Conformational Variability Assessment of the Mutation Sites for D614G, B.1.1.7, and B.1.351 using SSSCPreds

Hiroshi Izumi (izumi.h@aist.go.jp)
National Institute of Advanced Industrial Science and Technology (AIST) https://orcid.org/0000-0002-8369-3038

Laurence Nafie
Syracuse University https://orcid.org/0000-0002-3221-4332

Rina Dukor
BioTools Inc. https://orcid.org/0000-0002-1981-3381

Analysis

Keywords: Deep neural network, Conformational variability, Mutation, SARS-CoV-2, COVID-19

Posted Date: February 16th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-225013/v1

Abstract

Complementary techniques for the analysis of mutation sites at flexible regions, in which the position of atoms could not be determined by cryo-electron microscopy (Cryo-EM) such as the furin cleavage site of SARS-CoV-2, are necessary. The prediction data from SSSCPreds, a deep neural network-based prediction software of conformational flexibility or rigidity in proteins, can give insight into the conformational variability of mutation sites. We find that although the conformation of G614 is rigid, which is assigned as a left-handed (LH) α-helix-type one, that of D614 is flexible without the hydrogen bonding latch to T859. The rigidity of glycine which stabilizes the local conformation more effectively than that of aspartic acid with the latch, thereby contributes to the reduction of S1 shedding and increase in infectivity. Further it is predicted that no other amino acid allows the same conformation and stability as the glycine mutation in D614. The individual mutations in B.1.1.7 and B.1.351 have a lower effect and are not comparable to the overwhelming effectiveness of the D614G mutation. SSSCPreds provides important conformational flexibility insights into the deep neural network-based understanding of the current mutation sites and the potential for new ones in future.

Introduction

With the rapid expansion of the coronavirus disease 2019 (COVID-19) pandemic in 2020[1], a rate of 23.12 substitutions per year for SARS-CoV-2 is currently observed, and the evidence of possible reinfection with SARS-CoV-2 has been shown[2]. The infection has spread through the process of natural selection so that predicting and tracking the impact of spontaneous mutations is necessary. Further, complementary techniques for the analysis of mutation sites at flexible regions, in which the position of atoms could not be determined by cryo-electron microscopy (Cryo-EM) such as the furin cleavage site of SARS-CoV-2[3], are needed.

Especially, the possibility of the next pandemic of Variant of Concern 202012/01 (VOC-202012/01, B.1.1.7) identified in UK[4] and 501Y.v2 (B.1.351) having emerged in South Africa[5] are deeply concerning. After March 2020, only two mutations, RNA-directed RNA polymerase P323L (ORF1ab P4715L, ORF1b P314L) and spike protein D614G have nearly overwhelmed the original mutation sites (https://nextstrain.org/ncov/global?c=gt-S_614). As for the receptor-binding motif (RBM) of spike protein, before B.1.1.7 and B.1.351, the mutation of S477N has expanded in Australia and in Europe but has not led to an extension of the pandemic (https://nextstrain.org/ncov/global?c=gt-S_477).

Recently, we reported a deep neural network-based prediction program of conformational flexibility or rigidity in proteins (SSSCPreds)[6] using supersecondary structure code (SSSC)[7–9]. The sequence flexibility/rigidity map of SARS-CoV-2 RBD (receptor binding domain), obtained from SSSCPreds, resembles the sequence-to-phenotype maps of ACE2-binding (angiotensin-converting enzyme 2-binding) affinity and expression, which was experimentally obtained by the deep mutational scanning[10]. It suggests that the identical SSSC sequences among the ones predicted by three deep-neural-network-

based systems correlate well to the sequences with both lower ACE2-binding affinity and lower expression.

The frequency of mutations increases with the exponential increase in the number of infected people. The conformational flexibility of the protein sequences is deeply related to the ease of infection, and the accurate prediction is very important to make a countermeasure of COVID-19. In this paper, we report the conformational variability assessment of the mutation sites for D614G, and those of further strains B.1.1.7 and B.1.351.

Results

D614G mutation

As shown above, the D614G variant is now the dominant form worldwide[3]. Recently, Gobeil and coworkers described that Cryo-EM structures reveal altered RBD disposition; antigenicity and proteolysis experiments reveal structural changes and enhanced furin cleavage efficiency of the G614 variant[3]. However, the underlying factor of why glycine, and not other amino acids, can induce the effective strain replacement has not been explained.

