Studying folding kinetics of omicron to understand its hijack mechanism in human host cells

Saranya S
Bharathidasan University
Sangavai C
Bharathidasan University
Roja B
Bharathidasan University
Chellapandi P (pchellapandi@gmail.com)
Bharathidasan University

Research Article

Keywords: SARS-CoV-2, COVID-19, Folding rate, Virus maturation, Hijack mechanism, Mutations

Posted Date: December 22nd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2392479/v1

Abstract

Coronavirus disease (COVID-19) has rapidly expanded into a global pandemic caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Genetic drift in global SARS-CoV-2 isolates and protein evolution have an impact on their ability to escape from current antiviral therapeutics. Hence, our study aimed to reveal how mutations in the folding kinetics of assembly and maturation proteins drive the hijack ability to emerge SARS-CoV-2 variants in humans. In this study, we predicted the folding rate of these proteins using multiple regression analysis and validated the prediction accuracy using machine learning algorithms. Hybrid machine learning using linear regression, random forest, and decision tree was used to evaluate the predicted folding rates compared with other machine learning models. In SARS-CoV-2 variants, the sequence-structure-function-folding rate link stabilizes or retains the mutated residues, making stable near-native protein structures. The folding rates of these protein mutants were increased in their structural classes, particularly β-sheets, which accommodated the hijacking ability of new variants in human host cells. E484A and L432R were identified as potent mutations that resulted in drastic changes in the folding pattern of the spike protein. We conclude that receptor-binding specificity, infectivity, multiplication rate, and hijacking ability are directly associated with an increase in the folding rate of their protein mutants.

1. Introduction

Coronavirus disease (COVID-19) is a deadly infectious disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The spread rate and infectivity of this virus are continuously increasing owing to emerging variants, even though several vaccines have been prescribed globally. A variant is a viral genome that may contain one or more mutations and was designated by Public Health Organizations as a variant of concern or a variant of interest (Kandeel et al. 2022). Omicron (B.1.1.529) is currently a circulating variant of concern, which currently has the BA.1, BA.1.1, and BA.2 lineages. It is not yet clear whether omicron is more transmissible and infectious than other variants. According to the World Health Organization, there may be an increased risk of reinfection with omicron compared with other variants.

The SARS-CoV-2 genome assembly consists of 11 open reading frames (ORFs) with frequent mutations. The spike glycoprotein (S) facilitates attachment to host cell membrane receptors. Human TMPRSS2 has also been used for viral entry by priming human lung cells. The envelope protein (E) and membrane protein (M), which act as viroporins, are responsible for morphogenesis and assembly. The nucleocapsid protein (N) plays a role in viral entry and enhances the efficiency of the subgenomic viral RNA transcription/replication machinery. Nonstructural proteins (NSP1-16) are essential for viral replication, nucleic acid metabolism, and capsid formation (Chellapandi and Saranya, 2021). Viruses can cause many beneficial mutations that facilitate their ability to escape from antivirals and vaccines. Hence, studies of the mutational impacts on spike, assembly, and morphogenesis proteins are important to understand the genetic drifts across SARS-CoV-2 variants.

Protein folding information plays a key role in the therapeutic intervention for many viral diseases (Broglia et al. 2005). Proper folding is crucial to achieve a unique native functional structure. It also depends on the physicochemical characteristics of the residues in amino acid sequences and low Gibbs free energy (Hebert and Molinari, 2007). The amino acid swap in a protein significantly alters its folding, stability, specificity, and function (Gromiha et al. 2009; Pilipczuk et al. 2017). The deleterious mutation imposes a fitness penalty depending on the reduced folding rate of the protein (Starr et al. 2020). Protein evolution is a molecular drive that enhances the adaptation and survival of pathogens to resist new environmental niches (Seal et al. 2021). Folding landscape and structural integrity are key factors in protein evolution (Morcos et al. 2014). Information on residue-level co-evolution upon mutations is used to predict changes in protein folding rates (Mallik et al. 2016). The relationship between the sequence-structure-function-conformational dynamic space-folding rate has been reported as a molecular determinant of some virulence proteins in bacterial pathogens (Chellapandi 2014; Chellapandi et al. 2013; 2019;2018; Prathiviraj et al. 2015; Priscilla et al. 2016, 2017; Prathiviraj and Chellapandi, 2020a; 2020b).

The proper folding and assembly of proteins are critical for the infection, multiplication, and hijacking ability of SARS-CoV-2 in human hosts (Chellapandi and Saranya, 2021). Many theoretical models and experimental methods have been developed to predict and validate mutations in protein-folding rates (Raynes et al. 2014). In this study, we predicted and validated the folding rates of assembly and maturation proteins using multiple regression analysis followed by machine learning algorithms. The present study aimed to investigate how the folding kinetics of these proteins is influenced by SARS-CoV-2 variants compared with Wuhan-Hu-1. We also propose the hijack mechanism of a recently emerged variant called omicron, based on the predicted folding kinetics of these protein mutants.

2. Materials And Methods

Dataset

The mutational frequencies of the SARS-CoV-2 proteins were obtained from previous mutant libraries (Forster et al. 2020; Joshi et al. 2020; Mercatelli and Giorgi, 2020). A total of 2937 mutants were chosen from the spike, assembly, and morphogenesis proteins of SARS-CoV-2 genomic variants. Wuhan-Hu-1 proteins were used as a reference (wild-type) for this study. The mutated amino acid residue was introduced to replace the respective residues in the Wuhan-Hu-1 protein. For double mutations, two residues from the wild-type protein were replaced with mutated residues.

Prediction of folding kinetics

The folding rates of protein mutants were calculated using a FOLD-RATE server-based set of 49 amino acid properties (Pave) as follows (Gromiha et al. 2006).

$$P\left(ave\right)\left(i\right)=\sum_{j=1}^{N}P\left(j\right)/N$$

P(i) is a set of 49 amino acid property values for the jth residue in a protein (N). The predicted folding rate of each structural class (α, β, α + β, and α + β/α/β) was correlated with the experimental folding rate lnkf (i) using multiple regression. The folding kinetic orders of these proteins were computed from their 3D structures using the K-FOLD server (Capriotti and Casadio, 2007). The predicted folding rates of the protein mutants were compared to those with the respective wild-type proteins. Furthermore, discrimination analysis was used to evaluate the influence of mutations on the folding rate of SARS-CoV-2 proteins.

