to 90% more transmissible than the predecessor lineage based on statistical and dynamic modeling approaches[17]. The more worrying finding is that, an experiment suggests that mRNA-vaccine activity against SARS-CoV-2 variants that encode E484K-, N501Y- or K417N/E484K/N501-mutant S is reduced by a small—but significant—margin[18].

The structural basis is important to explain the epidemiological behavior of virus and search for the treatment. The spike protein of the coronavirus plays an essential role in receptor recognition, viral fusion and cell entry, serving as a promising target for drug and

vaccine development [19–21]. The process represents a complex mechano-chemical process, whereby during the entry into the host cell, the spike protein first binds to a cell receptor through the receptor-binding domain (RBD) and then begins the fusion process. (Figure 1.1) It is believed that the RBDs of different coronaviruses recognize different cell receptors [3, 22]. SARS-CoV, SARS-CoV-2 and HCoV-NL63 recognize angiotensin-converting enzyme 2 (ACE2) as their receptor in the human body, while MERS-CoV recognizes dipeptidyl peptidase 4 (DPP4) as its receptor [1, 23–25]. In order to successfully bind to the receptor, the spike protein of a specific coronavirus must maintain a receptor-accessible state with at least one RBD in upward conformation. This is because otherwise there would be steric clashes hindering the binding process [4]. In experimental work, this specific type of receptor-accessible state has been captured for MERS-CoV, SARS-CoV and SARS-CoV-2 [21, 26].

1.2 Motivation

Some recent research focuses on understanding and predicting epidemiological behavior of SARS-CoV-2 and its variants using different indicators from biological or medical aspect. By evaluating the binding energy of SARS-CoV-2 spike protein and the host cell receptor with a machine-learning model, researchers report that most likely future mutations will make SARS-CoV-2 more infectious[27]. It has also been shown that the viral load can be considered as an independent predictor of the transmissibility and mortality of SARS-CoV-2[28, 29]. Furthermore, with machine learning technique for natural language processing, the generated semantic landscapes for SARS-CoV-2 are able to predict viral escape mutations that produce sequences that are syntactically and/or grammatically correct but effectively different in semantics and thus able to evade the immune system[30].

What about the nano-mechanical aspect? Similar to many other biomaterials studied before, SARS-CoV-2 can also be considered as a type of biomaterial with structure, function and especially mechanical properties. The structure-function relationship has already been explored for biomaterials like intermediate filaments (IFs) and collagen

microfibrils[31, 32]. The hierarchical structure shown in Figure 1.3 illustrates how macroscale behavior of biomaterials can be linked to the nanoscale structure.

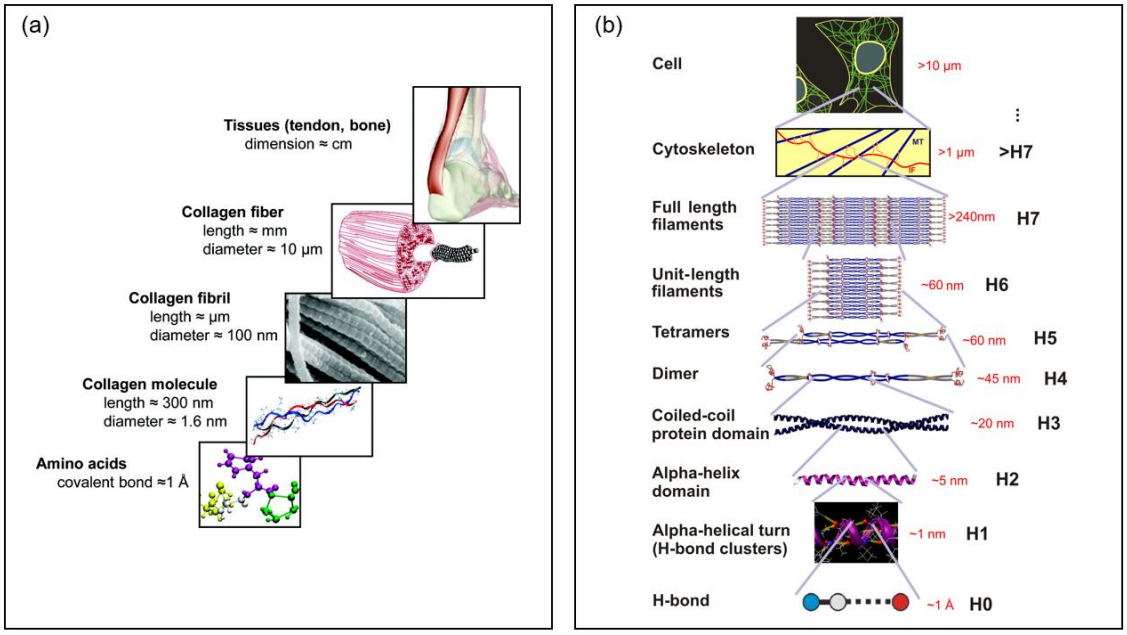

Figure 1. 3: The hierarchical structure of collagen microfibrils and intermediate filaments. [31, 32]

Down to the nanometer scale, the virus is not in the stationary state as shown in figures. In fact, all molecules at the nanoscale are continuously moving and vibrating. We hypothesize that the mechanics of the virus, and especially its motion, is key to understand it better. With that, we are motivated to investigate the correlation between nano-mechanical properties of coronavirus and large-scale population-level observation.

While the structure of coronavirus spike protein is well studied, much less attention has been focused on the connection between the mechanical features of the virus with virus lethality as well as infection rate. The structural basis of SARS-CoV viral infectivity has been explored to some extent, pointing out that the Trp-rich region of S protein is essential [33]. It has also been observed that cleavage of the spike protein of SARS-CoV is associated with viral infectivity [34]. However, there has been no lateral comparative study between similar coronaviruses on this type of connection. It remains a question which kind of mechanical and structural properties could possibly relate to the mortality and infection rate of the virus. If successful, the approach presented in this thesis may provide an

alternative or complementary way to screen viruses or mutations against large-scale epidemiological data, provide additional mechanistic insights into disease etiology, and offer potential targets for therapies or preventive measures.

1.3 Thesis Overview

Here, we provide a novel way towards better understanding the coronavirus spike proteins, connecting its nanomechanical features – especially their vibrational patterns - with virus lethality and infection rate, using a molecular modeling approach and a nanomechanical analysis. (Figure 1.4) ls. In a broader context, the mechanics of proteins has long been a subject of interest, and this study shows how it can be a useful tool to help us understand complex disease etiology by connecting nanoscopic physical features with epidemiological data [35–41].

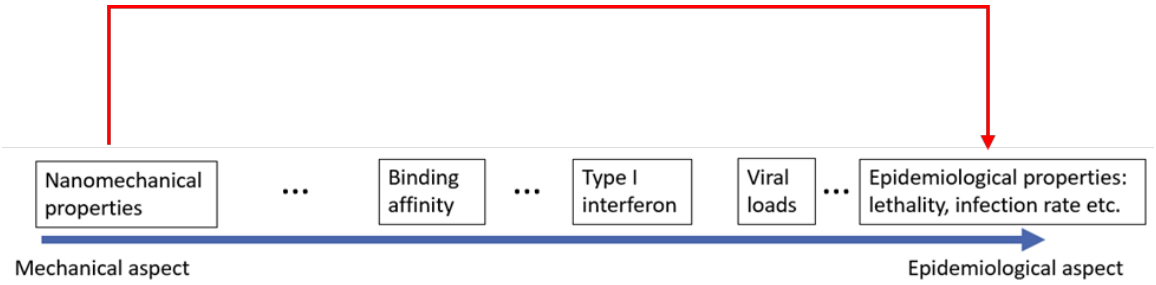

Figure 1. 4: Schematic diagram connecting mechanical and epidemiological aspects.

This thesis is organized as follows: Chapter 2 describes the data processing and simulation methods in detail; Chapter 3 shows the results of normal mode analysis on original beta-coronavirus spike proteins in both receptor-accessible and receptor-inaccessible conformations, and qualitatively correlates the vibrational and epidemiological features; A quantitative model in constructed in Chapter 4 to associate molecular motions and vibrational patterns of the virus spike protein with infectiousness and lethality, and predictions of several SARS-CoV-2 variants are presented.

2 Methods

2.1 Normal Mode Analysis

In protein science, normal mode analysis (NMA) has long been one of the most comprehensive yet efficient methods to calculate vibrational normal modes and analyze protein flexibility, which provides the rationale for use in this study[42]. Another reason behind the broad use of NMA is that the low-frequency modes elucidated by NMA could often describe the real-world motions of a protein, and often bear important functional significance [43].

Generally speaking, in NMA, the atoms are modeled as point masses, which are connected by springs that represents the interatomic force fields. After constructing the Hessian matrix based on the second-order partial derivatives of the potential energy function, the normal modes and corresponding frequencies can be directly obtained by diagonalizing the matrix and computing its eigenvalues [44].

