

D614G substitution at the hinge region enhances the stability of trimeric SARS-CoV-2 spike protein

Arangasamy Yazhini, Das Swayam Prakash Sidhanta & Narayanaswamy Srinivasan*

Molecular Biophysics Unit; Indian Institute of Science; Bangalore, Karnataka, 560012, India.* Tel: +91 80 22932837; Fax: +91 80 23600535; N Srinivasan - Email: ns@iisc.ac.in; AY - yazhini@iisc.ac.in ; DSPS - sidhantap@iisc.ac.in *Corresponding author

Received March 28, 2021; Revised March 30, 2021; Accepted March 30, 2021, Published March 31, 2021

DOI: 10.6026/97320630017439

Declaration on official E-mail:

The corresponding author declares that official e-mail from their institution is not available for all authors

Declaration on Publication Ethics:

The authors state that they adhere with COPE guidelines on publishing ethics as described elsewhere at <https://publicationethics.org/>. The authors also undertake that they are not associated with any other third party (governmental or non-governmental agencies) linking with any form of unethical issues connecting to this publication. The authors also declare that they are not withholding any information that is misleading to the publisher in regard to this article.

Abstract:

Mutations in the spike protein of SARS-CoV-2 are the major causes for the modulation of ongoing COVID-19 infection. Currently, the D614G substitution in the spike protein has become dominant worldwide. It is associated with higher infectivity than the ancestral (D614) variant. We demonstrate using Gaussian network model-based normal mode analysis that the D614G substitution occurs at the hinge region that facilitates domain-domain motions between receptor binding domain and S2 region of the spike protein. Computer-aided mutagenesis and inter-residue energy calculations reveal that contacts involving D614 are energetically frustrated. However, contacts involving G614 are energetically favourable, implying the substitution strengthens residue contacts that are formed within as well as between protomers. We also find that the free energy difference ($\Delta\Delta G$) between two variants is -2.6 kcal/mol for closed and -2.0 kcal/mol for 1-RBD up conformation. Thus, the thermodynamic stability has increased upon D614G substitution. Whereas the reverse mutation in spike protein structures having G614 substitution has resulted in the free energy differences of 6.6 kcal/mol and 6.3 kcal/mol for closed and 1-RBD up conformations, respectively, indicating that the overall thermodynamic stability has decreased. These results suggest that the D614G substitution modulates the flexibility of spike protein and confers enhanced thermodynamic stability irrespective of conformational states. This data concurs with the known information demonstrating increased availability of the functional form of spike protein trimer upon D614G substitution.

Keywords: SARS-CoV-2, COVID-19, spike protein, D614G variant, mutation, normal mode analysis, residue contacts, frustration index, protein stability

Background:

According to epidemiological surveillance of the disastrous COVID-19 pandemic, the causing agent severe acute respiratory

syndrome coronavirus-2 (SARS-CoV-2) virus harbours mutations and is linked to geographical-specific etiological effects [1,2]. Although vaccination has begun in many parts of the world, the

emergence of more infectious new variants of SARS-CoV-2 from different geographical locations such as B.1.1.7 from England, B.1.351 from South Africa and P.1 from Brazil continues to challenge our combat against COVID-19 [3]. These variants harbour several mutations in the genome and some of them present at the spike protein of the virus. Spike protein of SARS-CoV-2 is a 1273aa long transmembrane glycoprotein and comprises of three modules namely a large ectodomain that protrudes from the surface, a single-pass transmembrane anchor and a short intracellular tail. The ectodomain has S1 and S2 regions responsible for host cell binding and viral-host membrane fusion, respectively. At the junction of S1 and S2 regions, S1/S2 cleavage site is present and within the S2 region, S2' cleavage site is located. S1 region comprises of *N*-terminal domain (NTD, 27-305), receptor binding domain (RBD, 331-528), *C*-terminal domain 2 (CTD2, 529-590) and *C*-terminal domain 3 (CTD3, 591-685). Depending on RBD orientation in the S1 region, protomers in the functional form of spike protein trimer adopt closed or open conformation [4].