Validation of predicted data

Machine-learning models were captured using the sci-kit-learn machine-learning library in the Jupyter notebook. The performances of the resulting models were evaluated by comparing the trained dataset with the predicted results. Machine learning models were trained using the folding rates of the structural classes (α, β, α + β, and α + β/α/β) with 20 rows. These models were used to test 1202 rows of data, including the folding rates of all the structural classes predicted above. Accuracy, recall, and precision were used as metrics to evaluate this process (Dhasaradhan et al. 2021).

3. Results

As shown in Fig. 1A, our dataset contains 2937 protein mutations in the SARS-CoV-2 genome, the order of which was noted to be S > N > NSP4 > NSP6 > M > NSP8 > E > NSP7. D614G, S68F, T175M, RG203KR, F308Y, L37F, S25L, and M129I are the most prevalent mutations in S, E, M, N, NSP4, NSP6, NSP7, and NSP8 proteins of SARS-CoV-2 variants, respectively. The folding rate is a determining factor for protein multiplication, virulence, host specificity, and infectivity of SARS-CoV-2. Herein, the protein folding kinetics of SARS-CoV-2 variants involved in assembly and morphogenesis have been extensively studied (**Supplementary file**). The predicted fold rates ranged from to 20–93 molecules/sec. We scanned the machine-learning algorithms to calculate the accuracy of the predicted results (Fig. 1B). This shows that combined algorithms such as linear regression, random forest, and decision tree are required for the validation of the mutant library compared to other algorithms. These algorithms provided 100% accuracy for the predicted folding rate of proteins, suggesting that our computational predictions are more robust and reliable for driving a molecular hypothesis. Some rare mutational events, such as DF1041GL, A36V, T327N, I266L, and F15L, showed fast folding rates of S, E, NSP4, NSP6, and NSP8, respectively (Fig. 1C).

The influence of the mutational spectrum on the folding rates of different structural classes of proteins was analyzed, as shown in Fig. 2. Compared to the wild-type, fast-folding rates were predicted in E, M, NSP6, and NSP7 mutants, whereas slow folding rates were observed in S and NSP8 mutants. Mutation events are highly influenced by the β-class of many mutated proteins by increasing their folding rate. Interestingly, point mutations did not drastically change the folding rate of N but showed different folding patterns in structural classes in NSP7. The folding rate increased only in NSP7 mutants because of drastic changes in the α and β classes. Mutational events in amino acid residues involved in the β-class increase the folding rates of E, M, NSP4, and NSP6 mutants. No significant changes were observed in the

folding rates of the mixed and unknown structural classes of the assembly and morphogenesis proteins. Therefore, amino acid swapping can alter the folding kinetics of SARS-CoV-2 proteins during infection.

Generally, the folding rate of most viral proteins is increased by point mutations for accommodation in the host and escape from therapeutics. We predicted fast-folding rates (34 mol/sec) in ξ and γ variants, which can increase the multiplication rate of the virus (Fig. 3A). The folding rate of the α-class of proteins is increased in the α and ξ variants, whereas the folding rate of the β-class of proteins is increased in γ and κ variants. The folding rates of mixed structural classes decreased considerably in SARS-CoV-2 variants compared to the wild type. The Omicron, α, and λ variants show a slow folding rate in their proteins. This indicates low multiplication, assembly, and morphogenesis of virions in the host cells. The slow folding rate in omicron resulted from radical changes in unknown structural classes.

The folding rate is commonly increased in the most frequent mutations (E484K and L452R) among many variants. The folding rates of omicron (E484A) and τ (D253G) variants were higher than those of Wuhan-Hu-1 proteins (20.02 mol/sec) (Fig. 3B). The predicted folding rates of these variants were 35–39 mol/sec. The fast folding rate of the α-class was observed in γ (P681H), β-class in τ (D253G), mixed class in omicron (E484A), and unknown class in λ (N452Q) variants compared to the wild-type. It is important to note that the folding rate of omicron is significantly raised to 39.34 mol/sec when glutamate is mutated with alanine at the position of 484.

MukF_M, CaM_bind, TRPM_tetra, DUF3418, and MH1 were identified as common motifs in the spike glycoprotein of the most potent variants, of which four motifs were missing inδ, omicron and μ variants(Fig. 3C). MukF_M, CaM_bind, and TRPM_tetra motifs were found in Wuhan-Hu-1 cells. The DUF3418 and MH1 motifs were also found in αand γ variants, respectively. This suggests that the presence of CaM_bind in δ and MukF_M in omicron could determine their target specificity and infectivity in host cells.

The mass of the SARS-CoV-2 virion was 1fg with ~ 1000 nm and a volume of 107. This virion is enveloped by fully assembled mature proteins during the folding process (Fig. 4). Together with 2000 copies of NSP4 and 20 copies of NSP6, maturation, and envelop proteins, SARS-CoV-2 forms a viral coat during the infection and incubation processes. This protein assembly initiates the morphogenesis of 100 copies of spike glycoprotein within 10 min. The results of our study suggest that the folding kinetics of these proteins in newly emerging SARS-CoV-2 variants, particularly omicrons, could contribute to the replication, assembly, morphogenesis, and infection (attachment) of host cells. To execute this molecular process, SARS-CoV-2 can increase its genome plasticity and mutability in spike, assembly, and morphogenesis genes to conquer evolutionary fitness in the new environmental niche.

