Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
The D614G mutations in the SARS-CoV-2 spike protein: Implications for viral infectivity, disease severity and vaccine design

Danielle C. Groves, Sarah L. Rowland-Jones*, Adrienn Angyal

Florey Institute for Host-Pathogen Interactions and the Department of Infection, Immunity and Cardiovascular Science (IICD), Sheffield University Medical School, Beech Hill Road, Sheffield, S. Yorkshire, S102RX, United Kingdom

1. Introduction

Unlike many RNA viruses, coronaviruses possess a genetic proofreading mechanism due to the presence of a non-structural protein (nsp) with 3'-5' exoribonuclease activity (nsp14) [1,2]: therefore, the rate of variant accumulation is slower than for other RNA viruses such as HIV-1 or influenza A. However, antigenic drift is known to occur among the endemic human coronaviruses, and during the first SARS outbreak in 2003, a single amino acid mutation in SARS-CoV (D480 A/G) within the receptor binding domain (RBD) of the Spike (S) protein became the dominant variant due to its ability to escape from neutralising antibodies [3]. SARS-CoV-2 sequence diversity was initially thought to be very low [4], but the unprecedented level of viral sequence generation and sharing that has occurred in 2020 has led to an increasing number of variant sequences (over 148,000 by early October 2020) being deposited in the GISAID database (originally the Global Initiative on Sharing All Influenza Data, www.gisaid.org). (see Fig. 1)

It is important to monitor the mutations accumulating within the SARS-CoV-2 genome, not only to follow the geographical spread of the virus but also promptly to identify antigenic variation that may affect immune responses to the virus. The accessibility of whole viral genome sequences collected around the world in the GISAID database enabled Korber and colleagues to develop a pipeline to identify Spike variants that were increasing in frequency across different geographic locations: this study highlighted the increasing dominance of a point mutation, D614G, in the Spike protein in viral isolates from the USA and Europe [5]. More recent studies suggest that the D614G variant is close to reaching fixation around the world [6]. Herein we discuss the importance of the Spike D614G mutation in terms of the epidemiology of SARS-CoV-2 infection and its impact for the immune response and vaccine design.

2. Identification of the D614G mutation and its clinical associations

In February 2020 the first whole genome of the novel coronavirus, now known as SARS-CoV-2, was published, using a combination of Illumina and Nanopore sequencing [7]. Three complete genome sequences were submitted to GISAID (BetaCoV/Wuhan/IVDC-HB-01/2019, accession ID: EPI_ISL_402119; BetaCoV/Wuhan/IVDC-HB-04/2020, accession ID: EPI_ISL_402120; BetaCoV/Wuhan/IVDC-HB-05/2019, accession ID: EPI_ISL_402121). Building on this work, Kim et al. at the Institute for Basic Science, South Korea, generated a high resolution map of the SARS-CoV-2 genome, by combining Nanopore long-read RNA sequencing with DNA nanoball sequencing to characterise the complexities of the viral transcriptome. With this approach, they managed to analyse the entire length of the viral genome and produce accurate readings of
many short fragments of genomic and sub-genomic RNA [8].

Korber et al. made good use of the accumulating GISAID sequence data to develop a bioinformatics approach that could identify specific viral variants that were becoming increasingly common in particular geographic locations. This led to the identification of variants carrying the D614G mutation in the Spike protein that were rapidly becoming the dominant viral strains across the world, even in regions where the D614 strain had initially caused infection. They noted that this mutation is almost always accompanied by three other mutations: C241T is located in the 5′UTR region, there is a silent mutation, C3037T, and C14408T results in the P323L amino acid change in the RNA-dependent RNA polymerase (RdRp) [5]. Prominent amongst the UK sequencing data they analysed were SARS-CoV-2 sequences in one Northern UK city, Sheffield, where the initial presence of D614 strains had been superseded by G614 isolates.

The Sheffield Teaching Hospitals NHS Trust, in conjunction with the Sheffield Institute for Translational Neuroscience (SiTRAN), had produced the first two sequences in the UK of SARS-CoV-2 from patient samples in late March. Since then, the Sheffield COG-UK sequencing group has sequenced over 2000 genomes with >90% coverage for the COVID-19 Genomics UK (COG-UK) consortium (https://www.gov.uk/government/news/uk-launches-whole-genome-sequence-alliance-to-map-spread-of-coronavirus). All of the Sheffield COG-UK sequences have been generated by Oxford Nanopore technology [9], then the sequences were analysed by Read Assignment, Mapping, and Phylogenetic Analysis in Real Time (RAMPART) (https://github.com/artic-network/rampart) before uploading to GISAID.

