jz-2022-004693.R1

Name: Peer Review Information for "Omicron BA.2 (B.1.1.529.2): high potential to becoming the next dominating variant"

First Round of Reviewer Comments

Reviewer: 1

Comments to the Author

The authors carried out a timely artificial intelligence (AI)-based study of Omicron BA.2. Systematically, infectivity, vaccine-breakthrough potential, and antibody resistance of Omicron BA.2 variant are analyzed together with many other SARS-CoV-2 variants, including Alpha, Beta, Gamma, Delta, Lambda, Mu, BA.1, BA.2, and BA.3. The main predictions are that BA.2 is about 1.5 as contagious as BA.1 and 30% more capable than BA.1 to escape current vaccines. These predictions were confirmed in the past few days in the news. The team has made many successful contributions to COVID-19 studies in the past. The senior author, Wei, is a top expert in AI-based computational biology. This paper is well-written. I recommend its acceptance after some minor changes.

1) Sotrovimab developed by GlaxoSmithKline is regarded as not being affected by BA.2. But Figure 4 g2 shows that mutation G339D may have a considerable impact on Sotrovimab's efficacy. The authors need to explain their findings.

2) I understand that the manuscript was written when there was no experimental result. As more experimental data about BA.2 become available, the authors need to give a comparison of their predictions with newly available data.

3) It may be useful to explicitly give the mathematical expression of the binding free energy change.

Reviewer: 2

Comments to the Author

By applying deep learning model, the authors predicted the infectivity, vaccine breakthrough ability and antibody resistance of BA.2 and BA.3. Through comparative analysis of the current main variants, it is found that BA.2 is about 1.5 and 4.2 times as contagious as BA.1 and Delta, respectively. And the vaccine breakthrough ability is about 30% and 17-fold more capable than BA.1 and Delta, respectively. This work predicts that Omicron BA.2 will be the next dominant variant.

The base for analysis of the work is the machine learning of binding free energy, where the perdition error for each mutation of the BFE is not shown, so it is hard to say that total energy difference of 0.1-0.3 kcal/mol for Omicron BA.1, BA.2, and BA.3 is meaningful, which is 2.60, 2.98 and 2.88 kcal/mol. For SKEMPI 2.0 the reported standard errors in KD are around 0.25 kcal/mol. So, the author should prove that the prediction errors of 0.1-0.3 kcal/mol is suit for analysis and support the results.

The paper is underprepared, some figures mentioned in the text is now shown and some figures shown in the figure is not mentioned. So, it is hard to following. Some of them is listed in the following:

1. Figure 1, where is the Figure h?

2. Page 2, Section “Infectivity”. It should be stated first how the BFE data was obtained.

3. Page 2, line 45. "The larger the BFE change is, the higher infectivity will be." There is an ambiguity here. If the binding free energy decreases, the less infectivity is.

4. Page 2, line55. "Our model predicts that BA.2 is about 1.5 as contagious BA.2, which is the same as reported in an initial study." What the author should express here is that BA.2 is about 1.5 as contagious BA.1?

5. Page 3, line32. Why doesn't b3 in Figure 2 appear in the discussion?

6. Page 3, line 38. "BA.1 mutation G496S is also quite disruptive. BA.2 mutations T376A, D405N, and R408S may reduce the efficacy of many antibodies.". I don't think the G496S and D405N drop significantly.

7. Page 3, line 41. "Overall, Figure 2 shows more negative BFE changes than positive ones". It seems that only Figure2-b3 looks like this.

8. Page 4, line 49. "Gray color stands for no predictions due to incomplete structures". Can't find the Gray color in the figure, if there is, the author should use a more obvious color.

9. Page 5, Figure 3. What is the "4G" on the Figure3-f? And the figure f blocks part of the figure d2. In addition, as a comparison, why the x-axis labels of the a1, b1, c1, d1 graphs are not consistent

10. Page 7, line 49. "b2, b3 and b4 for Regeneron mAbs". "b2, b3" and "b4" have inconsistent formats.

Reviewer: 3

Comments to the Author

Chen and Wei in their work “Omicron BA.2 (B.1.1.529.2): high potential to become the next dominating variant” developed a deep learning model to predict vaccine infectivity and resistance. I consider it an interesting work, but I have some points that concern me in the methodology section (supplementary material). I would like to look at the revised work

1. In S5 Supplementary validation, I do not consider splitting 1539 and 1500 samples as suitable for validation because there is a high tendency to bias in the results. I suggest authors re-perform the model, selecting 70-80% of the dataset to train the model and the rest of the samples would be used for external validation.

2. Please provide also $R^2 = \{sum\ [(yipred - <y>)^2]\}/\{sum\ [(yi - <y>)^2\}$ for training and test sets. These data are critical for internal and external validations in statistical analysis.

3. At the beginning of the “S4 Supplementary machine learning methods” section, the authors used only 10-fold cross-validation. Due to the problem mentioned earlier, I consider it very poor to just use this method. In this case, the authors should also include the results of the leave-one-out, in addition to two or three different cases of n-fold cross-validation. I'll be happy if, and only if, all these results aren't too different for internal validation.