4. Discussion

Protein thermodynamic fitness accounts for mutational effects in RNA viruses, leading to the development of antiviral drug resistance and vaccine failure (Miller et al. 2014; Huchting, 2020). Protein folding is a change in the energetic coupling of amino acid residues to attain the native protein structure

of SARS-CoV-2 (Prithiviraj et al. 2021). Hence, our study indicates the influence of mutational events on the folding rates of mutated proteins in SARS-CoV-2 variants. The order of mutation frequency was S > N > NSP4 > NSP6 > M > NSP8 > E > NSP7, similar to previous studies (Forster et al. 2020; Mercatelli and Giorgi, 2020; Joshi et al. 2021). D614G, S68F, T175M, RG203KR, F308Y, L37F, S25L, and M129I were the most prevalent mutations, which may be associated with mild and severe clinical outcomes. This may be due to drastic changes in the folding dynamics and conformational stability of mutated protein structures while SARS-CoV-2 infected human lung cells (Ng et al. 2020; Nagy et al. 2020; Prithiviraj et al. 2021; Salpini et al. 2021). Additionally, we predicted fast-folding rates (20–93 molecules/sec) in DF1041GL, A36V, T327N, I266L, and F15L mutations. These mutations may have contributed to the conformational stability and folding kinetics of their structural classes.

Frequent changes and increased mutational rates of S, N, ORF8, and ORF1ab proteins are highly correlated with enhanced virulence and the spread of viral infection. However, infection rates are associated with the stabilization of hydrophobic interactions and globular conformation of these proteins. Mutational changes tend to destabilize folding by decreasing thermodynamic folding stability, similar to earlier studies (Miller et al. 2014; Prathiviraj et al. 2021; Baek et al. 2022). The folding rates of S and NSP8 decrease upon point mutation, which could influence viral spread and lethality (Cacore and Rabitz, 2020). Fast-folding rates were predicted in E, M, NSP6, and NSP7 mutants, suggesting accelerated assembly and morphogenesis processes to form a virion. In some mutational events, the folding rate was conferred to show the low stability of their protein structures, as described earlier for NSP6 (Benvenuto et al. 2020).

Functional innovation is more likely to evolve in proteinfolds that tolerate mutations (Tóth-Petróczy and Tawfik, 2014). Our analysis showed that mutational events brought about radical changes in the folding rates of unknown structural classes in many SARS-CoV-2 variants. Such changes in NSP7 and NSP8 might have influenced the stabilization of replication/transcription of the complexity, infectivity, and virulence of SARS-CoV-2, as described in previous studies (Rehman et al. 2020; Reshamwala et al. 2020). Core mutations drastically change the folding rates of β-sheets in assembly and morphogenesis proteins, but there are no significant changes in other structural classes. Chemically conserved residue changes in β-sheets, the establishment of β-strand conformation of the polypeptide chain, and contact orders are possible reasons for the changes in the folding dynamics of β-sheets (Lorch et al. 1999; Abrusán and Marsh, 2016). The contact order in α-helices is higher than that in β-sheets; therefore, α-helices can accumulate more mutations than strands without changing folding rates (Abrusán and Marsh, 2016).

Omicron, and δ have recently emerged as variants of concern that have accumulated high numbers of mutations and their ability to evade pre-existing immunity acquired through vaccination. These variants acquire hijacking ability by reducing the ability of antibodies to compete with receptor-binding domain (RBD)-receptor interaction and virus neutralization (Khandia et al. 2022). The results of our study indicated the occurrence of slow folding rates in the omicron, α, and λ variants upon mutation. This could favor the low multiplication, assembly, and morphogenesis of their virions in the host cells by amino acids adjacent to the predicted mutations. Omicron exhibits a mutation at E484A, which was observed to

increase the folding rate in RBI. The mutated position may be associated with reduced antibody recognition and antibody neutralization (Shah and Woo, 2022; Vogt et al. 2022). The D253G, N452Q, and E484A mutations in the RBD of spike proteins have been found to play a critical role in the stability of the RBD-ACE2 complex in human cells, similar to an earlier investigation (Nguyen et al. 2022). The omicron variant exhibited a polybasic cleavage site (P681H) alteration that determined its infectivity rate to be less than that of the variant. However, its transmissibility was increased by adding positive charges (other than P681H) to the region, accompanied by an increase in S1/S2 cleavage early in the viral life cycle (Saxena et al. 2022).

In this study, L452R was identified as a potent omicron mutation. It is associated with immune escape owing to its immediate proximity to the ACE2 interaction interface of the RBD. It is likely to increase viral transmissibility and infectivity and rescue fusogenicity (Tchesnokova et al. 2021). Hence, our study implied that the presence of E484A and L452R mutations could greatly enhance the ability of omicron to infect lung cells of ACE2 in humans.Protein functions often require the sequence conservation of short (6–12 nts.) motifs, which may correspond to binding sites. In this study, the MukF_M motif was only present in the spike protein of omicron and was devoid of other CaM_bind and TRPM_tetra motifs, similar to Wuhan-Hu-1. The loss of these motifs in omicron can lead to structural rearrangements to attain a near-native folding pattern that enhances host target specificity and infectivity. Our findings suggest that the folding rates of understudied proteins in newly emerging SARS-CoV-2 variants, particularly omicron, contribute to the viral life cycle in human cells. Our study also revealed that genome plasticity and mutability in spike, assembly, and morphogenesis genes in response to protein folding kinetics increased to conquer evolutionary fitness in the new environmental niche.

5. Conclusion

The thermodynamic stability of proteins accounts for a large fraction of the observed mutational effects. Our study shows that destabilizing mutations due to strong immunologic pressure in humans can inflict a fitness penalty (deleterious to beneficial) proportional to the reduction in the folded proteins of SARS-CoV-2 variants. It also revealed that the change in folding kinetics plays a vital role in determining the sequence-structure-function-virulence relationship of the SARS-CoV-2 proteome. The fast folding rates of their mutated spike, assembly and morphogenesis proteins could allow the virus to escape the responses of antivirals, vaccines, and host immunity. Consequently, SARS-CoV-2 variants are capable of accommodating increased virulence, infectivity, transmissibility, and evading immune responses to the host environment by modulating their protein folding rates in structural classes, particularly β-sheets. Studying viral genome variants and associated protein mutations can allow us to track evolution and spread, leading to the development of appropriate intervention strategies. The present study suggests that the development of protein-folding interdiction drugs based on the folding rates of spike, assembly, and morphogenesis proteins would likely be possible for COVID-19 caused by newly emerging SARS-CoV-2 variants.

