


 Cite this: *Soft Matter*, 2022, 18, 5833

## Nanomechanical analysis of SARS-CoV-2 variants and predictions of infectiousness and lethality†

 Yiwen Hu<sup>ab</sup> and Markus J. Buehler<sup>id</sup>\*<sup>abc</sup>

As variants of the pathogen that causes COVID-19 spread around the world, estimates of infectiousness and lethality of newly emerging strains are important. Here 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 most SARS-CoV-2 variants are predicted to be more infectious and less lethal compared to the original spike protein. However, lineage B.1.351 (Beta variant) is predicted to be less infectious and more lethal, and lineage B.1.1.7 (Alpha variant) is predicted to have both higher infectivity and lethality, showing the potential of the virus to mutate towards different performance regimes. The relatively more recent lineage B.1.617.2 (Delta variant), although contains a few key spike mutations other than D614G, behaves quite similar to the single D614G mutation in both vibrational and predicted epidemiological aspects, which might explain its rapid circulation given the prevalence of D614G. This work may 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 further open up the door towards better understanding other viruses.

 Received 13th August 2021,  
 Accepted 20th July 2022

DOI: 10.1039/d1sm01181b

[rsc.li/soft-matter-journal](https://rsc.li/soft-matter-journal)

### Introduction

As variants of the pathogen that causes COVID-19 spread around the world, estimates of infectiousness and lethality of newly emerging strains are important. Here we report, building off earlier work<sup>1</sup> in which we correlated infectiousness and lethality of different types of coronavirus pathogens such as Middle East Respiratory Syndrome (MERS), Severe Acute Respiratory Syndrome (SARS) and Coronavirus Disease 2019 (COVID-19), a predictive model that associates molecular motions and vibrational patterns of the virus spike protein with infectiousness and lethality.

First reported in China in 2019,<sup>2–4</sup> the novel coronavirus disease, COVID-19, has spread rapidly around the world, resulting in more than 180 million global cases as of July 2021.<sup>5,6</sup> SARS-CoV-2, the virus that causes COVID-19, can easily transmit from human to human, compared to previous two highly pathogenic human coronaviruses—namely, SARS-CoV and

MERS-CoV. Similar to other human coronaviruses, SARS-CoV-2 enters human cells through the interaction between the receptor-binding domain (RBD) of its spike protein and the human cell receptor, and it shares the same human receptor angiotensin-converting enzyme 2 (ACE2) with SARS-CoV.<sup>7</sup> Composed of an amino (N)-terminal S1 subunit and a carboxyl (C)-terminal S2 subunit, the spike protein of coronavirus bears functional significance in the infection process. While the S2 subunit participates more in the membrane fusion, the S1 subunit, consisting of an N-terminal domain (NTD), the receptor-binding domain (RBD) and two smaller intermediate subdomains, is responsible for recognizing and binding to the host cell receptor.<sup>8</sup> Previous literature reveals a prerequisite conformational state for receptor binding where at least one RBD should be in upward position, which makes it receptor-accessible, otherwise there would be steric clashes hindering the binding process.<sup>9</sup> In experimental work, this specific type of receptor-accessible state has been captured for MERS-CoV, SARS-CoV and SARS-CoV-2.<sup>7,8</sup> Given the essential role that the spike protein plays in receptor recognition, viral fusion and cell entry,<sup>7,10,11</sup> it is also a promising target for drug and vaccine development.

While vaccination progress has been progressing well in many countries across the world, an emergence of more transmissible SARS-CoV-2 variants has caused great concern.<sup>12,13</sup> SARS-CoV-2 D614G variant emerged in late January or early February 2020 and has quickly become the most prevalent form in the global pandemic.<sup>14</sup> Not only does the D614G variant adopts a range of

<sup>a</sup> *Laboratory for Atomistic and Molecular Mechanics (LAMM), Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139, USA. E-mail: mbuehler@mit.edu; Tel: +1-617-452-2750*

<sup>b</sup> *Department of Mechanical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139, USA*

<sup>c</sup> *Center for Computational Science and Engineering, Schwarzman College of Computing, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139, USA*

† Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d1sm01181b>



more open conformations than the original virus, but it also has added interactions to prevent premature dissociation of the trimer and increase the number of spikes that can facilitate infection.<sup>15,16</sup> A number of experimental and statistical research have also shown that D614G variant has enhanced infectivity and transmissibility compared to the wild-type virus.<sup>14,17,18</sup> The lineage B.1.1.7 (Alpha variant), first reported in the UK in December 2020, is predicted to be 43 to 90% more transmissible than the predecessor lineage based on statistical and dynamic modeling approaches.<sup>12</sup> Moreover, transmission of SARS-CoV-2 between human and animals has also been observed, with the most notable mink mutation reported in Denmark causing the culling of all farmed mink in the country.<sup>19</sup> Experimental evidence reveals that while mink-specific mutations do not increase entry into human cells, they do cause partial evasion from neutralization by a therapeutic antibody and convalescent plasma/sera.<sup>20</sup> A concerning finding is that, an experiment suggests that mRNA-vaccine activity against SARS-CoV-2 variants that encode E484K, N501Y- or K417N/E484K/N501Y-mutant spike may be reduced by a small—but significant—margin.<sup>21</sup>

Some recent research focuses on understanding and predicting epidemiological behavior of SARS-CoV-2 and its variants using different indicators from biological or clinical 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.<sup>22</sup> 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.<sup>23,24</sup> Furthermore, with machine learning technique for natural language processing, the generated semantic landscapes for SARS-CoV-2 are able to predict mutations that could potentially evade the immune system.<sup>25</sup>

The research reported here focuses on the spike protein of coronavirus, where it makes physical interaction with the human cell receptor ACE2 and then enters human body. We conduct a normal mode analysis (NMA) of different SARS-CoV-2

