Structural and Functional Impacts of SARS-CoV-2 Spike Protein Mutations: Insights from Predictive Modeling and Analytics

Edem K. Netsey, Samuel M. Naandam, Joseph Asante, Abraham E. Kuukua, Aayire C. Yadem, Gabriel Owusu, Jeffrey G. Shaffer, Sudesh K. Srivastav, Seydou Doumbia, Ellis Owusu-Dabo, Chris E. Morkle, Desmond Yemeh, Stephen Manortey, Ernest Yankson, Mamadou Sangare, Samuel Kakraba

Submitted to: JMIR Bioinformatics and Biotechnology
on: March 08, 2025

Table of Contents

Structural and Functional Impacts of SARS-CoV-2 Spike Protein Mutations: Insights from Predictive Modeling and Analytics

Edem K. Netsey[1*] BSc, MSc; Samuel M. Naandam[2*] PhD, MSc, BSc; Joseph Asante[3] BEd, PhD; Abraham E. Kuukua[4] BSc, MSc; Aayire C. Yadem[5] BSc, MSc, PhD; Gabriel Owusu[6] BSc, MS, PhD; Jeffrey G. Shaffer[6] BS, MS, PhD; Sudesh K. Srivastav[6] BSc, MS, PhD, MStat; Seydou Doumbia[7] MD, PhD; Ellis Owusu-Dabo[8] PhD, FWACP, FGCP, MSc, MBChB, BSc; Chris E. Morkle[9] M.Ed, B.Ed; Desmond Yemeh[6] PhamD; Stephen Manortey[10] Diploma, Msc, PhD; Ernest Yankson[2] BSc, MS, PhD; Mamadou Sangare[11] BSc, MSc, PhD; Samuel Kakraba[6*] B.Ed, MS, PhD

[1]Department of Mathematics and Information Communication Technology Dambai College of Education Dambai GH
[2] University of Cape Coast Cape Coast GH
[3] University of Arkansas for Medical Sciences Little Rock US
[4] Woodale High School Memphis US
[5] CytoAstra LLC Little Rock US
[6] Tulane University New Orleans US
[7] Malaria Research and Training Center, University of Sciences, Techniques and Technologies of Bamako Bamako ML
[8] Kwame Nkrumah University of Science and Technology Kumasi GH
[9] Suhum Senior High School Suhum GH
[10] Ensign Global University Tema GH
[11] 14 Laboratory of Malaria and Vector Research (LMVR), National Institute of Allergy and Infectious Diseases (NIAID) Rockville US
[*]these authors contributed equally

Corresponding Author:
Samuel Kakraba B.Ed, MS, PhD

Tulane University
1440 Canal St
1440 Canal St
New Orleans
US

Abstract

Background: The COVID-19 pandemic requires a deep understanding of SARS-CoV-2, particularly how mutations in the Spike Receptor Binding Domain (RBD) Chain E affect its structure and function. Current methods lack comprehensive analysis of these mutations at different structural levels.

Objective: To analyze the impact of specific COVID-19 associated point mutations (N501Y, L452R, N440K, K417N, E484A) on the SARS-CoV-2 Spike RBD structure and function using predictive modeling, including a graph-theoretic model, protein modeling techniques, and molecular dynamics simulations.

Methods: The study employed a multi-tiered graph-theoretic framework to represent protein structure across three interconnected levels. This model incorporated 19 top-level vertices, connected to intermediate graphs based on 6-angstrom proximity within the protein's 3D structure. Graph-theoretic molecular metrics were applied to weigh vertices and edges at all levels. The study also used Iterative Threading Assembly Refinement (I-TASSER) to model mutated sequences and molecular dynamic simulation (MD) tools to evaluate changes in protein folding and stability compared to the wildtype.

Results: Three distinct predictive modeling and analytical approaches successfully identified structural and functional changes in the SARS-CoV-2 Spike RBD (Chain E) resulting from mutations. The novel graph-theoretic model detected notable structural changes, with N501Y and L452R showing the most pronounced effects on conformation and stability. K147N and E484A mutations demonstrated less significant impacts. Ab initio modeling and MD simulation findings corroborated the graph-theoretic analysis. The multi-level analytical approach provided a comprehensive visualization of mutation effects, deepening our understanding of their functional consequences.

Conclusions: This study advanced our understanding of SARS-CoV-2 Spike RBD mutations and their implications. The multi-faceted approach characterized the effects of various mutations, identifying N501Y and L452R as having the most substantial impact on RBD conformation and stability. The findings have important implications for vaccine development, therapeutic design, and variant monitoring. This research underscores the power of combining multiple predictive analytical approaches in virology, contributing valuable knowledge to ongoing efforts against the COVID-19 pandemic and providing a framework for future studies on viral mutations and their impacts on protein structure and function.

(JMIR Preprints 08/03/2025:73637)
DOI: https://doi.org/10.2196/preprints.73637

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http
No. Please do not make my accepted manuscript PDF available to anyone.