The sequence flexibility/rigidity maps of all of the single amino acid mutations at the D614G mutation site using SSSCPreds indicate that only the mutation to glycine makes the other-type conformation (“T” conformation) rigid and reproduces the observed “T” conformations of Cryo-EM structures (Fig. 1a). On the other hand, although SSSCPred200 suggests the “T” conformation for D614, SSSCPred100 and SSSCPred predict the β-sheet-type conformations (“S” conformations). This means that the site of D614 is flexible without the hydrogen bonding latch between D614 and T859 (Fig. 1b,c).

Both observed “T” conformations of Cryo-EM structures for D614 and G614 (Protein Data Bank [PDB] ID code 6XR8 and 6XS6) are the same conformation[11], which is assigned as a left-handed (LH) α-helix-type one (Fig. 1b,c)[12]. In general, the LH α-helix is stabilized by only glycine because glycine does not have chirality. The rigidity of glycine which stabilizes the local conformation more effectively than that of aspartic acid with the latch, thereby contributes to the reduction of S1 shedding[13] and increase in infectivity without the latch between D614 and T859.

B.1.1.7 and B.1.351

As shown in Fig. 2 and Extended Data Fig. 1, the SSSCPreds data of the expanded S477N variant before B.1.1.7 and B.1.351 indicate that the S477N mutation increases the rigidity of the protein foundation GFNCYFPLQ. The foundation is located on the edge of flexible regions, in which the position of atoms could not be determined. The ratio of frequencies of the S477N mutation have gradually increased, but the mutation has not contributed to the pandemic. The SSSCPreds data of the N501Y mutation for B.1.1.7 show a similar increased stability of the foundation, but those of the E484K and N501Y mutations for B.1.351 suggest the equivalent rigidity of the original sequence (Fig. 2, Extended Data Fig. 1, and

Extended Data Fig. 5). The high ACE2-binding affinity of the single N501Y mutation has been reported[10]. Although the K417N mutation site can also contact with ACE2, the SSSCPreds data suggest flexibility of the nearby site (Extended Dataset 1 and Extended Data Fig. 4).

The mutations of HV69-70 deletion, Y144 deletion, and P681H for B.1.1.7 and D80A, D215G, and R246I for B.1.351 also correspond to the edge of flexible regions, in which the position of atoms could not be determined (Fig. 2, Extended Dataset 1, and Extended Data Fig. 2). The SSSCPreds data of P681H mutation indicate the stabilization of the sequence CASYQT with identical β-sheet-type conformations before the furin cleavage site (Extended Data Fig. 3).

From the SSSCPreds data, the mutations T716I and D1118H for B.1.1.7 (Fig. 2, Extended Dataset 1, and Extended Data Fig. 3) and the mutation A701V for B.1.351 (Extended Dataset 1 and Extended Data Fig. 5) increase the rigidity of the near sites. These mutation sites are located on the loop regions. The mutations T716I, S982A, and D1118H for B.1.1.7 and the mutation A701V for B.1.351 seem to contribute to the process that the transformation to post-fusion hairpin state in the membrane fusion proceeds smoothly.

Discussion

In this study, the conformational variability of the mutation sites for D614G, B.1.1.7, and B.1.351 has been evaluated by using SSSCPreds. The overwhelming D614G mutation is rationalized by the more rigid conformation of glycine than that of aspartic acid, which is assigned as a LH α-helix-type one. In view of the conformational variability except the strength of binding affinity, individual mutations comparable to the overwhelming D614G mutation are not found in B.1.1.7 and B.1.351. As of September, the frequencies of N477 against S477 (https://nextstrain.org/ncov/global?c=gt-S_477) was similar to the present frequencies of Y501 against N501 (https://nextstrain.org/ncov/global?c=gt-S_501). B.1.1.7 has Q27stop of ORF8 (https://nextstrain.org/ncov/global?c=gt-ORF8_27). The mutation Q27stop of ORF8 seems to be correlated with milder disease[14]. Another factor that leads to a new more virulent form is the P681H mutation at the furin cleavage site. The P681H mutation has appeared as of August (https://nextstrain.org/ncov/global?c=gt-S_681), but no report of a highly virulent form exists. The E484K mutation of B.1.351 (https://nextstrain.org/ncov/global?c=gt-S_484) has already appeared as of September, but the toxicity of B.1.351 seems to be equivalent to that of the original sequence. The S24L mutation of ORF8 is now going around in US (https://nextstrain.org/ncov/global?c=gt-ORF8_24) but are not overwhelming. It is suggested with the nextstrain data that the integrated effect of mutations contributes to the expansion of B.1.1.7 and B.1.351 but are not overwhelming.