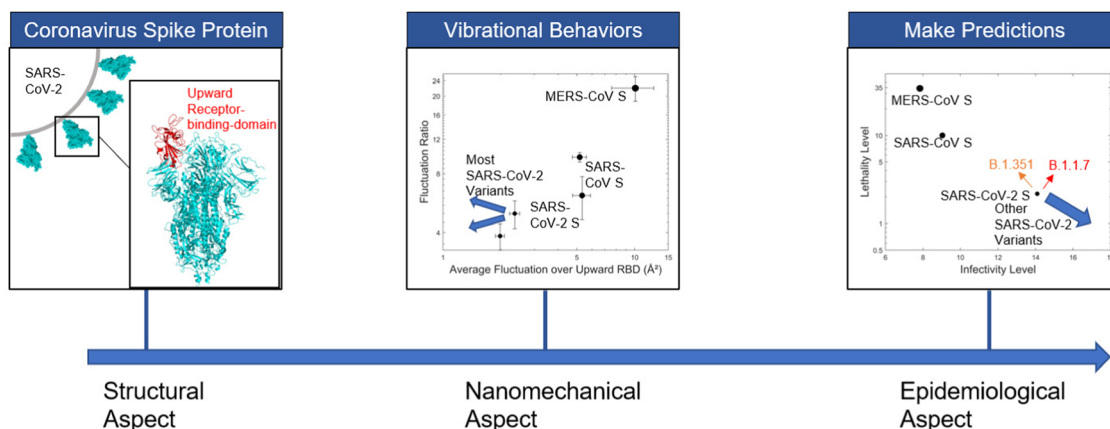
variants in receptor-accessible conformation with one upward RBD, analyze their vibrational behaviors, and utilize the correlation that we reported in earlier work<sup>1</sup> to make predictions on the infectiousness and lethality (Fig. 1).

Compared to previous predictive models,<sup>22–25</sup> our presented predictive model is the first one to make predictions based on pure nanomechanical properties, *i.e.* the average fluctuation level and mobility ratio of receptor-binding domain that we identified in previous work.<sup>1</sup> This work may 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.

## Materials and methods

### Data preparation

The beta-coronavirus spike proteins that we consider in this paper are all coronavirus spike proteins with one upward receptor-binding domain (RBD), and they are MERS-CoV S (PDB ID: 5X5F), SARS-CoV S (PDB ID: 6CRZ, 6CRW), SARS-CoV-2 S (PDB ID: 6Z97, 7DDN). All coronavirus spike protein structures considered above, accessible from the Protein Data Bank,<sup>26</sup> contain complete RBD structure, so that we can evaluate the largest flexibility in the extended loop of upward RBD based on our previous results.<sup>1</sup> 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, and implement the spike mutations of different variants using PyMOL. A few mutations are not implemented due to missing regions in templet protein structure (Table 1). The atomistic models of all coronavirus spike proteins, including the wild-type and the mutants, are prepared with Visual Molecular Dynamics (VMD).<sup>27</sup> The spike protein structures are immersed in a TIP3P water box with a minimum distance of 20 Å to the nearest protein atom, followed by ionization with Sodium and Chlorine



**Fig. 1** Overview of the work reported in this paper. Our analysis focuses on the mechanics of the spike protein of coronaviruses, where it makes physical contact with the human cell receptor ACE2. Vibrational behaviors of SAR-CoV-2 variants are investigated, and the differences induced by mutations are analyzed. Building off earlier work in which we correlated infectiousness and lethality of different types of coronavirus pathogens, the vibrational feature map of SARS-CoV-2 variants can then be utilized to make predictions about the epidemiological properties.



**Table 1** Different SARS-CoV-2 variants considered in this study, and a description of the spike mutations of these variants. Given the protein structure template we use for the original SARS-CoV-2 spike protein (PDB ID: 6Z97), there are a few mutations we are unable to implement to the spike protein because of missing regions. Future work may explore these mutations with similar methods as used here

Variant name	WHO label	First reported	Spike mutations	Mutations implemented in this work
D614G	—	China	D614G	D614G
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 (501Y.V1)	Alpha	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)	Beta	South Africa	L18F, D80A, D215G, R246I, K417N, E484K, N501Y, D614G, A701V	D215G, K417N, E484K, N501Y, D614G, A701V
P.1 (501Y.V3)	Gamma	Brazil	L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, D614G, H655Y, T1027I	D138Y, R190S, K417T, E484K, N501Y, D614G, H655Y, T1027I
B.1.617.2	Delta	India	T19R, G142D, 156del, 157del, R158G, L452R, T478K, D614G, P681R, D950N	G142D (chain A/C), L452R, T478K, D614G, D950N
Double mutant	—	India	L452R, E484Q, D614G	L452R, E484Q, D614G

atoms. Then, we carry out 10 000 steps of conjugate gradient energy minimization and equilibrate the system for 80 ns with temperature controlled at 300 K based on Nanoscale Molecular Dynamics (NAMD).<sup>28</sup> The simulation is performed with CHARMM36 all-atom additive protein force field,<sup>29</sup> and both electrostatic and van der Waals calculations for non-bonded interactions are truncated smoothly at cutoff distance 12 Å using smooth switching functions. Nosé-Hoover Langevin piston pressure control,<sup>30,31</sup> together with Langevin dynamics, are employed to simulate the *NPT* ensemble. 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.

