

SARS-CoV-2 viral spike G614 mutation exhibits higher case fatality rate

Manuel Becerra-Flores¹ and Timothy Cardozo²

¹NYU Langone Health

²Affiliation not available

April 28, 2020

Abstract

Aim: The COVID pandemic is caused by infection with the SARS-CoV-2 virus. The major mutation detected to date in the SARS-CoV-2 viral envelope spike protein, which is responsible for virus attachment to the host and is also the main target for host antibodies, is a mutation of an aspartate (D) at position 614 found frequently in Chinese strains to a glycine (G). We sought to infer health impact of this mutation. **Result:** Increased case fatality rate correlated strongly with the proportion of viruses bearing G614 on a country by country basis. The amino acid at position 614 occurs at an internal protein interface of the viral spike, and the presence of G at this position was calculated to destabilize a specific conformation of the viral spike, within which the key host receptor binding site is more accessible. **Conclusion:** These results imply that G614 is a more pathogenic strain of SARS-CoV-2, which may influence vaccine design. The prevalence of this form of the virus should also be included in epidemiologic models predicting the COVID-19 health burden and fatality over time in specific regions. Physicians should be aware of this characteristic of the virus to anticipate the clinical course of infection. What is known about this topic? Nothing is known about the health significance of the D614G SARS-CoV-2 variant. What does this article add? A molecular clue to viral molecular pathogenesis of COVID-19 disease.

Introduction

In December 2019, an outbreak of pneumonia, which was later deduced to be due to a coronavirus, was reported in the Wuhan, Hubei Province by Chinese health officials. On January 11, 2020, the genetic sequence of SARS-CoV-2, the coronavirus that causes disease, which was named COVID-19, was published. The first outbreak in the United States of America (USA) was in Washington state on the USA West Coast and was attributed to Chinese origin of the virus, while recent studies have determined that the most intense outbreak in New York state on the USA East Coast is from SARS-CoV-2 viruses of European origin. The only dominant variation in the SARS-CoV-2 viral envelope spike protein, which executes viral glycoprotein-mediated binding to host cells and subsequent fusion of virus and host cell membranes, is mutation of an aspartate (D) at position 614 found in nearly all Chinese strains to a glycine (G) enriched in European strains. To form a better view of the global distribution and CFRs of SARS-CoV-2 variants, we calculated a linear regression of global, country-by-country CFR vs country-by-country D614 percentage and analyzed structural differences of the D to G mutation using published cryo-electron microscopy 3D structures of the SARS-CoV-2 viral spike.

Methods

Deaths and confirmed cases data from the European CDC (<https://www.ecdc.europa.eu/en/covid-19-pandemic>)¹, which is updated daily, was accessed on April 6, 2020. SARS-CoV-2 viral spike sequences were accessed from the GISAID database (<https://www.gisaid.org>) on April 6, 2020. Assuming that a patient was tested and accounted as a confirmed case 11 days before death, the case fatality rate (CFR) was calculated from total

cases and total deaths per day in each country, considering an average of 11 days from hospitalization to death (medRxiv 2020.01.29.20019547). The analysis was restricted to a time window consisting of the last 8 days from the data accession date (3/30/2020-4/6/2020). Both the average and median CFR were calculated for analysis. Standard linear regression was performed for CFR vs D614 percentage from this data using GraphPad Prism, including and excluding the data from the United Kingdom, which exhibited an unusually low number of cases due to an unusually low level of testing/diagnosis and an unusually high level of death reporting.

For structural analysis of the impact of 614 D/G identity, we used published cryo-electron microscopy structures of the SARS-CoV-2 trimeric viral spike in its “down”, or unliganded, state, wherein the host angiotensin converting enzyme 2 (ACE-2) binding site is buried and inaccessible (PDB ID 6VXX), and its “up” state, wherein this host receptor binding site that is necessary for viral infection is exposed (PDB ID 6VYB, 6VSB)^{2,3}. A single monomer (chain A) in the trimer was mutated *in silico* to G614. For both 614 identities, the neighboring sidechains (10.0 Å) were minimized using the Biased-Probability Monte Carlo algorithm for 10⁶ steps⁴. Van der Waals energy, vacuum Coulomb electrostatics, solvation electrostatics, hydrogen bonding, torsional energy and entropy were calculated for each conformation searched and the lowest energy conformation was identified. The free energy change caused by the mutation of D614 to G was then calculated as: $\Delta\Delta G = \Delta G_{G-Mutant} - \Delta G_{D-WildType}$. Thus, a positive value indicates destabilization of the mutant G relative to the wild-type D, since ΔG was a negative value for both forms. All biophysical calculations, molecular modeling and molecular graphics were performed using ICM-Pro (Molsoft LLC, La Jolla, CA).

