Emerging mutation patterns in SARS-CoV-2 variants

David A. Ostrov [a, *], Glenn W. Knox [b]

[a] Department of Pathology, Immunology and Laboratory Medicine, University of Florida College of Medicine, Gainesville, FL, USA
[b] Department of Surgery, University of Florida College of Medicine, Jacksonville, FL, USA

A R T I C L E   I N F O

A B S T R A C T

There is an urgent need to understand the functional effects of mutations in emerging variants of SARS-CoV-2. Variants of concern (alpha, beta, gamma and delta) acquired four patterns of spike glycoprotein mutations that enhance transmissibility and immune evasion: 1) mutations in the N-terminal domain (NTD), 2) mutations in the Receptor Binding Domain (RBD), 3) mutations at interchain contacts of the spike trimer, and 4) furin cleavage site mutations. Most distinguishing mutations among variants of concern are exhibited in the NTD, localized to sites of high structural flexibility. Emerging variants of interest such as mu, lambda and C.1.2 exhibit the same patterns of mutations as variants of concern. There is a strong likelihood that SARS-CoV-2 variants will continue to emerge with mutations in these defined patterns, thus providing a basis for the development of next line antiviral drugs and vaccine candidates.

© 2021 Published by Elsevier Inc.

1. Introduction

The predominating sequences of the SARS-CoV-2 spike glycoprotein have shifted significantly since 2019. A major global transition occurred after the D614G mutation emerged in February 2020. SARS-CoV-2 614G viruses have predominated worldwide since April 2020 [1]. This single mutation, D614G, was demonstrated to increase transmissibility by mechanisms unrelated to ACE2 binding. The SARS-CoV-2 spike protein forms a trimeric structure in which position 614 is located at a site of intermolecular contact between adjacent chains. The dramatic emergence of the D614G mutation worldwide (Ro approximately 3) suggests additional mutations of this type (at spike protein interchain contact sites) could lead to new variants of interest, variants of concern and, possibly, variants of high consequence.

The SARS-CoV-2 alpha variant (B.1.1.7), which spread rapidly in 2020, demonstrated fast replication and increased transmissibility (Ro 4–5) resulting from distinguishing mutations in the RBD, interchain contacts within the spike trimer, at the furin cleavage site separating the S1 and S2 subunits [2]. The alpha variant spike protein also exhibited distinguishing mutations in the NTD of the spike protein. Although the functional consequences of mutations in the NTD on fitness are not known, SARS-CoV-2 elicits neutralizing antibodies that bind multiple epitopes on the spike protein, including sites on the RBD (which binds ACE2 as a host cell receptor) and the NTD [3]. These data suggest that the NTD exhibits functional binding properties important for SARS-CoV-2 infection.

The SARS-CoV-2 delta variant, continuing to spread rapidly in 2021, shows enhanced transmissibility (Ro 5–8), presumably resulting from distinguishing mutations. We mapped distinguishing mutations in each variant of concern to understand how patterns of mutations are presented on the structure of the SARS-CoV-2 spike protein trimer. Defining patterns of emerging SARS-CoV-2 mutations may provide the basis for development of targeted antiviral drugs and new vaccine candidates.