4. I consider it very important to have a comparison with experimental data. After re-perform the model, the authors must recalculate the Pearson correlation coefficient and R^2, using predicted data (by the new model) and the 35 experimental cases from reference 28. Furthermore, I consider it essential for the authors to develop a regression model: IC50 = ypred + b; where IC50 is a parameter that is the experimental measurement and ypred is the parameter predicted by the model. This may be important for other works to use the model to predict IC50, which is the most commonly used quantitative parameter for measuring antibody titers that recognize RBD.

Reviewer: 4

Comments to the Author

This paper provides a timely investigation on the potential VOC omicron BA.2a a subvariant BA.1 and implications on BA.3. The results are of high interest and cover many different areas of research using machine learning (ML) or deep learning (DL) since no experimental data are currently available. The results obtained are convincing and supported by evidences from a combination of overlapping scientific disciplines including biomedical, mathematical, computational sciences using state-of-the-art methods.

I am quite impressed by the 4 figures displayed in the main text. They are very creative and yet clearly convey a massive collection of data points described in Section 3. Such a large collection of data covering many aspects of research in unprecedented.

The methods used in this study for ML and DL are truly outstanding involving frontier areas of mathematics and topology but also down to atomic details. This is seldom done in computational biomolecular science, as such the results reported are way ahead of other groups.

The supplementary information (SI) contains many additional useful data to the research community. It is so useful and detailed that when combined with the main text can be presented as a review article.

Some minor points for authors to consider:

1. Please include some data on quantum chemical calculations in using ML or DL that may enhance or speed up the prediction.

2. The quantitative binding free energy (BFE) is generally obtained from MD simulations. Would it be possible to use alternative concepts based on interatomic bonding on fixed structure models to validate the BFE data used in ML?

3. Please check for typo errors in the manuscript and figure captions carefully. There are several of them I can spot. 5. In Fig.2, the position of b2 and b3 should be moved down a little to avoid confusion.

4. It is not clear where the data mentioned S2.1 and S2.2 in the SI can be easily located.

5. On page 8, abbreviation for ‘PPI’ should specify its full name when first mentioned. On the other hand, abbreviations such as ACE2, mAbs need not be repeated in later parts once defined.

6. Section 3 Materials and Methods is overly condensed and referred mostly to SI. The three subsections in S3.2 should be discussed in more detail, especial step 2 on feature generation or identification since this contains the “meat” of ML protocol. Merely reference to author’s previous publications lender the reading less smooth or coherent.

7. Need to add more details on the description on Residue level section in S3.2.2. The distance of 10Å mentioned appears to be large for interatomic interactions based on bond pairs.

8. Detailed description of atomic-level interactions is desirable. Are they obtained by actual quantum chemical calculations or obtained from ML exercise with different training sets?

9. The conclusion section could be more succinct or just list as few bullets.

I recommend this paper be accepted by Journal of Physical Chemistry Letter with minor modification.

Author's Response to Peer Review Comments:

Dr Guowei Wei
Department of Mathematics
Michigan State University
East Lansing MI 48824
Email: wei@math.msu.edu
Phone: (517) 353 4689

March 25, 2022

Prof. Editor
Senior Editor
The Journal of Physical Chemistry Letters

Manuscript ID: jz-2022-004693

Dear Professor Editor,

Thank you very much for your mail of March 8, 2022 and referees' reports concerning our paper entitled: "Omicron BA.2 (B.1.1.529.2): High Potential to Becoming the Next Dominating Variant". manuscript number, jz-2022-004693. We have carefully gone through referees' comments, and have accordingly revised the manuscript accordingly. The changes are marked in red color.

I would like to point out that essentially, all predictions made in this work have been *near perfectly* confirmed by experiments or the World Health Organization. I believe that the revised version is acceptable for publication in the Journal of Physical Chemistry Letters.

Thank you very much for your assistance.

Sincerely yours,

Guo-Wei Wei
MSU Foundation Professor of
Mathematics,
Biochemistry and Molecular Biology,
Electrical and Computer Engineering

Answers to Reviewer 1' Comments

Comments: *The authors carried out a timely artificial intelligence (AI)-based study of Omicron BA.2. Systematically, infectivity, vaccine-breakthrough potential, and antibody resistance of Omicron BA.2 variant are analyzed together with many other SARS-CoV-2 variants, including Alpha, Beta, Gamma, Delta, Lambda, Mu, BA.1, BA.2, and BA.3. The main predictions are that BA.2 is about 1.5 as contagious as BA.1 and 30% more capable than BA.1 to escape current vaccines. These predictions were confirmed in the past few days in the news. The team has made many successful contributions to COVID-19 studies in the past. The senior author, Wei, is a top expert in AI-based computational biology. This paper is well-written. I recommend its acceptance after some minor changes.*

Answer: We thank the reviewer for the condensed summary and the positive comments.