Upon open conformation, RBD exposes angiotensin-converting enzyme 2 (ACE2) receptor binding region and interacts with a peptidase domain of the ACE2 receptor. This primary step clasps the virus onto the host surface [4]. Subsequent proteolysis at the S1/S2 cleavage site sheds the S1 region from the spike protein and cleavage at the S2' site near fusion peptide causes a large conformational change in the S2 region. Such conformational change leads to the insertion of fusion peptide to the host membrane and a six-helix bundle formation. At this state, spike protein bridges the viral envelope and host membrane. Hairpin-like bend in the S2 region brings both membranes to close proximity for fusion, following which genetic material gets injected into the cytoplasm of the host cell [5]. It is also noted that due to the multi-basic nature of the S1/S2 cleavage site, the SARS-CoV-2 spike protein can be preactivated by the furin enzyme during viral packaging [6]. In contrast to SARS-CoV infection, this process reduces the dependence of SARS-CoV-2 on target cell proteases for the succeeding infection.

Therefore, mutations in the spike protein that influence the initial steps for viral infection are associated with altered virus transmissibility and pathogenicity [1,7]. Early in the pandemic, the emergence of a new variant having D614G substitution in the spike protein was identified in China. Subsequently, the ancestral variant with aspartate at 614th position (S^{D614}) has been asynchronously superseded by the glycine variant (S^{G614}) worldwide [8]. This substitution is retained in the newly emerged variants (B.1.1.7, B.1.351 and P.1), indicating its positive selection [9]. Several studies in the past few months demonstrate that the D614G substitution

enhances the infectivity and efficient replication of the virus [8,10-12]. The substitution disrupts a salt bridge interaction between aspartate at the 614th position and threonine at the 859th position of an adjacent protomer, thereby alters local inter-residue interactions [12]. It prevents premature shedding of the S1 region while promotes protease cleavage at the S1/S2 site and increases the conformational sampling of open conformation [13,14]. Therefore, it is of interest to document the effect of D614G substitution on the structural flexibility, inter-residue interaction energies and thermodynamic stability of the spike protein trimer.

Materials and Methods:

Gaussian network model-based normal mode analysis:

To identify regions that precisely act as hinges for domain-domain motions facilitating the transition between closed and open conformations, we performed Gaussian network model (GNM)-based normal mode analysis (NMA). GNM-NMA is a robust method to accurately predict hinges [15] and the calculations were performed for closed (protein data bank or PDB code: 6VXX) and 1-RBD up conformations (PDB code: 6VYB). The slowest normal mode was considered for analyzing Gaussian dynamics of the spike protein trimer.

3-D structural model for D614G variant of the spike protein trimer:

We generated an *in silico* model for the D614G variant of spike protein trimer using structure editing tool in UCSF chimera with default parameters [16]. Sidechains were optimized using SCWRL 4.0 program [17]. Two D614G variant models were generated corresponding to closed and 1-RBD up conformations of the spike protein trimer based on the reference cryo-EM structures available in the PDB entries 6VXX and 6VYB, respectively. Besides, spike protein structures with D614G substitution have been released at the time of our study [14] and hence were included in this analysis.

Calculation of frustration index in the local residue contacts:

The effect of D614G substitution on local interaction energies was examined using Frustratometer algorithm [18], which follows the principle that a native protein comprises several conflicting residue contacts resulting in local frustration. Based on C β distance, the residue contacts are categorized as short range (<6.5Å), long range (6.5-9.5Å) and water-mediated (long range and exposed to solvent). The algorithm computes the frustration index for a given residue or residue contact. The value below -1 indicates that the interacting pair is highly frustrated, while the index between -1 to 0.78 or above 0.78 indicates that the interacting pair is neutrally or minimally frustrated, respectively. The frustration index is calculated at three levels namely 'mutational', 'configurational' and

'single-residue'. In the mutational frustration, other residue types replace residue type, while in the configurational frustration, all possible interaction types between the native residue pairs were sampled through altering residue configuration. In the case of single-residue level, only other residue types for frustration index

calculation replace a single residue site. Here, we analyzed all categories of frustration indices for the ancestral and dominant variants of spike protein trimer (SD^{614} and SG^{614}) in closed and 1-RBD up conformations.