### Normal mode analysis

Normal mode analyses (NMA) of coronavirus spike proteins in receptor-accessible state with one upward RBD is conducted to assess the molecular mechanics and vibrational properties from an atom-by-atom perspective. Here, we use the same approach as presented in our previous paper.<sup>1</sup> We employ a coarse-grained elastic network model (ENM) available in the Bio3D<sup>32–34</sup> package in *R* to analyze the normal modes of coronavirus spike protein structures and calculate atomic displacements for temperature 300 K to construct the fluctuation profile of each spike protein considered. The coarse-graining model<sup>35</sup> is constructed with three to five atoms per residue, where the N, CA, C atoms represent the protein backbone and zero to two selected side chain atoms represent the side chain. The side chain atoms are select based on side chain size and the distance to CA. The energy function is defined by

$$U_{ij}(\mathbf{r}) = k \left( \left\| \mathbf{r}_{ij}^0 \right\| \right) \left( \left\| \mathbf{r}_{ij} \right\| - \left\| \mathbf{r}_{ij}^0 \right\| \right)^2 \quad (1)$$

where  $\mathbf{r}$  is the current protein conformation,  $\mathbf{r}^0$  represents the equilibrium conformation, and  $\left\| \mathbf{r}_{ij} \right\|$  is the distance between atoms  $i$  and  $j$ . The force constant  $k$  implemented here, with cutoff distance set to be 15 Å, is obtained by fitting to a local energy minimum of a crambin model derived from the

AMBER99SB force field,<sup>32</sup> leading to a pair force constant function dependent on the equilibrium distance of two interacting atoms

$$k(\left\| \mathbf{r}_{ij}^0 \right\|) = \min \{7424 \cdot \left\| \mathbf{r}_{ij}^0 \right\|^{-6}, 1500\} \quad (2)$$

with the unit given in kcal mol<sup>-1</sup> Å<sup>-2</sup> for all non-covalent interactions, and specific force constants are applied to covalent and intra-residue atom pairs.<sup>35</sup> The mean square fluctuations of coronavirus spike proteins are calculated based on the first 20 non-trivial normal modes. The fluctuation profiles of wild-type spike proteins as well as SARS-CoV-2 variants of concern are 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.

We continue to use the similar two important factors that we identify in previous work,<sup>1</sup> *i.e.* flexibility level and flexibility ratio, to describe the vibrational behavior, and they are defined as:

$$\text{Average fluctuation of upward RBD} = \sum_{i=1}^n F_i/n \quad (3)$$

$$\text{Fluctuation ratio} = \text{mean}(F_1, F_2, \dots, F_n) / \text{mean}(f_1, f_2, \dots, f_m) \quad (4)$$

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. Here, considering the variation with molecular dynamics (MD) sampling, mean value instead of maximum value is used to define fluctuation ratio. Note that the coronavirus spike trimer in one-upward RBD conformation has two RBDs in downward position, and we calculate  $f_i$  based on the more active downward RBD because it has larger possibility to change to a standing position.



## Epidemiological properties and correlation model

The global confirmed case number and the case fatality rate are updated as of Jul 1, 2022. Korber *et al.*<sup>14</sup> have developed a web tool<sup>36</sup> to track mutations of SARS-CoV-2 based on updated GISAID SARS-CoV-2 sequence database. 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

$$\text{Effective total case number} = \sum_i p_i N_i \quad (5)$$

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.

The logarithm of global infection number and the logarithm of case fatality rate in percentage are defined as the infectivity level and lethality level in this paper, 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:

$$y = a_0 + a_1 f_1(x_1) + a_2 f_2(x_2) \quad (6)$$

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  $R^2 > 0.99$ .

## Structural analysis

Several important angles and distances are analyzed using VMD. (see Fig. 3(a)) Given the coronavirus spike trimer and the previous finding that the S2 subunit of coronavirus spike protein is quite rigid, we then define the central axis as the perpendicular line that passes the center of the S2 subunit central triangle, which consists of geometrical centers of the S2 subunits in three chains in coronavirus spike protein. The orientation of a certain domain is defined as the orientation of the maximum diameter of the domain, and the angle of a domain is the orientation with respect to the horizontal plane defined by the central axis. The distance of a domain is defined as the distance of the mass center from the central axis, indicating how tightly a domain packs against the central axis. We investigate the influence of various structural factors using a simple linear model:

$$Y = a_0 + a_1 \theta_{\text{RBD-B}} + a_2 d_{\text{RBD-B}} + a_3 \theta_{\text{RBD-C}} + a_4 d_{\text{RBD-C}} + a_5 \theta_{\text{NTD-A}} + a_6 \theta_{\text{NTD-B}} + a_7 \theta_{\text{NTD-C}} + a_8 d_{\text{NTD}} \quad (7)$$

where  $Y$  is the model target,  $a_i$  are coefficients to be solved,  $\theta$  is the structural angle with subscripts denoting the

corresponding domain and chain, and  $d$  is the structural distance with subscripts denoting the corresponding domain and chain. Note that because the distances of NTD in different chains are quite similar to each other, only the average value of the distance of NTDs is taken into consideration in the model.