Results

SARS-CoV-2 viral spike amino acid sequences from strains infecting patients in Europe exhibited a predominance of glycine at amino acid position 614 in the viral spike (G614), while countries in the Far East exhibited a high percentage of D614 viral spikes. Based on case fatality rate (CFR) and SARS-CoV-2 viral spike sequence data available on April 6, 2020, both the average and median CFR correlate strongly ($p < 0.02$) with the proportion of viruses in the same geographic region bearing aspartate at position 614 (D614) in the viral spike (**Figure 1**). For example, the proportion of recorded viruses in China exhibiting D614 was nearly 100% and this country exhibited the second lowest average or median CFR.

As of the time of writing, the complete daily data necessary to calculate the CFR for US states was not available, so this correlation could not be plotted for US states, however, the percentage of viruses in each state exhibiting aspartate (D) instead of glycine (G) at amino acid residue number 614 in the viral spike (S) protein can be assessed (**Table 1**), revealing dramatically lower G614 in Western USA states.

Modeling of the molecular impact of aspartate (D) or glycine (G) at position 614 in the 3D structure of the viral envelope spike trimer revealed that G is a less stable occupant from a biophysical point of view specifically in the “up” state of the viral spike, in which the surface of the viral spike that binds human ACE-2 is exposed and accessible to this host cell surface protein. The “down” or unliganded form shows no or mildly the opposite effect. The model predicts the G mutation to be destabilizing relative to D for the local area around the amino acid 614 position, apparently from loss of packing with the side-chain of threonine 858 from the helical core stalk of the viral spike of an adjacent monomer (**Figure 2**), creating an unstable cavity in the protein. Cavities in the protein core or at protein interfaces are well known to destabilize protein tertiary and quaternary structure⁵. Thus, the structures predict a higher energy barrier for rearrangement of the viral spikes harboring G614 to their infectious form (i.e. the conformation of the viral spike that is most optimal for binding host ACE-2, which is the first animal/human receptor to which the virus binds in order to infect and cause disease).

Discussion

Recent prior bioinformatics analyses have not evaluated the SARS-CoV-2 viral spike in isolation from the rest of the viral genome. The viral spike protein is relatively invariant, when compared to influenza or HIV, across worldwide strains, with essentially only one missense variant so far in the viral spike at amino acid

position 614 (D/G), despite global spread. We here report that the G614 variant correlates strongly with case fatality rate in a global survey of countries, suggesting that a higher CFR possibly emerging in the USA East Coast or other regions affected later in the pandemic may be at least partly a consequence of this virus-autonomous factor.

Molecular modeling suggests that G614 destabilizes the infectious form of the viral spike protein, thereby favoring the ground-state, unliganded, “down” SARS-CoV-2 viral forms that would be expected to infect *less* readily, due to masking of the host receptor binding site. This suggests that the mechanism by which G614 viruses causes greater fatality is immunologic rather than virologic: namely, that the form that binds the receptor less well is also better shielded from host immune system attack and/or elicits harmful anti-viral-spike antibodies or other harmful immune responses. Indeed, there are several precedents in other viral diseases: the viral spikes of both HIV and RSV exhibit unliganded states that do not elicit protective antibodies via vaccination, and indeed, may elicit antibodies that enhance disease^{6,7}. Furthermore, disease-enhancing, immunodominant, antibody-targeted epitopes in the SARS viral spike have been definitively identified⁸, and a recent report emphasized that key cryptic antibody-targeted neutralization epitopes in the SARS family of viral spikes are accessible only in the conformation in which the ACE-2 binding site is exposed⁹.

There are several potential limitations and/or confounding issues in the data we used to calculate country-by-country CFR. First, the data relies heavily on diagnostic testing and fatality reporting, which have both been highly variable in penetrance and accuracy across the countries studies, at present. For example, China recently revised their fatality count upward and the United Kingdom is an outlier with an unusually high CFR, and when included, the correlation weakens (**Figure S1**). This outlier may be caused by the unusually low incidence of active cases being reported in the UK during the 2 weeks under study. Selection bias is unlikely to be a confounder in this analysis, however, since the same reporting bias would have to occur in the same direction and for the same reason across all of the global regions, which is unlikely. Social distancing measures have been implemented in a highly variable fashion across the countries studied in this report, but the case fatality rate should be unaffected by social distancing, as it does not affect the clinical course of viral disease, only acquisition of virus. Measuring this correlation at this early time point in the pandemic may be optimal to avoid the confounder of successful medical treatment to artificially reduce the CFR.

Based on our results, short-term epidemiologic models of the pandemic that attempt to predict the incidence and prevalence of severe disease and fatalities over time that do not take the D614 percentage of viruses in a region into account may be inaccurate. The presence of G614 in a virus detected in a patient may also have prognostic significance.

Acknowledgements

We thank Jennifer Fuller for editing and writing assistance. T.C. conceived the study, analyzed the data, visualized the data and wrote the manuscript. M.B-F. edited the manuscript and performed the study, including data acquisition, data analysis design and data visualization. No funding sources or conflicts of interest encumber this work.