Declarations

Acknowledgments

The authors would like to thank the Science and Engineering Research Board, Department of Science and Technology (MSC/2020/000438), New Delhi, India, for financial support.

Conflict of interest

The authors confirm that this article has no conflicts of interest.

References

1. Abrusan, G., Marsh, J.A., 2016. Alpha Helices Are More Robust to Mutations than Beta Strands. PLoSComput. Biol. 9, 12
2. Baek, K.T., Mehra, R., Kepp, K.P., 2022. Stability and expression of SARS-CoV-2 spike-protein mutations.BioRxiv. 03.21.485157
3. Benvenuto, D., Giovanetti, M., Ciccozzi, A., Spoto, S,,Angeletti, S., Ciccozzi, M., 2020. The 2019-new coronavirus epidemic: Evidence for virus evolution. J. Med. Virol. 92, 455-459.
4. Broglia, R. A., Tiana, G., Sutto, L., Provasi, D., Simona, F., 2005. Design of HIV-1-PR Inhibitors that do not Create Resistance: Blocking the Folding of Single Monomers. Prot. Sci. 14, 2668–2681.
5. Capriotti, E., Casadio, R., 2007. K-Fold: a tool for the prediction of the protein folding kinetic order and rate. Bioinformatics.23, 385-6.
6. Chellapandi, P, Sakthi, S.S, Bharathi, M., 2013.Phylogenetic approach for inferring the origin and functional evolution of bacterial ADP-ribosylation superfamily. Protein and Peptide Letters 20, 1054-1065.
7. Chellapandi, P., 2014. Structural-functional integrity of hypothetical proteins identical to ADP-ribosylation Superfamily upon point mutations.Protein and Peptide Letters. 21, 722-735.
8. Chellapandi, P., Prathiviraj, R., Prisilla, A., 2019. Deciphering structure, function and mechanism of Plasmodium IspD homologs from their evolutionary imprints. Journal of Computer-Aided Molecular Design 33, 419-436.
9. Chellapandi, P., Prathiviraj, R., Sangavai, C., Bharathi, M., 2018.*Methanobacteriumformicicum* as a target rumen methanogen for the development of new methane mitigation interventions. Veterinary and Animal Science 6, 86-94.
10. Chellapandi, P., Saranya, S., 2021.A perspective on cordycepin and its nucleoside analogs for systemic COVID-19 infection.Coronaviruses. e100921196360
11. Dhasaradhan, K., Jaichandran, R., Shunmuganathan, K.L., UshaKiruthika, S., Rajaprakash, S., 2021. Hybrid Machine Learning Model Using Decision Tree and Support Vector Machine for Diabetes Identification. In: Bhateja, V., Satapathy, S.C., Travieso-González, C.M., Aradhya, V.N.M. (eds) Data

Engineering and Intelligent Computing. Advances in Intelligent Systems and Computing, vol 1407.Springer, Singapore. https://doi.org/10.1007/978-981-16-0171-2_28.