## 2. Materials and methods

2.1. SARS-CoV-2 mutations in variants of concern and variants of interest

The Wuhan-Hu-1 sequence was used as a reference (GENBANK accession number MN908947). The Centers for Disease Control and Prevention definition of mutations that distinguish variants of concern and variants of interest were used (https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html).
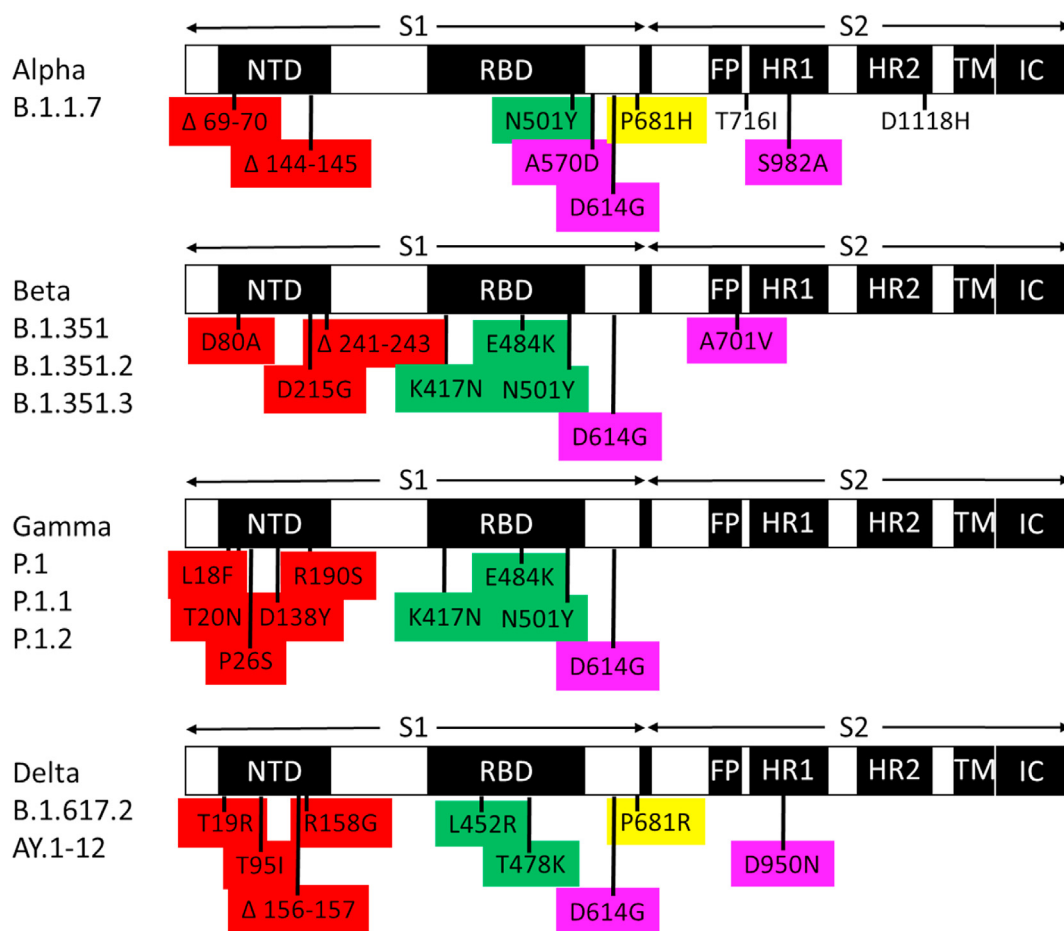
Fig. 1. SARS-CoV-2 variants of concern exhibit four patterns of mutations in the spike glycoprotein. NTD, N-terminal domain; RBD, receptor binding domain; FP, fusion peptide; HR1, heptad repeat 1; HR2, heptad repeat 2; TM, transmembrane anchor; IC, intracellular tail. Mutations that distinguish variants of concern are colored based on each pattern: red, mutations in the NTD likely to evade binding of neutralizing antibodies, green, RBD mutations likely to evade binding of neutralizing antibodies, yellow, fusion cleavage site mutations, magenta, interchain contact mutations. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

2.2. Mapping mutations on the structure of the SARS-CoV-2 spike glycoprotein trimer

Three cryoEM structures of the SARS-CoV-2 spike protein were used for mapping mutations; the RBD/ACE2-B0AT1 complex [4] (PDB 6M17), the prefusion SARS-CoV-2 spike glycoprotein with a single receptor-binding domain up [5] (PDB 6VSB), and P17—H014 Fab cocktail in complex with the trimeric SARS-CoV-2 spike protein (PDB 7CWN). SSM in COOT [6] was used to superimpose the spike protein based on Cα atoms. PyMOL (https://pymol.org/2/) was used to generate molecular graphic images.