Based on harmonic approximation, the normal modes of a system can be determined by solving the following equation[45]

$$\boldsymbol{K}\boldsymbol{x} = \omega^2 \boldsymbol{M}\boldsymbol{x} \tag{2.1}$$

where

$$\boldsymbol{K} = \begin{bmatrix} k_{11} & k_{12} & \cdots & k_{1N} \\ k_{21} & k_{22} & \cdots & k_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ k_{N1} & k_{N2} & \cdots & k_{NN} \end{bmatrix};\ \boldsymbol{M} = \begin{bmatrix} m_1 & 0 & \cdots & 0 \\ 0 & m_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & m_N \end{bmatrix};\ \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \tag{2.2}$$

In the equation, x_i is the Cartesian displacement of the i^{th} atom with respect to its equilibrium positions; m_i is the atomic mass; $k_{ij} = \partial^2 V / (\partial x_i \partial x_j)$ is the second

derivative of the potential energy calculated for equilibrium positions; ω denotes an eigen angular frequency; N represents the number of degrees of freedom of the system.

Multiply both sides in equation (2.1) by $\boldsymbol{M}^{-1/2}$ and we can get the transformed eigenvalue equation

$$\boldsymbol{H}\boldsymbol{q} = \omega^2\boldsymbol{q} \tag{2.3}$$

where $\boldsymbol{H}$ is the Hessian matrix defined by $\boldsymbol{H} = \boldsymbol{M}^{-1/2}\boldsymbol{K}\boldsymbol{M}^{-1/2}$ and $\boldsymbol{q}$ is the mass-weighted Cartesian coordinate vector defined by $\boldsymbol{q} = \boldsymbol{M}^{1/2}\boldsymbol{x}$.

After diagonalizing the Hessian matrix and solving its eigenvalue problem, the atomic fluctuations can be obtained with[46]

$$\langle x_i^2 \rangle = \frac{k_B T}{m_i} \sum_{j=1}^{N_v} \frac{a_{ij}^2}{\omega_j^2} \tag{2.4}$$

where $f_j = \omega_j/2\pi$ is the frequency of the j^{th} normal mode and a_{ij} is the corresponding coordinate displacement of the i^{th} atom. N_v denotes the total number of normal modes considered. Previous research has shown that low-frequency normal modes of proteins are actually responsible for most of their atomic displacements[47, 48].

2.2 Molecular Dynamics Simulation

Molecular dynamics (MD) simulations has developed rapidly in recent years and can now be used effectively to capture the behavior of proteins and other biomolecules in full atomic detail and understand macromolecular structure-to-function relationships[49, 50]. While the most basic application of MD simulation is to evaluate the mobility of various region of a biomolecule, it can also be used to build or refine structural models based on experimental data, and to determine the response of certain biomolecular system to some perturbation[49, 51, 52].

The basic idea behind MD simulation can be summarized with Newton's laws and simple update rules:

$$\boldsymbol{a}_i = \frac{\boldsymbol{F}_i}{m_i} = -\frac{1}{m_i}\frac{\partial E_{pot}}{\partial \boldsymbol{x}_i} \tag{2.5}$$

$$\boldsymbol{v}_i(t + dt) = \boldsymbol{v}_i(t) + \boldsymbol{a}_i dt \tag{2.6}$$

$$\boldsymbol{x}_i(t + dt) = \boldsymbol{x}_i(t) + \boldsymbol{v}_i dt \tag{2.7}$$

where E_{pot} denotes the potential energy; t is simulation time; m_i, $\boldsymbol{F}_i$, $\boldsymbol{a}_i$, $\boldsymbol{v}_i$ and $\boldsymbol{x}_i$ denote the mass, force vector, acceleration, velocity and position of the i^{th} atom. Force fields currently used in atomistic molecular simulations differ in the way they are parameterized. Parameters are not necessarily interchangeable, and not all force-fields allow to represent all molecule types, but simulations conducted using modern force-fields are normally equivalent[50, 53, 54].

2.3 Data Processing

To assess the molecular mechanics from an atom-by-atom perspective, we conduct normal mode analyses (NMA) of coronavirus spike proteins in receptor-accessible state with one upward RBD as well as receptor-inaccessible state where all three RBDs are in downward position. We access the desired three-dimensional protein structures from the Protein Data Bank [55] and prepare the atomistic models with Visual Molecular Dynamics (VMD) [2, 56].

Before normal mode analysis, 10,000 steps of conjugate gradient energy minimization are performed using Nanoscale Molecular Dynamics (NAMD) [57] in order to relax the original coronavirus spike protein structure [58]. No further MD simulation is performed since the protein structure from Protein Data Bank is already experimentally equilibrated.

In regards to SARS-CoV-2 variants, we use the SARS-CoV-2 spike protein structure with PDB ID 6Z97 as our base protein structure, implement the spike mutations of different variants using PyMOL, and prepare the atomistic models with Visual Molecular Dynamics (VMD)[59]. The spike protein structures are immersed in a TIP3P water box with a minimum distance of 20Å to the nearest protein atom, followed by ionization with Sodium and Chlorine atoms. Then, we carry out 10000 steps of conjugate gradient energy minimization and equilibrate the system for 80 ns with temperature controlled at 300K based on NAMD[57]. After equilibration, we compute the average root mean square derivations (RMSD) based on the last 20 ns equilibrium period and pick out 10 frames with the nearest RMSD from the trajectory file in order to perform normal mode analysis on them.

A coarse-grained elastic network model (ENM) available in the Bio3D package in R is employed to analyze the normal modes of coronavirus spike protein structures [42, 45, 46]. This model uses N, CA, C atoms to represent the protein backbone and selects 0 to 2 significant side chains based on their size and distance to CA atoms, proved to have comparable accuracy with all-atom ENM. Here, the atomic displacements are scaled for temperature 300K. While the fluctuation profile of original coronavirus spike proteins is the vibrational fluctuation distribution of energy-minimized protein structure, the fluctuation profile of SARS-CoV-2 variants of concern is computed as the average fluctuation distribution over the 10 configurations that we picked out from the trajectory file. Equilibration by simulation would result in some local unfolding occurs at the end of each chain, which induces abnormally high fluctuations in the fluctuation profile. Considering that the terminal regions are far from the RBD, we neglect the effects of the unfolding behavior and set fluctuations of the terminal 5-8 residues to zero.

After normal mode analysis, we identify two important factors in the fluctuation profile of coronavirus spike protein, i.e. flexibility level and flexibility ratio, to describe the vibrational behavior, and they are defined as:

$$Average\ fluctuation\ of\ upward\ RBD = \sum_{i=1}^{n} F_i/n \tag{2.8}$$

$$Fluctuation\ ratio = max(F_1, F_2 \dots F_n)/max(f_1, f_2 \dots f_m) \tag{2.9}$$

where *m, n* denotes the number of residues in downward and upward RBD and f_i, F_i represents fluctuation of the i^{th} residue in downward and upward RBD respectively.

The global confirmed case number and the case fatality rate used in this thesis are updated as of Apr 1, 2021. Korber et al. have developed a web tool to track mutations of SARS-CoV-2 based on updated GISAID SARS-CoV-2 sequence database[60][61]. We calculate the portion of D614 on a monthly basis using this tool and evaluate the effective total global confirmed case number for original SARS-CoV-2 with D614

$$Effective\ total\ case\ number = \sum_{i} p_i N_i \tag{2.10}$$

where i denotes the number of months, p_i denotes the portion of D614 sequences in that month, and N_i stands for the newly confirmed global case number in that month. Since it is difficult to differentiate the case fatality rate for virus with D614 and G614, we directly use the general global case fatality rate as the mortality rate for the original SARS-CoV-2.

3 Comparative Analysis of MERS-CoV, SARS-CoV and SARS-CoV-2

In this chapter, we provide a novel way in understanding coronavirus spike proteins, connecting their nanomechanical features – specifically its vibrational spectrum and quantitative measures of mobility – with virus lethality and infection rate. The key result of our work is that both, the overall flexibility of upward RBD and the mobility ratio of RBDs in different conformations, represent two significant factors that show a positive scaling with virus lethality and an inverse correlation with the infection rate[62].

3.1 Different Conformations and Global Movements

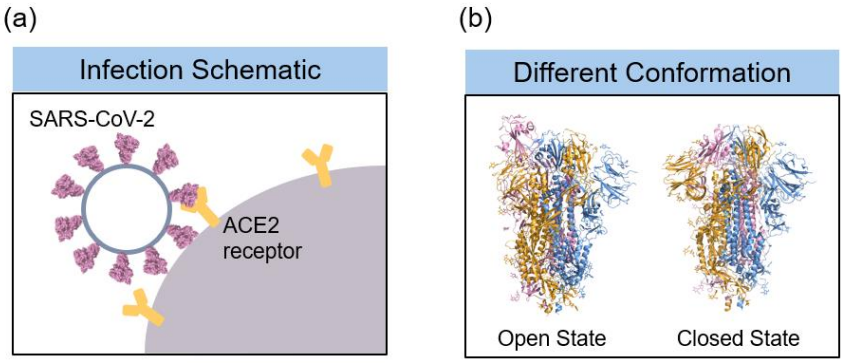

Figure 3. 1: Infection schematic and different conformations of SARS-CoV-2.