It was possible to link the Sheffield SARS-CoV-2 sequences with clinical data for 999 patients, extracted from electronic patient records, as well as from the clinical virology laboratories where the initial sample was analysed. This analysis showed a correlation between the D614G mutation and the cycle threshold (CT) values from the real-time polymerase chain reaction (RT-PCR) used for clinical diagnosis, suggesting that the variant is associated with increased viral load - this could suggest that the D614G mutation makes the virus more infectious. However, analysis of clinical data from the Sheffield cohort showed no relationship between the D614G mutation and disease severity (such as the need for hospital admission or transfer to the Intensive Care Unit) [5]. Further reports from other patient cohorts described similar findings: in the Washington State outbreak, G614 replaced the original CoV-2 strain expressing D614 over time, which was associated with increased CT values but no evidence for more severe disease. Similarly, studies in a cohort in Chicago showed that strains expressing G614 were associated with higher airway viral loads but not with worse disease outcomes [10]. Nevertheless, a study looking at the reported case fatality rate (cfr) of Covid-19 in different countries found a significant correlation between cfr and the relative frequency of the G614 variant [11], so further studies are probably warranted. In a recent paper deposited on BioRxiv, the impact of the G614 variant on viral load was confirmed in vivo using engineered whole SARS-CoV-2 variants, differing only at position 614, in a hamster infection model [12].

3. Impact of the D614G mutation on spike protein structure and interactions with host cells

The Sars-CoV-2 Spike protein is a class I fusion protein that forms trimers on the viral surface: it is heavily glycosylated, which enables entry into host cells [13–15]. The target receptor for entry into the host cell is the angiotensin-converting enzyme 2 (ACE2), which is highly expressed throughout the body. Receptor binding occurs through the receptor binding domain (RBD), ultimately leading to the fusion of the viral and host cell membranes [16].

Each Spike proteome protein consists of S1 and S2 subunits and a single transmembrane (TM) anchor [15]. Korber et al. used the available structures to map the D614G substitution to the surface of S1 in the spike proteome, where cryo-electron microscopy studies have showed that it forms a hydrogen bond with the T859 residue on S2 of the neighbouring proteome. They suggested that the G614 mutation would disrupt this bond and could potentially also affect glycosylation in the adjacent N616 site [5].
Korber and colleagues showed that in vitro the G614 mutation in spike-pseudotyped virus generated higher titres of infectious virus than virus expressing the D614 spike [5], consistent with the clinical data suggesting the G614-expressing strains are more infectious than the ancestral variant. Further studies by Li et al. examined the impact of a range of mutations in the Sars-CoV-2 Spike protein on infectivity [18]. They selected three groups of naturally occurring variants and experimental mutants and constructed pseudotyped viruses in order to study the effect of the mutations in vitro. Their results showed that pseudotyped viruses expressing either the D614G single mutation or a combination of mutations that included D614G are more infectious than the reference strain, whereas no difference was found between single D614G and D614G combination variants, which suggests that the enhanced infectivity is most likely due to the presence of D614G itself. They also pointed out that mutations affecting glycosylation of viral proteins could significantly affect virus-host interactions.

The Sars-CoV2 Spike protein is heavily glycosylated, with 22 putative glycosylation sites [14], but only a few of them are documented as sites of mutations in the GISAID database to date (N74K, N149H, and T719A). Experimental double deletions of glycosylation sites in the RB domain of the spike protein led to a drastic reduction in viral infectivity [19].