## 2.3. Modeling interactions between CEA and the SARS-CoV-2 spike protein NTD

The crystal structure of mouse coronavirus receptor-binding domain [7] (chain B from PDB 3R4D) complexed with its murine receptor, CEA-related cell adhesion molecule 1, isoform 1/2S, was used to superimpose galectin-like domains in mouse hepatitis coronavirus and SARS-CoV-2 spike protein NTDs (chains A, B and C of the trimeric SARS-CoV-2 spike protein, PDB 7CWN). PyMol was used to generate molecular graphic images showing CEA as oriented by SSM in COOT (https://pymol.org/2/).

3. Results

3.1. The majority of distinguishing mutations in variants of concern alter the NTD of the SARS-CoV-2 spike protein

The spike protein mutations that distinguish variants of concern (alpha, beta, gamma, delta) were mapped on the trimeric spike protein structure (PDB 7CWN). The majority of distinguishing mutations alter the NTD of the spike protein (Figs. 1 and 2A). Data suggests that acquisition of amino acid substitutions and deletions in the NTD contributes to evasion of neutralizing antibody responses.

The NTD of SARS-CoV-2 exhibits a galectin-like fold, structurally similar to the NTD of Mouse Hepatitis coronavirus (MHV) [7]. During maturation, the spike protein of SARS-CoV-2 and MHV are cleaved into a receptor-binding subunit S1 and a membrane fusion subunit S2 that associate through noncovalent interactions. In MHV, the spike protein NTD serves as a virus receptor binding CEACAM1, and the MHV RBD binds ACE2, similar to SARS-CoV-2 RBD. SARS-CoV-2 mutations in variants of concern are located in the NTD at sites predicted to interact with CEACAM1.
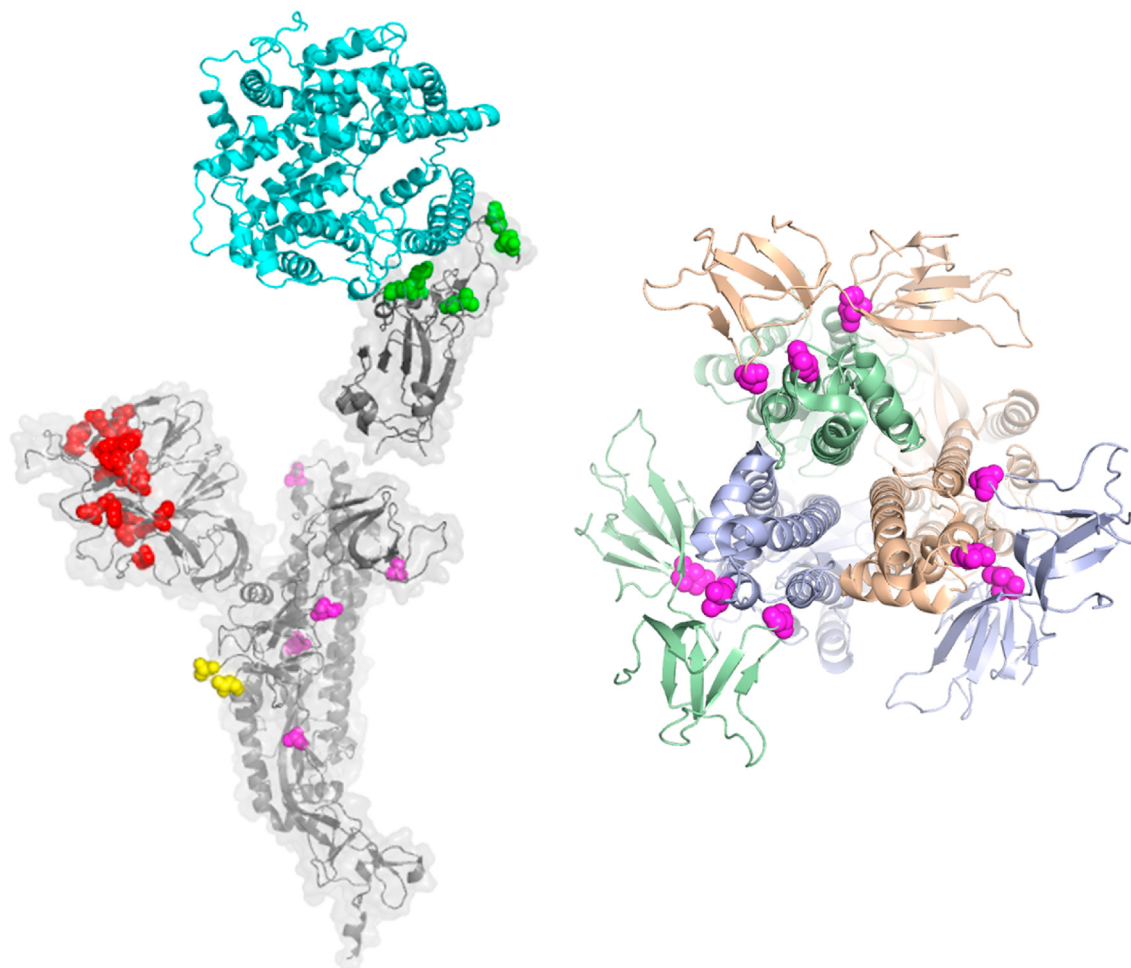
Fig. 2. The structure of mutation patterns in variants of concern. A. The structure of the SARS-CoV-2 spike protein (PDB 7CWN) is shown with spheres indicating mutations that distinguish variants of concern: red, NTD, green, RBD, yellow, furin cleavage site, magenta, interchain contact mutations. ACE2 is shown in blue, as modeled based on the interaction of ACE2 and the RBD in PDB 6M17. CEA is shown in orange, as modeled based on the interaction of Mouse Hepatitis Virus NTD complexed to murine CEA (PDB 3R4D). B. SARS-CoV-2 variants of concern exhibit distinguishing mutations at sites of intermolecular contact between adjacent chains of the spike trimer. Top view showing ribbon diagram of the trimeric spike protomer (PDB 6VSB) with the NTD and RBD omitted for clarity. Chains are colored beige, violet and green. Magenta spheres depict the location of mutations in variants of concern that alter intermolecular contacts between adjacent chains of the trimer. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