SARS-CoV-2 enters human cells through the interaction between the receptor-binding domain (RBD) of its spike protein and the cell receptor ACE2, as is depicted in the schematic shown in Figure 3.1(a). Our study hence focuses on the spike protein, which is essential for the infection to take place. The spike protein of coronavirus is composed of an amino (N)-terminal S1 subunit and a carboxyl (C)-terminal S2 subunit. The S1 subunit, which consists of an N-terminal domain (NTD) and three C-terminal domains (CTD), is responsible for recognizing and binding to the host cell receptor. It has been reported that for the beta-coronavirus that utilizes CTD1 of its S1 subunit as RBD, there is a prerequisite conformational state for receptor binding where at least one RBD should be in upward

position [4, 21]. In this thesis, we refer to this receptor-accessible state as “open state” and receptor-inaccessible state as “closed state”. The SARS-CoV-2 spike protein in these two conformations is shown in Figure 3.1(b).

In the following parts of chapter, we conduct a normal mode analysis (NMA) [43, 44, 63] of different spike proteins in open and closed states respectively and analyze their detailed structures and flexibility variations (details see Chapter 2). Given the fact that only MERS-CoV, SARS-CoV and SARS-CoV-2 spike proteins have been captured in open sates, our open-state analysis is limited to these three highly pathogenic beta-coronaviruses. For closed-state exploration, since there have been no reports about closed conformational state of the MERS-CoV spike protein, we choose the HCoV-NL63 spike protein for the comparative study because it shares the same relevant cell receptor ACE2 with SARS-CoV and SARS-CoV-2 spike proteins.

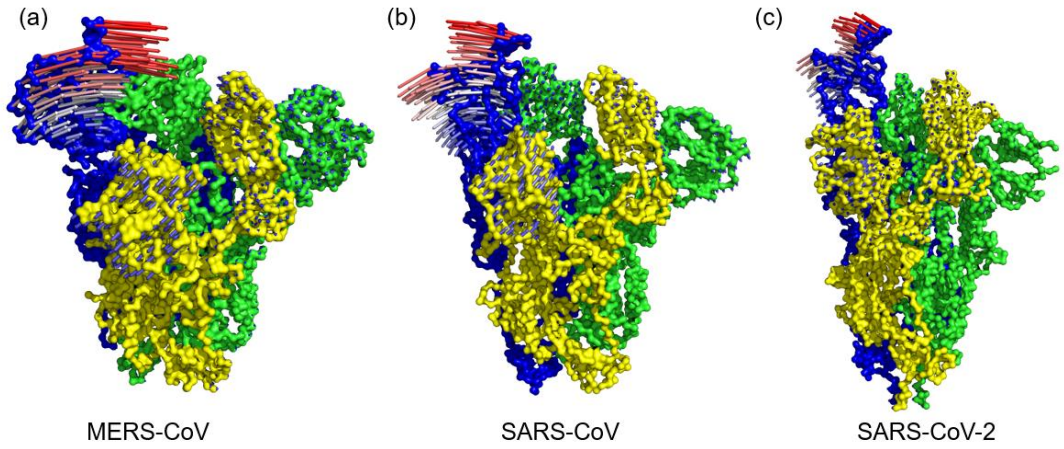

Figure 3. 2: Depiction of the lowest normal mode movement: swing of upward receptor-binding domain (RBD). (a) MERS-CoV spike protein (PDB: 5X5F); (b) SARS-CoV spike protein (PDB: 6CRZ); (c) SARS-CoV-2 spike protein (PDB: 6Z97). The three protomers of the above spike proteins are colored in blue, yellow and green, with the arrows denoting the corresponding vector fields.

Figure 3.2 shows that the lowest-frequency normal modes of MERS-CoV, SARS-CoV and SARS-CoV-2 spike proteins are all associated with a swing motion of upward receptor-binding domain (RBD) to different extents. This type of global movements corresponds well with the molecular motions directly reported in experiments[4, 21]. Considering the required receptor-accessible state for receptor binding, the swing of RBD is of functional

significance since it is the likely way by which a spike protein changes from closed state to open state to facilitate the binding of target receptors.

From a lateral aspect, this observed shared type of lowest-frequency normal mode movements indicates the structural similarity of MERS-CoV, SARS-CoV and SARS-CoV-2 spike proteins. According to our analysis, while the sequence identity of SARS-CoV S and SARS-CoV-2 S is as high as 78.8%, MERS-CoV S has only about 30% sequence identity with SARS-CoV S and SARS-CoV-2 S. This indicates that a small portion of the whole sequence of beta-coronavirus would largely determine the general structure topology and thus the overall global motion. (details see Appendix A) Previous literature[64] reveals that the percentage identity in NTD and RBD, which reflect the major parts participating in the lowest normal mode movement, is about 22%. Compared with 30% sequence identity of the whole spike protein, this partial percentage identity is lower and confirms the concept that probably only a few sequence pieces ultimately determine the shared global movement of coronavirus spike protein in open state. We further note that the sequence similarity in the S2 subunit could play an important role, as it may contributes to the comparatively higher rigidness of the S2 subunit.

3.2 Fluctuation Profiles and Structural Basis

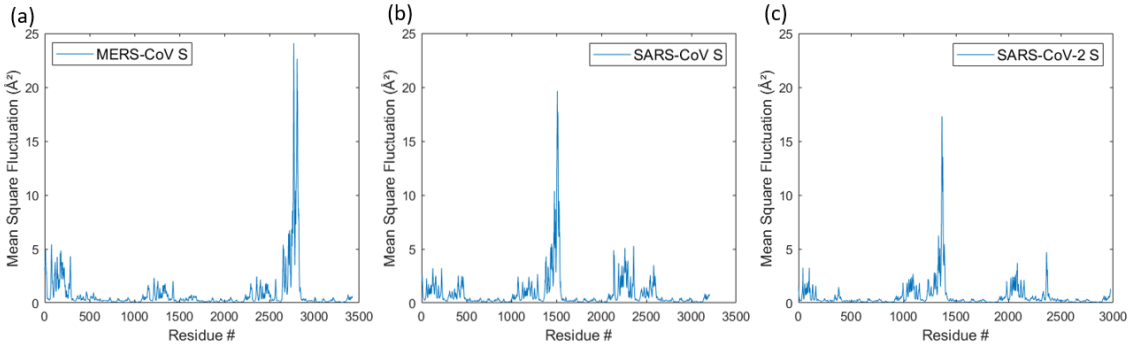

Figure 3. 3: Open-state fluctuation profiles of MERS-CoV, SARS-CoV and SARS-CoV-2 spike protein. Panels (a) to (c): Fluctuation profiles over the lowest 20 normal modes for MERS-CoV

(PDB: 5X5F), SARS-CoV (PDB: 6CRZ) and SARS-CoV-2 spike protein (PDB: 6Z97), respectively.

While MERS-CoV, SARS-CoV and SARS-CoV-2 spike proteins share the same type of lowest-frequency normal mode movements, their fluctuation profiles differ, as is illustrated by Figure 3.3. Generally speaking, the S1 subunit of coronavirus spike protein is much more flexible than its S2 subunit. Among these beta-coronaviruses, the active RBD of MERS-CoV spike enjoys super mobility and only SARS-CoV-2 spike protein exhibits a comparatively flexible downward RBD. Interestingly, when we focus on fluctuations over the upward RBD, the figure reveals that for all these spike proteins the fluctuations first slowly build up to its maximum and then decrease sharply, which demonstrates that the appearance of large flexibility of upward RBD is based on some common detailed structures of these beta-coronaviruses.