Original Manuscript

Original Paper

Structural and Functional Impacts of SARS-CoV-2 Spike Protein Mutations: Insights from Predictive Modeling and Analytics

Edem K. Netsey[1, †], Samuel M. Naandam[2], Joseph Asante[3], Kuukua E. Abraham[4], Aayire C. Yadem[5], Gabriel Owusu[6], Jeffrey G. Shaffer[7], Sudesh K. Srivastav[7] Seydou Doumbia[8,9], Ellis Owusu-Dabo[10], Chris E. Morkle[11], Desmond Yemeh[12], Stephen Manortey[13], Ernest Yankson[2], Mamadou Sangare[14] and Samuel Kakraba [7,12, †, *]

[1] Department of Mathematics and Information Communication Technology, Dambai College of Education, Dambai, 00233, Ghana; edem.netsey@stu.ucc.edu.gh (E.K.N.)
[2] Department of Mathematics, University of Cape Coast, Cape Coast, 00233, Ghana; snaandam@ucc.edu.gh (S.M.N.); ernest.yankson@ucc.edu.gh (E.Y.)
[3] Department of Geriatrics, University of Arkansas for Medical Sciences, Little Rock, AR 72205, USA; jasante@uams.edu (J.A.J.)
[4] Department of Mathematics, Woodale High School, 5151 Scottsdale Ave, Memphis, TN 38118, USA; abrahamk1@scsk12.org (K.E.A.)
[5] CytoAstra LLC, Little Rock, AR, USA; clement@cytoastra.com (A.C.Y.)
[6] Office of Research Proposal Development, Tulane University Celia Scott Weatherhead School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA 70112, USA; gowusu1@tulane.edu (G.O.)
[7] Department of Biostatistics and Data Science, Tulane University Celia Scott Weatherhead School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA 70112, USA; jshaffer@tulane.edu (J.G.S.); ssrivas@tulane.edu (S.K.S.); skakraba@tulane.edu (S.K.)
[8] Malaria Research and Training Center, University of Sciences, Techniques and Technologies of Bamako, Bamako, 00223 Mali; sdoumbi@gmail.com (S.D.)
[9] University Clinical Research Center, University of Sciences, Techniques and Technologies of Bamako, Bamako, 0223, Mali
[10] Department of Global and International School of Public Health, Kwame Nkrumah University of Science and Technology, Kumasi, 00233, Ghana; eowusu-dabo.chs@knust.edu.gh (E.O.D.)
[11] Department of Mathematics, Suhum Senior High Technical School, 00233, Ghana; morklechris@gmail.com (C.E.M.)
[12] Tulane Center for Aging, School of Medicine, Tulane University, 1430 Tulane Ave. 8513. New Orleans, LA 70112, USA; dyemeh@tulane.edu (D.Y.); skakraba@tulane.edu (S.K.)
[13] Ensign Global University, Tema Akosombo Highway, Kpong(E/R), 00233, Ghana; steve.manortey@ensign.edu.gh (S.M.)
[14] Laboratory of Malaria and Vector Research (LMVR), National Institute of Allergy and Infectious Diseases (NIAID), Rockville, Maryland, MD 20852, USA; mamadou.sangare@nih.gov (M.S.)
[*] Correspondence: Samuel Kakraba, Ph.D.; skakraba@tulane.edu;Tel.: +1-504-988-2475
[†] These authors contributed equally to this work.

Structural and Functional Impacts of SARS-CoV-2 Spike Protein Mutations: Insights from Predictive Modeling and Analytics

Abstract

Background: The COVID-19 pandemic requires a deep understanding of SARS-CoV-2, particularly how mutations in the Spike Receptor Binding Domain (RBD) Chain E affect its

structure and function. Current methods lack comprehensive analysis of these mutations at different structural levels.

Objective: To analyze the impact of specific COVID-19 associated point mutations (N501Y, L452R, N440K, K417N, E484A) on the SARS-CoV-2 Spike RBD structure and function using predictive modeling, including a graph-theoretic model, protein modeling techniques, and molecular dynamics simulations.

Methods: The study employed a multi-tiered graph-theoretic framework to represent protein structure across three interconnected levels. This model incorporated 19 top-level vertices, connected to intermediate graphs based on 6-angstrom proximity within the protein's 3D structure. Graph-theoretic molecular metrics were applied to weigh vertices and edges at all levels. The study also used Iterative Threading Assembly Refinement (I-TASSER) to model mutated sequences and molecular dynamic simulation (MD) tools to evaluate changes in protein folding and stability compared to the wildtype.

Results: Three distinct predictive modeling and analytical approaches successfully identified structural and functional changes in the SARS-CoV-2 Spike RBD (Chain E) resulting from mutations. The novel graph-theoretic model detected notable structural changes, with N501Y and L452R showing the most pronounced effects on conformation and stability. K147N and E484A mutations demonstrated less significant impacts. Ab initio modeling and MD simulation findings corroborated the graph-theoretic analysis. The multi-level analytical approach provided a comprehensive visualization of mutation effects, deepening our understanding of their functional consequences.

Conclusions: This study advanced our understanding of SARS-CoV-2 Spike RBD mutations and their implications. The multi-faceted approach characterized the effects of various mutations, identifying N501Y and L452R as having the most substantial impact on RBD conformation and stability. The findings have important implications for vaccine development, therapeutic design, and variant monitoring. This research underscores the power of combining multiple predictive analytical approaches in virology, contributing valuable knowledge to ongoing efforts against the COVID-19 pandemic and providing a framework for future studies on viral mutations and their impacts on protein structure and function.

Keywords: COVID-19 (Coronavirus Disease 2019); SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2); Graph-theoretic model; 6M0J; Predictive modeling; Covid-19 associated mutations (N501Y, E484A, L452R, N440K, K417N); Unsupervised machine learning