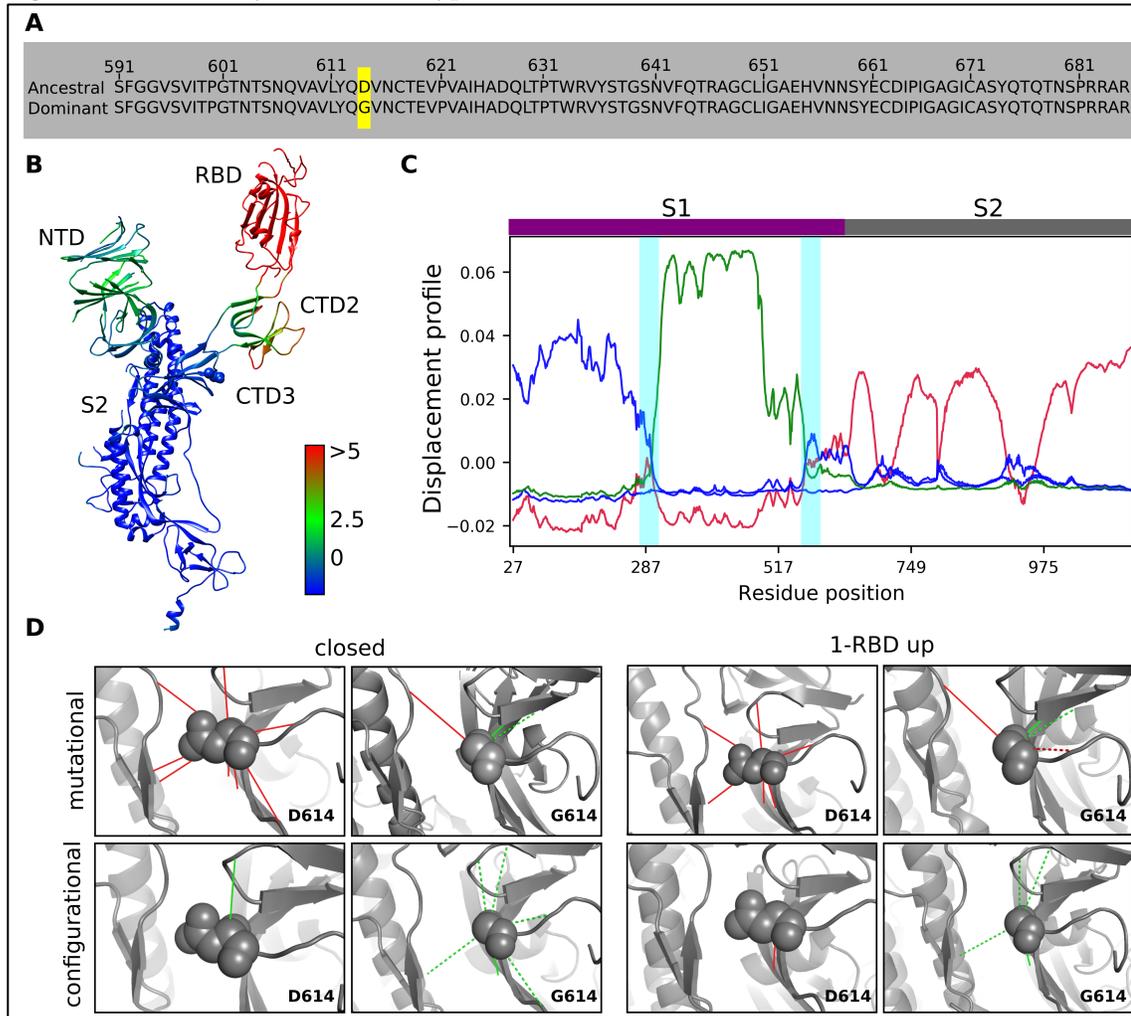


Figure 1. The significance of D614G substitution on the protein flexibility and interaction energy of residue contacts. A) Sequence alignment of CTD3 domain (591-685) between the ancestral (D614) and glycine (G614) variants of the spike protein. A yellow background highlights the substituted site. B) Cartoon representation of spike protein in RBD up conformation with domains labelled and colored based on the $C\alpha$ deviation with respect to closed conformation. Color scale blue to red indicates $C\alpha$ deviation (in Å) from low to high. The position 614 is highlighted by the sphere representation of the residue. C) Shown as a line plot is the displacement profile of protomers in the closed (crimson) and 1-RBD up (green- open; blue- closed) conformations. Hinges in the S1 region are highlighted with vertical cyan bars. S1 (magenta) and S2 (grey) regions are indicated by horizontal bars at the top of the plot. D) Shown is the mutational (top panel) and