## Results and discussion

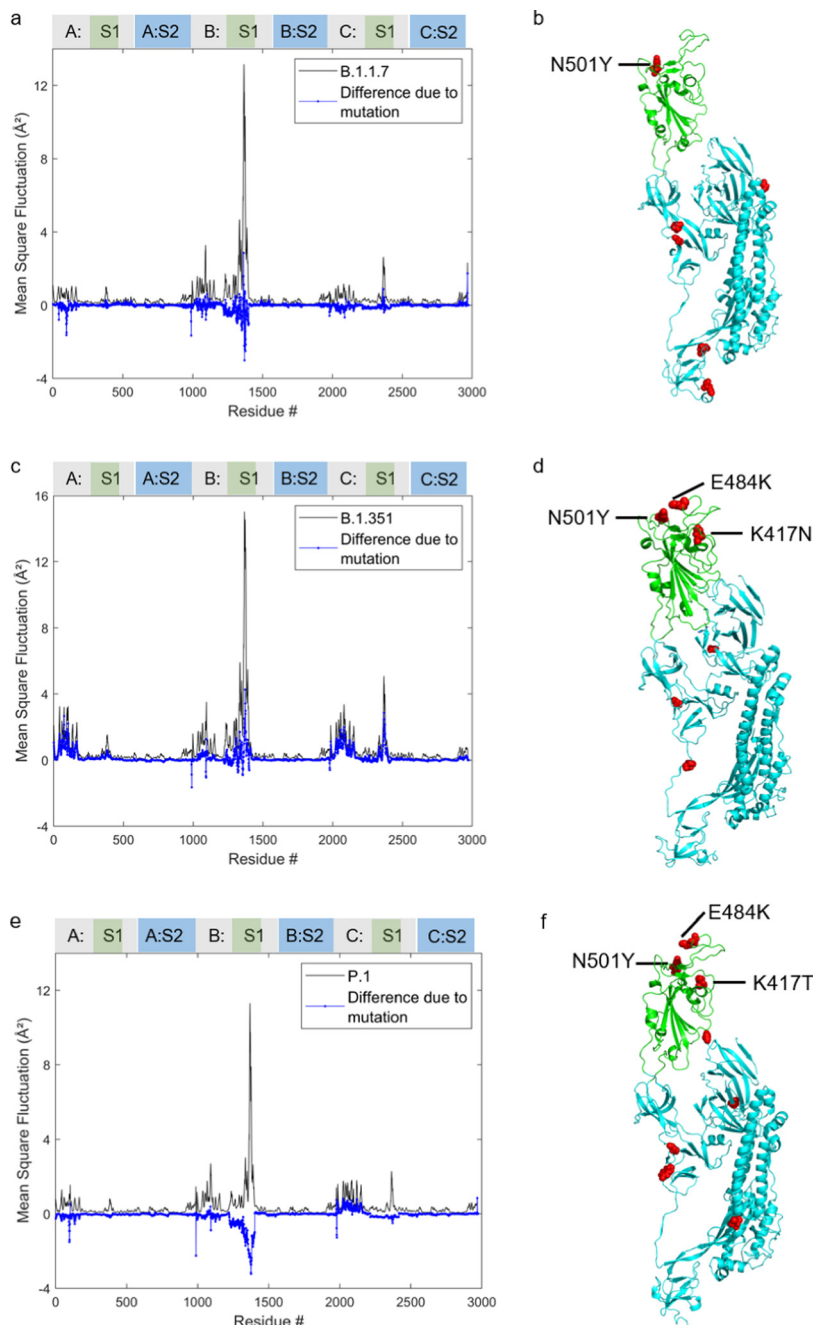
Our study focuses on the spike protein of beta-coronavirus, which is the part that interacts with the human cell receptor and is essential for the infection to take place. In this paper, we mainly study and make predictions for the following SARS-CoV-2 variants: (1) the D614G variant that has quickly become the most prevalent form in the global pandemic since its emergence;<sup>14</sup> (2) mink associated mutations including the “Cluster 5” variant identified by Statens Serum Institut (SSI) in Denmark and single spike substitution F486L;<sup>19,37</sup> (3) the B.1.1.7 variant (Alpha variant) that was first reported in the UK; (4) the B.1.351 variant (Beta variant) that arose in South Africa, sharing the same mutation N501Y with B.1.1.7 (Alpha variant); (5) the P.1 variant (Gamma variant) that caused widespread infection in Brazil, containing spike substitution E484K and N501Y that present in B.1.351 (Beta variant); (6) the two key mutations that are referred as “double mutant” in B.1.617 variant (parent lineage of Kappa and Delta variant), which emerged in India in late 2020, and also the sub-variant B.1.617.2 (Delta variant).<sup>38</sup> (Table 1) Among these variants of interest, the mutation D614G and variants B.1.1.7 (Alpha variant), B.1.351 (Beta variant), P.1 (Gamma variant) and B.1.617.2 (Delta variant) are still circulating globally. In this paper, all the spike mutations are implemented based on the template SARS-CoV-2 spike protein structure in one RBD upward conformation (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.

The fluctuation profile of coronavirus spike proteins, obtained through the use of normal mode analysis, reveals the vibrational flexibility of the spike. In previous work, 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.<sup>1</sup>

Given the small number of spike mutations compared to the total residue number near 3000, will the mutations make a difference to the flexibility profile? The answer is yes. In Fig. 2, we show the flexibility profiles of three notable SARS-CoV-2 variants that are still circulating globally—B.1.1.7, B.1.351 and P.1, identified by WHO as Variants of Concern (VOC) and labelled as Alpha, Beta and Gamma variant respectively. The fluctuation profiles of these three 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





**Fig. 2** Flexibility profile and spike mutation spots of the three notable SARS-CoV-2 variants that are still circulating globally—B.1.1.7 (Alpha variant), B.1.351 (Beta variant) and P.1 (Gamma variant). (a) Fluctuation profile of SARS-CoV-2 B.1.1.7 (Alpha variant) spike protein and the flexibility difference between this variant and the original spike protein (PDB ID: 6Z97). The flexibility difference is calculated by subtracting the fluctuation of wild-type virus from that of the mutant. Labels on the top indicate the corresponding subunits S1 and S2 in chain A, B and C. The green labels located in the middle of S1 subunit labels demonstrate the locations of receptor binding domain (RBD). (b) equilibrated structure of SARS-CoV-2 B.1.1.7 (Alpha variant) spike protein, showing only the chain with upward 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. The standard deviation (RMSD) that occurs during molecular dynamics (MD) relaxation is about 4.94 Å. (c and d) Fluctuation profile and equilibrated structure of SARS-CoV-2 B.1.351 (Beta variant) spike protein respectively, with RMSD around 4.51 Å during MD relaxation. (e and f) fluctuation profile and equilibrated structure of SARS-CoV-2 P.1 (Gamma variant) spike protein respectively, with RMSD around 3.85 Å during MD relaxation. Although B.1.1.7 (Alpha variant), B.1.351 (Beta variant) and P.1 (Gamma variant), also known as 501Y.V1, 501Y.V2 and 501Y.V3, share the important spike mutation N501Y, their fluctuation profiles behave differently. Compared with B.1.1.7 (Alpha variant) and P.1 (Gamma variant), B.1.351 mutant (Beta variant) has more flexible upward RBD and downward RBD. While spike substitutions generally decrease the mobility of RBDs, several residues in the upward and downward RBDs of B.1.351 (Beta variant) obtain increased flexibility due to mutations.