Introduction

Background

The COVID-19 pandemic, caused by SARS-CoV-2, had a devastating global impact, with hundreds of millions of confirmed cases and millions of deaths worldwide [1-4]. The virus's spike protein shares significant structural and sequence similarities with the SARS virus from 2003, including the use of the ACE2 receptor for cell entry. The SARS-CoV-2 spike protein's receptor-binding domain (RBD) has undergone various mutations, each with distinct impacts on viral behavior and infectivity [5]. The wild-type phenotype serves as a baseline, exhibiting no significant changes or enhanced

impacts[1]. Protein folding, maintenance, mutation, aggregation, neurodegenerative diseases, and COVID-19 are interconnected through complex biological processes[6-8].Except intrinsically disordered proteins, proteins must fold correctly to function properly, but mutations can disrupt this process, leading to misfolding and aggregation. Cells have protective mechanisms to maintain protein homeostasis, but when these fail, aggregates can form and contribute to neurodegenerative diseases like Alzheimer's (AD) and Parkinson's (PD) [6-10]. COVID-19 has been linked to AD, PD and other neurodegenerative diseases through various mechanisms[11-13]. The SARS-CoV-2 spike protein contains aggregation-prone regions that can form amyloid aggregates, potentially contributing to neurological complications[14-17]. Additionally, the virus's 3CL protease has been shown to induce tau protein aggregation, a hallmark of neurodegenerative diseases. Genetic studies have revealed a causal association between COVID-19 hospitalization and increased risk of AD[18]. The virus may accelerate neurodegeneration through inflammation, microvascular injury, and prion-like spread of misfolded protein. These aggregates were observed *in-vitro* using various experimental techniques, including fluorescence spectroscopy and electron microscopy [19-21].

Specific mutations have emerged that alter the virus's characteristics in different ways. Mutations such as K417N, associated with the Beta, Gamma, and Omicron variants, demonstrate moderate severity due to their role in immune evasion [22-24]. These mutations reduce antibody neutralization, potentially compromising the effectiveness of vaccines and natural immunity [25]. Similarly, N440K and S477N mutations, linked to localized outbreaks, show mild impacts by enhancing binding to the ACE2 receptor, which increases infectivity without causing major antigenic alterations [26, 27]. More concerning are mutations like T478K and L452R, found in Delta and Omicron variants, which display moderate to severe consequences. L452R, in particular, is associated with immune evasion and increased transmission due to its resistance to neutralizing antibodies [28]. The E484 mutations (A, K, Q), present in Beta, Gamma, and other variants, exhibit moderate to severe immune escape capabilities, significantly reducing neutralization by vaccine-induced or convalescent sera [29, 30]. Of note is the profound N501Y mutation, observed in Alpha, Beta, and Omicron variants. This mutation is classified as very severe due to its dual role in enhancing binding affinity to the ACE2 receptor and aiding in immune escape [31]. N501Y has been a pivotal factor in increasing transmissibility and contributing to widespread outbreaks [32]. These diverse mutations in the SARS-CoV-2 RBD demonstrate varying degrees of severity and impact, ranging from mild infectivity increases to severe immune evasion and transmissibility. The evolving nature of these mutations underscores the critical importance of continuous genomic surveillance and adaptive vaccine strategies in combating the COVID-19 [33-35]. As SARS-CoV-2 has spread, it has evolved into various strains, with the D614G mutation becoming nearly ubiquitous. Several "variants of concern" (VOCs) have emerged, characterized by mutations that may affect viral behavior and immune evasion [36-40]. These VOCs, identified in different regions, include B.1.1.7 (UK), P.1 and P2 (Brazil), B.1.351 (South Africa), B.1.617 (India), and B.1.526 (US), among others [41]. The emergence of these variants has significant implications for treatment strategies and vaccine efficacy. Some variants, such as B.1.351, have shown resistance to neutralization by convalescent plasma and certain monoclonal antibody treatments [42]. Additionally, *in-vitro* studies suggest that sera from vaccinated individuals may have reduced neutralizing capacity against variants with specific mutations, like E484K and N501Y [24, 43, 44]. While VOCs are defined by specific mutation patterns, the interplay between these mutations in affecting viral behavior is not fully understood [45-48]. Analysis of a large dataset of SARS-CoV-2 spike sequences revealed hundreds of amino acid variants, with a significant number occurring in the receptor-binding domain [49, 50]. Machine learning algorithms have revolutionized data analysis by uncovering hidden patterns in datasets and addressing critical questions across various discipline including disease diagnostics among others [51-54]. Unsupervised learning techniques like clustering and dimensionality reduction reveal intrinsic structures in data[55], enabling applications such as customer segmentation, anomaly

detection, and bioinformatics[56].To better understand the emergence and spread of new variants, novel bioinformatics approaches are being developed to identify spatially and temporally correlated mutations [57, 58]. The evolving nature of SARS-CoV-2 suggests that future vaccine design may need to be tailored to address the specific strain ensembles prevalent in different regions [59-61]. This approach could enhance vaccine efficacy against locally dominant variants and potentially provide broader protection against emerging strains.

Prior Work

Recent advancements in graph-theoretic modeling have provided valuable insights into the complex relationship between point mutations and disease phenotypes[62, 63]. Researchers have developed various graph-based approaches to model and analyze the intricate connections between genes, mutations, and phenotypes. These methods leverage the ability of graphs to represent complex biological relationships and interactions [64-77]. This approach has been particularly effective in elucidating the structural and functional impacts of genetic variations in proteins associated with hereditary disorders. One application of graph theory in mutation analysis involves the use of protein-protein interaction (PPI) networks to identify discriminative subnetworks associated with specific diseases[9, 78-86] allowing researchers to uncover patterns of mutations that may collectively contribute to phenotypes, going beyond the limitations of single-gene analyses.