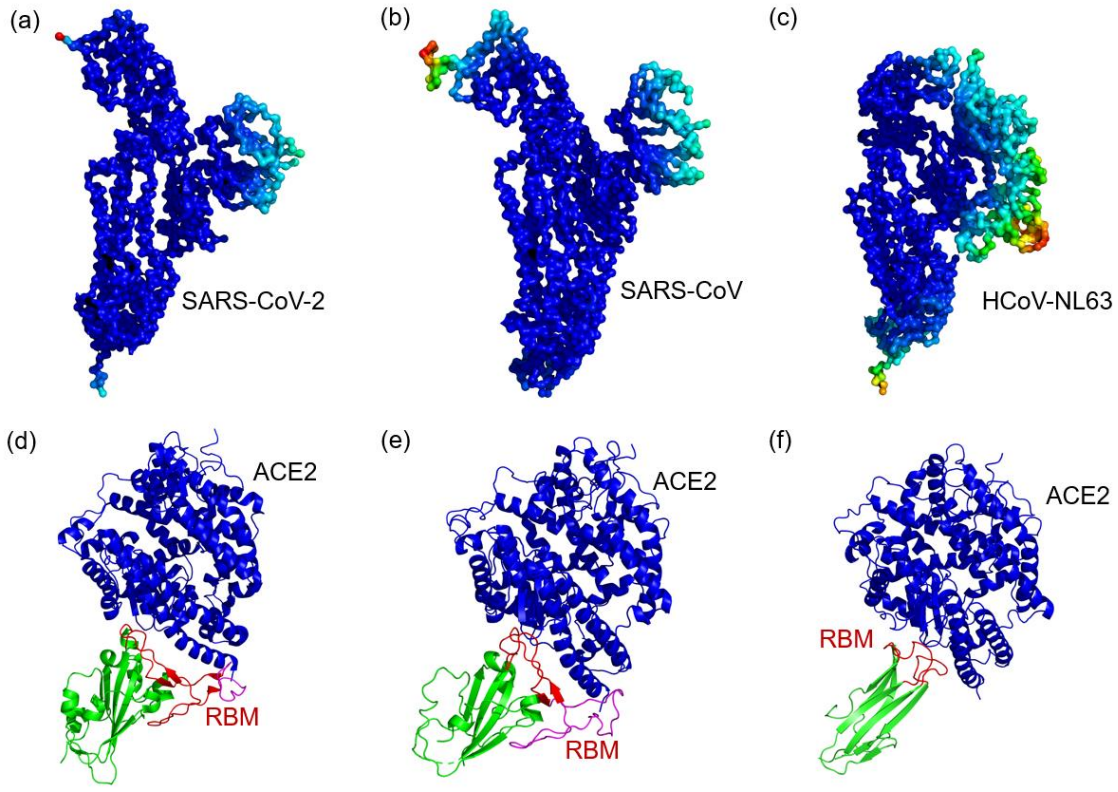

Figure 3. 4: Fluctuation visualization of the closed-state spike protein protomer and structural basis for flexibility variance. Panels (a) to (c): Fluctuation visualization of a protomer in closed-state SARS-CoV-2 (PDB: 6VXX), SARS-CoV (PDB: 5XLR) and HCoV-NL63 (PDB: 5SZS) spike protein, respectively. (d) to (f): The receptor binding domain of the above virus complexed with their shared receptor human ACE2 (PDB: 6M0J, 2AJF, 3KBH), where the receptor binding motif

(RBM) is colored in red and magenta. The loop colored in magenta is where the largest fluctuation occurs. The rest part of the receptor binding domain is colored in green, and the human ACE2 receptor is colored in blue.

We further explore the structural basis for the flexibility variation using normal mode analysis on coronavirus spike proteins in closed conformations. Panels (a) to (c) in Figure 3.4 show different flexibility variations in SARS-CoV-2, SARS-CoV and HCoV-NL63 spike protein, which share the same human receptor ACE2. Among them, there exists a sharp increase and large variation in flexibility in RBD of SARS-CoV-2 S, while SARS-CoV S appears to have more flexible structural regions. HCoV-NL63, the only one in this comparative study that is classified as alpha-coronavirus and cause only mild symptoms, has a different S1 subunit topology, bringing more flexibility to NTD rather than CTD1, where its RBD is situated. The result of closed-state fluctuation analysis shows that the similar structure in S1 subunit, especially CTD1, is the reason for the common flexibility distribution in beta-coronavirus spike proteins.

Panels (d) to (f) in Figure 3.4 provide detailed structures of the RBD of the above coronavirus complexed with their shared receptor human ACE2. The CTD1 in S1 subunit of coronavirus often contains a core structure, which is composed of several antiparallel β-sheet and short connecting α-helices, and one or more extended loops. These extended loops, referred to as receptor binding motif (RBM), is located at the edge of the core structure and is usually responsible for realizing the interactions with the receptor if the virus uses CTD1 as its RBD. Beta-coronaviruses such as SARS-CoV and SARS-CoV-2 have a unique long-extended loop as their RBM, accounting for the most flexible region of their RBD in both open and closed states. As for HCoV-NL63, there are three separated short RBMs, which are quite restricted and unable to generated large mobility. These insights may also explain its lower-affinity interaction with ACE2, at least to some extent.

3.3 Correlation of Vibrational and Epidemiological Properties

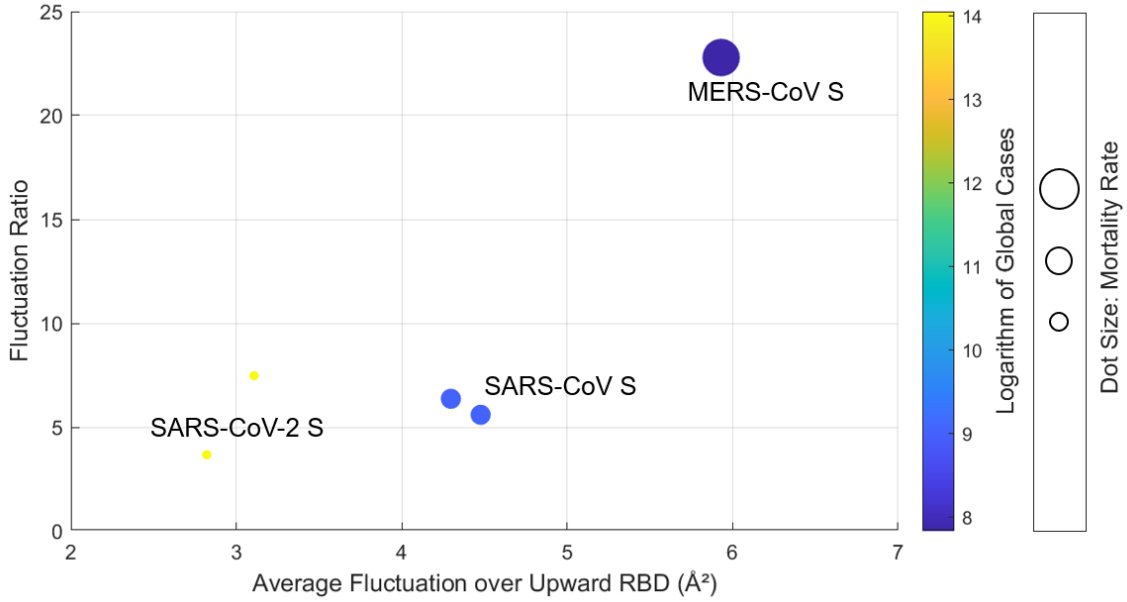

Figure 3. 5: Correlation diagram for SARS-CoV-2, SARS-CoV and MERS-CoV spike protein, linking nanomechanical, molecular vibrational measures with epidemiological data. The fluctuation ratio is calculated as the ratio of maximum fluctuations over upward and downward RBDs; the dot size denotes the mortality rate and the color bar shows the logarithm of global confirmed case numbers. Hence, the two smallest dots represent SARS-CoV-2 S (PDB: 6Z97, 7DDN), the two middle-sized dots represent SARS-CoV S (PDB: 6CRZ, 6CRW), and the largest sphere denotes MERS-CoV S (PDB: 5X5F).

Using the fluctuation profile data, two significant mechanical factors are identified, which are: (1) the overall flexibility of upward RBD and (2) the mobility ratio of RBDs in different conformations. Here, we show that these two vibrational properties can be linked directly to epidemiological virus characteristics. Figure 3.5 provides a correlation diagram for MERS-CoV, SARS-CoV and SARS-CoV-2 spike protein, where the overall flexibility of upward RBD is evaluated by the average fluctuation of open-state RBD and the mobility ratio is quantified as the ratio of maximum fluctuations over upward and downward RBDs. The data shows that both factors have positive correlation with case fatality rate and inverse relationship with the virus infectivity. We find that for the flexibility ratio, the smaller it is, the larger the mobility of downward RBD is compared to upward one, which could indicate a larger possibility toward a second standing RBD. This would make it even easier for the

spike protein to bind to the host cell receptor and thereby increasing the virus infectivity. On the other hand, if the flexibility of downward RBD is not large enough to generate conformational change, as flexibility ratio decreases, it becomes more difficult for the receptor to bind with the right RBD since the downward one is quite active. This may provide an explanation for the positive correlation between flexibility ratio and virus lethality that we see in the epidemiological data.

One possible reason for the positive relationship between overall flexibility of upward RBD and the mortality rate could be that the flexible upward RBD is more active when binding to the receptor, and may hence benefit the subsequent membrane fusion process. Even though there is limited available empirical data at this point, there is actually an intrinsic negative relationship between the mortality rate and infection rate of MERS-CoV, SARS-CoV and SARS-CoV-2, which could help explain why the infectivity is inversely correlated with the general flexibility of upward RBD. While there are many other factors situated between nanomechanical and epidemiological aspects, such as binding affinities [65] and dysregulation of type I interferon responses [66], the influence of which on different epidemiological characteristics of coronavirus has not been fully explained, and our analysis points out the direct correlation between nanomechanical features and the lethality and infection rate of coronavirus. The goal is to attempt to improve our understanding of the direct relationship between the nanoscale and epidemiological level, not considering internal relationships. As the results show, this perspective provides useful insights into the mechanics of disease relationships (Figure 1.4).

4 Nanomechanical Analysis of SARS-CoV-2 Variants

In this chapter, we report a predictive model that associates molecular motions and vibrational patterns of the virus spike protein with infectiousness and lethality. The key finding is that all the SARS-CoV-2 variants discussed in this thesis are more infectious and a bit less lethal than the original SARS-CoV-2.

4.1 Spike Mutations

As variants of the pathogen that causes COVID-19 spread around the world, estimates of infectiousness and lethality of newly emerging strains are important. Considering that spike protein is the part that SARS-CoV-2 and other similar coronaviruses use to bind to its human receptor and enter human cell [1, 21, 23, 67], we focus our analysis on the spike mutations of different SARS-CoV-2 variants.

Table 4. 1: Description of different SARS-CoV-2 variants of concern.