3.2. Mutations in the NTD accumulate at sites of high flexibility in the SARS-CoV-2 spike protein

Mutations that distinguish SARS-CoV-2 variants of concern have been acquired at sites in the NTD that exhibit high levels of structural flexibility (Supplemental Fig. 1). We compared the location of distinguishing mutations in variants of concern with thermal factor (B factor, indicating uncertainty of atom positions due to disorder) values from a cryo-EM structure of the SARS-CoV-2 trimer (PDB 7CWN). Variants of concern acquired mutations at positions with high B factors, > 120 Å^3. These data suggest that acquired mutations in the NTD may enhance overall virus fitness by modulating critical ligand binding properties (e.g., evasion of neutralizing antibody binding, enhancement of host cell CEACAM1 binding, host cell sialic acid binding [8]).

3.3. Mutations in the RBD alter antibody and/or ACE2 binding

RBD mutations that distinguish variants of concern have the potential to influence neutralizing antibody binding (such as solvent exposed residues K417 N, L452R, E484K, Figs. 1 and 2A). One

RBD mutation, N501Y, was shown to increase the affinity of the spike protein for the host receptor ACE2 [9]. These data suggest that advantageous mutations have been acquired in the spike protein S1 NTD and RBD for immune evasion, and possibly for enhanced receptor/co-receptor binding.