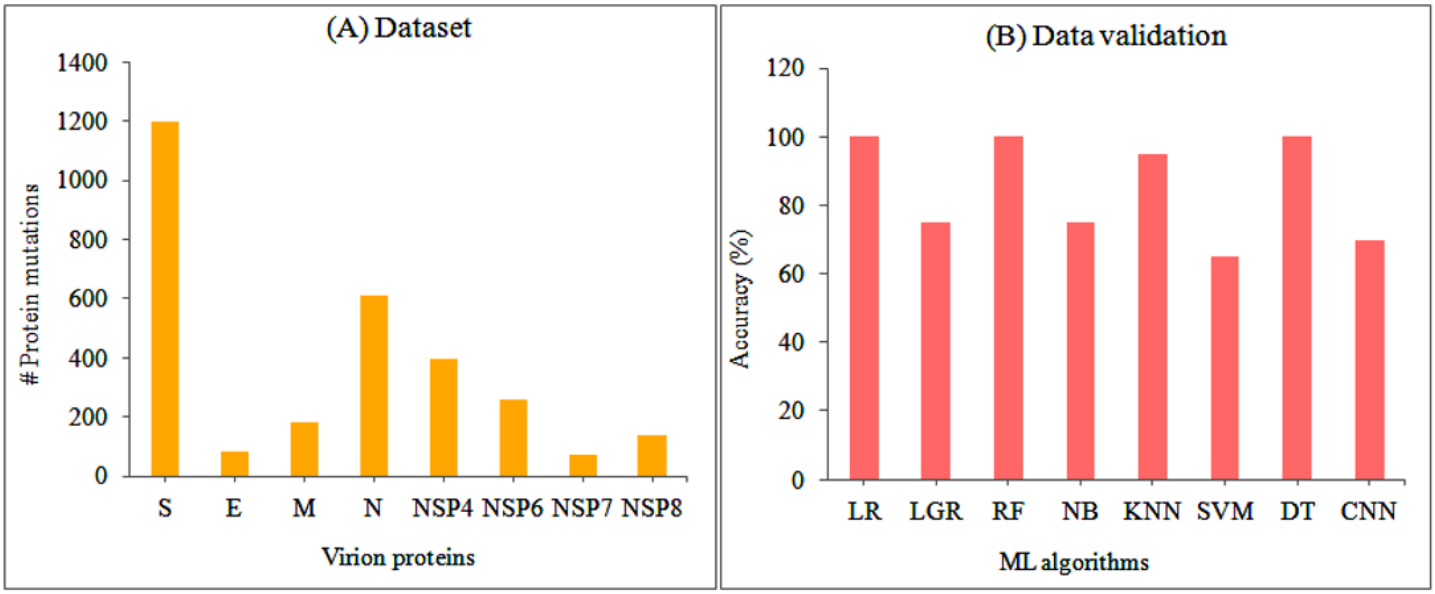
12. Forster, P.M., Forster, H.I., Evans, M.J., Gidden, M.J., Jones, C.D., Keller, C.A., Lamboll, R.D., Le Quéré, C., Rogelj, J., Rosen, D., Schleussner, C-F., Richardson, T.B., Smith, C.J., Turnock , S.T., 2020. Current and future global climate impacts resulting from COVID-19. Nat. Clim. Chang. 10, 913–919.
13. Gromiha, M.M., Yabuki, Y., Suresh, M.X., Thangakani, A.M., Suwa, M., Fukui, K., 2009. TMFunction: database for functional residues in membrane proteins. Nucleic Acids Res.**37**, 201-D204.
14. Hebert, D.N., Molinari, M., 2007. In and out of the ER: protein folding, quality control, degradation, and related human diseases. Physiol. Rev. 87, 1377-1408.
15. Huchting, J., 2020. Targeting viral genome synthesis as broad-spectrum approach against RNA virus infections.Antivir. Chem. Chemother. 28, 2040206620976786.
16. Joshi, M., Puvar, A., Kumar, D., Ansari, A., Pandya, M., Raval, J., Patel, Z., Trivedi, P., Gandhi, M., Pandya, L., Patel, K., Savaliya, N., Bagatharia, S., Kumar, S., Joshi, C., 2021. Genomic Variations in SARS-CoV-2 Genomes From Gujarat: Underlying Role of Variants in Disease Epidemiology. Front. Genet.12, 586569.
17. Kandeel, M., Mohamed, M.E.M., Abd El-Lateef, H.M., Venugopala, K.N., El-Beltagi, H.S., 2022.Omicron variant genome evolution and phylogenetics. J. Med. Virol. 94, 1627-1632.
18. Khandia, R., Singhal, S., Alqahtani, T., Kamal, M.A., El-Shall, N.A., Nainu, F., Desingu, P.A., Dhama, K., 2022. Emergence of SARS-CoV-2 Omicron (B.1.1.529) variant, salient features, high global health concerns and strategies to counter it amid ongoing COVID-19 pandemic. Environ. Res. 209, 112816.
19. Lorch, M., Mason, J.M., Clarke, A.R., Parker, M.J., 1999. Effects of core mutations on the folding of a beta-sheet protein: implications for backbone organization in the I-state. Biochemistry. 38, 1377-85.
20. Mallik, S., Das, S., Kundu, S., 2016. Predicting protein folding rate change upon point mutation using residue-level coevolutionary information.Proteins. 84, 3-8.
21. Mercatelli, D., Giorgi, F.M., 2020. Geographic and Genomic Distribution of SARS-CoV-2 Mutations. Front. Microbiol.11, 1800.
22. Miller, C.R., Lee, K.H., Wichman, H.A., Ytreberg, F.M., 2014. Changing folding and binding stability in a viral coat protein: a comparison between substitutions accessible through mutation and those fixed by natural selection.PLoS One.9, 112988.
23. Morcos, F., Schafer, N.P., Cheng, R.R., Onuchic, J.N., Wolynes, P.G., 2014.Coevolutionary information, protein folding landscapes, and the thermodynamics of natural selection. Proc. Natl. Acad. Sci. USA. 111, 12408-12413.
24. Nagy, C., Maitra, M., Tanti, A., Suderman, M., Théroux, J.F., Davoli, M.A., Perlman, K., Yerko, V., Wang, Y.C., Tripathy, S.J., Pavlidis, P., Mechawar, N., Ragoussis, J., Turecki, G., 2020. Single-nucleus transcriptomics of the prefrontal cortex in major depressive disorder implicates oligodendrocyte precursor cells and excitatory neurons. Nat. Neurosci. 23, 771-781.
25. Nguyen, T.T., Pathirana, P.N., Nguyen, T., Nguyen, O.V.H., Bhatti, A., Nguyen, D.C., Nguyen, D.T., Nguyen, N.D., Creighton, D., Abdelrazek, M., 2021.Genomic mutations and changes in protein secondary

structure and solvent accessibility of SARS-CoV-2 (COVID-19 virus).Sci. Rep. 11, 3487.

26. Pilipczuk, J., Piatek Z.B., Bruzdziak, P., Czub, J., Wieczor, M., Olszewski, M., Wanarska, M., Nowicki, B., Nowacka, A.D., Piatek, R., 2017.Role of the disulfide bond in stabilizing and folding of the fimbrial protein DraE from uropathogenic Escherichia coli.J. Biol. Chem. 292, 16136-16149.
27. Prathiviraj, R., Chellapandi, P., 2020a. Evolutionary genetic analysis of unassigned peptidase clan-associated microbial virulence and pathogenesis.Biologia (Poland) 75, 2083–2092.
28. Prathiviraj, R., Chellapandi, P., 2020b. Comparative genomic analysis reveals starvation survival systems in *Methanothermobacter thermautotrophicus* ΔH. Anaerobe64, 102216.
29. Prathiviraj, R., Prisilla, A., Chellapandi, P., 2015. Structure-function discrepancy in *Clostridium botulinum* C3 toxin for its rational prioritization as a subunit vaccine.Journal of Biomolecular Structure and Dynamics34, 1317-1329.
30. Prathiviraj, R., Saranya, S., Bharathi, M., Chellapandi, P., 2021. A hijack mechanism of Indian SARS-CoV-2 isolates for relapsing contemporary antiviral therapeutics. Computers in Biology and Medicine 132, 104315.
31. Prisilla, A., Prathiviraj, R., Sasikala, R., Chellapandi, P., 2016. Structural constraints-based evaluation of immunogenic avirulent toxins from *Clostridium botulinum* C2 and C3 toxins as subunit vaccines.Infection, Genetics and Evolution 44, 17-27.
32. Prisilla, A., Prathiviraj, R., Chellapandi, P., 2017. Molecular evolutionary constraints that determine the avirulence state of *Clostridium botulinum* C2 toxin. Journal of Molecular Evolution 86, 174-186.
33. Raynes, Y., Sniegowski, P.D., 2014. Experimental evolution and the dynamics of genomic mutation rate modifiers. Heredity. 113, 375-380.
34. Rehman, S., Mahmood, T., Aziz, E., Batool, R., 2020. Identification of novel mutations in SARS-COV-2 isolates from Turkey. Arch. Virol. 165, 2937-2944.
35. Reshamwala, S.M.S., Likhite, V., Degani, M.S., Deb, S.S., Noronha, S.B., 2021. Mutations in SARS-CoV-2 nsp7 and nsp8 proteins and their predicted impact on replication/transcription complex structure. J. Med. Virol. 93, 4616-4619.
36. Salpini, R., Alkhatib, M., Costa, G., Piermatteo, L., Ambrosio, F.A., Maio, D.V.C., Scutari, R., Duca, L., Berno, G., Fabeni, L., Alcaro, S., Silberstein, C.F., Artese, A., Svicher V., 2021.Key genetic elements, single and in clusters, underlying geographically dependent SARS-CoV-2 genetic adaptation and their impact on binding affinity for drugs and immune control. J. Antimicrob. Chemother. 76, 396-412.
37. Saxena, S.K., Kumar, S., Ansari, S., Paweska, J.T., Maurya, V.K., Tripathi, A.K., Abdel-Moneim, A.S., 2022. Characterization of the novel SARS-CoV-2 Omicron (B.1.1.529) variant of concern and its global perspective. J. Med. Virol. 94, 1738-1744.
38. Seal, S., Dharmarajan, G., Khan, I., 2021. Evolution of pathogen tolerance and emerging infections: A missing experimental paradigm. ELife.10, e68874.
39. Shah, M., Woo, H.G., 2022. Omicron: A Heavily Mutated SARS-CoV-2 Variant Exhibits Stronger Binding to ACE2 and Potently Escapes Approved COVID-19 Therapeutic Antibodies. Front. Immunol.12, 830527.