Declarations

ONLINE METHODS

The FASTA-format files containing the amino acid sequences and SSSCs of protein subunits were obtained from the observed Protein Data Bank (PDB) files[15] by using the SSSCview program (available online at https://staff.aist.go.jp/izumi.h/SSSCPreds/index-e.html)[9].

The original and mutation sequences of protein subunits were converted to the predicted SSSCs by using the SSSCPreds program (available online at https://staff.aist.go.jp/izumi.h/SSSCPreds/index-e.html)[6].

Data availability

The reference models and the original amino acid sequences were downloaded from the PDB. The mutation sequences were obtained based on References 4 and 5.

ACKNOWLEDGEMENTS

This work was supported partly by JSPS KAKENHI Grant Number JP19K05431.

Author contributions

H.I. designed methodology and software and performed research; L.A.N. and R.K.D. organized research; and H.I., L.A.N., and R.K.D. wrote the paper.

Competing interests

The authors declare no competing financial interest.

ORCID

Hiroshi Izumi https://orcid.org/0000-0002-8369-3038

Laurence A. Nafie https://orcid.org/0000-0002-3221-4332

Rina K. Dukor https://orcid.org/0000-0002-1981-3381

References

1. Grubaugh, N. D., Hanage, W. P. & Rasmussen, A. L. Making sense of mutation: What D614G means for the COVID-19 pandemic remains unclear. *Cell* **182**, 794–795 (2020).
2. Tillett, R. L. *et al.* Genomic evidence for reinfection with SARS-CoV-2: a case study. *Lancet Infect. Dis.* **21**, 52–58 (2020).
3. Gobeil, S. M-C. *et al.* D614G mutation alters SARS-CoV-2 spike conformation and enhances protease cleavage at the S1/S2 junction. *Cell Rep.* **34**, 108630 (2021)
4. Wise, J. Covid-19: New coronavirus variant is identified in UK. *The BMJ* **371**, m4857 (2020).
5. Tegally, H. *et al.* Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *Medrxiv*

https://doi.org/10.1101/2020.12.21.20248640.

6. Izumi, H., Nafie, L. A. & Dukor, R. K. SSSCPreds: Deep neural network-based software for the prediction of conformational variability and application to SARS-CoV-2. *ACS Omega* **5**, 30556–30567 (2020).
7. Izumi, H., Ogata, A., Nafie, L. A. & Dukor, R. K. A revised conformational code for the exhaustive analysis of conformers with one-to-one correspondence between conformation and code: Application to the VCD analysis of (*S*)-ibuprofen. *J. Org. Chem.* **74**, 1231–1236 (2009).
8. Izumi, H., Wakisaka, A., Nafie, L. A. & Dukor, R. K. Data mining of supersecondary structure homology between light chains of immunogloblins and MHC molecules: absence of the common conformational fragment in the human IgM rheumatoid factor. *J. Chem. Inf. Model.* **53**, 584–591 (2013).
9. Izumi, H. Homology searches using supersecondary structure code. *Methods Mol. Biol.* 1958, 329–340 (2019).
10. Starr, T. N. *et al.* Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell* **182**, 1295–1310 (2020).
11. Yurkovetskiy, L. *et al.* Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. *Cell* **183**, 739–751 (2020).
12. Mensch, C., Barron, L. D. & Johannessen, C. Ramachandran mapping of peptide conformation using a large database of computed Raman and Raman optical activity spectra. *Phys. Chem. Chem. Phys.* **18**, 31757–31768 (2016).
13. Zhang, L. *et al.* SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. *Nat. Commun.* **11**, 6013 (2020).
14. Young, B. E. *et al.* Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study. *Lancet* **396**, 603–611 (2020).
15. Touw, W. G. *et al.* A series of PDB-related databanks for everyday needs. *Nucleic Acids Res.* **43**, D364–D368 (2015).