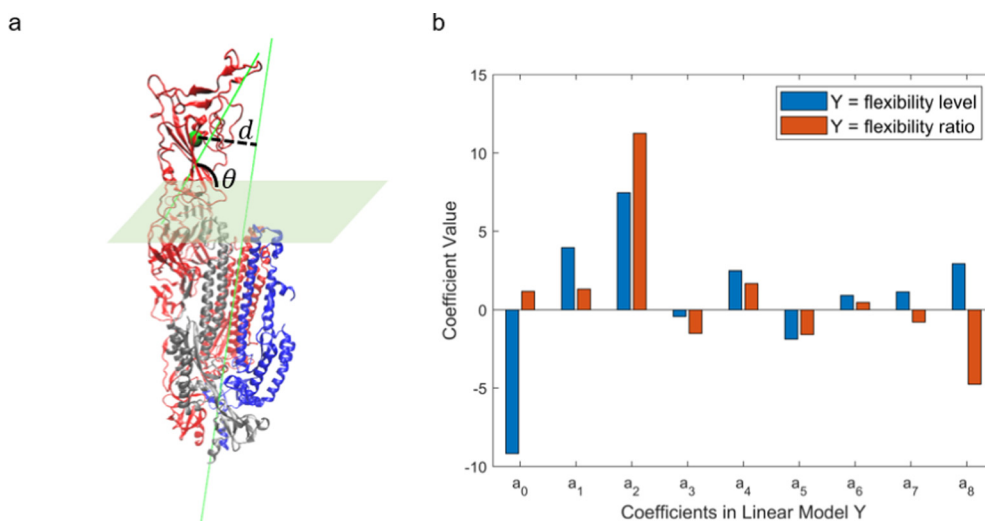
largest flexibility locating in its upward RBD. Note that the dynamics of stalk hinge (central helix and HR in S2 subunit), which is reported in previous literature,<sup>39</sup> is not observed in our analysis because the stalk region is not included in our model. On the other hand, the mutations generally 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 retain rigidity. Comparing B.1.1.7 (Alpha variant), B.1.351 (Beta variant) and P.1 (Gamma variant), which also known as 501Y.V1, 501Y.V2 and 501Y.V3 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.1.1.7 (Alpha variant) is smaller than the difference due to mutations in B.1.351 (Beta variant) and P.1 (Gamma variant), suggesting B.1.1.7 (Alpha 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 (Alpha variant) and P.1 (Gamma variant), B.1.351 (Beta 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 (Beta variant) obtain increased flexibility due to mutations, showing some specialty of lineage B.1.351 (Beta variant).

We further investigate important angles and distances in the coronavirus spike protein structure to understand the structural parameters that could influence the flexibility level and mobility ratio, the two significant vibrational factors that we identified in previous work.<sup>1</sup> According to Fig. 3, among all structural parameters concerned, the factors associated with

the upward RBD show a significant positive relationship with the flexibility level. The further upward RBD is from the central axis, the larger its standing angle is, the more “open” the spike conformation is, and thus the higher the vibrational flexibility level is. In terms of the flexibility ratio, the distance of upward RBD and the average distance of NTDs both make significant contributions. While the flexibility ratio increases with the distance of upward RBD, it scales negatively with the average distance of NTDs. Therefore, considering that the flexibility ratio is quantified as the ratio of RBDs in upward and downward conformation, the more tightly the NTDs pack against the central axis, the larger the flexibility ratio is, the more distinct the RBDs in different positions are.

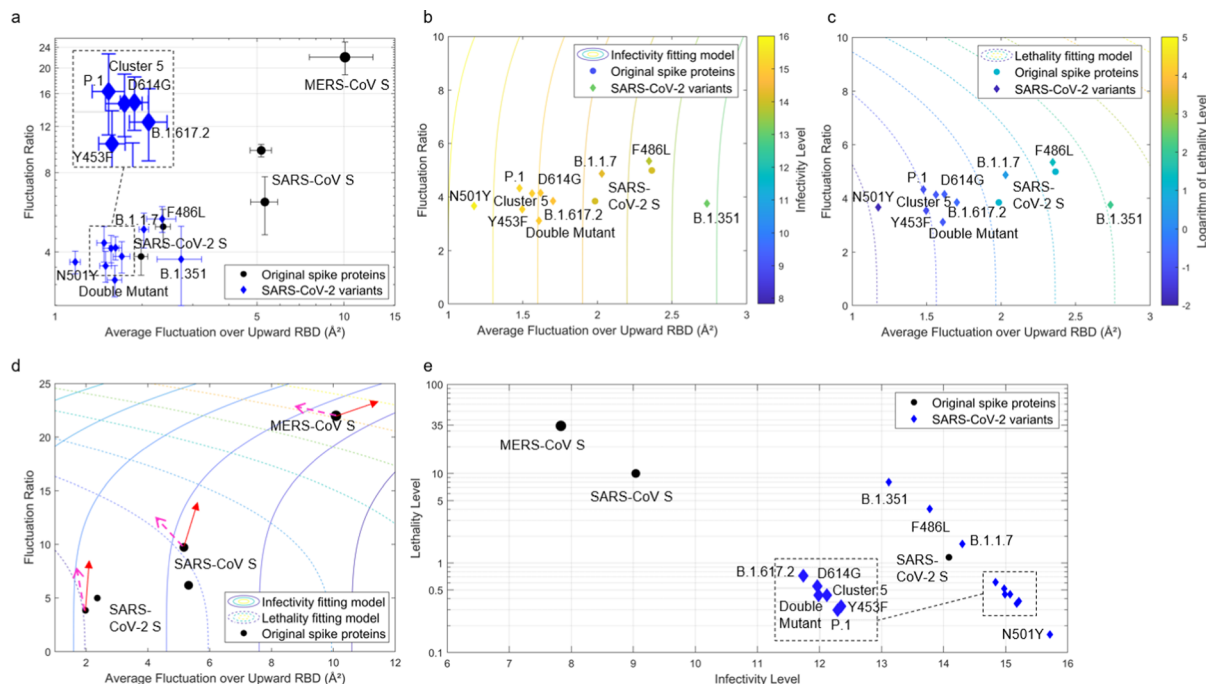
Fig. 4 depicts the quantitative model to predict the infectivity and lethality level of SARS-CoV-2 variants of interest. The nanomechanical vibrational feature map of coronavirus spike proteins and SARS-CoV-2 variants is shown in panel (a). It is demonstrated that generally all variants of concern except B.1.351 (Beta variant) have decreased average fluctuation over upward RBD, which corresponds with the trend we observe with detailed fluctuation profiles in Fig. 2. 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.