Also, other studies exemplify the power of this methodology in understanding the molecular basis of point mutation-associated diseases like cystic fibrosis (CF), sickle cell disease (SCD), among others. Kakraba et al. [62, 87], developed a hierarchical graph-theoretic model to investigate the effects of point mutations on the NBD2 domain of the CFTR protein, which is implicated in CF. By constructing a multi-level graph (nested graph) representation of interacting amino acid residues, they developed a nested graph capable of quantifying both local and global structural changes resulting from virtual point mutations. This innovative method enabled the differentiation of mild mutations (such as Y1219G and G1271E) from severe ones (like N1303K) when compared to the wildtype, demonstrating the sensitivity and relevance of their graph-theoretic methods and resulting molecular descriptors (graph invariants) in analyzing complex biological networks. Building on this framework, Edem et al. [88], applied similar graph-theoretic techniques to explore the impact of point mutations on the hemoglobin protein (1A3N) in SCD. Utilizing author-adopted and molecular descriptors from the database established by Kakraba et al. [62, 87], Edem et al. successfully captured the structural effects of various mutations, including E6V, V23I, and K82N. Their analysis not only distinguished mild mutations from the wild type, but also highlighted the significant devastation caused by the severe E6V mutation, further validating the utility of graph-theoretic molecular nested graphs in understanding disease mechanisms. These studies collectively demonstrate the power of graph-theoretic modeling in bridging the gap between genetic mutations and their phenotypic manifestations in complex diseases. In recent years, the fields of *ab initio* modeling and molecular simulation dynamics have revolutionized our approach to studying intricate biological systems and networks[89-92]. These advanced computational methods enable researchers to create detailed models of biological entities, ranging from individual proteins to entire cellular structures, with unprecedented atomic-level precision[93]. By combining these techniques with systems biology principles, researchers can now explore the complex relationships between genetic alterations and disease manifestations. These sophisticated MD simulations provide a unique window into the molecular consequences of mutations, allowing scientists to track cascading effects from the smallest cellular components up to organism-wide changes[89-92]. This multi-tiered modeling strategy offers valuable insights into disease mechanisms, illuminating how specific genetic changes can influence protein behavior, disrupt cellular functions, and ultimately result in diverse clinical presentations. The fusion of structural data with network analysis significantly enhances our ability to predict and understand the connections between genetic makeup and disease outcomes. This

integrated approach provides a more holistic view of the intricate relationships between genotypes and phenotypes in complex disorders, paving the way for more targeted and effective therapeutic interventions[94]. By providing a computational framework for quantifying structural changes at multiple levels, this approach offers a promising avenue for predicting disease severity and potentially informing therapeutic strategies.

Goal of This Study

In this study, we hypothesized that specific point mutations in the SARS-CoV-2 spike protein significantly alter its structural conformation, dynamic behavior, and stability, thereby influencing its binding affinity, immune escape potential, and overall viral fitness. To test this hypothesis, we employed a multi-faceted computational approach combining graph-theoretic modeling with *ab initio* protein structure prediction and molecular dynamics simulations to create a comprehensive analysis framework. Using the Iterative Threading Assembly Refinement (I-TASSER)[95] platform, we generated detailed 3D models of mutated spike proteins, which allowed us to visualize how point mutations affect protein structure, including potential changes in binding sites and overall conformation. We then conducted MD simulations to study the dynamic behavior of these mutated proteins, revealing how mutations impact protein movement and stability compared to the wild type. By integrating these computational methods with our graph-based approach, we developed a holistic understanding of mutation effects, bridging the gap between genetic changes and their structural consequences. This comprehensive approach provided insights into both sequence-level alterations and their macromolecular impacts, enhancing our understanding of SARS-CoV-2's evolution and potential behavioral changes. Through the utilization of these complementary methods, we were able to construct a more complete picture of how spike protein mutations influence the virus's structure and function.

Methods

2.1. Graph-theoretic Model of SARS-CoV-2 Spike RBD (Chain E)

This work extends our previous graph-theoretic modeling approaches [62, 88, 96] by focusing specifically on the chain E chain of the SARS-CoV-2 spike protein and expanding combinatorial descriptors to include edge-weight assignments among others. The study incorporates molecular indices from earlier research [62, 88, 96] and introduces additional graph invariants. As a result, the molecular descriptors for subdomain graphs, the chain E of the SARS-CoV-2 spike protein, and the examined mutations are tailored to chain E of the SARS-CoV-2 spike RBD protein.

2.1.1. Subsequence Partition of SARS-CoV-2 Spike Protein Chain E

The tertiary crystal structure of SARS-CoV-2 spike RBD bound with ACE2 [97] was retrieved from the Protein Data Bank [98](PDB ID: 6M0J) and visualized using UCSF Chimera [99], as illustrated in **Figure 1**.