configurational (bottom panel) frustrations that exist in the inter-residue contacts formed by aspartate or glycine at position 614. Green and red lines indicate minimally and highly frustrated interactions, respectively. Water-mediated interactions are represented as dashed lines and the variant residue is shown as a sphere. Results from closed and 1-RBD up conformations were shown only for a protomer (chain ID: A) within the trimer since similar patterns are observed for the other two protomers (Supplementary Table S1).

Calculation of thermodynamic stability of S^{D614} and S^{G614} variants:

To study the effect of D614G variation on the thermodynamic stability of the spike protein trimer, we calculated free energy changes upon aspartate to glycine substitution using buildmodel function in FoldX [19]. Five iterations of free energy calculations were carried out to obtain converged results. The free energy changes were also calculated for the reverse scenario where glycine was substituted to aspartate in cryo-EM structures of S^{G614} (PDB codes: 7KDK and 7KDL). Inferences of the results were derived from closed and 1-RBD up conformations of the spike protein trimer.

Results and Discussion:

The D614G substitution is present in the CTD3 domain and is highlighted in the sequence alignment between ancestral and dominant variants (Fig. 1A). Comparison of spike protein in closed and 1-RBD up conformations shows that besides RBD undergoing a large displacement between these two states, the NTD and CTD2 domains have considerable $C\alpha$ deviations with RMSD values 1.6Å and 3.4Å, respectively (Fig. 1B). However, CTD3 superposes well (RMSD 0.6Å) and hence likely facilitates domain-domain motions. To understand the association of position 614 to protein flexibility,

we analyzed displacement profiles of the slowest Gaussian mode of spike protein trimer derived from GNM-NMA. The crossover of displacement profiles from negative to positive or *vice versa* signifies the presence of a hinge. We observe that the crossover occurs consistently around Lys310-Phe318 (hinge-1) and Gly593-Val618 (hinge-2) regions in the CTD3 of both conformations and hence they serve as hinges (Fig. 1C). The hinge-1 mediates NTD-RBD motions while the hinge-2 mediates RBD-S2 motions (Fig. 1C). Given that glycine is a highly flexible residue, the substitution potentially influences the flexibility of hinge-2, thereby modulates RBD-S2 motions essential for closed to open transition. We analyzed the effect of D614G substitution on the interaction energies by computing the frustration index of residues in the S^{D614} and S^{G614} . As shown in Table 1, the frustration index of aspartate at 614th position in the S^{D614} is less than -1 for three protomers in closed and 1-RBD up conformations. Hence, in both the states, aspartate is highly frustrated. Conversely, the glycine in S^{G614} is neutrally frustrated, with the index values ranging from -1 to 0.78. This result was affirmed by similar observations when we analyze the cryo-EM structures of S^{G614} (released at the time of our analysis) [14]. It means that the frustrated residue at position 614 has become neutral upon the glycine substitution.

Table 1: Single-residue level frustration index of aspartate and glycine in the S^{D614} and S^{G614} , respectively. Results are shown for closed and 1-RBD up conformations.

Frustration index of the residue at 614 th position	S^{D614}		S^{G614} (in silico models)		S^{G614} (Cryo-EM structures)	
	Closed	1-RBD up	Closed	1-RBD up	Closed	1-RBD up
Protomer 1	-1.25	-1.24	-0.48	-0.50	-0.55	-0.75
Protomer 2	-1.25	-1.31	-0.42	-0.35	-0.19	-0.39
Protomer 3	-1.30	-1.28	-0.46	-0.37	-0.88	-0.31

Further, we calculated the mutational frustration index of residue contacts. In both conformations of S^{D614} , aspartate forms several intra- and inter-protomer contacts through direct, long-range electrostatic or water-mediated interactions (Table S1A), of which 8 are highly frustrated (Fig. 1D, top panel). In the case of S^{G614} in closed conformation, the glycine has two minimally frustrated contacts with Leu611 and Cys649 of the same protomer and a highly frustrated contact with Pro862 of an adjacent protomer (Fig. 1D, top left panel). In the 1-RBD up conformation of S^{G614} , the

glycine has the same contacts, as observed in the closed conformation, in addition to another contact with Asn616. The contacts with Leu611 and Cys649 are minimally frustrated (Fig. 1D, top right panel). Similar observations are seen for the cryo-EM structures of S^{G614} (Table S1A). Therefore, the overall number of highly frustrated contacts is reduced upon aspartate to glycine substitution.