40. Starr, T.N., Greaney, A.J., Hilton, S.K., Ellis, D., Crawford, K.H.D., Dingens, A.S., Navarro, M.J., Bowen, J.E., Tortorici, M.A., Walls, A.C., King, N.P., Veesler, D., Bloom, J.D., 2020. Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. Cell.182, 1295-1310.e20.
41. Tchesnokova, V., Kulakesara, H., Larson, L., Bowers, V., Rechkina, E., Kisiela, D., Sledneva, Y., Choudhury, D., Maslova, I., Deng, K., Kutumbaka, K., Geng, H., Fowler, C., Greene, D., Ralston, J., Samadpour, M., Sokurenko, E., 2021. Acquisition of the L452R mutation in the ACE2-binding interface of Spike protein triggers recent massive expansion of SARS-Cov-2 variants. J. Clin. Microbio.59, e00921-21.
42. Toth-Petroczy, A., Tawfik, D.S., 2014.The robustness and innovability of protein folds.Curr.Opin.Struct. Biol. 26, 131-138.
43. Vogt, A.S., Augusto, G., Martina, B., Chang, X., Nasrallah, G., Speiser, D.E., Vogel, M., Bachmann, M.F., Mohsen, M.O., 2022. Increased Receptor Affinity and Reduced Recognition by Specific Antibodies Contribute to Immune Escape of SARS-CoV-2 Variant Omicron. Vaccines (Basel).10, 743.

Figures



(C) Fast and slow folding rates

S		E		M		N		NSP4		NSP6		NSP7		NSP8	
Wuhan-Hu-1	20.02	Wuhan-Hu-1	73.41	Wuhan-Hu-1	73.41	Wuhan-Hu-1	73.41	Wuhan-Hu-1	20.02	Wuhan-Hu-1	20.02	Wuhan-Hu-1	93.8	Wuhan-Hu-1	40.49
DF1041GL	54.05	A36V	89.56	D209H	53.843	D377Y	35.97	T327N	53.912	I266L	77.655	V33F	92	F15L	47.31
SN968LH	50.81	V52E	87.12	S99F	53.3	R209I	35.97	W392C	50.087	L259Q	70.84	D77N	91.3	Q24R	45.02
R21T	48.23	P71L	83.72	H148Q	51.432	S180C	35.97	L111S	50.057	R233H	70.26	N78S	88.7	S7N	41.722
L1200F	47.87	A41T	83.37	K15R	49.574	S202G	35.97	G309C	49.874	W31S	69.65	A48V	88.01	A191T	40.62
W258C	45.85	V17A	82.77	T127N	47.78	S183F	35.97	P187S	48.76	G177R	68.64	E50G	87.32	V34F	39.7
C1247Y	44.59	P71H	82.29	N117H	47.087	S183T	35.97	T143A	47.898	M126V	68.4	A42S	84.7	I156L	38.91
G72E	43.84	G10V	80.66	N43S	46.905	S232R	35.97	I23F	45.103	S32P	66.52	K7R	84.2	S7G	38.33
K1191Q	43.63	T9I	79.46	C64F	45.424	S235P	35.97	V373Q	44.622	V3M	65.547	R79S	83.9	K127T	37.735
C136R	43.6	R61H	78.32	E167Q	44.987	S255A	35.97	S34F	44.241	G188V	64.8	V22L	80.64	G144S	37.73
R1091C	43.19	L73F	78.02	P165W	44.76	S310I	35.97	R112G	44.185	H11Y	63.34	T45N	80.4	L122V	37.72
D614G	**35.3**	**S68F**	**75.39**	**T175M**	**36.83**	**RG203KR**	**6.91**	**F308Y**	**38.459**	**L37F**	**63.11**	**S25L**	**63.887**	**M129I**	**31.05**
M697T	11.26	E7Q	59.13	V10I	17.621	R92S	-3.35	P166Q	17.846	N264K	38.2	C32Y	64.71	S173I	18.13
VL860QK	11.2	P71S	58.85	Q19E	17.502	E253A	-5.02	Q200E	17.713	D133N	37.48	S15L	63.887	K61N	17.662
R847K	10.38	D72H	58.46	L206S	17.148	E290Q	-5.02	W365S	16.942	Y80C	36.06	M75T	59.25	T84I	17.37
R567K	10.38	S50G	57.62	L176P	17.125	T325A	-7.13	S81G	16.901	E195D	35.51	V16A	58.95	N176T	17.26
H519P	10.35	C44Y	57.51	I161M	15.79	S180I	-11.1	C150W	16.804	A88G	34.6	M75L	57.57	A14T	15.46
M902V	10.27	R69I	56.69	V70A	14.043	D216A	-11.1	M368L	12.298	Q257R	34.09	L76V	57.4	T141M	14.33
N914Y	9.69	V29A	55.77	A68T	12.736	M210I	-11.1	I11N	8.882	V289L	30.81	K70N	56.4	M55L	13.65
A522P	9.43	N45R	53.97	S4A	12.394	G97S	-11.1	S76N	8.042	L259R	30.58	L56F	55.2	L169P	4.39
Q954K	5.43	L73V	39.97	C33S	10.764	L407W	-11.1	N4S	7.178	T222S	28.27	E47K	54.2	S4L	3.74
F92L	3.867	D72Y	33.63	I217L	9.728	K248N	-12.2	L436R	4.101	W140R	25.13	S61P	54.09	S11L	3.74