Comments: *1) Sotrovimab developed by GlaxoSmithKline is regarded as not being affected by BA.2. But Figure 4 g2 shows that mutation G339D may have a considerable impact on Sotrovimab's efficacy. The authors need to explain their findings.*

Answer: In Figure 4g2, it shows the BFE changes of S309. It is the parent antibody for Sotrovimab. The final structure of Sotrovimab is not accessible. Therefore, we cannot make a very strong statement about its efficacy. From the Figure 4g2, it is also noticed that G339D has a BFE change of -0.4kcal/mol, and except that the overall impacts is not as large as other antibodies.

Comments: *2) I understand that the manuscript was written when there was no experimental result. As more experimental data about BA.2 become available, the authors need to give a comparison of their predictions with newly available data.*

Answer: We thank the reviewer for pointing out this concern. As we submit the manuscript, we collected experimental data and added a section of "Note added in proof".

Comments: *3) It may be useful to explicitly give the mathematical expression of the binding free energy change.*

Answer: We thank the reviewer for useful suggestion. We add the mathematical expression in the Supporting Information.

Answers to Reviewer 2' Comments

Comments: *By applying deep learning model, the authors predicted the infectivity, vaccine breakthrough ability and antibody resistance of BA.2 and BA.3. Through comparative analysis of the current main variants, it is found that BA.2 is about 1.5 and 4.2 times as contagious as BA.1 and Delta, respectively. And the vaccine breakthrough ability is about 30% and 17-fold more capable than BA.1 and Delta, respectively. This work predicts that Omicron BA.2 will be the next dominant variant. The base for analysis of the work is the machine learning of binding free energy, where the perdition error for each mutation of the BFE is not shown, so it is hard to say that total energy difference of 0.1-0.3 kcal/mol for Omicron BA.1, BA.2, and BA.3 is meaningful, which is 2.60, 2.98 and 2.88 kcal/mol. For SKEMPI 2.0 the reported standard errors in KD are around 0.25 kcal/mol. So, the author should prove that the prediction errors of 0.1-0.3 kcal/mol is suit for analysis and support the results.*

Answer: This is a comparative study of a variety of SARS-COV-2 variants, including Alpha, Beta, Gamma, Delta, Lambda, Mu, Omicron BA.1, BA.2, and BA.3 on exactly the same setting. Therefore, the normal error analysis of theoretical predictions against experimental results does not applied. The error in SKEMPI 2.0 is for experimental data collected over different a large number of labs. In our case, our predicted is based on a single high-resolution x-ray structure. Since, all variants are studied in the same setting, the relative error in our predictions about these variants is reflected in our use of significant numbers.

The difference in the BFE changes for BA.1 and BA.2 was reported as 0.38 kcal/mol, giving rise to 0.46 times higher infectivity. This result perfectly matches the experimental results about the relative infectivities between BA.1 and BA.2 in the literature.

Comments: *The paper is underprepared, some figures mentioned in the text is now shown and some figures shown in the figure is not mentioned. So, it is hard to following. Some of them is listed in the following:*

1) Figure 1, where is the Figure h?

Answer: The figure reference is updated. All figures are mentioned.

Comments: *2) Page 2, Section "Infectivity". It should be stated first how the BFE data was obtained.*

Answer: We thank the reviewer for useful suggestions. As we discuss the structure information, we added description about the BFE data.

Comments: *3) Page 2, line 45. "The larger the BFE change is, the higher infectivity will be." There is an ambiguity here. If the binding free energy decreases, the less infectivity is.*

Answer: We thank the reviewer for pointing this confusion. We updated the sentence.

Comments: *4) Page 2, line55. "Our model predicts that BA.2 is about 1.5 as contagious BA.2,* *which is the same as reported in an initial study." What the author should express here is that BA.2 is about 1.5 as contagious BA.1?*

Answer: We thank the reviewer for pointing this confusion. We update the sentence.

Comments: *5) Page 3, line32. Why doesn't b3 in Figure 2 appear in the discussion?*

Answer: We thank the reviewer. It is updated.

Comments: *6) Page 3, line 38. "BA.1 mutation G496S is also quite disruptive. BA.2 mutations T376A, D405N, and R408S may reduce the efficacy of many antibodies.". I don't think the G496S and D405N drop significantly.*

Answer: We update the description. G496S and D405N are compared BA.1 unique mutations and BA.2 unique mutations respectively.

Comments: *7) Page 3, line 41. "Overall, Figure 2 shows more negative BFE changes than positive ones". It seems that only Figure2-b3 looks like this.*

Answer: Please note that the scale for negative BFE change, it is from -4 kcal/mol to 0 kcal/mol, which is a large range compared to positive BFE.

Comments: *8) Page 4, line 49. "Gray color stands for no predictions due to incomplete structures". Can't find the Gray color in the figure, if there is, the author should use a more obvious color.*

Answer: We thank the reviewer for pointing out the typo. It is updated.

Comments: *9) Page 5, Figure 3. What is the "4G" on the Figure3-f? And the figure f blocks part of the figure d2. In addition, as a comparison, why the x-axis labels of the a1, b1, c1, d1 graphs are not consistent*

Answer: This is should be figure typesetting issue. We have revised it. The "4G" should be D614G. The differences of x-axis ranges are according the prediction values for each antibody. Omicron variant and subvariants have dramatic impacts, while Delta variant has less.