In a recent study using ACE2 orthologues, Yurkovetskiy et al. showed that the increased infectivity of the D614G variant is not specific for the human ACE2 receptor but also increases the ability of the D614G strain to enter cells expressing equivalent receptors from a variety of mammalian species, suggesting that the mutation has not been selected by the spread of the virus within humans [6]. They demonstrated that the mutation had no impact on spike protein synthesis, processing or incorporation into viral particles, nor did it lead to higher affinity binding to the ACE2 receptor. When comparing the tertiary structures of the two variants using cryo-electron microscopy, their atomic model showed that the mutation has two consequences. Firstly, the D to G substitution at position 614 within the Spike protein disrupts the inter-protomer hydrogen bond with T859, thereby weakening the stability of the trimer (which they described as “loosening the latch” that secures the two protomers together). Secondly, the intra-protomer distance between the backbone amide of residue 614 and the backbone carbonyl group of residue 647 is shortened, thereby stabilising the C-terminal domain of the protein. They went on to show that there was a substantial difference between the two variants in the presentation of the protomers in the “open” conformation that allows ACE2 binding, with D614G protomers being much more likely to assume this “open” conformation than the D614 variants [6]. Taken together these data suggest that the main effect of the D614G mutation is to increase the availability of spike trimer components in the conformation that permits the most efficient binding of the spike protein to ACE2.

4. The effects of the D614G mutation on the immune response towards the virus

The structural studies by Yurkovetskiy and colleagues showed that the gain of infectivity provided by the G614 mutation correlated with a higher proportion of spike proteins in an open conformation. Similar data were generated by molecular dynamic simulations: this study also showed that the RBD was more exposed in spikes expressing G614, which could affect the vulnerability of the virus to antibody-mediated neutralisation [19]. Korber and colleagues examined the ability of D614 and G614-expressing pseudoviruses to be neutralised by a small panel of patient-derived polyclonal sera and found little difference between the two variants [5]. Similarly, in an extensive analysis of the antigenicity of spike mutations expressed in pseudoviruses, there was little evidence that the D614G substitution significantly affected neutralisation by a panel of monoclonal neutralising antibodies [6]. However, in a recent manuscript deposited on medRxiv, Weissman et al. report that the D614G spike mutation increases the susceptibility of SARS-CoV-2 to neutralisation [20]. They reported that the G614-bearing pseudovirus used for their in vitro studies was more susceptible to neutralisation by monoclonal antibodies specific for the RBD, as well as by convalescent sera from people infected with either the G614 or G614 forms of the virus (identified from amongst the Sheffield cohort). Similarly, in the engineered whole virus studies using the hamster model described earlier, sera from D614-infected animals consistently showed higher neutralisation titres against G614 than D614 viruses [12].

Another aspect to consider is whether the G614D mutation could affect cellular immune responses that are mounted against the virus. In addition to inducing the production of neutralising antibodies, a successful anti-viral vaccine candidate is likely also to need to stimulate a cellular response, which early data suggest would be more durable than antibody responses to SARS-CoV-2. Through T-cell epitope mapping using convalescent patient samples, Peng et al. identified a total of 41 peptides containing SARS-CoV-2 T cell epitope regions, 18 of which were derived from the Spike protein [21]. One of these peptides, S-34 (CTFEYSQPFLM-DLE), containing both CD4+ and CD8+ T cell epitopes, was recognised by 29% of the participants, while the peptides S-151 (NLLLQYGSFCTQNR) and S-174 (TDEMIAQYTSALLAG) were recognised by predominantly by CD4+ T cells in 24% and 18% of the participants, respectively. None of these epitopes spans the D614G position which might suggest that the mutation does not induce T cell escape, but more studies are needed to map T-cell epitopes in Spike in different populations with distinct HLA repertoires which may restrict epitopes in the vicinity of D614G.

5. Summary

Despite the relatively low rates of mutation described for coronaviruses, a mutation in the S1 sub-unit of the Spike protein of SARS-CoV-2 emerged and has become the dominant strain worldwide within a matter of months. Studies to date suggest that the mutation is associated with higher viral loads in patients and animal models, probably because it leads to a more open conformation adopted by individual spike protomers, enhancing the binding of the virus spike to the ACE2 receptor: however, the mutation does not appear to lead to worse disease outcomes in most clinical studies. Although initial studies suggested that the mutation had little impact on antibody recognition, more recent data imply that the G614 variant may be more susceptible to neutralisation. These important findings emphasise the value of generating and sharing real-time viral sequence data on a worldwide scale, which has been one of the most impressive features of the scientific efforts to combat the Covid-19 pandemic in 2020.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank Dr Thushan de Silva for his suggestions on the manuscript and for leading the Sheffield Covid-19 Genomic Group.

This work did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.