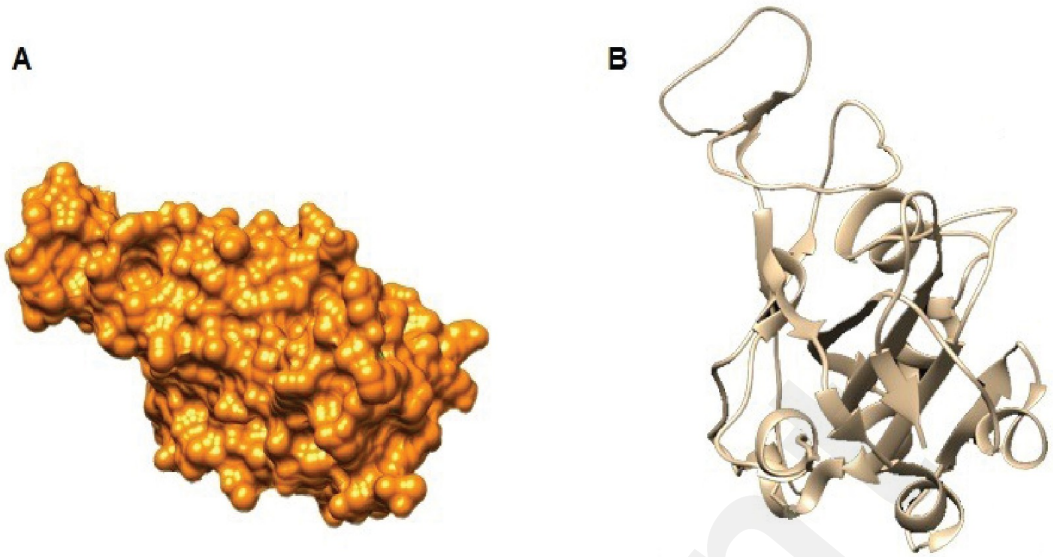


Figure 1: Structure of SARS-CoV-2 spike receptor-binding domain (RBD). **A)** Quaternary structure of the SARS-CoV-2 spike RBD (Chain E) in complex with ACE2. The RBD forms a concave outer surface that accommodates the N-terminal helix of ACE2, illustrating the key interaction interface for viral entry. The RBD is shown in one color, while ACE2 is depicted in a contrasting color to highlight the binding interface. **B)** Secondary structure of the SARS-CoV-2 spike RBD. The RBD features a twisted five-stranded antiparallel β sheet (β1, β2, β3, β4, and β7) with short connecting helices and loops. Four disulfide bonds are present: three in the core (Cys336–Cys361, Cys379–Cys432, and Cys391–Cys525) stabilizing the β sheet structure, and one (Cys480–Cys488) connecting loops at the distal end of the receptor-binding motif (RBM).

To provide a more comprehensive understanding of the SARS-CoV-2 Spike RBD structure, we generated a network graph using Cytoscape[100]. Figure 2 illustrates this network visualization, offering valuable insights into the RBD's intricate architecture and its potential functional significance. This graphical representation elucidates the complex relationships between various structural elements within the RBD, enhancing our understanding of its overall organization and possible mechanistic implications.

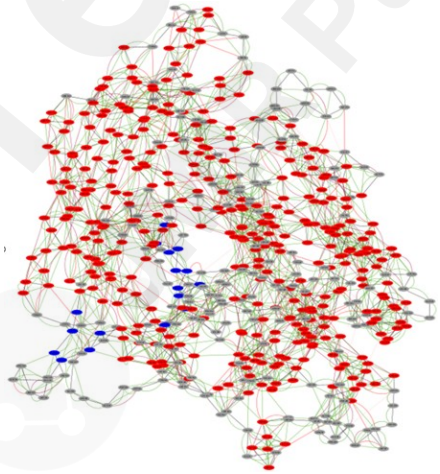


Figure 2: Network Graph of SARS-CoV-2 Spike Receptor-Binding Domain (RBD). This figure illustrates a network graph representation of the SARS-CoV-2 Spike protein's RBD, derived from the crystal structure 6M0J obtained from the Protein Data Bank. Each vertex in the graph represents an individual amino acid residue of the RBD. Edges connecting these vertices were established using a 6-angstrom proximity threshold, where two residues are connected if any of their atoms are within 6Å of each other. This representation provides a comprehensive view of the RBD structure, from individual amino acids to their network interactions.

The RBD, corresponding to chain E of the structure [97], was meticulously divided into nineteen subsequences (G1 to G19) while preserving the integrity of crucial biological information within the protein's secondary structures. This partitioning approach, following our previously published method [62, 88, 96], enables a more detailed analysis of the RBD's structure, which in turn informs the graph-theoretic modeling of the subsequence. During the

partitioning for analyzing the SARS-CoV-2 RBD (chain E), we followed three key principles to maintain structural integrity and facilitate detailed analysis. We preserved binding sites and secondary structures, isolated different structural elements into separate subsequences, and limited each subsequence to 13 amino acid residues. To account for protein complexity, loop regions were given more flexibility, allowing for the inclusion of turns, 3/10-helices, and short alpha helices. This careful approach provided a comprehensive view of the RBD's structure, enabling more in-depth analysis and modeling. The core structure of the SARS-CoV-2 RBD consists of a twisted five-stranded antiparallel β sheet (β1, β2, β3, β4, and β7), interconnected by short helices and loops. This detailed structural breakdown sets the foundation for further study of the RBD's composition and function. [97, 101]. **Table 1** provides a comprehensive overview of the subsequence partition, including subsequence identifiers (G1 to G19), corresponding amino acid residues, secondary structure classification, corresponding subdomains, and reason for the partitions. This detailed structural breakdown facilitates in-depth analysis of the SARS-CoV-2 spike RBD (chain E).

Table 1. **Subsequence partition of SARS-CoV-2 Spike RBD (Chain E)**

Subdomain graph	Subsequence	Amino acid sequence	Structural and or functional regions
G1	TNLCP	333 - 337	Coil, turn
G2	FGEVFNA	338 - 344	Alpha helix, turn, genome variant site (339), Mutagenesis (343)
G3	TRFASVYA	345 - 352	Alpha helix, coil, genome variant site (346)
G4	WNRKRISNCV	353 - 362	Beta sheet, coil.
G5	ADYSVLYNSAS	363 - 373	Alpha helix, coil, genome variant site (371, 373)
G6	FSTFKCYG	374 - 381	Beta Sheet, coil, genome variant site (375).
G7	VSPTKLNDLCF	382 - 392	Coil, Alpha helix.
G8	TNVYADSFVIR	393 - 403	Beta sheet, coil.
G9	GDEVRQIAPG	404 - 413	Alpha helix, coil.
G10	QTGKIADYNYKLPDD	414 - 428	Alpha Helix, coil, genome variant site (417)
G11	FTGCVIAWNS	429 - 438	Beta sheet, coil.
G12	NNLDSKVGGNY	439 - 449	Alpha helix, coil, turn, genome variant site (440, 446)
G13	NYLYRLFRK	450 - 458	Beta sheet, coil, genome variant site (452, 453), Mutagenesis (452, 453)
G14	SNLKPFERDISTEIY	459 - 473	Coil, turns.
G15	QAGSTPCNGVEGFN	474 - 487	Coil, turns, genome variant site (477, 478, 484), Mutagenesis (475, 483).
G16	CYFPLQSYGF	488 - 497	Beta sheet, coil, genome variant site (490, 493, 496), Mutagenesis (490,