Comments: *10) Page 7, line 49. "b2, b3 and b4 for Regeneron mAbs". "b2, b3" and "b4" have inconsistent formats.*

Answer: It is corrected. Thanks.

Answers to Reviewer 3' Comments

Comments: *Recommendation: This paper may be publishable, but major revision is needed; I would like to be invited to review any future revision.*

Comments:
Chen and Wei in their work "Omicron BA.2 (B.1.1.529.2): high potential to become the next dominating variant" developed a deep learning model to predict vaccine infectivity and resistance. I consider it an interesting work, but I have some points that concern me in the methodology section (supplementary material). I would like to look at the revised work

Comments: *1) In S5 Supplementary validation, I do not consider splitting 1539 and 1500 samples as suitable for validation because there is a high tendency to bias in the results. I suggest authors re-perform the model, selecting 70-80% of the dataset to train the model and the rest of the samples would be used for external validation.*

Answer: We thank the reviewer for the useful suggestion. We will implement it for our next generation platform work. The current paper focus on the comparative analysis of BA.2 over other variants. On the other hand, as the high tendency to have bias, what we tested here is leave-one-dataset-out.

Comments: *2) Please provide also $R^2 = sum[(yipred- < y >)]^2/sum[(yi- < y >)^2$ for training and test sets. These data are critical for internal and external validations in statistical analysis.*

Answer: We thank the Reviewer for the suggestion. The goal of this work to present a rapid report about BA.2 variant's potential to become a dominating variant. In fact, this prediction has been confirmed by the WHO on March 22, indicating the predictive value of our work. We will further improve our deep learning models and implement suggested validations in our future work.

Comments: *3) At the beginning of the "S4 Supplementary machine learning methods" section, the authors used only 10-fold cross-validation. Due to the problem mentioned earlier, I consider it very poor to just use this method. In this case, the authors should also include the results of the leave-one-out, in addition to two or three different cases of n-fold cross-validation. I'll be happy if, and only if, all these results aren't too different for internal validation.*

Answer: In the paper, Chen, et al., Mutations Strengthened SARS-CoV-2 Infectivity, *Journal of Molecular Biology*, 432, 5212-5226, 2020, we implemented the leave-one-out in the Supporting Information. The results outperformed other methods'. The validity of our approach is now confirmed by its near perfect predictions against experimental results.

Comments: *4) I consider it very important to have a comparison with experimental data. After re-perform the model, the authors must recalculate the Pearson correlation coefficient and R^2, using predicted data (by the new model) and the 35 experimental cases from reference 28. Furthermore, I consider it essential for the authors to develop a regression model: IC50 = ypred + b; where IC50 is a parameter that is the experimental measurement and ypred is the parameter predicted by the model.*

This may be important for other works to use the model to predict IC50, which is the most commonly used quantitative parameter for measuring antibody titers that recognize RBD.

Answer: Thanks for the suggestion. As mentioned, this is not a methodology paper. Its goal is to rapid communicate the threat of BA.2 to the society. All of our predictions, namely, the infectivity, vaccine breakthrough potential, and antibody disruption, as well as becoming the dominating variant, have been perfectly confirmed by experiments or the WHO.

Answers to Reviewer 4' Comments

Comments: *Recommendation: This paper is publishable subject to minor revisions noted. Further review is not needed.*

Comments:
This paper provides a timely investigation on the potential VOC omicron BA.2a a subvariant BA.1 and implications on BA.3. The results are of high interest and cover many different areas of research using machine learning (ML) or deep learning (DL) since no experimental data are currently available. The results obtained are convincing and supported by evidences from a combination of overlapping scientific disciplines including biomedical, mathematical, computational sciences using state-of-the-art methods.

I am quite impressed by the 4 figures displayed in the main text. They are very creative and yet clearly convey a massive collection of data points described in Section 3. Such a large collection of data covering many aspects of research in unprecedented. The methods used in this study for ML and DL are truly outstanding involving frontier areas of mathematics and topology but also down to atomic details. This is seldom done in computational biomolecular science, as such the results reported are way ahead of other groups. The supplementary information (SI) contains many additional useful data to the research community. It is so useful and detailed that when combined with the main text can be presented as a review article.

Answer: We thank the reviewer for the condensed summary and the positive comments.

Comments: *1) Please include some data on quantum chemical calculations in using ML or DL that may enhance or speed up the prediction.*

Answer: Our deep learning model built from a variety of theoretical and computational methods. For example, electrostatic features from the Poisson Boltzmann (PB) model involve quantum chemical calculations of force fields.

Comments: *2) The quantitative binding free energy (BFE) is generally obtained from MD simulations. Would it be possible to use alternative concepts based on interatomic bonding on fixed structure models to validate the BFE data used in ML?*

Answer: BFE calculations in our work do not directly use MD simulations. We use algebraic topology, deep learning, electrostatics from the PB model, and co-evolution information from BLAST.

