Received: 17 April 2020 | Revised: 30 April 2020 | Accepted: 2 May 2020

DOI: 10.1111/ijcp.13525

SHORT REPORT

INFECTIOUS DISEASES

THE INTERNATIONAL JOURNAL OF CLINICAL PRACTICE WILEY

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

Manuel Becerra-Flores | Timothy Cardozo

Department of Biochemistry and Molecular Pharmacology, NYU Langone Health, New York, NY, USA

Correspondence
Timothy Cardozo, NYU Langone Health, 550 First Avenue, New York, NY 10016, USA.
Email: cardot01@nyumc.org

Funding information
No funding sources encumber this work.

Abstract

Aim: The COVID-19 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 destabilise 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.

1 | INTRODUCTION

In December 2019, an outbreak of pneumonia 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 significant 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 compare case fatality rates (CFRs) of SARS-CoV-2 variants, we calculated a linear regression of global, country-by-country CFR versus country-by-country G614 percentage and analysed structural differences of the D to G mutation using published cryo-electron microscopy 3D structures of the SARS-CoV-2 viral spike.

2 | 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 CFR was calculated from total cases and total deaths per day in each country, considering an average of 11 days from hospitalisation 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 (March 30, 2020-April 6, 2020). Both the average and median CFR were calculated for analysis. Standard linear regression was performed for CFR versus G614 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, or 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 neighbouring sidechains (10.0 Å) were minimised using the Biased-Probability Monte Carlo algorithm for 10^6 steps.[4] 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: $\Delta\Delta G = \Delta G_{G\text{-}Mutant} - \Delta G_{D\text{-}WildType}$. Thus, a positive value indicates destabilisation of the mutant G relative to the wild-type D, since ΔG was a negative value for both forms. All biophysical calculations, molecular modelling and molecular graphics were performed using ICM-Pro (Molsoft LLC, La Jolla, CA).

What's known

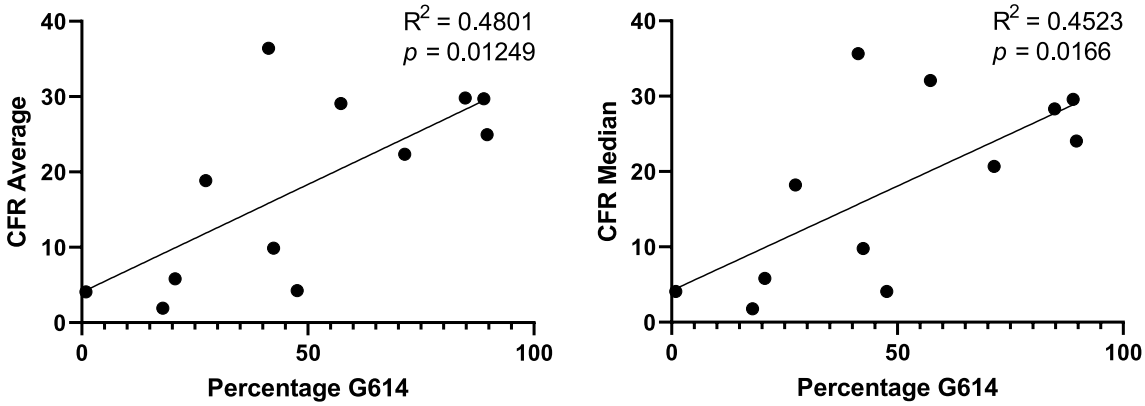
- The health significance of the D614G SARS-CoV-2 variant is unknown.

What's new

- A molecular clue to viral molecular pathogenesis of COVID-19 disease.

3 | RESULTS

SARS-CoV-2 viral spike amino acid sequences from strains infecting patients in Europe exhibited a predominance of glycine at amino acid

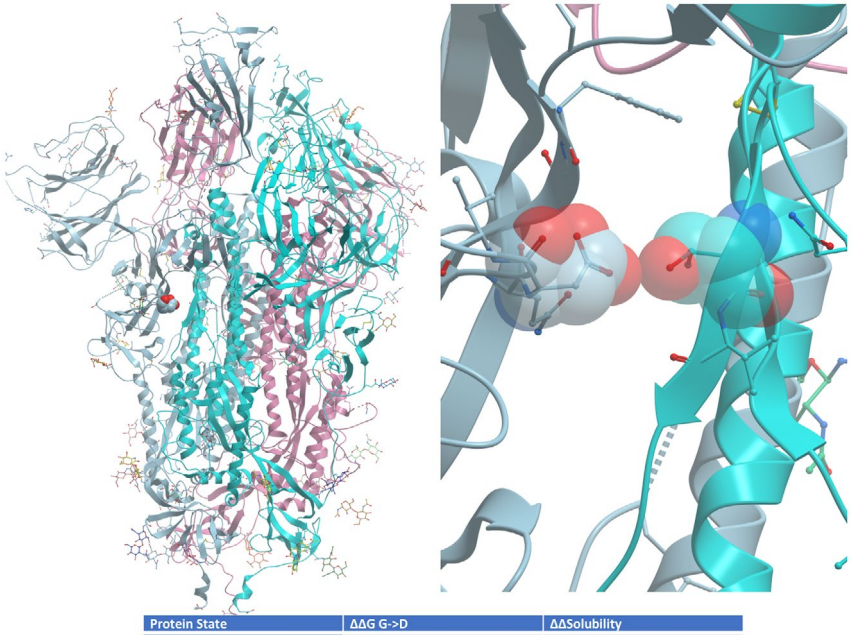


Country	Percentage G614 Mutation	CFR Average	CFR Median
Belgium	88.89	29.73	29.57
France	84.82	29.83	28.32
China	0.89	4.07	4.07
Germany	47.62	4.25	4.08
Netherlands	57.30	29.07	32.07
Brazil	71.43	22.36	20.67
Canada	42.40	9.89	9.76
Italy	89.66	24.93	24.04
Australia	17.91	1.92	1.77
Japan	20.62	5.82	5.82
Spain	41.30	36.43	35.66
US	27.42	18.85	18.20