The configurational frustration index, indicating how favourable a native contact relative to other possible contacts between interacting residue pairs, further shows that in the closed conformation, aspartate (S^{D614}) has one minimally frustrated contact with Arg646 (Fig. 1D, left bottom panel). In contrast, glycine (S^{G614}) has six minimally frustrated contacts (Fig. 1D, left bottom panel). Similarly, aspartate in 1-RBD up conformation of S^{D614} has a highly frustrated contact with Gly593, while upon glycine substitution (S^{G614}), this contact becomes minimally frustrated. In addition, this residue forms three other minimally frustrated contacts with Thr645, Arg646 and Thr859 (Fig. 1D, right bottom panel). Moreover, configurational indices calculated from the cryo-EM structures of S^{G614} indicate that glycine has three minimally frustrated contacts in both conformations and corroborate with the results from *in silico* models (Table S1B). Hence, glycine has more favourable contacts than aspartate between S1 and S2 regions of the same and adjacent protomers. Overall, our calculations of single residue, mutational and configurational frustrations reveal that glycine substitution modifies the local interaction energy in a favourable direction. Further, our calculations of the free energy changes upon D614G substitution show that the total free energy differences ($\Delta\Delta G$) between two variants of spike protein trimer are -2.6 kcal/mol and -2.0 kcal/mol for closed and 1-RBD up conformations, respectively. The differences are higher than the reasonable threshold ± 0.5 kcal/mol and hence the substitution stabilizes the spike protein [20]. When we perform reverse mutation in S^{G614} structures, the differences have positive values *i.e.*, 6.6 kcal/mol and 6.3 kcal/mol for closed and 1-RBD up conformations respectively, implying destabilizing effect. Together, it suggests that the glycine substitution creates a favourable local environment and enhances the overall stability of the spike protein trimer. Such effect may confer the increased availability of a functional form of spike protein, resulting in higher infectivity than the S^{D614} as reported by recent experimental studies [8,10,13].

Conclusions:

The D614G substitution in the SARS-CoV-2 spike protein, which is being intensively studied across the globe for COVID-19 prophylaxis and treatment, is under positive selection in the ongoing pandemic. This computational study demonstrates that D614G substitution occurs at the hinge region and potentially influences the conformational transition essential for human ACE2 receptor binding. Glycine changes the energetically frustrated local environment into favourable conditions for contacts present between S1 and S2 regions of the same protomer as well as from adjacent protomers. Consequently, the free energy of S^{G614} is lower than that of S^{D614} , indicating that the local changes in the interaction energies at the 614th position in each protomer have a significant

effect on the thermodynamic stability of the spike protein trimer. Our computational work based on a theoretical framework provides for the first time the protein flexibility and residue interaction energy-based rationale for the enhanced stability of the spike protein. These new findings add to the existing knowledge on the mechanism of increased transmissibility of S^{G614} and would help further investigations on the influence of additional substitutions acquired in the newly emerged SARS-CoV-2 variants of concern.

Acknowledgements:

This research is supported by the Department of Biotechnology, Government of India through the IISc-DBT partnership program and Bioinformatics and computational biology centre. This research is also supported by FIST program sponsored by the Department of Science and Technology, Government of India. Support from UGC, India - Centre for Advanced Studies and Ministry of Human Resource Development, India is gratefully acknowledged. DSPS acknowledges the financial support from CSIR, Government of India towards his research fellowship; NS is a J. C. Bose National Fellow.

Author contributions:

NS conceived the project and supervised the work. AY and DSPS performed the analysis. AY wrote the draft version of the manuscript and finalized it with the help of DSPS and NS. All authors read and approved the manuscript.