Comments: *3) Please check for typo errors in the manuscript and figure captions carefully. There are several of them I can spot. 5. In Fig.2, the position of b2 and b3 should be moved down a little to avoid confusion.*

Answer: We thank the reviewer for pointing out those typos. It is updated.

Comments: *4) It is not clear where the data mentioned S2.1 and S2.2 in the SI can be easily located.*

Answer: We have improved our description to make it easily located.

Comments: *5) On page 8, abbreviation for 'PPI' should specify its full name when first mentioned. On the other hand, abbreviations such as ACE2, mAbs need not be repeated in later parts once defined.*

Answer: It is updated. Thanks!

Comments: *6) Section 3 Materials and Methods is overly condensed and referred mostly to SI. The three subsections in S3.2 should be discussed in more detail, especial step 2 on feature generation or identification since this contains the "meat" of ML protocol. Merely reference to author's previous publications lender the reading less smooth or coherent.*

Answer: Thanks for comments. As mentioned, the goal of this paper to rapidly report the threat of BA.2 to the human health. At this point, it is clear that our predictions are near prefect.

Comments: *7) Need to add more details on the description on Residue level section in S3.2.2. The distance of 10Å mentioned appears to be large for interatomic interactions based on bond pairs.*

Answer: We have improved our presentation in the Supporting Information. However, we will publish a separately methodology paper in the future to improve our methods.

Comments: *8) Detailed description of atomic-level interactions is desirable. Are they obtained by actual quantum chemical calculations or obtained from ML exercise with different training sets?*

Answer: We represent atomic-level interactions by algebraic topology. Some electrostatic interactions are described by the Poisson-Boltzmann model.

Comments: *9) The conclusion section could be more succinct or just list as few bullets.*

Answer: We followed the practice of most papers in Journal of Physical Chemistry Letters.

Comments: *I recommend this paper be accepted by Journal of Physical Chemistry Letter with minor modification.*

Answer: We thank the reviewer for the positive comments, again.

We thank all reviewers again!

jz-2022-004693.R2

Name: Peer Review Information for "Omicron BA.2 (B.1.1.529.2): high potential to becoming the next dominating variant"

Second Round of Reviewer Comments

Reviewer: 3

Comments to the Author

I do not recommend publishing the current manuscript because I believe that internal and external validations are key to generating good predictive models. The fact of showing that the models predict little WHO data does not imply that the model is good.

Some published works show the importance of internal and external validation to generate good prediction models. The authors of these works are experienced in statistical studies in the most diverse areas of knowledge:

Golbraikh A, Tropsha A. Beware of q2! J Mol Graph Model. 2002 Jan;20(4):269-76. doi: 10.1016/s1093-3263(01)00123-1. PMID: 11858635.

Cabitza F, Campagner A, Soares F, García de Guadiana-Romualdo L, Challa F, Sulejmani A, Seghezzi M, Carobene A. The importance of being external. methodological insights for the external validation of machine learning models in medicine. Comput Methods Programs Biomed. 2021 Sep;208:106288. doi: 10.1016/j.cmpb.2021.106288. Epub 2021 Jul 22. PMID: 34352688.

Plante TB, Blau AM, Berg AN, Weinberg AS, Jun IC, Tapson VF, Kanigan TS, Adib AB. Development and External Validation of a Machine Learning Tool to Rule Out COVID-19 Among Adults in the Emergency Department Using Routine Blood Tests: A Large, Multicenter, Real-World Study. J Med Internet Res. 2020 Dec 2;22(12):e24048. doi: 10.2196/24048. PMID: 33226957; PMCID: PMC7713695.

Ramspek CL, Jager KJ, Dekker FW, Zoccali C, van Diepen M. External validation of prognostic models: what, why, how, when and where? Clin Kidney J. 2020 Nov 24;14(1):49-58. doi: 10.1093/ckj/sfaa188. PMID: 33564405; PMCID: PMC7857818.

Steyerberg EW, Harrell FE Jr. Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol. 2016 Jan;69:245-7. doi: 10.1016/j.jclinepi.2015.04.005. Epub 2015 Apr 18. PMID: 25981519; PMCID: PMC5578404.

Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE, Moons KG. Internal and external validation of predictive models: a simulation study of bias and precision in small samples. J Clin Epidemiol. 2003 May;56(5):441-7. doi: 10.1016/s0895-4356(03)00047-7. PMID: 12812818.

Dangeti, Pratap. Statistics for machine learning. Packt Publishing Ltd, 2017.

Esposito, F., Malerba, D., Semeraro, G. and Kay, J., 1997. A comparative analysis of methods for pruning decision trees. IEEE transactions on pattern analysis and machine intelligence, 19(5), pp.476-491.

Therefore, for me to accept the publication, it is necessary that the authors adequadally validate the models, as mentioned, otherwise I tend to believe that the model has a bias and will be inadequate to predict the variables of interest.

Reviewer: 2

Comments to the Author

My comments have been addressed.