Figure 1

Computational validation of predicted folding kinetics of spike, assembly, and maturation proteins in SARS-CoV-2 variants using machine learning algorithms

(LR: Linear regression, LGR: Logistic regression, RF: Random forest, NB: Naive Bayes, KNN: K-Nearest Neighbor, SVM: Support vector machine, DT: Decision tree, CNN: Convolutional Neural Network)

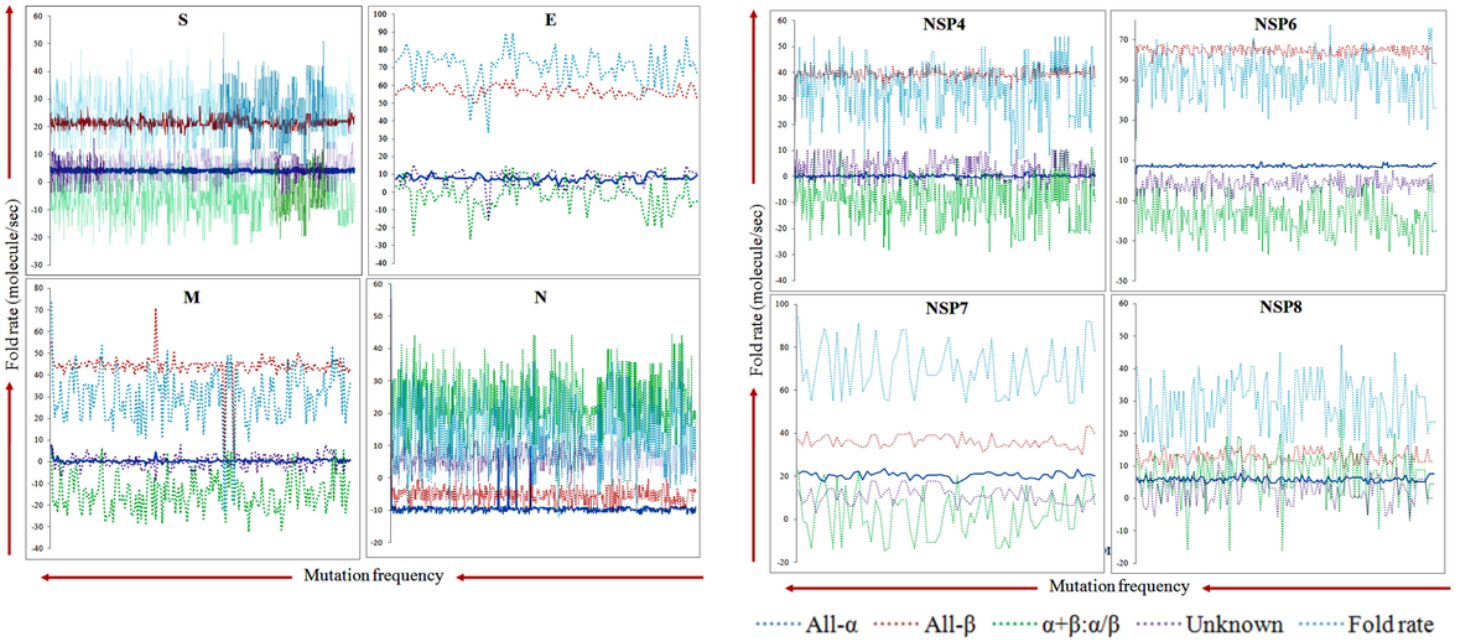


Figure 2

Impact of point mutations on folding kinetics of structural classes in spike, assembly, and maturation proteins in SARS-CoV-2 variants

(A) SARS-CoV-2 variants

Classification	First identified	WHO label	All-α	All-β	α+β:α/β	Unknown	Fold rate ln(kf)
	China	Wuhan-Hu-1	3.4	21.5	-11	6.12	20.02
Variant of Concern	India	Delta	4.49	21.5	-5.87	6.44	26.56
	South Africa	Omicron	2.88	21	-13	0.877	13.757
Variants Being Monitored	UK	Alpha	4.69	19.9	-15.7	6.93	15.82
	South Africa	Beta	4.31	23.6	-3.13	5.8	30.58
	Brazil	Gamma	4.28	25.6	-3.9	8.9	34.88
	USA	Epsilon	4.79	24.1	-1.45	7.1	34.54
		Eta	3.59	21.4	-6.65	4.04	22.38
		Iota	3.04	23.8	-4.33	3.55	26.06
	India	Kappa	4.35	25.5	-11	9.84	28.69
	Colombia	Mu	4.19	23.4	-10.6	12.8	29.79
	Brazil	Zeta	4.02	22	-8.88	10.8	27.94
Variant of Interest	Peru	Lambda	3.46	17.2	-7.53	6.47	19.6
	Philippines	Theta	3.95	24.6	-4.56	4.66	28.65