3.4. Transmissibility mutations occur at interchain contacts in the spike protein trimer

Strikingly, there are specific acquired mutations in the spike protein that distinguish variants of concern located at interfaces between subunits of the trimeric protomer [2]. As shown in Fig. 2B, A570D, D614G, A701V, D950 N, and S982A are located at interchain contact sites. The substitutions at spike trimer interfaces likely reduce intermolecular binding affinity. These acquired mutations likely destabilize the spike protein in a manner that enhances dynamic virus processes that include spike protein cleavage, structural rearrangement and host cell fusion mechanisms.
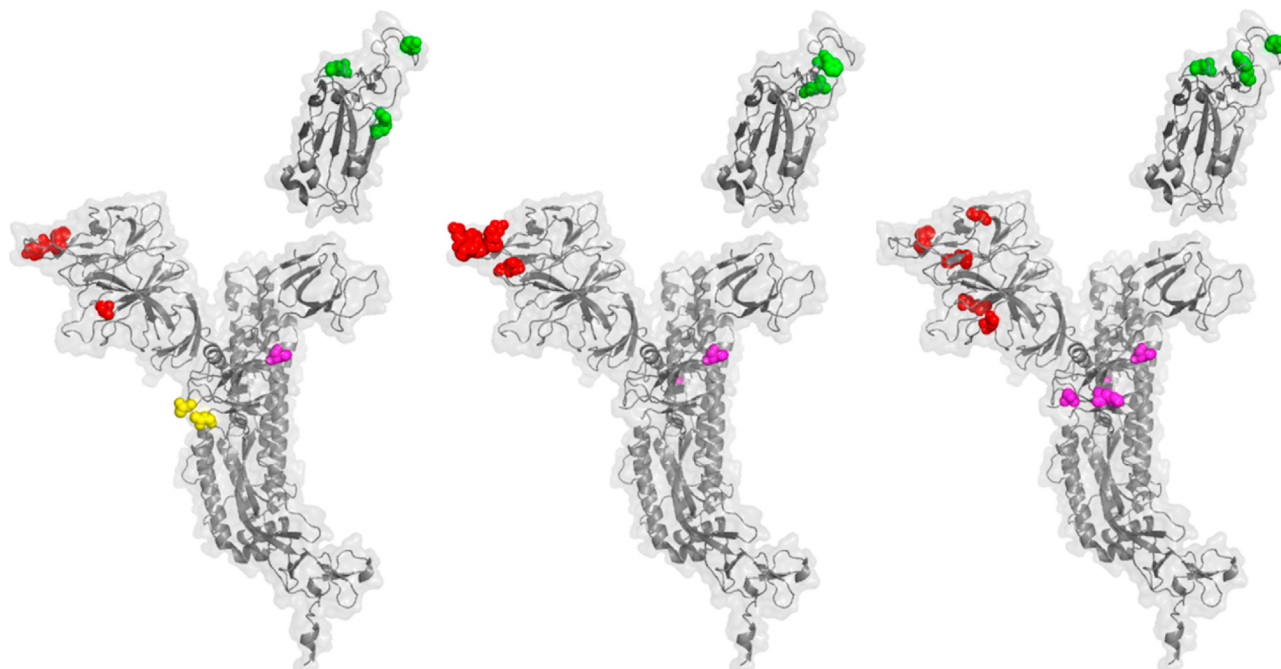
Fig. 3. Emerging variants of interest exhibit distinguishing mutations in the same patterns as variants of concern. A. A ribbon diagram and transparent surface are shown for the SARS-CoV-2 spike protein (PDB 7CWN) with spheres indicating sites of mutation that distinguish the Mu variant, B.1.621. Red spheres indicate mutations in the NTD; T95I, Y144T, Y144S. Green spheres indicate mutations in the RBD; E484K, N501Y. Magenta spheres indicate an interchain contact mutation; D614G. Yellow spheres indicate a mutation in the fusion cleavage site; P681H. B. The SARS-CoV-2 spike protein is shown with spheres indication the positions