Variant Name	First Reported	Spike Mutations	Mutations Implemented in This Work
D614G	Europe	D614G	D614G
B.1.177 (20A.EU1)	Spain	D614G, A222V	D614G, A222V
CLUSTER 5	Denmark	del69-70, Y453F, D614G, I692V, M1229I	del69 (chain A), Y453F, D614G, I692V
OTHER MINK ASSOCIATED MUTATIONS	Europe	F486L	F486L
B.1.1.7	UK	del69-70, del Y144, N501Y, A570D, D614G, P681H, T716I, S982A, D1118H	del69 (chain A), N501Y, A570D, D614G, T716I, S982A, D1118H
B.1.351 (501Y.V2)	South Africa	L18F, D80A, D215G, R246I, K417N, E484K, N501Y, D614G, A701V	D215G, K417N, E484K, N501Y, D614G, A701V

In Table 4.1, we list all the SARS-CoV-2 variants that are taken into consideration in this chapter. Among them, the mutation D614G and variants B.1.1.7 and B.1.351 are still circulating globally. In this chapter, all the spike mutations are implemented based on the template SARS-CoV-2 spike protein structure (PDB: 6Z97), so a few mutations are not implemented due to missing regions. Future work may explore these mutations with similar methods as used here.

4.2 Fluctuation Profiles

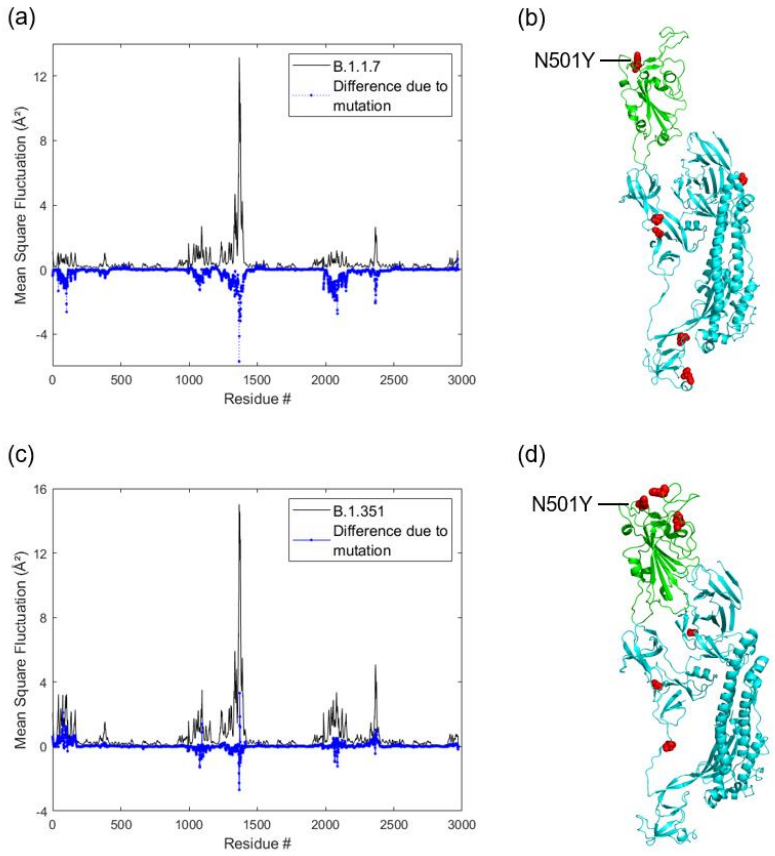

Figure 4. 1: Flexibility profile and spike mutation spots of the two most notable new SARS-CoV-2 variants that are still circulating globally—B.1.1.7 and B.1.351. (a) fluctuation profile of SARS-CoV-2 B.1.1.7 spike protein and the flexibility difference between this variant and the original spike protein (PDB: 6Z97). (b) equilibrated structure of SARS-CoV-2 B.1.1.7 spike protein, showing only the chain with upward receptor binding domain (RBD). The upward RBD is colored in green and other parts are colored in cyan. The mutation spots in this chain are colored in red and

shown as spheres. (c)(d) fluctuation profile and equilibrated structure of SARS-CoV-2 B.1.351 spike protein, respectively.

The fluctuation profile of coronavirus spike proteins, obtained through the use of normal mode analysis, reveals the vibrational flexibility of the spike. In previous chapter, we have shown that while having similar protein structure and sharing the same type of global movement—the swing of upward receptor-binding domain (RBD), MERS-CoV, SARS-CoV and SARS-CoV-2 spike proteins have quite different fluctuation profile.

Given the small number of spike mutations compared to the total residue number near 3,000, will the mutations make a difference to the flexibility profile? The answer is yes. In Figure 4.1, we show the flexibility profiles of two most notably new SARS-CoV-2 variants that are still circulating globally—B.1.1.7 and B.1.351. The fluctuation profiles of these two SARS-CoV-2 variants also follow the general characteristics that we discovered for MERS-CoV, SARS-CoV and SARS-CoV-2 spike proteins.

We find that they all have a much more flexible S1 subunit and a comparatively rigid S2 subunit, with the slowly built-up largest flexibility locating in its upward RBD. On the other hand, the mutations generally decrease the mobility of N-terminal domains (NTD) and RBDs. No obvious flexibility changes to other regions, which indicates that the rigid regions are pretty stable and remain its rigidity. Comparing B.1.1.7 and B.1.351 variants, which also known as 501Y.V1 and 501Y.V2 due to the shared N501Y mutation in RBD, there are some differences in their fluctuation profile. In general, the fluctuation difference induced by mutations in B.1.351 is smaller than the difference due to mutations in B.1.1.7, suggesting B.1.351 variant behaves more similar to the original SARS-CoV-2 spike protein. It is also important to note that compared with B.1.1.7, B.1.351 variant has larger mobility in both upward and downward RBD. While spike substitutions generally decrease the mobility of RBDs, several residues in the RBD of B.1.351 obtain increased flexibility due to mutations, showing some specialty of variant B.1.351.

4.3 Prediction of Infectivity and Lethality

Based on the correlation diagram of MERS-CoV, SARS-CoV and SARS-CoV-2 that links the epidemiological virus properties directly to pure nanomechanical, vibrational aspects, we construct a quantitative model to predict the infectivity and lethality level of beta-coronaviruses and their variants.

In our model, the infectivity level and lethality level are defined as the logarithm of global infection number and the logarithm of case fatality rate in percentage, ensuring that negative values are also meaningful. When constructing the quantitative correlation model, we investigate the accuracy of polynomial model, mixture model of polynomial and exponential function and exponential model. Given the limited data, only linear dependency is explored, in the form:

$$y = a_0 + a_1 f_1(x_1) + a_2 f_2(x_2) \tag{4.1}$$

where y is the model target, a_i are constants to be solved, x_i are factors denoting vibrational behavior, and $f_i(x_i)$ are trial functions. Although appropriate mixture model of polynomial and exponential function will slightly increase the model accuracy, the pure polynomial model is good enough to describe the relationship between vibrational and epidemiological properties with a correlation coefficient $R^2 > 0.99$.

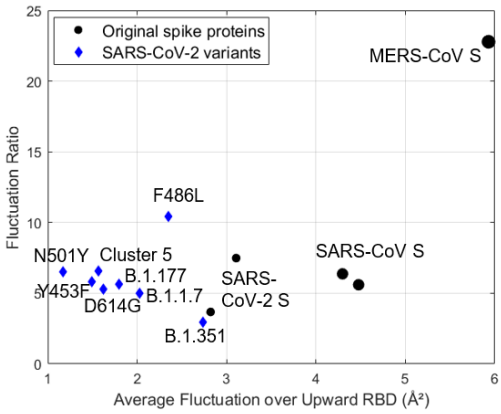

Figure 4. 2: Nanomechanical vibrational feature map of coronavirus spike proteins and SARS-CoV-2 variants. The largest sphere denotes MERS-CoV S (PDB ID: 5X5F); two middle-size

spheres denote SARS-CoV S (PDB ID: 6CRZ and 6CRW); two small spheres denote SARS-CoV-2 S (PDB ID: 6Z97 and 7DDN). Points in diamond shape denote spike proteins of different SARS-CoV-2 variants.

Figure 4.2 depicts the nanomechanical vibrational feature map of coronavirus spike proteins and SARS-CoV-2 variants of concern. It is demonstrated that generally all variants of concern have decreased average fluctuation over upward RBD, which corresponds with the trend we observe with detailed fluctuation profiles in Figure 4.1. Compared with the original SARS-CoV-2 spike protein, the variants are moving towards the bottom left side of the diagram, suggesting that their behavior differs even more from the behavior of SARS-CoV and MERS-CoV spike proteins.