(B) Potent mutants

WHO label	Variant	Mutation	All-α	All-β	α+β:α/β	Unknown	Fold rate ln(kf)
Wuhan-Hu-1			3.4	21.5	-11	6.12	20.02
Delta	B.1.617.2	K417N	3.83	21.8	-0.549	2.75	27.831
		T478K	4.51	20.2	-4.96	7.31	27.06
		E484K	3.06	17.5	-14.9	6.85	12.51
Omicron	B.1.1.529	E484A	4.45	23.6	9.7	1.59	39.34
Alpha	B.1.1.7	L452R	4.71	23.9	8.18	1.21	38
Beta	B.1.351	L18F	4.51	20	-10.2	11	25.31
Gamma	P.1	P681H	4.94	18.7	-8.1	4.82	20.36
Iota	B.1.526	D253G	4.65	24.9	-0.137	5.89	35.303
Mu	B.1.621	R346K	3.27	19.8	-14.2	1.51	10.38
		N501Y	4.32	20.3	-10.7	-4.23	9.69
Lambda	C.37	N452Q	4.18	24.3	-5.26	10.6	33.82

(C) Functional motifs

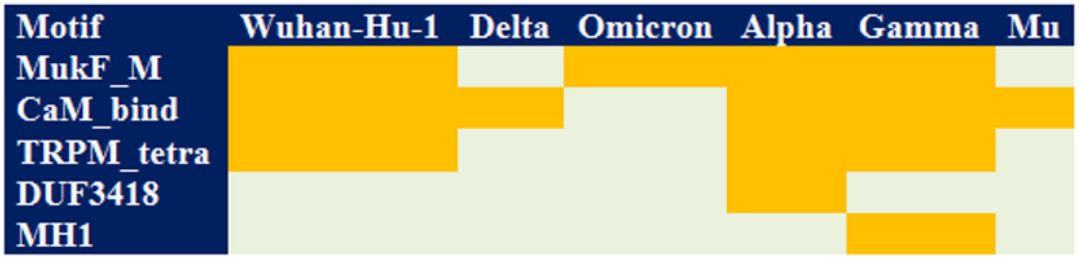


Figure 3

Impact of the potent mutations on folding kinetics of the SARS-CoV-2 variants classified by the World Health Organization (https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/)

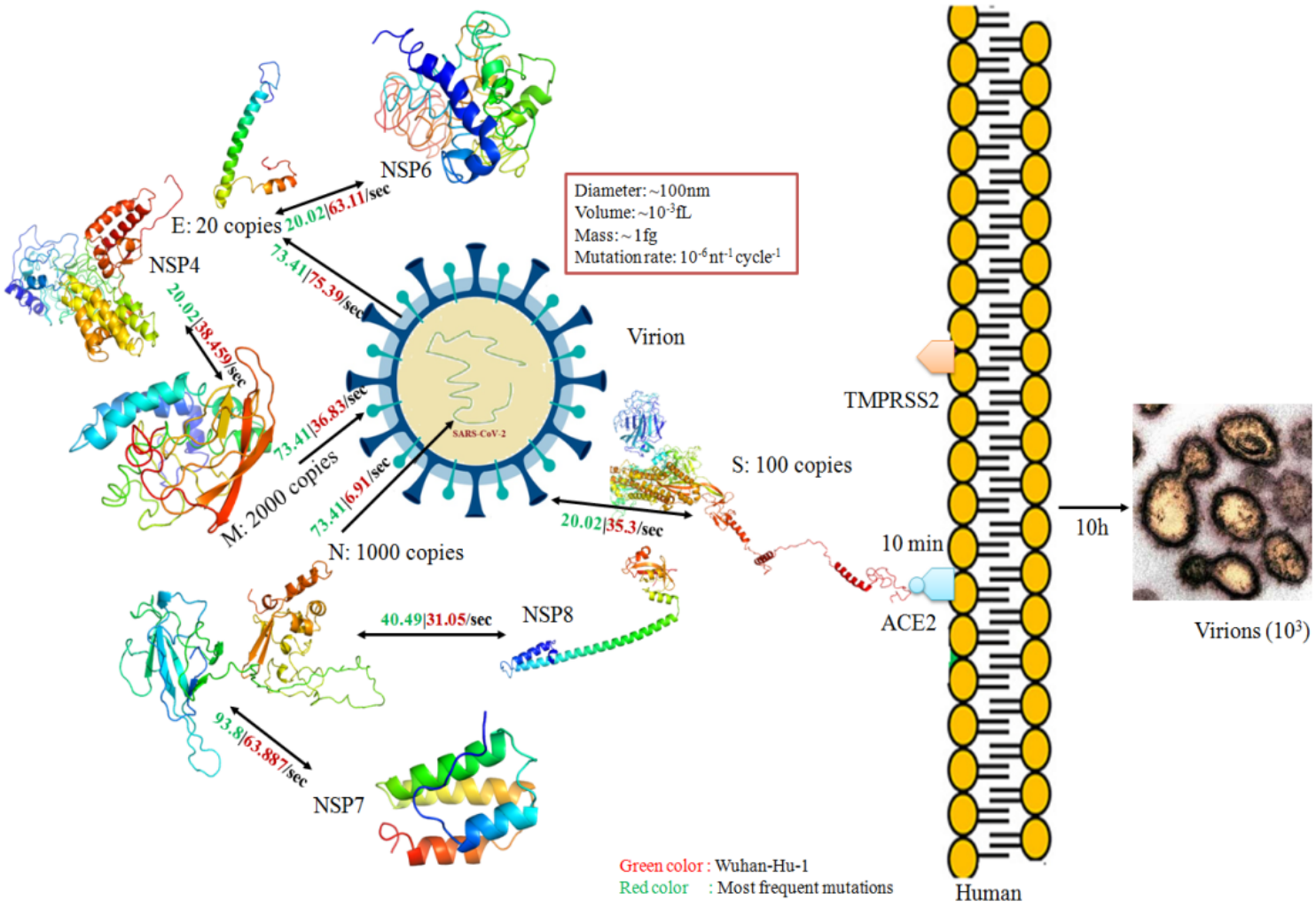


Figure 4

The proposed hijack mechanism of SARS-CoV-2 in host cells using the folding kinetics of spike, assembly, and maturation protein mutants.

Supplementary Files

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

- Supplementaryfile.xlsx
- Graphicalabstract.png