			493).
G17	QPTNGVGYQ	498 - 506	Alpha helix, coil, genome variant site (501, 505), Mutagenesis (501).
G18	PYRVVVLSFELLH A	507-520	Beta sheet, coil, Mutagenesis (519)
G19	PATVCG	521-526	Coil

2.1.2. Subdomain Graphs (Corresponding to Subsequence Partitions) of SARS-CoV-2 Spike Protein chain E

After partitioning into subsequences as shown in **Table 1**, we applied a sophisticated multi-step graph-theoretic modeling approach. Using I-TASSER[95, 102], a state-of-the-art protein structure prediction tool, we generated *ab initio* models for each subsequence, employing a proximity threshold of 6 angstroms and determining endpoints based on each amino acid residue's center of mass. These structural predictions were visualized using Cystoscape[100], a powerful network visualization software, resulting in 19 comprehensive subdomain graphs. Each subdomain graph, corresponding to a previously classified subsequence in **Table 1**, offers a detailed visual representation of the structural and interaction patterns within each RBD subdomain.

Figure 3 illustrates the subdomain graphs for subsequences G13 and G17 of the SARS-CoV-2 Spike protein's RBD in Chain E. These specific subdomains are of particular interest as they contain the locations of severe mutations N501Y and L452R, respectively. The graphical representation provides a detailed view of the structural context surrounding these critical mutation sites, offering insights into how these changes might affect the protein's function and interactions.

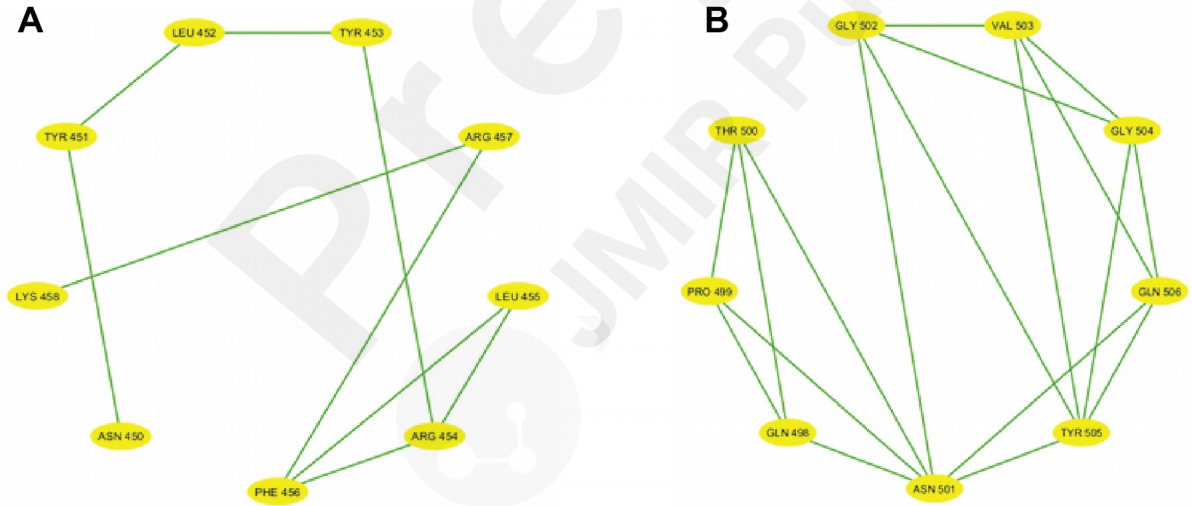


Figure 3: Subdomain graphs of SARS-CoV-2 Spike protein RBD (Chain E) where N501Y and L452R mutations occur. A) Subdomain graph for G13 containing N501Y mutation. B) Subdomain graph for G17 containing L452R mutation. *Ab initio* models were generated for each subsequence, employing a proximity threshold of 6 angstroms and determining endpoints based on each amino acid residue's center of mass. These graphs illustrate the local structural context surrounding the mutation sites, with nodes representing amino acid residues and edges indicating spatial proximity. The central nodes (N501Y in A and L452R in B) highlight the locations of the severe mutations, while surrounding nodes depict neighboring residues that may influence or be affected by these mutations. This representation aids in visualizing potential structural and functional impacts of these critical mutations on the Spike protein's RBD.

This methodical approach not only enables a more nuanced understanding of the RBD's structure but also lays the groundwork for further analysis of how mutations might affect these interactions and,

consequently, the entire protein. The supplementary materials (**see supplementary materials**) encompass all remaining subdomain graphs featured in this study, excluding G13 and G17.

2.1.3. Graph-Theoretic Model of SARS-CoV-2 Spike RBD (Chain E)

Next, we modeled SARS-CoV-2 Spike RBD corresponding to chain E of the structure protein to consist of three distinct levels, with each level offering a unique perspective on the protein's structure and interactions. At the foundation, we have the lowest level, comprising 20 vertex-weighted amino acids representing the essential building blocks of the protein. Moving up, the middle level features 19 distinct vertex-weighted subgraphs (subdomain graphs), each corresponding to a specific subsequence of SARS-CoV-2 Spike Receptor-Binding Domain (RBD) corresponding to chain E (**See Table 1 and Supplementary figures**) of the structure protein. In these subgraphs or subdomain graphs, individual amino acids are represented as vertices, with weights assigned based on graph invariants derived from the lower-level graphs. The top level of our model presents a more condensed view, where each subdomain graph from the middle level is consolidated into a single weighted vertex. The weights of these vertices are determined by molecular descriptors calculated from their respective subdomain graphs. To establish connections between vertices in the SARS-CoV-2 RBD corresponding to chain E nested domain graph, we employed a proximity threshold of 6 angstroms between adjacent residues. A 6-angstrom proximity threshold is often used to define connections in the SARS-CoV-2 Spike RBD graph because it captures meaningful atomic interactions, such as hydrogen bonds and van der Waals forces, essential for protein structure and function. This threshold ensures accurate representation of residue connectivity, aiding in understanding how the RBD stabilizes its structure and interacts with the ACE2 receptor[103]. Additionally, it accommodates the dynamic flexibility of the RBD, reflecting both stable and transient interactions critical for receptor binding and regulation[104, 105]. This balance of specificity and inclusivity makes it effective for modeling structural and functional relationships. **Figure 4** provides a visual representation of our graph-theoretic model for SARS-CoV-2 Spike RBD corresponding to chain E, illustrating the hierarchical structure and interactions captured by this approach.