Conflict of interest:

There was no conflict of interest.

Supplementary table legend

Supplementary Table S1: The table contains details of frustration index of inter-residue contacts present at the position 614 of spike protein trimer in closed and 1-RBD up conformations. Table S1A in sheet 1 provides the mutational frustration index of contacts present at the 614th position in the ancestral (D614) and dominant (G614) variants of the spike protein trimer. Table S1B in sheet 2 provides the configurational frustration index of those contacts in the ancestral (D614) and dominant (G614) variants. Frustration state namely minimally, neutral or highly represents that the contact is energetically favourable, neutral or unfavourable, respectively. Three categories of residue contacts are considered as short range (below 6.5Å), long range (6.5-9.5Å) and water-mediated (long range and solvent exposed) based on the $C\beta$ distance between residue pairs.

References:

- [1] Brufsky A. *J. Med. Virol.* 2020. p. 1386–90. [PMCID: PMC7264516]
- [2] Mercatelli D & Giorgi FM. *Front Microbiol.* 2020;11. [PMID: 32793182]
- [3] Luan B *et al.* *FEBS Lett.* 2021. [PMID: 33728680]
- [4] Yan R *et al.* *Science.* 2020;367:1444–8. [PMID: 32132184]
- [5] Belouzard S *et al.* *Viruses.* 2012. p. 1011–33. [PMID: 22816037]
- [6] Shang J *et al.* *Proc Natl Acad Sci USA.* 2020;117:11727–11734. [PMID: 32376634]
- [7] Li Q *et al.* *Cell.* 2020;182:1284–1294.e9. [PMID: 32730807]
- [8] Korber B *et al.* *Cell.* 2020;182:812–827.e19. [PMID: 32697968]
- [9] Volz E *et al.* *Cell.* 2021;184:64–75.e11. [PMID: 33275900]
- [10] Plante JA *et al.* *Nature.* 2020;1–9. [PMID: 33106671]
- [11] Hou YJ *et al.* *Science.* 2020;370:1464–8. [PMID: 33184236]
- [12] Yurkovetskiy L *et al.* *Cell.* 2020;183:739–751.e8. [PMID: 32991842]
- [13] Zhang L *et al.* *Nat Commun.* 2020;11:1–9. [PMID: 33243994]
- [14] Gobeil SMC *et al.* *Cell Rep.* 2020;108630. [PMID: 33417835]
- [15] Haliloglu T *et al.* *Physical Rev Lett.* 1997;79:3090.
- [16] Pettersen EF *et al.* *J Comput Chem.* 2004;25:1605–12. [PMID: 15264254]
- [17] Krivov GG *et al.* *Proteins Struct Funct Bioinforma.* 2009;77:778–95. [PMID: 19603484]
- [18] Parra RG *et al.* *Nucleic Acids Res.* 2016;44:W356–60. [PMID: 27131359]
- [19] Schymkowitz *et al.* *Nucleic Acids Res.* 2005;1:W382–388. [PMID: 15980494]
- [20] Bromberg Y & Rost B. *BMC Bioinformatics.* 2009;10. [PMID: 19758472]

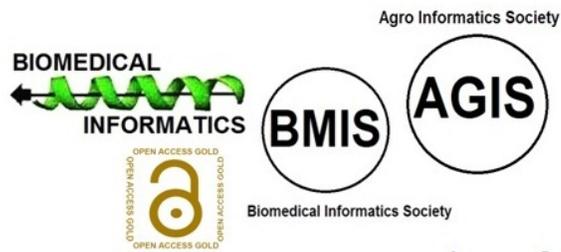
Edited by P Kanguane**Citation:** Yazhini *et al.* *Bioinformation* 17(3): 439–445 (2021)

License statement: This is an Open Access article which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. This is distributed under the terms of the Creative Commons Attribution License

Articles published in BIOINFORMATION are open for relevant post publication comments and criticisms, which will be published immediately linking to the original article for FREE of cost without open access charges. Comments should be concise, coherent and critical in less than 1000 words.

BIOINFORMATION

Discovery at the interface of physical and biological sciences



since 2005

BIOINFORMATION

Discovery at the interface of physical and biological sciences

indexed in