mutations that distinguish the lambda variant, C.37. Red spheres; G75V, T76I, 246–253 RSYLTPGD. Green spheres; L452Q, F490S. Magenta spheres indicate interchain contact mutations; D614G, T859 N. C. The SARS-CoV-2 spike protein is shown with spheres indication the positions mutations that distinguish the variant of interest C.1.2. Red spheres; C136F, R190S, D215G, Δ144, Δ242-243. Green spheres; Y449H, E484K, N501Y. Magenta spheres indicate interchain contact mutations; D614G, H655Y, N679K, T859 N. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)
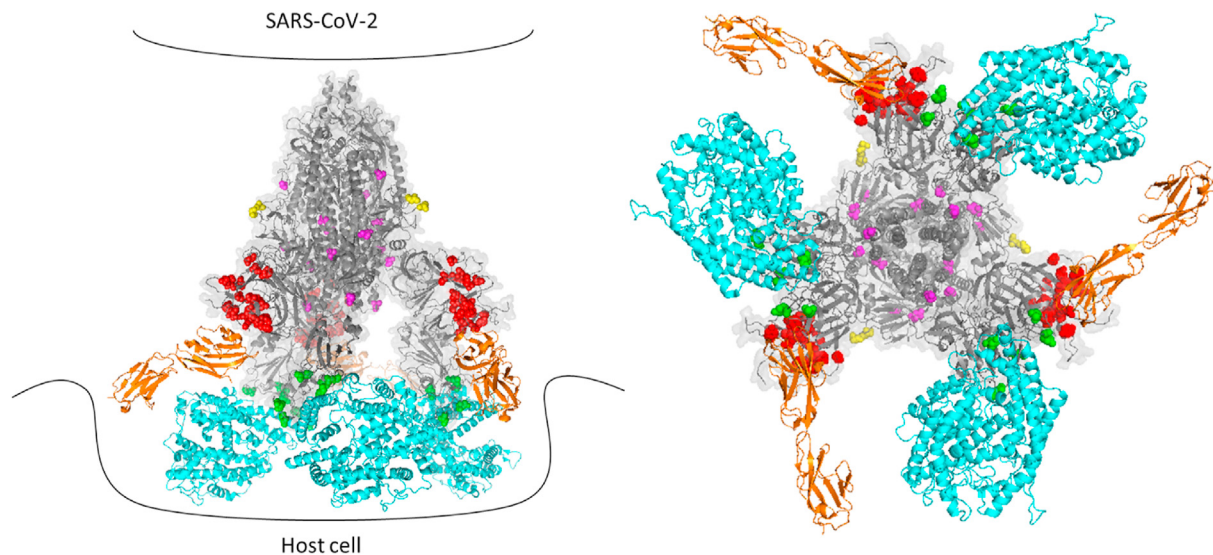
Fig. 4. Model of interactions between SARS-CoV-2 spike protein variants and host receptors. A. Side view cartoon of potential spike protein/host receptor interactions showing the structure of the trimeric SARS-CoV-2 spike protomer (PDB 7CWN), with spheres indicating mutations that distinguish variants of concern: red, NTD, green, RBD, yellow, furin cleavage site, magenta, interchain contact mutations. ACE2 is shown in blue, as modeled based on the interaction of ACE2 and the RBD in PDB 6M17. CEA is shown in orange, as modeled based on the interaction of Mouse Hepatitis Virus NTD complexed to murine CEA (PDB 3R4D). B. Top view of interactions between the trimeric spike protein protomer (gray) with ACE2 (blue) and CEA (orange).