Aside from the qualitative conclusion drawn from panel (a), we establish our quantitative model to further make predictions about the infectivity and lethality level of SARS-CoV-2 variants of interest based on the relationship observed between



**Fig. 3** Quantitative structural analysis of SARS-CoV-2 variants. (a) Definition of important angles and distances in the spike protein structure. Three chains in the spike protein are colored in blue, red and grey. S1 subunits of chain A and C are hidden to provide a clear view. The central axis of the structure and the maximum diameter of upward receptor-binding domain (RBD) are drawn in green. (b) Coefficients in the linear fitting model for flexibility level and ratio. The linear model is defined as  $Y = a_0 + a_1\theta_{\text{RBD-B}} + a_2d_{\text{RBD-B}} + a_3\theta_{\text{RBD-C}} + a_4d_{\text{RBD-C}} + a_5\theta_{\text{NTD-A}} + a_6\theta_{\text{NTD-B}} + a_7\theta_{\text{NTD-C}} + a_8d_{\text{NTD}}$ . The angle  $\theta$  is the normalized angle between the maximum diameter of important domain, *i.e.* RBD and NTD, in each chain and the horizontal plane defined by central axis. The distance  $d$  is the normalized distance of the mass center of certain domain from the central axis. The subscript denotes the corresponding domain and chain. The figure shows that when evaluating the flexibility level, the structural parameters concerning the upward RBD play significant roles. The further upward RBD is from the central axis, the larger its standing angle is, the higher the fluctuation level is. As for the flexibility ratio, not only the distance of upward RBD, but also the average distance of NTDs have significant impact. The flexibility ratio increases with the distance of upward RBD, and has a negative correlation with the average distance of NTDs.





**Fig. 4** Nanomechanical vibrational feature map of coronavirus spike proteins and SARS-CoV-2 variants, and analyses of correlations with viral infectivity and lethality, including the B.1.617.2 mutant (Delta variant). (a) 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. (b) Polynomial fitting model linking fluctuation level and ratio to infectivity, where infectivity level is calculated as the logarithm of global confirmed case number. (c) Polynomial fitting model relating fluctuation level and ratio to lethality, where lethality level is calculated as the logarithm of mortality rate in percentage. (d) 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. (e) Prediction on epidemiological features of SARS-CoV-2 variants. This analysis shows that, generally, most variants of concern are predicted to be more infectious and less lethal than the original SARS-CoV-2. Lineage B.1.351 (Beta variant) and mutant F486L are predicted to be less infectious and more lethal, while lineage B.1.1.7 (Alpha variant) is anticipated to have both higher infectivity and lethality. All containing N501Y substitution in spike, B.1.1.7 (Alpha variant), B.1.351 (Beta variant) and P.1 (Gamma variant) are predicted to have quite different epidemiological features compared to variant with single N501Y mutation. The result suggests that other spike mutations, which have not been discussed in previous literature, also have significant influence on the infectivity and lethality of SARS-CoV-2 variants. The emerging lineage B.1.617.2 (Delta variant), although contains a few key spike mutations other than D614G, behaves quite similar to the single D614G mutation in both vibrational and predicted epidemiological aspects, which might explain its fast circulation given the prevalence of D614G. From panel (d) we can see that the possibility of both more lethal and more infectious variants is quite small for SARS-CoV-2, but it changes with the vibrational features and may become a serious concern when both general flexibility level and fluctuation ratio of the coronavirus spike protein are large.

vibrational and epidemiological properties.<sup>1</sup> As shown in panel (b and c), the polynomial fitting model is able to describe the quantitative relationship appropriately, with correlation coefficient  $R^2 = 0.991$  and  $R^2 = 0.998$  for infectivity and lethality fitting model respectively. The analysis indicates that most variants of interest are predicted to be more infectious and less lethal than the original SARS-CoV-2. Lineage B.1.351 (Beta variant) and mutant F486L are predicted to be both less infectious and more lethal, while lineage B.1.1.7 (Alpha variant) is anticipated to achieve higher infectivity and higher lethality simultaneously. Our model predictions on lineage B.1.1.7 (Alpha variant) not only correspond with its greater transmissibility reported in previous studies,<sup>12</sup> but also agree with the results in literature where analysis on positive SARS-CoV-2 community tests show that B.1.1.7 may lead to increased mortality.<sup>40</sup> Within the range of error bars, the qualitative predictions of infectiousness and

lethality of most SARS-CoV-2 variants remain the same. The only exception would be mutant F486L, which lies in between the two different performance regions and could result in different qualitative predictions.

Focusing on the mutations associated with mink, we find that single mutation Y453F and cluster 5 variant, which contains Y453F and other two spike substitutions, have similar vibrational behavior and predicted epidemiological properties, and this proves the stability of our model to some extent. However, another mink-associated spike mutation, namely F486L, behaves distinctively from the previous two mutants, which might result from its special location at the receptor binding motif (RBM) of RBD.

Here, we can also further compare the predicted epidemiological characteristics of B.1.1.7 (Alpha variant), B.1.351 (Beta variant), P.1 (Gamma variant) and their shared N501Y spike



substitution. It is shown that the four 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.<sup>41</sup> Notably, we emphasize here that other spike mutations, which have not been discussed in previous 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.

Both containing spike substitution in residue 484, B.1.351 (Beta variant) and the “double mutant” variant that has two key mutations in B.1.617 are predicted to have smaller fluctuation ratio than the original SARS-CoV-2 spike protein. Considering the comparatively higher fluctuation ratio of B.1.617.2 (Delta variant), which lacks the E484Q mutation present in the “double mutant”, we suggest that the spike substitution in residue 484 may be associated with lower fluctuation ratio in SARS-CoV-2 spike protein, leading to more similar RBDs in different conformations and larger possibility toward multiple receptor-accessible RBDs.