Figures

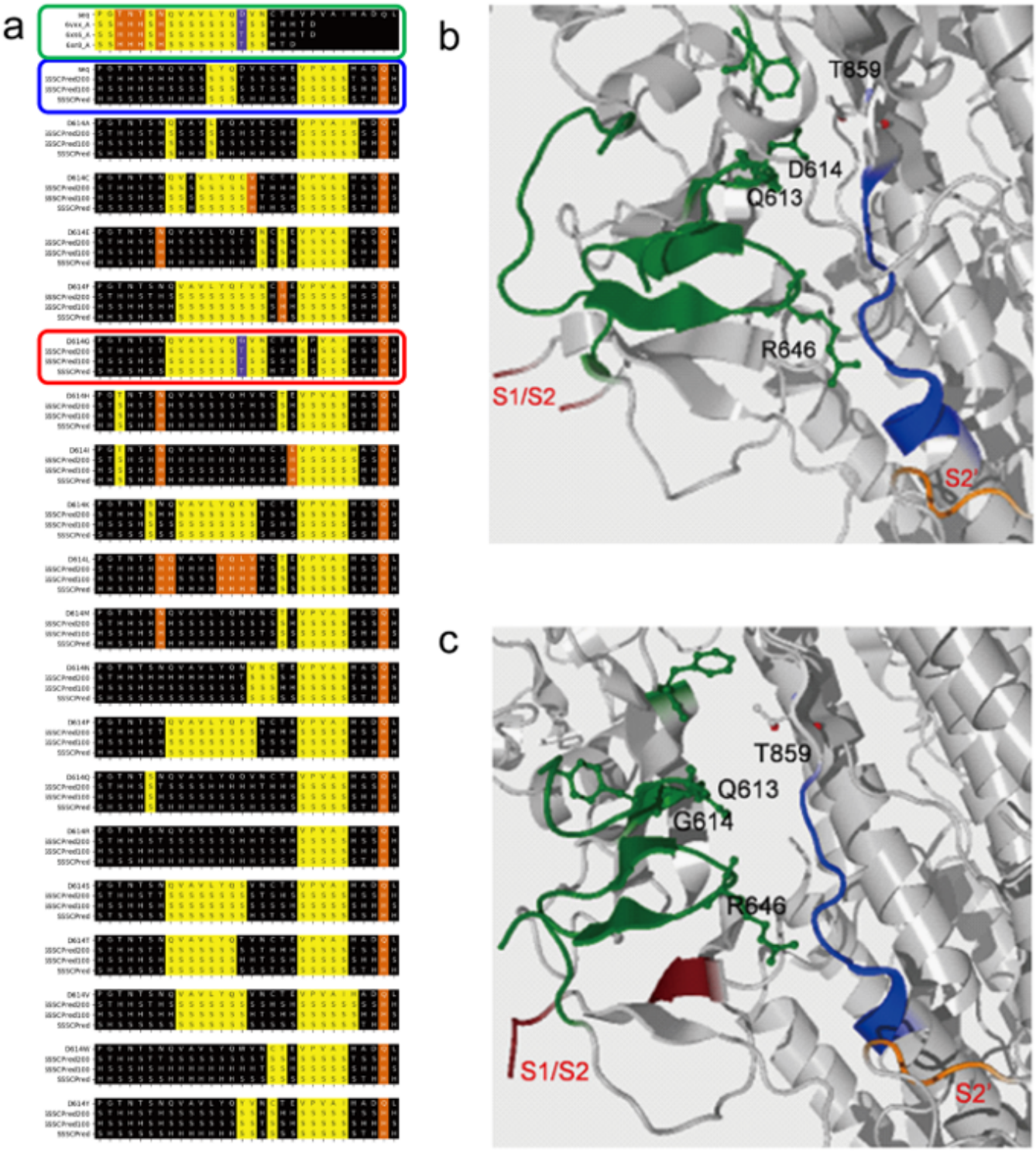

Figure 1

Conformational variability of D614G mutation site. a, Sequence flexibility/rigidity maps of all of the single amino acid mutations at the D614G mutation site (orange, identical α-helix-type conformations; yellow, identical β-sheet-type conformations; purple, identical other-type conformations; black, flexible conformations; green frame, observed sequences; blue frame, original sequences; red frame, mutation sequences). b,c, Cryo-EM structures of D614G mutation sites (green, SD2; blue, undeformable motif of

S26; magenta, S1/S2 cleavage site; orange, S2’ cleavage site). D614 (PDB ID code 6XR8) (b) and G614 (PDB ID code 6XS6) (c).