References

1. Roser M, Ritchie H, Ortiz-Ospina E. Coronavirus Disease (COVID-19) – Statistics and Research. *Our-WorldInData.org*.2020;<https://ourworldindata.org/coronavirus>(Online Resource).
2. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*. 2020.
3. Wrapp D, Wang N, Corbett KS, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*.2020;367(6483):1260-1263.

4. Abagyan R, Totrov M. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *Journal of molecular biology*. 1994;235(3):983-1002.
5. Xue M, Wakamoto T, Kejlberg C, et al. How internal cavities destabilize a protein. *Proceedings of the National Academy of Sciences of the United States of America*. 2019;116(42):21031-21036.
6. Acosta PL, Caballero MT, Polack FP. Brief History and Characterization of Enhanced Respiratory Syncytial Virus Disease. *Clinical and vaccine immunology : CVI*. 2015;23(3):189-195.
7. Cardozo T, Shmelkov S, Carr K, Rotrosen J, Mateu-Gelabert P, Friedman S. Cocaine and HIV Infection. In: Montoya ID, ed. *Biologics to Treat Substance Use Disorders* : Springer International Publishing; 2016:75-103.
8. Wang Q, Zhang L, Kuwahara K, et al. Immunodominant SARS Coronavirus Epitopes in Humans Elicited both Enhancing and Neutralizing Effects on Infection in Non-human Primates. *ACS infectious diseases*. 2016;2(5):361-376.
9. Yuan M, Wu NC, Zhu X, et al. A highly conserved cryptic epitope in the receptor-binding domains of SARS-CoV-2 and SARS-CoV. *Science*. 2020.

Table I: Percentage of viruses with D614 in US states

| State | Total Seq | D Mutants | G Mutants | % D | Total Deaths (04/09/20) |
|-------------|-----------|-----------|-----------|------|-------------------------|
| Washington | 387 | 324 | 55 | 83.7 | 431 |
| Oregon | 4 | 3 | 1 | 75 | 38 |
| California | 30 | 22 | 8 | 73.3 | 507 |
| Connecticut | 84 | 44 | 40 | 52.3 | 335 |
| New York | 40 | 7 | 33 | 17.5 | 6268 |
| Virginia | 28 | 9 | 19 | 32.1 | 75 |
| Wisconsin | 29 | 9 | 20 | 31.0 | 103 |
| Utah | 22 | 10 | 12 | 45.4 | 13 |

Figure Legends

Figure 1 : A) Linear regression of average case fatality rate (CFR Average; Y-axis) with percentage of viruses exhibiting an aspartate (D) at amino acid position 614 in the viral envelope spike protein (Percentage D614; X-axis). B) Same linear regression as A but using median CFR. C) Table of underlying data used in the regression.

Figure 2 : A) Location of D614 (CPK depiction; arrow) in viral envelope trimeric spike structure. B) close-up of D614 showing optimal packing with T858 in adjacent monomer helical stalk, which would result in a cavity upon mutation to G614, which has no side chain. C) Energy calculation for D to G mutation for different 3D structural conformations of the SARS-CoV-2 viral spike.

Figure S1: Linear regression of average and median CFR analysis with United Kingdom included (red dot).

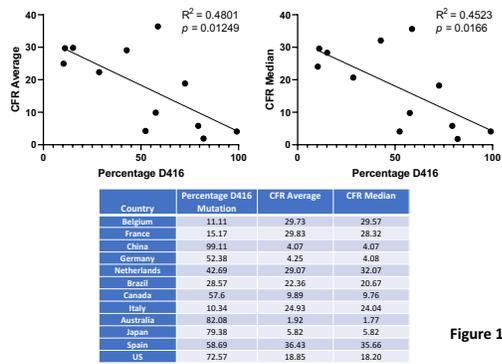


Figure 1

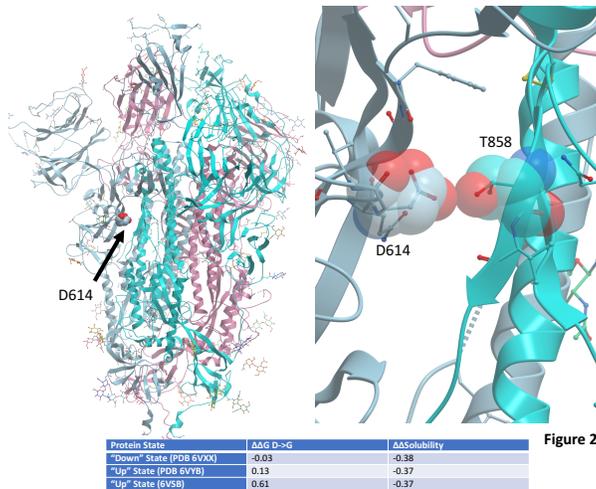


Figure 2