While all current variants of concern except lineage B.1.1.7 (Alpha variant) show single trend of either higher infectivity and lower lethality or lower infectivity and higher lethality, we demonstrate in panel (d) that the possibility for the variants to be both more lethal and more infectious could change with different vibrational properties. The possibility is larger for MERS-CoV S compared with that of SARS-CoV S and SARS-CoV-2 S. In the top region of Fig. 4(d), the nonlinearity of epidemiological prediction becomes significant, and thus the likelihood of more lethal and more infectious variants could be quite large. Therefore, dangerous mutations could be a serious concern with future coronavirus whose spike protein has comparatively large flexibility level and fluctuation ratio.

Furthermore, noticing that three of the WHO identified Variants of Concern (VOC), namely B.1.1.7 (Alpha variant), P.1 (Gamma variant) and B.1.617.2 (Delta variant), have similar nanomechanical and epidemiological behavior compared with the single D614G variant, which has quickly become the most prevalent form since it occurred, we suggest that aside from the both more lethal and more infectious variants, there might be some area centering the D614G variant in the nanomechanical vibrational feature map where SARS-CoV-2 variants could be quite dangerous and pose serious threat to global public health.

## Conclusions

A predictive model that associates molecular motions and vibrational patterns of the virus spike protein with

infectiousness and lethality is constructed and analyzed in this paper. Our findings suggest the possibility to correlate nanomechanical features of spike protein vibration patterns, and how they change due to mutations, with large-scale population-level observation. This enables us now to explore a variety of mechanistic approaches to understand variants, and the future evolution of variants, based on nanomechanical concepts.

As shown in Fig. 4, most variants of concern are predicted to be more infectious and less lethal than the original SARS-CoV-2. Lineage B.1.351 (Beta variant) and mutant F486L are predicted to be less infectious and more lethal, while lineage B.1.1.7 (Alpha variant) is anticipated to have both higher infectivity and lethality. The different behavior of B.1.1.7 (Alpha variant), B.1.351 (Beta variant) and P.1 (Gamma variant) 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. Among all variants of interest, the emerging lineage B.1.617.2 (Delta variant), although contains a few key spike mutations other than D614G, behaves quite similar to the single D614G mutation in vibrational feature map and is predicted to have similar infectivity and lethality, which may help explain its fast circulation considering how quickly D614G has become the prevalent form since its occurrence.

For the SARS-CoV-2 virus, its higher infectivity and lethality performance regime appears to be located in the small area above it, as shown in Fig. 4(d). Given the connection between spike mutation at residue 484 and the lower vibrational fluctuation ratio revealed in Fig. 4, we suggest that spike mutations at residue 484 might help drive SARS-CoV-2 variants towards the special performance regime. What's more, the distinct epidemiological predictions found here for B.1.351 (Beta variant) and P.1 (Gamma variant) implies that the combination of spike mutations K417N, E484K and N501Y might be quite dangerous and could possibly lead to higher lethality. Although only lineage B.1.1.7 (Alpha variant), among all variants of interest, is predicted to have higher infectivity and lethality, the possibility of both more lethal and more infectious variants could change with vibrational features and may indeed become a serious concern when both general flexibility level and fluctuation ratio of coronavirus spike protein are large. This can be studied in more detail in future work, perhaps with the combination of advanced protein folding tools, molecular modeling, and experimental assays.

Overall, this work may provide a tool to estimate the epidemiological effects of new variants, offer a pathway to screen mutations against high threat levels and help design possible inhibitors that compensates for the motions induced by mutations. Moreover, the nanomechanical approach, as a novel tool to predict virus-cell interactions, may offer a tool to better understand other viruses. Considering the limitation of implementing mutations based on the current SARS-CoV-2 spike protein templet with several missing regions, future work may analyze certain missing spike mutations with more complete spike protein structures as templets. 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.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

We acknowledge support from the MIT-IBM AI Lab, Google Cloud Computing, ONR, and NIH.