FIGURE 1 A, Linear regression of average case fatality rate (CFR Average; *Y*-axis) with percentage of viruses exhibiting a glycine (G) at amino acid position 614 in the viral envelope spike protein (Percentage G614; *X*-axis). B, Same linear regression as A but using median CFR. C, Table of underlying data used in the regression

TABLE 1 Percentage of viruses with G614 in US states (April 6, 2020)

USA State	Total seq	D mutants	G mutants	%G	Total deaths (April 9, 2020)
Washington	379	324	55	16.3	431
Oregon	4	3	1	25	38
California	30	22	8	26.7	507
Connecticut	84	44	40	47.7	335
New York	40	7	33	82.5	6268
Virginia	28	9	19	67.9	75
Wisconsin	29	9	20	69	103
Utah	22	10	12	54.6	13



Protein State	ΔΔG G->D	ΔΔSolubility
"Down" State (PDB 6VXX)	-0.03	-0.38
"Up" State (PDB 6VYB)	0.13	-0.37
"Up" State (6VSB)	0.61	-0.37

FIGURE 2 A, Location of D614 (CPK depiction; arrow) in viral envelope trimeric spike structure. B, Close-up of D614 showing optimal packing with T859 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

position 614 in the viral spike (G614), while countries in the Far East exhibited a high percentage of D614 viral spikes. Based on CFR and SARS-CoV-2 viral spike sequence data available on April 6, 2020, both the average and median CFR correlate strongly ($P < .02$) with the proportion of viruses in the same geographic region bearing glycine at position 614 (G614) in the viral spike (Figure 1). For example, the proportion of recorded viruses in China exhibiting G614 was less than 1% 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 glycine (G) instead of aspartate (D) at amino acid residue number 614 in the viral spike (S) protein can be assessed (Table 1), revealing dramatically lower G614 and absolute number of fatalities in Western USA states.

Modelling 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 destabilising relative to D for the local area around the amino acid 614 position, apparently from loss of packing with the side-chain of threonine 859 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 destabilise protein tertiary and quaternary structure.[5] Thus, the structures predict a higher energy barrier for rearrangement of the viral spikes harbouring G614 to their infectious form (ie, 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).

4 | DISCUSSION

Recent molecular epidemiology 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 correlated strongly with CFR during an early time window in the pandemic 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 modelling suggests that G614 destabilises the infectious form of the viral spike protein, thereby favouring 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 identified,[8] and a recent report emphasised that key cryptic antibody-targeted neutralisation epitopes in the SARS family of viral spikes are accessible only in the conformation in which the ACE-2 binding site is exposed.[9]

There are several potential limitations and/or confounding issues in the data we used to calculate country-by-country CFR. First, the data assumes unbiased comorbidities, diagnostic testing and fatality reporting, which were highly variable in prevalence and accuracy across countries during the time window studied. For example, large volume testing took place in South Korea during the window studied, and the United Kingdom is an outlier with an unusually high CFR, which weakens the correlation when included (Figure S1). This outlier may be caused by the unusually low incidence of active cases being reported in the United Kingdom during the 2 weeks under study. Such selection bias is unlikely to be a confounder in this analysis, however, since the same reporting or incidence bias would have to occur in the same direction and for the same reason across diverse countries. 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 a limited window in the pandemic may be optimal to detect this correlation and avoid the confounder of successful medical treatment that may artificially reduce the organic CFR risk of the virus.

Based on our results, short-term epidemiologic models of the pandemic aiming to predict the incidence and prevalence of severe disease and fatalities over time, which do not take the G614 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, analysed the data, visualised the data and wrote the manuscript. M.B-F. edited the manuscript and performed the study, including data acquisition, data analysis design and data visualisation.

DISCLOSURES

No conflicts of interest encumber this work.

ORCID

Timothy Cardozo https://orcid.org/0000-0002-0643-4497

REFERENCES

1. Roser M, Ritchie H, Ortiz-Ospina E. Coronavirus Disease (COVID-19)—Statistics and research. *OurWorldInDataorg*. 2020. https://ourworldindata.org/coronavirus (Online Resource). Accessed April 6, 2020.
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;181:281-292.
3. Wrapp D, Wang N, Corbett KS, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*. 2020;367:1260-1263.
4. Abagyan R, Totrov M. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *J Mol Biol*. 1994;235:983-1002.
5. Xue M, Wakamoto T, Kejlberg C, et al. How internal cavities destabilize a protein. *Proc Natl Acad Sci USA*. 2019;116:21031-21036.
6. Acosta PL, Caballero MT, Polack FP. Brief history and characterization of enhanced respiratory syncytial virus disease. *Clin Vaccine Immunol*. 2015;23:189-195.
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*. Basel, Switzerland: 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 Infect Dis*. 2016;2:361-376.
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;368:630-633.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Becerra-Flores M, Cardozo T. SARS-CoV-2 viral spike G614 mutation exhibits higher case fatality rate. *Int J Clin Pract*. 2020;74:e13525. https://doi.org/10.1111/ijcp.13525