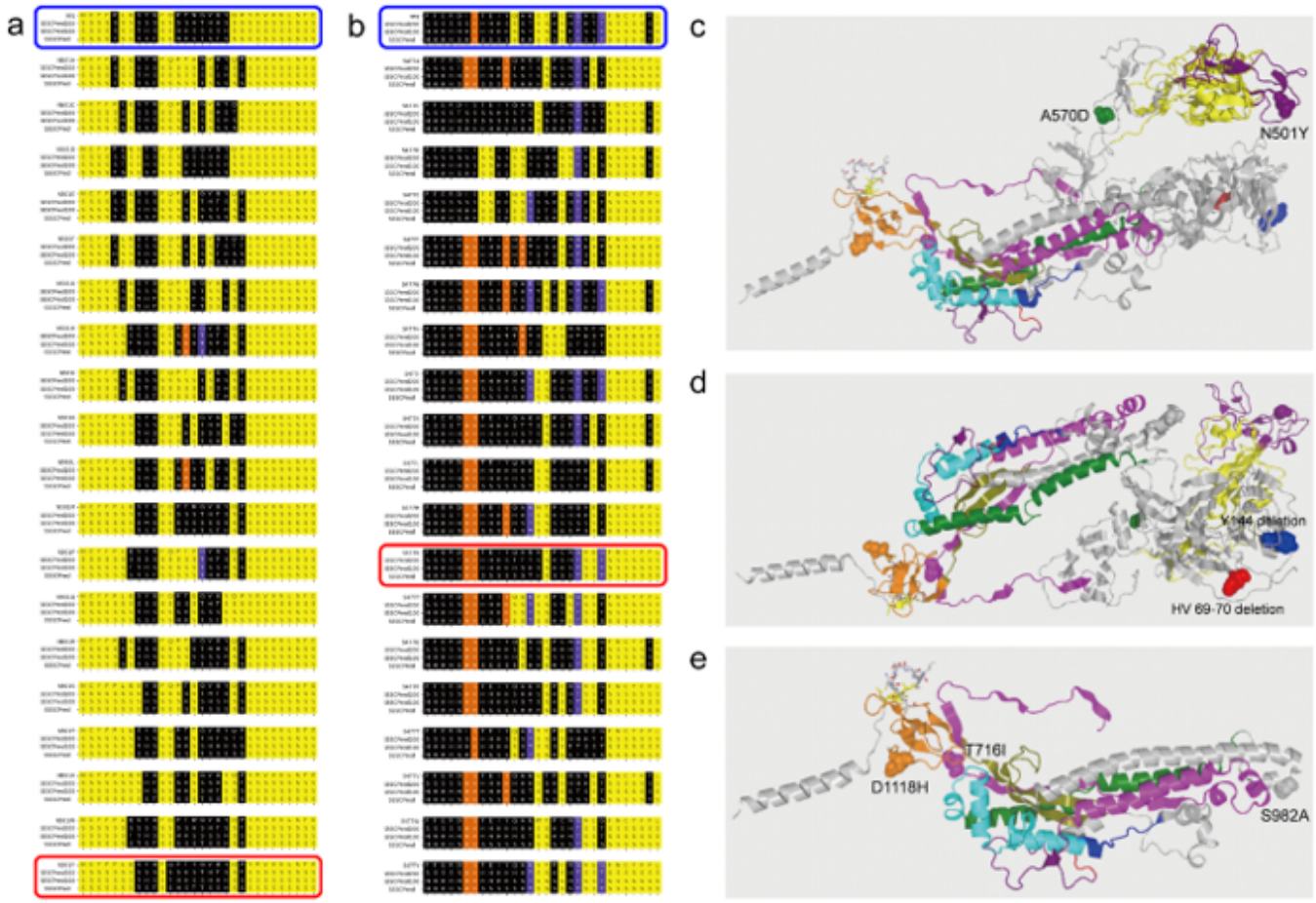

Figure 2

Conformational variability of N501Y and S477N mutation sites. a,b, Sequence flexibility/rigidity maps of all of the single amino acid mutations at the N501Y (a) and S477N (b) mutation sites (orange, identical α-helix-type conformations; yellow, identical β-sheet-type conformations; purple, identical other-type conformations; black, flexible conformations; blue frame, original sequences; red frame, mutation sequences). c-e, Cryo-EM structures of B.1.1.7 mutation sites (PDB ID code 6XR8). N501Y at the RBD and A570D at the SD1 (c), HV69-70 deletion and Y144 deletion at the NTD (d), and T716I, S982A, and D1118H at the S2 subunit (e).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SI.pdf