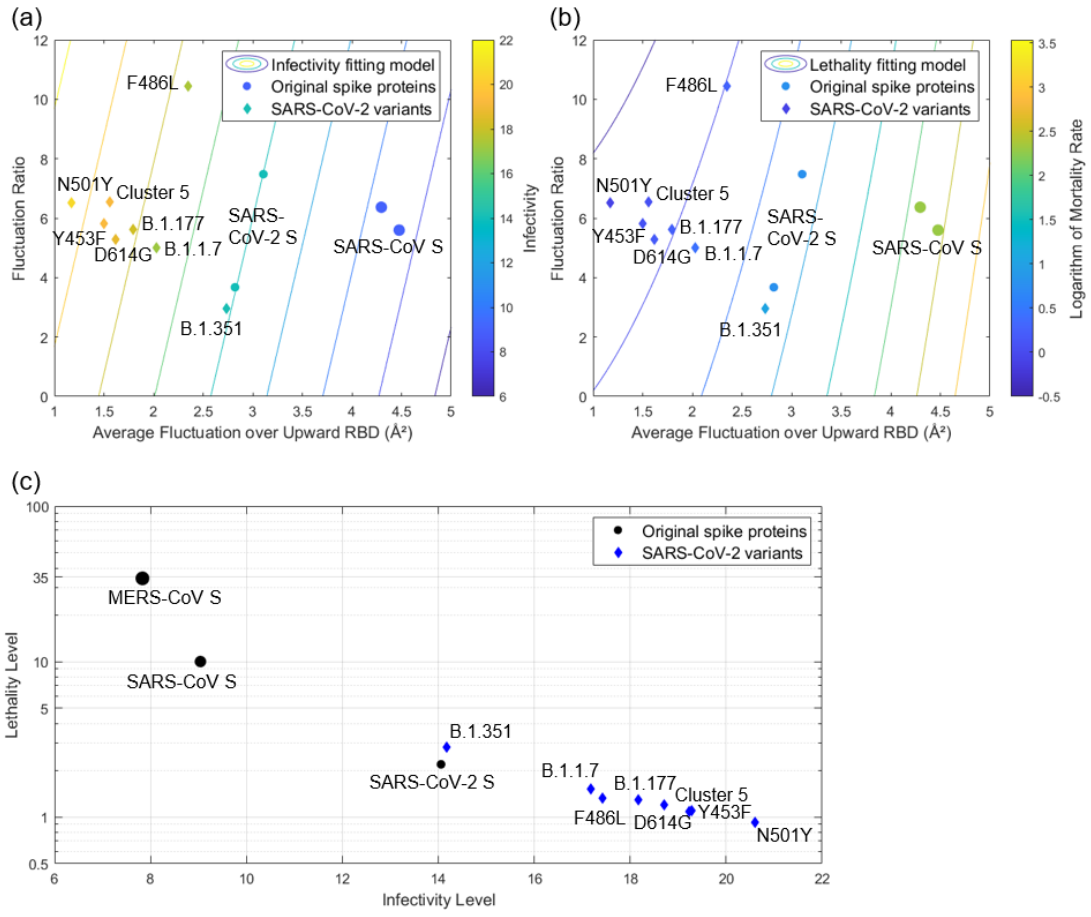

Figure 4. 3: Quantitative model predicting the infectivity and lethality level of SARS-CoV-2 variants of concern. (a) Linear fitting model linking fluctuation level and ratio to infectivity, where infectivity level is calculated as the logarithm of global confirmed case number. (b) Quadratic fitting model relating fluctuation level and ratio to lethality, where lethality level is calculated as

the logarithm of mortality rate in percentage. (c) Prediction on epidemiological features of SARS-CoV-2 variants.

Aside from the qualitative conclusion drawn from Figure 4.2, we establish our quantitative model to further make predictions about the infectivity and lethality level of SARS-CoV-2 variants of concern based on the relationship between vibrational and epidemiological properties[62]. As shown in Figure 4.3, the polynomial fitting model is able to describe the quantitative relationship appropriately, with correlation coefficient $R^2 = 0.990$ and $R^2 = 0.993$ for infectivity and lethality fitting model respectively. The analysis indicates that all variants of concern are predicted to be more infectious and less lethal than the original SARS-CoV-2, with B.1.351 lineage having the most similar performance to the original virus.

Focusing on the mutations associated with mink, we find that although the single mutation F486L seems to be significantly different from the other two variants in the vibrational feature map, their predicted lethality levels are quite similar. The similar prediction result for single mutation Y453F and cluster 5 variant, which contains Y453F and other two spike mutations, proves the stability of our model to some extent.

Here, we can also further compare the predicted epidemiological characteristics of B.1.351, B.1.1.7 variants and their shared N501Y spike substitution. The analysis shows that the three of them behave quite differently, with single mutation N501Y predicted to be the most infectious and the least lethal. The different behavior can be considered as the result of the competition between N501Y mutation and other spike mutations— N501Y pushes the variant to be more infectious and less lethal, and other mutations together drag the variant towards lower infectivity and higher lethality. This analysis result agrees with previous findings revealing that N501Y is associated with increased infectivity and virulence of SARS-CoV-2 in a mouse model[68]. Notably, we emphasize here that other spike mutations, which have not been discussed in previous research work, could also have significant impact on infectivity and lethality of SARS-CoV-2 variants, providing new aspects for further research on SARS-CoV-2 variants.

In Figure 4.3(c), it is shown that only B.1.351 variant is predicted to be more infectious and more lethal compared to the original spike protein, and all other variants of concern follow a trend of higher infectivity and less lethality. We demonstrate in Figure 4.4 that the possibility for the variants to be both more lethal and more infectious could change with different vibrational properties. The possibility is a bit larger for SARS-CoV-2 S compared with that of SARS-CoV S and MERS-CoV S. In the bottom left region of Figure 4.4, the nonlinearity of lethality prediction becomes significant, and thus the likelihood of more lethal and more infectious variants is quite large. Therefore, dangerous mutations could be a serious concern with future coronavirus whose spike protein has comparatively small flexibility level and fluctuation ratio.

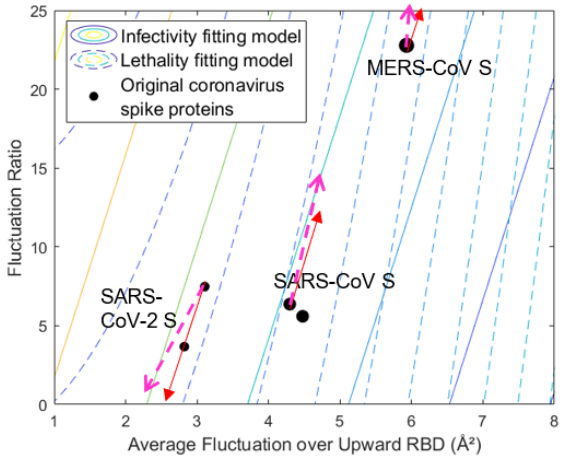

Figure 4. 4: Nanomechanical vibrational feature map indicating the possibility of more infectious and more lethal variant. Arrow in red indicates the direction where variants are of the same infectivity but more lethal; arrow in magenta indicates the direction where variants are of the same lethality but more infectious. Small area between two arrows is the phase space to generate both more infectious and more lethal variant.

5 Conclusion

We reported an analysis linking key nanomechanical vibrational features of various coronavirus spike proteins with epidemiological data. As shown in Figure 3.2, structural similarity and major movement associated with the swing of upward RBD is seen throughout this family of viruses, representing a sort of universal feature of this class of pathogens. The molecular modeling results show that the general motion corresponds with experimental observation and have functional importance. We find that the active RBD of MERS-CoV enjoys super mobility, whereas only SARS-CoV-2 shows a comparatively flexible downward RBD.

As shown in Figure 3.5, the general fluctuation profiles of upward RBD and the associated fluctuation ratio have a positive correlation with case fatality rate and inverse relationship with the virus infectivity. Our results offer two different explanations for the effects of the flexible downward RBD: (1), the possibility towards a second standing RBD if the mobility is large enough, and (2) that it could indicate difficulty for the receptor to bind with the right RBD if no conformational change happens. We hypothesize that there may be a possible threshold between these two effects, which could be studied in future research.

These insights offer several possible applications, including a search for inhibitors that could bind to downward RBD may provide a viable strategy. We find a sharp increase and large variation in flexibility in SARS-CoV-2 S, whereas more flexible structural regions are present in the closed-state SARS-CoV S. The long-extended loop is unique for beta-coronaviruses, and also accounts for the most flexible region in open states. As for possible applications, we may target the RBM to identify new inhibitors that lock the closed S-protein conformation. This opens a question whether perhaps, we can we utilize the significant flexibility difference for potential inhibitor screening for the development of novel treatment methods or design future experiments for gain-of-function research that may help us in preventing future outbreaks, or to understand evolutionary principles behind the change of various coronaviruses and when they transition across species [69, 70].

Moreover, a predictive model that associates molecular motions and vibrational patterns of the virus spike protein with infectiousness and lethality was constructed and analyzed in this thesis. As shown in Figure 4.3, almost all variants of concern are predicted to have higher infectivity and lethality level than the original SARS-CoV-2, with only B.1.351 lineage predicted to be more infectious and more lethal than the original virus. The different behavior of B.1.351 and B.1.1.7 variants suggests that spike mutations other than N501Y, which hasn't been discussed in previous literature, also have significant influence on the infectivity and lethality of SARS-CoV-2 variants. While generally, the possibility of both more lethal and more infectious variants is quite small, it may indeed become a serious concern when both general flexibility level and fluctuation ratio of coronavirus spike protein is small. This model could provide a tool to estimate the epidemiological effects of new variants, and offer a pathway to screen mutations against high threat levels. Moreover, the nanomechanical approach, as a novel tool to predict virus-cell interactions, may offer a tool to better understand other viruses. Other future work may assess such potential mutations from an evolutionary or genetic perspective, to assess the likelihood of their occurrence, or to identify the conditions under which they may emerge.