## References

- 1 Y. Hu and M. J. Buehler, Comparative Analysis of Nanomechanical Features of Coronavirus Spike Proteins and Correlation with Lethality and Infection Rate, *Matter*, 2021, **4**(1), 265–275.
- 2 J. F. W. Chan, S. Yuan and K. H. Kok, *et al.*, A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster, *Lancet*, 2020, **395**(10223), 514–523.
- 3 C. Huang, Y. Wang and X. Li, *et al.*, Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China, *Lancet*, 2020, **395**(10223), 497–506.
- 4 Y. Wan, J. Shang, R. Graham, R. S. Baric and F. Li, Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus, *J. Virol.*, 2020, **94**(7), 1–9.
- 5 COVID-19 Map – Johns Hopkins Coronavirus Resource Center.
- 6 R. Verity, L. C. Okell and I. Dorigatti, *et al.*, Estimates of the severity of coronavirus disease 2019: a model-based analysis, *Lancet Infect. Dis.*, 2020, **20**(6), 669–677.
- 7 D. Wrapp, N. Wang and K. S. Corbett, *et al.*, Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation, *Science*, 2020, **367**(6483), 1260–1263.
- 8 Y. Yuan, D. Cao and Y. Zhang, *et al.*, Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains, *Nat. Commun.*, 2017, **8**, 15092.
- 9 M. Gui, W. Song and H. Zhou, *et al.*, Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding, *Cell Res.*, 2017, **27**(1), 119–129.
- 10 R. N. Kirchdoerfer, C. A. Cottrell and N. Wang, *et al.*, Prefusion structure of a human coronavirus spike protein, *Nature*, 2016, **531**(7592), 118–121.
- 11 R. Yan, Y. Zhang, Y. Li, L. Xia, Y. Guo and Q. Zhou, Structural basis for the recognition of the SARS-CoV-2 by full-length human ACE2, *Science*, 2020, **367**(6485), 1444–1448.
- 12 N. G. Davies, S. Abbott and R. C. Barnard, *et al.*, Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England, *Science*, 2021, **372**(6538), eabg3055.
- 13 E. Volz, V. Hill and J. T. McCrone, *et al.*, Evaluating the Effects of SARS-CoV-2 Spike Mutation D614G on Transmissibility and Pathogenicity, *Cell*, 2021, **184**(1), 64–75.e11.
- 14 B. Korber, W. M. Fischer and S. Gnanakaran, *et al.*, Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus, *Cell*, 2020, **182**(4), 812–827.e19.
- 15 D. J. Benton, A. G. Wrobel and C. Roustan, *et al.*, The effect of the D614G substitution on the structure of the spike glycoprotein of SARS-CoV-2, *Proc. Natl. Acad. Sci. U. S. A.*, 2021, **118**(9), 2–5.
- 16 J. Zhang, Y. Cai and T. Xiao, *et al.*, Structural impact on SARS-CoV-2 spike protein by D614G substitution, *Science*, 2021, **372**(6541), 525–530.
- 17 L. Yurkovetskiy, X. Wang and K. E. Pascal, *et al.*, Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant, *Cell*, 2020, **183**(3), 739–751.e8.
- 18 Y. J. Hou, S. Chiba and P. Halfmann, *et al.*, SARS-CoV-2 D614G variant exhibits enhanced replication *ex vivo* and earlier transmission *in vivo*, *Science*, 2020, **370**(6523), 1464–1468.
- 19 B. B. O. Munnink, R. S. Sikkema and D. F. Nieuwenhuijse, *et al.*, Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans, *Science*, 2021, **371**(6525), 172–177.
- 20 M. Hoffmann, L. Zhang and N. Krüger, *et al.*, SARS-CoV-2 mutations acquired in mink reduce antibody-mediated neutralization, *Cell Rep.*, 2021, **35**(3), 109017.
- 21 Z. Wang, F. Schmidt and Y. Weisblum, *et al.*, mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants, *Nature*, 2021, **592**(7855), 616–622.
- 22 J. Chen, R. Wang, M. Wang and G. W. Wei, Mutations Strengthened SARS-CoV-2 Infectivity, *J. Mol. Biol.*, 2020, **432**(19), 5212–5226.
- 23 A. Goyal, D. B. Reeves, E. Fabian Cardozo-Ojeda, J. T. Schiffer and B. T. Mayer, Viral load and contact heterogeneity predict sars-cov-2 transmission and super-spreading events, *eLife*, 2021, **10**, 1–63.
- 24 E. Pujadas, F. Chaudhry and R. McBride, *et al.*, SARS-CoV-2 viral load predicts COVID-19 mortality, *Lancet Respir. Med.*, 2020, **8**(9), 70.
- 25 B. Hie, E. D. Zhong, B. Berger and B. Bryson, Learning the language of viral evolution and escape, *Science*, 2021, **371**(6526), 284–288.
- 26 H. M. Berman, J. Westbrook and Z. Feng, *et al.*, The Protein Data Bank, *Nucleic Acids Res.*, 2000, **28**(1), 235–242.
- 27 W. Humphrey, A. Dalke and K. Schulten, VMD: Visual molecular dynamics, *J. Mol. Graphics*, 1996, **14**(1), 33–38.
- 28 M. T. Nelson, W. Humphrey and A. Gursoy, *et al.*, NAMD: A parallel, object-oriented molecular dynamics program, *Int. J. High Perform. Comput. Appl.*, 1996, **10**(4), 251–268.
- 29 J. Huang and A. D. Mackerell, CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data, *J. Comput. Chem.*, 2013, **34**(25), 2135–2145.
- 30 G. J. Martyna, D. J. Tobias and M. L. Klein, Constant pressure molecular dynamics algorithms, *J. Chem. Phys.*, 1994, **101**(5), 4177–4189.



- 31 S. E. Feller, Y. Zhang, R. W. Pastor and B. R. Brooks, Constant pressure molecular dynamics simulation: The Langevin piston method, *J. Chem. Phys.*, 1995, **103**(11), 4613–4621.
- 32 K. Hinsen, A. J. Petrescu, S. Dellerue, M. C. Bellissent-Funel and G. R. Kneller, Harmonicity in slow protein dynamics, *Chem. Phys.*, 2000, **261**(1-2), 25–37.
- 33 P. Durand, G. Trinquier and Y.-H. Sanejouand, A new approach for determining low-frequency normal modes in macromolecules, *Biopolymers*, 1994, **34**(6), 759–771.
- 34 F. Tama, F. X. Gadea, O. Marques and Y. H. Sanejouand, Building-block approach for determining low-frequency normal modes of macromolecules, *Proteins: Struct., Funct., Genet.*, 2000, **41**(1), 1–7.
- 35 X. Q. Yao, L. Skjærven and B. J. Grant, Rapid Characterization of Allosteric Networks with Ensemble Normal Mode Analysis, *J. Phys. Chem. B*, 2016, **120**(33), 8276–8288.
- 36 COVID-19 Viral Genome Analysis Pipeline.
- 37 L. van Dorp, C. C. S. Tan and S. D. Lam, *et al.* No evidence for increased transmissibility from recurrent mutations in SARS-CoV-2, *Nat. Commun.*, 2020, **11**, 5986.
- 38 P. Mlcochova, S. Kemp and M. S. Dhar, *et al.* SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion, *Nature*, 2021, **599**(7883), 114–119.
- 39 P. V. Raghuvamsi, N. K. Tulsian and F. Samsudin, *et al.*, SARS-CoV-2 S protein:ACE2 interaction reveals novel allosteric targets, *eLife*, 2021, **10**(M), 1–47.
- 40 N. G. Davies, C. I. Jarvis and K. van Zandvoort, *et al.*, Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7, *Nature*, 2021, **593**(7858), 270–274.
- 41 H. Gu, Q. Chen and G. Yang, *et al.*, Adaptation of SARS-CoV-2 in BALB/c mice for testing vaccine efficacy, *Science*, 2020, **369**(6511), 1603–1607.