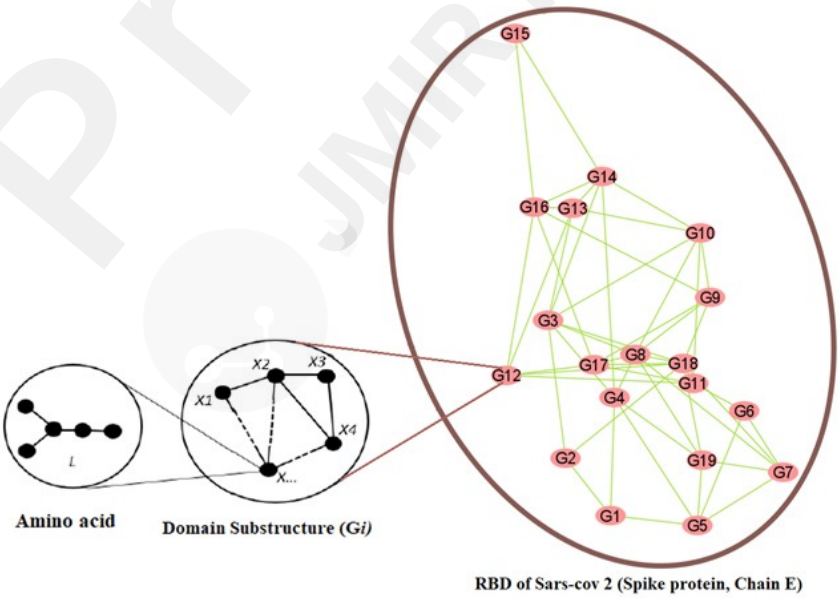


Figure 4: Hierarchical Graph Model of SARS-CoV-2 Spike Receptor-Binding Domain (RBD) Chain E. This figure depicts a three-level hierarchical graph model of the SARS-CoV-2 Spike RBD (Chain E), illustrating the protein's structure and interactions at different scales. The lower level shows 20 vertex-weighted amino acids, the middle level presents 19 vertex-weighted subdomain graphs, and the top level displays a condensed view with single weighted vertices representing each subdomain. Edges between vertices are established using a 6-angstrom proximity threshold, providing a comprehensive view from individual amino acids to subdomain interactions.

Building on our previous work[62, 87, 88], we utilized Cytoscape[100] to calculate the change in molar mass between adjacent vertices per average degree of the subdomain graphs (mid-level graphs), as shown in Table 2 and based on Equation 1. Specifically, we estimated the change in molar mass per average degree (ΔMd) along the edge connecting vertices R_i and R_j within these mid-level graphs using our previously established equation (Equation 1).

$$\Delta Md = \left|\frac{R_i - R_j}{\bar{d}}\right| \quad \text{kg/deg} \quad \text{.....(1)} \quad [88]$$

where ΔMd denotes the displaced mass of one amount of substance (in kilogram per average degree of graph), $|R_i - R_j|$ denotes the absolute difference in molar mass of adjacent residues and d is the average degree of the top-level graph.

Table 2. **Molar Mass as Weighted Degrees of Subdomain**

Subdomain	Molar mass in g/mol	Subdomain	Molar mass in g/mol
G1	618.7	G11	1259.4
G2	890.9	G12	1360.4
G3	1040.1	G13	1416.7
G4	1437.6	G14	2064.2
G5	1369.4	G15	1614.7
G6	1078.3	G16	1386.7
G7	1416.6	G17	1107.2
G8	1464.5	G18	1877.1
G9	1203.3	G19	636.7
G10	1993.2		

2.1.3. Virtual Mutations in SARS-CoV-2 Spike RBD (Chain E)

To evaluate the impact of single point mutations on the entire SARS-CoV-2 Spike RBD corresponding to chain E of the structural protein, we selected five (5) prevalent mutations associated with mild or severe Covid-19 from existing literature. Table 3 provides a comprehensive view of how each mutation influences the SARS-CoV-2 Spike RBD corresponding to chain E of the structural protein.

Table 3. **Mutation phenotypes of SARS-CoV-2 Spike RBD[106-110]**

Mutation Phenotype	Strain	Phenotype	Impact
Wildtype	-----	Typical or normal form of a species as it occurs in nature; baseline for comparison with mutated or altered forms	No impact/baseline
K417N	Beta/B.1.351, Gamma/P.1, Omicron/BA.1, Omicron/BA.2	Moderate	Immune evasion, associated with reduced neutralization by

			antibodies.
N440K	Omicron/BA.1, Omicron/BA.2	Mild	Potential immune escape, enhances binding to ACE2 receptor but not a major antigenic change.
L452R	Delta (B.1.617.2 and AY lineages), Kappa (B.1.617.1)	Severe	Immune evasion & transmission, linked to resistance to neutralizing antibodies and higher infectivity.
E484A	Omicron	Moderate	Immune evasion, reduces neutralization by convalescent and vaccine-induced sera.
N501Y	Alpha, Beta, Omicron, Mu	Very severe	Increased binding affinity & immune escape, enhances ACE2 binding and associated with immune evasion.