Other future work may also address additional influences of temperature dependence (to explore whether seasonal variations of temperature and other environmental factors can be linked to nanoscopic phenomena). Additional aspects may include a more detailed analysis of intermediate steps in the mechanistic hierarchical progression as schematically outlined in Figure 1.4, including aspects of the strong age-dependence of COVID-19 disease progression.

Appendix A – Sequence Alignment

Figure A. 1: Sequence alignment of MERS-CoV S, SARS-CoV S and SARS-CoV-2 S.

References

1. Wu, K., Li, W., Peng, G., & Li, F. (2009). Crystal structure of NL63 respiratory coronavirus receptor-binding domain complexed with its human receptor. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(47), 19970–19974. https://doi.org/10.1073/pnas.0908837106
2. Humphrey, W., Dalke, A., & Schulten, K. (1996). Sartorius products. *Journal of molecular graphics*, *14*(October 1995), 33–38. https://doi.org/10.1016/0263-7855(96)00018-5
3. Wan, Y., Shang, J., Graham, R., Baric, R. S., & Li, F. (2020). Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. *Journal of Virology*, *94*(7), 1–9. https://doi.org/10.1128/jvi.00127-20
4. Gui, M., Song, W., Zhou, H., Xu, J., Chen, S., Xiang, Y., & Wang, X. (2017). Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding. *Cell Research*, *27*(1), 119–129. https://doi.org/10.1038/cr.2016.152
5. Van Der Hoek, L., Pyrc, K., & Berkhout, B. (2006). Human coronavirus NL63, a new respiratory virus. *FEMS Microbiology Reviews*, *30*(5), 760–773. https://doi.org/10.1111/j.1574-6976.2006.00032.x
6. Lau, S. K. P., Woo, P. C. Y., Yip, C. C. Y., Tse, H., Tsoi, H. W., Cheng, V. C. C., … Yuen, K. Y. (2006). Coronavirus HKU1 and other coronavirus infections in Hong Kong. *Journal of Clinical Microbiology*, *44*(6), 2063–2071. https://doi.org/10.1128/JCM.02614-05
7. Gaunt, E. R., Hardie, A., Claas, E. C. J., Simmonds, P., & Templeton, K. E. (2010). Epidemiology and clinical presentations of the four human coronaviruses 229E, HKU1, NL63, and OC43 detected over 3 years using a novel multiplex real-time PCR method. *Journal of Clinical Microbiology*, *48*(8), 2940–2947. https://doi.org/10.1128/JCM.00636-10
8. De Wit, E., Van Doremalen, N., Falzarano, D., & Munster, V. J. (2016). SARS and MERS: Recent insights into emerging coronaviruses. *Nature Reviews Microbiology*, *14*(8), 523–534. https://doi.org/10.1038/nrmicro.2016.81
9. Zaki, A. M., Van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. M. E., & Fouchier, R. A. M. (2012). Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *New England Journal of Medicine*, *367*(19), 1814–1820. https://doi.org/10.1056/NEJMoa1211721
10. Su, S., Wong, G., Liu, Y., Gao, G. F., Li, S., & Bi, Y. (2015). MERS in South Korea and China: A potential outbreak threat? *The Lancet*, *385*(9985), 2349–2350. https://doi.org/10.1016/S0140-6736(15)60859-5
11. Chan, J. F. W., Yuan, S., Kok, K. H., To, K. K. W., Chu, H., Yang, J., … Yuen, K. Y. (2020). A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *The Lancet*, *395*(10223), 514–523. https://doi.org/10.1016/S0140-6736(20)30154-9
12. Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., … Cao, B. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The*

Lancet, *395*(10223), 497–506. https://doi.org/10.1016/S0140-6736(20)30183-5
13. COVID-19 Map - Johns Hopkins Coronavirus Resource Center. (n.d.).
14. Verity, R., Okell, L. C., Dorigatti, I., Winskill, P., Whittaker, C., Imai, N., … Ferguson, N. M. (2020). Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet Infectious Diseases*, *20*(6), 669–677. https://doi.org/10.1016/S1473-3099(20)30243-7
15. Yurkovetskiy, L., Wang, X., Pascal, K. E., Tomkins-Tinch, C., Nyalile, T. P., Wang, Y., … Luban, J. (2020). Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. *Cell*, *183*(3), 739-751.e8. https://doi.org/10.1016/j.cell.2020.09.032
16. Hou, Y. J., Chiba, S., Halfmann, P., Ehre, C., Kuroda, M., Dinnon, K. H., … Baric, R. S. (2020). SARS-CoV-2 D614G variant exhibits enhanced replication ex vivo and earlier transmission in vivo. *Science*, *370*(6523), 1464–1468. https://doi.org/10.1126/science.abe8499
17. Davies, N. G., Abbott, S., Barnard, R. C., Jarvis, C. I., Kucharski, A. J., Munday, J. D., … Edmunds, W. J. (2021). Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science*, *372*(6538), eabg3055. https://doi.org/10.1126/science.abg3055
18. Wang, Z., Schmidt, F., Weisblum, Y., Muecksch, F., Barnes, C. O., Finkin, S., … Nussenzweig, M. C. (2021). mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *Nature*, *592*(April). https://doi.org/10.1038/s41586-021-03324-6
19. Kirchdoerfer, R. N., Cottrell, C. A., Wang, N., Pallesen, J., Yassine, H. M., Turner, H. L., … Ward, A. B. (2016). Pre-fusion structure of a human coronavirus spike protein. *Nature*, *531*(7592), 118–121. https://doi.org/10.1038/nature17200
20. Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., & Zhou, Q. (2020). Structural basis for the recognition of the SARS-CoV-2 by full-length human ACE2. *Science (New York, N.Y.)*, *367*(6485), 1444–1448. https://doi.org/10.1126/science.abb2762
21. Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C.-L., Abiona, O., … McLellan, J. S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science (New York, N.Y.)*, *367*(6483), 1260–1263. https://doi.org/10.1126/science.abb2507
22. Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., … Li, F. (2020). Structural basis of receptor recognition by SARS-CoV-2. *Nature*, *581*(7807), 221–224. https://doi.org/10.1038/s41586-020-2179-y
23. Li, F., Li, W., Farzan, M., & Harrison, S. C. (2005). Structural biology: Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science*, *309*(5742), 1864–1868. https://doi.org/10.1126/science.1116480
24. Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., … Wang, X. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*, *581*(7807), 215–220. https://doi.org/10.1038/s41586-020-2180-5
25. Lu, G., Hu, Y., Wang, Q., Qi, J., Gao, F., Li, Y., … Gao, G. F. (2013). Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26. *Nature*, *500*(7461), 227–231. https://doi.org/10.1038/nature12328
26. Yuan, Y., Cao, D., Zhang, Y., Ma, J., Qi, J., Wang, Q., … Gao, G. F. (2017). Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the

dynamic receptor binding domains. *Nature Communications*, *8*(China CDC), 1–9. https://doi.org/10.1038/ncomms15092
27. Chen, J., Wang, R., Wang, M., & Wei, G. W. (2020). Mutations Strengthened SARS-CoV-2 Infectivity. *Journal of Molecular Biology*, *432*(19), 5212–5226. https://doi.org/10.1016/j.jmb.2020.07.009
28. Goyal, A., Reeves, D. B., Fabian Cardozo-Ojeda, E., Schiffer, J. T., & Mayer, B. T. (2021). Viral load and contact heterogeneity predict sars-cov-2 transmission and super-spreading events. *eLife*, *10*, 1–63. https://doi.org/10.7554/eLife.63537
29. Pujadas, E., Chaudhry, F., McBride, R., Richter, F., Zhao, S., Wajnberg, A., … Cordon-Cardo, C. (2020, September 1). SARS-CoV-2 viral load predicts COVID-19 mortality. *The Lancet Respiratory Medicine*. Lancet Publishing Group. https://doi.org/10.1016/S2213-2600(20)30354-4
30. Hie, B., Zhong, E. D., Berger, B., & Bryson, B. (2021). Learning the language of viral evolution and escape. *Science*, *371*(6526), 284–288. https://doi.org/10.1126/science.abd7331
31. Gautieri, A., Vesentini, S., Redaelli, A., & Buehler, M. J. (2011). Hierarchical structure and nanomechanics of collagen microfibrils from the atomistic scale up. *Nano Letters*, *11*(2), 757–766. https://doi.org/10.1021/nl103943u
32. Qin, Z., Kreplak, L., & Buehler, M. J. (2009). Hierarchical structure controls nanomechanical properties of vimentin intermediate filaments. *PLoS ONE*, *4*(10). https://doi.org/10.1371/journal.pone.0007294
33. Lu, Y., Neo, T. L., Liu, D. X., & Tam, J. P. (2008). Importance of SARS-CoV spike protein Trp-rich region in viral infectivity. *Biochemical and Biophysical Research Communications*, *371*(3), 356–360. https://doi.org/10.1016/j.bbrc.2008.04.044
34. Du, L., Kao, R. Y., Zhou, Y., He, Y., Zhao, G., Wong, C., … Zheng, B. J. (2007). Cleavage of spike protein of SARS coronavirus by protease factor Xa is associated with viral infectivity. *Biochemical and Biophysical Research Communications*, *359*(1), 174–179. https://doi.org/10.1016/j.bbrc.2007.05.092
35. Arslana, M., Qin, Z., & Buehler, M. J. (2011). Coiled-coil intermediate filament stutter instability and molecular unfolding. *Computer Methods in Biomechanics and Biomedical Engineering*, *14*(5). https://doi.org/10.1080/10255842.2011.560147
36. Buehler, M. J., & Yung, Y. C. (2009). Deformation and failure of protein materials in physiologically extreme conditions and disease. *Nature Materials*, *8*(3), 175–188. https://doi.org/10.1038/nmat2387
37. Keten, S., Xu, Z., & Buehler, M. J. (2011). Triangular core as a universal strategy for stiff nanostructures in biology and biologically inspired materials. *Materials Science and Engineering C*, *31*(4). https://doi.org/10.1016/j.msec.2011.01.004
38. Buehler, M. J. (2020). Nanomechanical sonification of the 2019-nCoV coronavirus spike protein through a materiomusical approach. Retrieved from http://arxiv.org/abs/2003.14258
39. Tarakanova, A., Yeo, G. C., Baldock, C., Weiss, A. S., & Buehler, M. J. (2018). Molecular model of human tropoelastin and implications of associated mutations. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(28), 7338–7343. https://doi.org/10.1073/pnas.1801205115