Author's Response to Peer Review Comments:

Dr Guowei Wei
Department of Mathematics
Michigan State University
East Lansing MI 48824
Email: wei@math.msu.edu
Phone: (517) 353 4689

April 8, 2022

Prof. **Editor**
Senior Editor
The Journal of Physical Chemistry Letters

Manuscript ID: jz-2022-004693.R1

Dear Professor **Editor**,

Thank you very much for your mail of April 8, 2022 and referees' reports concerning our paper entitled: "Omicron BA.2 (B.1.1.529.2): High Potential to Becoming the Next Dominating Variant". manuscript number, jz-2022-004693.R1. We have carefully gone through referees' comments and modified our manuscript to address Reviewer 3's concerns.

We have also added 3 related references from the Journal of Physical Chemistry Letters.

I believe that the revised version is acceptable for publication in the Journal of Physical Chemistry Letters.

Thank you very much for your assistance.

Sincerely yours,

Guo-Wei Wei
MSU Foundation Professor of
Mathematics,
Biochemistry and Molecular Biology,
Electrical and Computer Engineering

We thank Reviewer 1 and Reviewer 4 who have accepted our paper.

Answers to Reviewer 2' Comments

Comments: *Recommendation: This paper represents a significant new contribution and should be published as is.*

Answer: We thank the reviewer for enthusiastic support.

Answers to Reviewer 3' Comments

Comments: *I do not recommend publishing the current manuscript because I believe that internal and external validations are key to generating good predictive models. The fact of showing that the models predict little WHO data does not imply that the model is good.*

Some published works show the importance of internal and external validation to generate good prediction models. The authors of these works are experienced in statistical studies in the most diverse areas of knowledge:

Golbraikh A, Tropsha A. Beware of q2! J Mol Graph Model. 2002 Jan;20(4):269-76. doi: 10.1016/s1093-3263(01)00123-1. PMID: 11858635.

Cabitza F, Campagner A, Soares F, García de Guadiana-Romualdo L, Challa F, Sulejmani A, Seghezzi M, Carobene A. The importance of being external. methodological insights for the external validation of machine learning models in medicine. Comput Methods Programs Biomed. 2021 Sep;208:106288. doi: 10.1016/j.cmpb.2021.106288. Epub 2021 Jul 22. PMID: 34352688.

Plante TB, Blau AM, Berg AN, Weinberg AS, Jun IC, Tapson VF, Kanigan TS, Adib AB. Development and External Validation of a Machine Learning Tool to Rule Out COVID-19 Among Adults in the Emergency Department Using Routine Blood Tests: A Large, Multicenter, Real-World Study. J Med Internet Res. 2020 Dec 2;22(12):e24048. doi: 10.2196/24048. PMID: 33226957; PMCID: PMC7713695.

Ramspek CL, Jager KJ, Dekker FW, Zoccali C, van Diepen M. External validation of prognostic models: what, why, how, when and where? Clin Kidney J. 2020 Nov 24;14(1):49-58. doi: 10.1093/ckj/sfaa188. PMID: 33564405; PMCID: PMC7857818.

Steyerberg EW, Harrell FE Jr. Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol. 2016 Jan;69:245-7. doi: 10.1016/j.jclinepi.2015.04.005. Epub

2015 Apr 18. PMID: 25981519; PMCID: PMC5578404.

Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE, Moons KG. Internal and external validation of predictive models: a simulation study of bias and precision in small samples. J Clin Epidemiol. 2003 May;56(5):441-7. doi: 10.1016/s0895-4356(03)00047-7. PMID: 12812818.

Dangeti, Pratap. Statistics for machine learning. Packt Publishing Ltd, 2017.

Esposito, F., Malerba, D., Semeraro, G. and Kay, J., 1997. A comparative analysis of methods for pruning decision trees. IEEE transactions on pattern analysis and machine intelligence, 19(5), pp.476-491.

Therefore, for me to accept the publication, it is necessary that the authors adequadally validate the models, as mentioned, otherwise I tend to believe that the model has a bias and will be inadequate to predict the variables of interest.

Answer: We fully agree with the reviewer that "internal and external validations are key to generating good predictive models". Our work goes further beyond "showing that the models predict little WHO data". As shown in the section of "Materials and methods", an early version of deep learning model (i.e., TopNetTree), published in Nature Machine Intelligence, 2116-123, 2020 (Ref. [33]), was extensive validated in the benchmark data SKEMPT 2.0. Our model significantly outperforms all other competing methods for the AB-Bind S645 set in the terms of correlation R_p (Table 1 of Ref. [33]):

Method	R_p
TopNetTree	0.65/0.68*
TopGBT	0.56
mCSM-AB	0.53/0.56*
TopCNN	0.53
Discovery Studio	0.45
mCSM-PPI	0.31
FoldX	0.34
STATIUM	0.32
DFIRE	0.31
bASA	0.22
dDFIRE	0.19
Rosetta	0.16

Our TopNetTree result is also the best for the SKEMPI dataset of 1,131 mutations in the terms of correlation R_p (Table 2 of Ref. [33]):

Method	R_p
TopNetTree	0.850
BindProfX	0.738
Profile-score+FoldX	0.738
Profile-score	0.675
SAAMBE	0.624
FoldX	0.457
BeAtMuSic	0.272
Dcomplex	0.056

To the best of our knowledge, none has reported a model that outperforms ours.