3.5. Variants of concern with the highest levels of transmissibility exhibit mutations in the furin cleavage site

Mutations in position 681 of the spike protein distinguish the highly transmissible alpha and delta variants, but not less transmissible variants of concern, beta and gamma (Fig. 1). Position 681 is located adjacent to the RRAR proprotein convertase motif (furin cleavage site) considered a hallmark of high pathogenesis (**P**RRAR in Wuhan-Hu-1 [10], **H**RRAR in alpha, **R**RRAR in delta). Since endosomal S1/S2 cleavage occurs in an acidified

environment, a protonated histidine at position 681 of the alpha variant has the potential to influence the rate of spike protein cleavage and subsequent membrane fusion mechanisms to gain cell entry. Positively charged amino acids at position 681 in highly transmissible variants (H in alpha, R in delta) suggests an emerging pattern of concerning mutations.

3.6. Emerging variants exhibit mutations that represent the same patterns as variants of concern

To determine if emerging variants of interest exhibit mutations in the same patterns as variants of concern, we mapped mutations that distinguish mu (B.1.621) (https://www.medrxiv.org/content/10.1101/2021.05.08.21256619v1.full.pdf), lambda (C.37) ("Spike Variants: Lambda variant, aka B.1.1.1". covdb.stanford.edu. Stanford University Coronavirus Antiviral & Resistance Database. 1 July 2021.) and C.1.2. (doi: https://doi.org/10.1101/2021.08.20.21262342). As shown in Fig. 3, these variants of interest exhibit mutations in patterns common to variants of concern: 1) localized NTD mutations, 2) RBD mutations near the ACE2 binding interface, and 3) interchain contact mutations. Mu, unlike lambda and C.1.2, exhibits a furin cleavage site mutation at position 681: **H**RRAR. These data show that a subset of variants of interest (e.g., mu) exhibit the same patterns of mutations that distinguish highly transmissible variants of concern (alpha and delta). The presence of multiple mutations in these patterns may drive high levels of virus transmissibility.

4. Discussion

Although it is clear why mutations in the RBD have the potential to change virus fitness (by altering neutralizing antibody and host cell ACE2 binding), it is not clear why advantageous distinguishing mutations are emerging in the NTD.

Since SARS-CoV-2 elicits neutralizing antibodies that bind multiple epitopes on the spike protein, including the NTD [3], these data suggest that the NTD participates in an important but unknown function related to virus fitness (e.g., co-receptor binding). Based on similarity to the NTD of other coronaviruses, the SARS-CoV-2 may interact with a co-receptor such as CEACAM1 [7] (Fig. 4), as MHV, or carbohydrate, as for coronavirus Transmissible Gastroenteritis Virus (TGEV), in which the spike protein NTD binds host cell sialic acid [8]. An alternative possibility is that host cell restriction factors (e.g., Interferon-induced transmembrane proteins, IFITMs) block virus entry by binding the SARS-CoV-2 NTD unless mutations are acquired to evade restriction factor binding [11]. The identity of host cell ligands (co-receptor or restriction factor) for the NTD represents a significant unanswered question with potentially large impact on drug development and COVID treatment strategies.

Patterns of mutations are present in the most highly transmissible variants of concern, alpha and delta: 1) clustered mutations in the NTD, 2) mutations near the RBD/ACE2 interface (Figs. 2A), 3) interchain contact mutations in the spike trimer (Figs. 2B), 4) furin cleavage site mutations (Figs. 1 and 2A). Variants of interest exhibit these patterns, but differ in the number of mutations in each pattern. For example, lambda, C.37, exhibits distinguishing NTD, RBD and interchain contact mutations, but lacks cleavage site mutations. It seems reasonable to expect new mutations to be acquired in each pattern, with neutral, positive or negative effects on virus fitness.

In the United Kingdom, a sublineage of the delta variant, AY.4.2, is currently increasing in frequency. This variant exhibits distinguishing spike mutations Y145H and A222V. Position 145 is located in the NTD pattern of SARS-CoV-2 mutations, at a site

mutated in the alpha variant (deletion). If the number and biochemical characteristics of mutations in each pattern drives fitness, C.1.2 is a particularly concerning variant of interest. C.1.2 exhibits 5 distinguishing mutations in the NTD, 3 in the RBD, and 4 interchain contact mutations. An additional mutation in the furin cleavage site (e.g., P681 to H or R) of C.1.2 could generate a high fitness SARS-CoV-2 virus with mutations in more pattern positions than any other variant to date.

These patterns may be useful for artificial intelligence prediction of mutations with a strong likelihood of emergence, providing a basis for preparation of new vaccine strategies and antiviral drugs. New sequence data from COVID-19 patients can be used to refine the pattern definitions and clarify mechanisms of immune evasion and transmissibility advantage.

Since SARS-CoV-2 vaccine efficacy has drifted as mutations have been acquired in variants of concern, SARS-CoV-2 may exhibit a phenomenon similar to "influenza mismatch" caused by major and minor mutations of circulating viruses. As a result, the virus contained in the vaccine did not match the circulating strain, determining a reduction in the effectiveness of influenza vaccines. The recurring mutations of influenza strains prompted the introduction of a quadrivalent inactivated vaccine, the composition of which is determined on the basis of the most frequent strains isolated in the previous season during continuous surveillance.