40. Knowles, T. P. J., & Buehler, M. J. (2011). Nanomechanics of functional and pathological amyloid materials. *Nature Nanotechnology*, *6*(8). https://doi.org/10.1038/nnano.2011.102
41. Yavuz, B., Morgan, J. L., Herrera, C., Harrington, K., Perez-Ramirez, B., LiWang, P. J., & Kaplan, D. L. (2019). Sustained release silk fibroin discs: Antibody and protein delivery for HIV prevention. *Journal of Controlled Release*, *301*, 1–12. https://doi.org/10.1016/j.jconrel.2019.03.001
42. Tama, F., Gadea, F. X., Marques, O., & Sanejouand, Y. H. (2000). Building-block approach for determining low-frequency normal modes of macromolecules. *Proteins: Structure, Function and Genetics*, *41*(1), 1–7. https://doi.org/10.1002/1097-0134(20001001)41:1<1::AID-PROT10>3.0.CO;2-P
43. Bahar, I., Lezon, T. R., Yang, L. W., & Eyal, E. (2010). Global dynamics of proteins: Bridging between structure and function. *Annual Review of Biophysics*, *39*(1), 23–42. https://doi.org/10.1146/annurev.biophys.093008.131258
44. Qin, Z., & Buehler, M. J. (2019). Analysis of the vibrational and sound spectrum of over 100,000 protein structures and application in sonification. *Extreme Mechanics Letters*, *29*, 100460. https://doi.org/10.1016/j.eml.2019.100460
45. Durand, P., Trinquier, G., & Sanejouand, Y. -H. (1994). A new approach for determining low-frequency normal modes in macromolecules. *Biopolymers*, *34*(6), 759–771. https://doi.org/10.1002/bip.360340608
46. Hinsen, K., Petrescu, A. J., Dellerue, S., Bellissent-Funel, M. C., & Kneller, G. R. (2000). Harmonicity in slow protein dynamics. *Chemical Physics*, *261*(1–2), 25–37. https://doi.org/10.1016/S0301-0104(00)00222-6
47. Brooks, B., & Karplus, M. (1985). Normal modes for specific motions of macromolecules: application to the hinge-bending mode of lysozyme. *Proceedings of the National Academy of Sciences of the United States of America*, *82*(15), 4995–4999. https://doi.org/10.1073/pnas.82.15.4995
48. Levy, R. M., Perahia, D., & Karplus, M. (1982). Molecular dynamics of an -helical polypeptide: Temperature dependence and deviation from harmonic behavior. *Proceedings of the National Academy of Sciences*, *79*(4), 1346–1350. https://doi.org/10.1073/pnas.79.4.1346
49. Scott A. Hollingsworth, & Ron O. Dror. (2018). Molecular dynamics simulation for all. *Neuron*, *99(6)*(1), 1129–1143. https://doi.org/10.1016/j.neuron.2018.08.011.Molecular
50. Hospital, A., Goñi, J. R., Orozco, M., & Gelpí, J. L. (2015). Molecular dynamics simulations: Advances and applications. *Advances and Applications in Bioinformatics and Chemistry*, *8*(1), 37–47. https://doi.org/10.2147/AABC.S70333
51. Brunger, A. T., & Adams, P. D. (2002). Molecular dynamics applied to X-ray structure refinement. *Accounts of Chemical Research*, *35*(6), 404–412. https://doi.org/10.1021/ar010034r
52. Delemotte, L., Tarek, M., Klein, M. L., Amaral, C., & Treptow, W. (2011). Intermediate states of the Kv1.2 voltage sensor from atomistic molecular dynamics simulations. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(15), 6109–6114. https://doi.org/10.1073/pnas.1102724108
53. Rueda, M., Ferrer-Costa, C., Meyer, T., Pérez, A., Camps, J., Hospital, A., … Orozco, M. (2007). A consensus view of protein dynamics. *Proceedings of the*

National Academy of Sciences of the United States of America, *104*(3), 796–801. https://doi.org/10.1073/pnas.0605534104
54. Pérez, A., Lankas, F., Luque, F. J., & Orozco, M. (2008). Towards a molecular dynamics consensus view of B-DNA flexibility. *Nucleic Acids Research*, *36*(7), 2379–2394. https://doi.org/10.1093/nar/gkn082
55. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., … Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Res*, *28*(1), 235–242.
56. Berman, H. M., Battistuz, T., Bhat, T. N., Bluhm, W. F., Bourne, P. E., Burkhardt, K., … Zardecki, C. (2002). The protein data bank. *Acta Crystallographica Section D: Biological Crystallography*, *58*(6 I), 899–907. https://doi.org/10.1107/S0907444902003451
57. Nelson, M. T., Humphrey, W., Gursoy, A., Dalke, A., Kale, L. V., Skeel, R. D., & Schulten, K. (1996). NAMD: A parallel, object-oriented molecular dynamics program. *International Journal of High Performance Computing Applications*, *10*(4), 251–268. https://doi.org/10.1177/109434209601000401
58. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., … Schulten, K. (2005). Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, *26*(16), 1781–1802. https://doi.org/10.1002/jcc.20289
59. Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, *14*(1), 33–38. https://doi.org/10.1016/0263-7855(96)00018-5
60. Korber, B., Fischer, W. M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., … Montefiori, D. C. (2020). Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell*, *182*(4), 812-827.e19. https://doi.org/10.1016/j.cell.2020.06.043
61. COVID-19 Viral Genome Analysis Pipeline. (2020).
62. Hu, Y., & Buehler, M. J. (2021). Comparative Analysis of Nanomechanical Features of Coronavirus Spike Proteins and Correlation with Lethality and Infection Rate. *Matter*, *4*(1), 265–275. https://doi.org/10.1016/j.matt.2020.10.032
63. Rodrigues, C. H. M., Pires, D. E. V., & Ascher, D. B. (2018). DynaMut: Predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic Acids Research*, *46*(W1), W350–W355. https://doi.org/10.1093/nar/gky300
64. Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., … Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research*, *47*(W1), W636–W641. https://doi.org/10.1093/nar/gkz268
65. Shang, J., Wan, Y., Luo, C., Ye, G., Geng, Q., Auerbach, A., & Li, F. (2020). Cell entry mechanisms of SARS-CoV-2. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(21). https://doi.org/10.1073/pnas.2003138117
66. Acharya, D., Liu, G. Q., & Gack, M. U. (2020). Dysregulation of type I interferon responses in COVID-19. *Nature Reviews Immunology*, *20*(7), 397–398. https://doi.org/10.1038/s41577-020-0346-x
67. Nishiga, M., Wang, D. W., Han, Y., Lewis, D. B., & Wu, J. C. (2020, September). COVID-19 and cardiovascular disease: from basic mechanisms to clinical perspectives. *Nature Reviews Cardiology*. Nature Research.

https://doi.org/10.1038/s41569-020-0413-9
68. Gu, H., Chen, Q., Yang, G., He, L., Fan, H., Deng, Y., … Zhou, Y. (2020). Adaptation of SARS-CoV-2 in BALB/c mice for testing vaccine efficacy. *Science*, *369*(6511), 1603–1607.
69. Sirotkin, K., & Sirotkin, D. (2020). Might SARS-CoV-2 Have Arisen via Serial Passage through an Animal Host or Cell Culture?: A potential explanation for much of the novel coronavirus' distinctive genome. *BioEssays*, *42*(10), 1–7. https://doi.org/10.1002/bies.202000091
70. Evans, N. G. (2018). Ethical and Philosophical Considerations for Gain-of-Function Policy: The Importance of Alternate Experiments. *Frontiers in Bioengineering and Biotechnology*, *6*(FEB), 11. https://doi.org/10.3389/fbioe.2018.00011