As described in our paper, our current deep learning model for SARS-CoV-2 was further trained and validated with SARS-CoV-2 deep mutational experimental dataset of spike RBD and CTC-445.2 complex (Figure 17 of Ref. [25] published in Chemical Science 12, 6929 - 6948 (2021)):

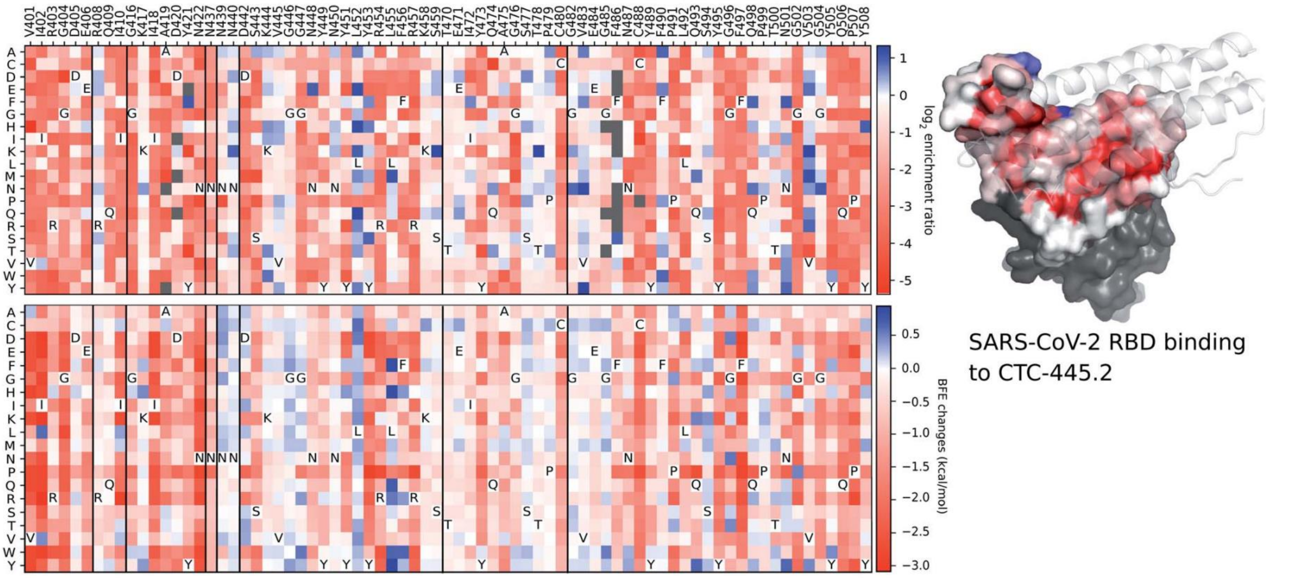

Fig. 17 A comparison between experimental deep mutation enrichment data and TopNetTree predictions for the SARS-CoV-2 S protein RBD and CTC-445.2 complex (7KL9 (ref. 89)). Top left: deep mutational scanning heatmap showing the average effect on the enrichment for single site mutants of the RBD when assayed by yeast display for binding to CTC-445.2.[89] Top right: the RBD colored by average enrichment at each residue position bound to CTC-445.2. Bottom: machine learning predicted BFE changes for the CTC-445.2 and S protein complex induced by single site mutations on the RBD.

SARS-CoV-2 deep mutational experimental dataset of spike RBD and ACE2 complex (Figure 21 of Ref. [31] published in Journal of Molecular Biology, 433, 167155 (2021)):