Methods are established to rapidly develop a quadrivalent SARS-CoV-2 mRNA vaccine containing the sequences of all variants of concern (alpha, beta, gamma, delta). Based on rapid the spreading of alpha and delta in 2021, there could be significant advantages to updating SARS-CoV-2 sequences in vaccines every 6 months to keep up with changes in the virus, similar to common influenza vaccines.

As SARS-CoV-2 circulates, new mutations are likely to be observed in the described patterns associated with evasion of the immune system or virus transmissibility. Although existing variants of interest may transform into variants of concern, current methods permit straightforward vaccine updating, for example, with SARS-CoV-2 mRNA sequences of the spike protein from mu, lambda and C.1.2.

5. Conclusion

We found structural patterns of mutations that distinguish SARS-CoV-2 variants of concern. Most distinguishing mutations in variants of concern cluster to a pattern in the galectin-like NTD of the spike protein. A separate pattern of mutations in variants of concern is localized to the RBD/ACE2 interface, primarily for evasion of neutralizing antibodies. A striking pattern of mutations was identified at interchain contact sites within the spike protein trimer. The most highly transmissible variants of concern, alpha and delta, exhibit an additional pattern: mutations in the furin cleavage site. Variants of interest exhibit the same patterns of distinguishing mutations as variants of concern. Emerging mutations in these distinguishing patterns are expected to increase the fitness of SARS-CoV-2.

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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbrc.2021.11.059.

References

[1] B. Korber, W.M. Fischer, S. Gnanakaran, H. Yoon, J. Theiler, W. Abfalterer, et al., Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus, Cell 182 (4) (2020 Aug 20) 812–827, e19.

[2] D.A. Ostrov, Structural consequences of variation in SARS-CoV-2 B.1.1.7, J. Cell Immunol. 3 (2) (2021) 103–108.

[3] M. McCallum, A. De Marco, F.A. Lempp, M.A. Tortorici, D. Pinto, A.C. Walls, et al., N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2, Cell 184 (9) (2021 Apr 29) 2332–2347, e16.

[4] R. Yan, Y. Zhang, Y. Li, L. Xia, Y. Guo, Q. Zhou, Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2, Science 367 (6485) (2020 Mar 27) 1444–1448.

[5] D. Wrapp, N. Wang, K.S. Corbett, J.A. Goldsmith, C.L. Hsieh, O. Abiona, et al., Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation, Science 367 (6483) (2020 Mar 13) 1260–1263.

[6] P. Emsley, B. Lohkamp, W.G. Scott, K. Cowtan, Features and development of coot, Acta Crystallogr. D Biol. Crystallogr. 66 (Pt 4) (2010 Apr) 486–501.

[7] G. Peng, D. Sun, K.R. Rajashankar, Z. Qian, K.V. Holmes, F. Li, Crystal structure of mouse coronavirus receptor-binding domain complexed with its murine receptor, Proc. Natl. Acad. Sci. U. S. A. 108 (26) (2011 Jun 28) 10696–10701.

[8] C. Schwegmann-Wessels, G. Herrler, Identification of sugar residues involved in the binding of TGEV to porcine brush border membranes, Methods Mol. Biol. 454 (2008) 319–329.

[9] T.N. Starr, A.J. Greaney, S.K. Hilton, D. Ellis, K.H.D. Crawford, A.S. Dingens, et al., Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding, Cell 182 (5) (2020 Sep 3) 1295–1310, e20.

[10] J. Shang, G. Ye, K. Shi, Y. Wan, C. Luo, H. Aihara, et al., Structural basis of receptor recognition by SARS-CoV-2, Nature 581 (7807) (2020 May) 221–224.

[11] G. Shi, A.D. Kenney, E. Kudryashova, A. Zani, L. Zhang, K.K. Lai, et al., Opposing activities of IFITM proteins in SARS-CoV-2 infection, EMBO J. 40 (3) (2021 Feb 1), e106501.