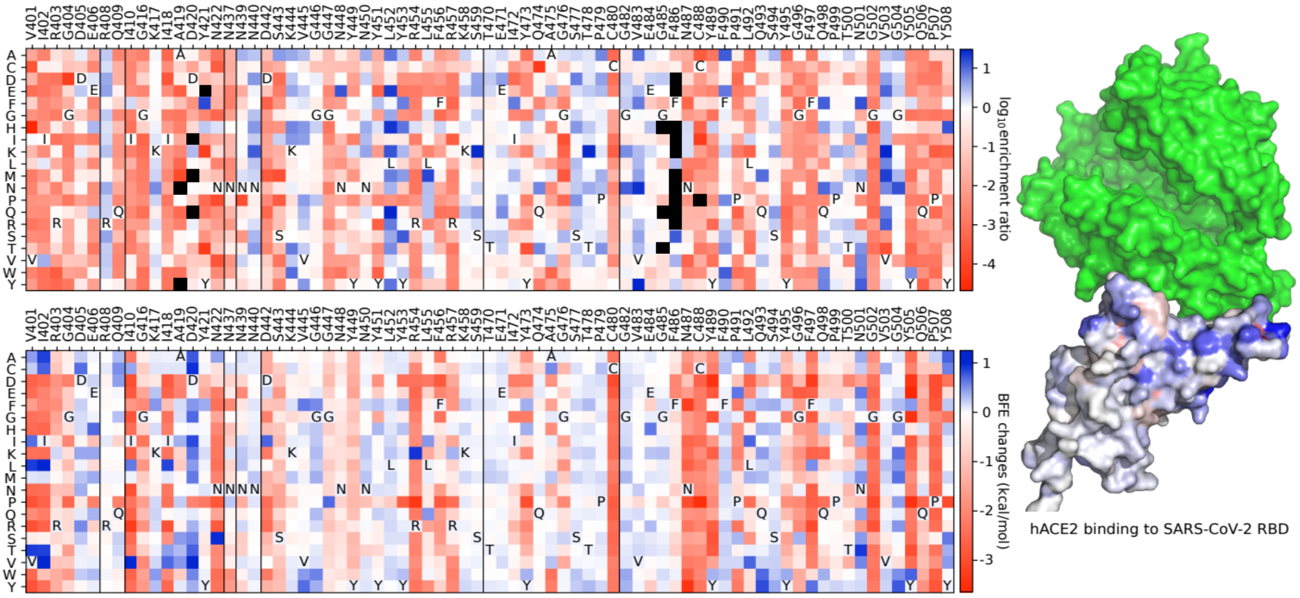

Figure 21: A comparison between experimental RBD deep mutation enrichment data and predicted BFE changes for SARS-CoV-2 RBD binding to ACE2 (6M0J) [27]. **Top left**: deep mutational scanning heatmap showing the average effect on the enrichment for single-site mutants of RBD when assayed by yeast display for binding to the S protein RBD [27]. **Right**: RBD colored by average enrichment at each residue position bound to the S protein RBD. **Bottom left**: machine learning predicted BFE changes for single-site mutants of the S protein RBD.

and experimental mutations in the Alpha variant (N501Y) and the Delta variant (L452R and T478K) (Figure 5 of Ref. [40] published in ACS Infectious Diseases, 8, 3, 546–556 (2022)):

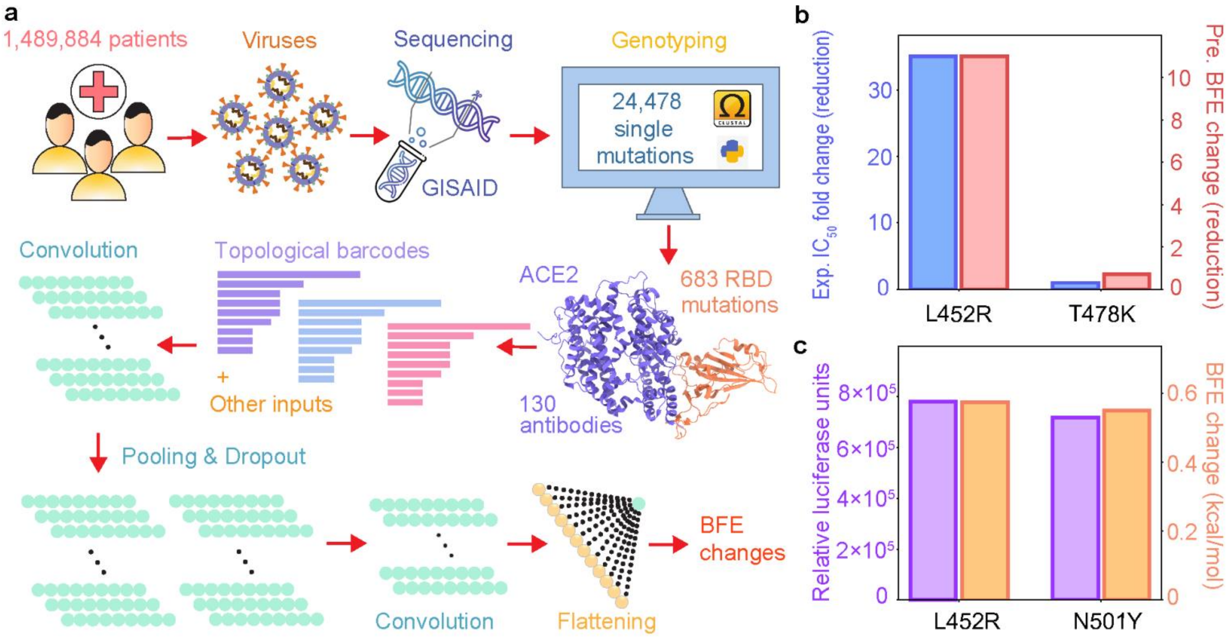

Figure 5. (a) Illustration of genome sequence data preprocessing and BFE change predictions. (b) Comparison of experimental CT-P59 IC_{50} fold change (reduction)[35] and predicted BFE changes induced by mutations L452R and T478K. (c) Comparison of predicted BFE changes and relative luciferase units[25] for pseudovirus infection changes of ACE2 and S protein complex induced by mutations L452R and N501Y.

These validations were described in our paper. However, we cannot republish these tables and figures in the current paper.

Additionally, in our manuscript, validation with experimental data, including internal validation among different variants, was also given in Figures 3 and 5. The nearly perfect confirmations of our predictions by lately reported experimental results speak for reliability and significance of our deep learning model predictions.