Subsequently, the changes in molar mass between adjacent vertices per average degree (as shown in Table 2), derived from the weighted network interaction data, were assigned as vertex weights for the subdomain graphs within the top-level graph, G. This step enabled the generation of molecular descriptors (graph invariants) based on the weighted network interaction data in Cytoscape [95], representing the edge-interaction weights of the top-level graph. The resulting molecular database of graph invariants represented the wildtype graph (no mutation). For each of the five mutations, a virtual mutation process was performed. This involved identifying the specific amino acid in the relevant subdomain graph, mutating it via substitution or deletion, and creating a "mutant-specific vertex-weighted graph" for the affected subdomain by submitting the mutated FASTA sequence to I-TASSER [93, 94] for *ab initio* modeling (see supplementary material for mutant-specific vertex-weighted graphs). New graph-theoretic molecular descriptors were then computed for each mutated subdomain graph and applied to the top-level graph. Subsequently, the same molecular descriptors previously computed for the top-level graph of the wildtype were recalculated to capture the mutation's impact. This approach allowed us to observe both local (subdomain) and global (entire SARS-CoV-2 Spike RBD), corresponding to chain E of the structural protein) effects of each point mutation on the SARS-CoV-2 Spike protein. Table 4 presents a comprehensive set of molecular descriptors for the wildtype SARS-CoV-2 Spike RBD and its various mutations. These graph invariants offer quantitative insights into how mild and severe mutations affect the RBD's structural and functional characteristics. Through analysis of these metrics across various variants, we elucidated the distinct effects of each mutation on the Spike protein's characteristics and functional implications.

Table 4. **Edge (interaction) weights of mutation phenotypes for SARS-CoV-2 Spike RBD (Chain E).**

Subdomain	Wildtype	K417N	N440K	L452R	N501Y	E484A

Interaction						
G14 (meta) G15	0.08990	0.08990	0.08990	0.08990	0.08990	0.10150
G13 (meta) G14	0.12950	0.12950	0.12950	0.12090	0.12950	0.12950
G12 (meta) G13	0.01126	0.01126	0.00844	0.01986	0.01126	0.01126
G11 (meta) G12	0.02020	0.02020	0.02302	0.02020	0.02020	0.02020
G7 (meta) G18	0.09210	0.09210	0.09210	0.09210	0.09210	0.09210
G2 (meta) G18	0.19724	0.19724	0.19724	0.19724	0.19724	0.19724
G11 (meta) G18	0.12354	0.12354	0.12354	0.12354	0.12354	0.12354
G12 (meta) G18	0.10334	0.10334	0.10052	0.10334	0.10334	0.10334
G9 (meta) G18	0.13476	0.13476	0.13476	0.13476	0.13476	0.13476
G4 (meta) G14	0.12532	0.12532	0.12532	0.12532	0.12532	0.12532
G3 (meta) G13	0.07532	0.07532	0.07532	0.08392	0.07532	0.07532
G3 (meta) G18	0.16740	0.16740	0.16740	0.16740	0.16740	0.16740
G3 (meta) G4	0.07950	0.07950	0.07950	0.07950	0.07950	0.07950
G2 (meta) G3	0.02984	0.02984	0.02984	0.02984	0.02984	0.02984
G3 (meta) G14	0.20482	0.20482	0.20482	0.20482	0.20482	0.20482
G11 (meta) G17	0.03044	0.03044	0.03044	0.03044	0.02062	0.03044
G12 (meta) G17	0.05064	0.05064	0.05346	0.05064	0.04082	0.05064
G9 (meta) G17	0.01922	0.01922	0.01922	0.01922	0.00940	0.01922
G17 (meta) G18	0.15398	0.15398	0.15398	0.15398	0.14416	0.15398
G8 (meta) G9	0.05224	0.05224	0.05224	0.05224	0.05224	0.05224
G8 (meta) G18	0.08252	0.08252	0.08252	0.08252	0.08252	0.08252
G3 (meta) G8	0.08488	0.08488	0.08488	0.08488	0.08488	0.08488
G4 (meta) G8	0.00538	0.00538	0.00538	0.00538	0.00538	0.00538
G8 (meta) G17	0.07146	0.07146	0.07146	0.07146	0.06164	0.07146
G8 (meta) G7	0.00958	0.00958	0.00958	0.00958	0.00958	0.00958
G6 (meta) G7	0.06766	0.06766	0.06766	0.06766	0.06766	0.06766
G6 (meta) G11	0.03622	0.03622	0.03622	0.03622	0.03622	0.03622
G8 (meta) G19	0.16556	0.16556	0.16556	0.16556	0.16556	0.16556
G7 (meta) G19	0.15598	0.15598	0.15598	0.15598	0.15598	0.15598
G4 (meta) G19	0.16018	0.16018	0.16018	0.16018	0.16018	0.16018
G19 (meta) G18	0.24808	0.24808	0.24808	0.24808	0.24808	0.24808
G1 (meta) G2	0.05444	0.05444	0.05444	0.05444	0.05444	0.05444
G1 (meta) G4	0.16378	0.16378	0.16378	0.16378	0.16378	0.16378
G5 (meta) G19	0.14654	0.14654	0.14654	0.14654	0.14654	0.14654
G5 (meta) G4	0.01364	0.01364	0.01364	0.01364	0.01364	0.01364
G1 (meta) G5	0.15014	0.15014	0.15014	0.15014	0.15014	0.15014
G5 (meta) G6	0.05822	0.05822	0.05822	0.05822	0.05822	0.05822
G5 (meta) G7	0.00944	0.00944	0.00944	0.00944	0.00944	0.00944
G16 (meta) G15	0.04560	0.04560	0.04560	0.04560	0.04560	0.03400
G12 (meta) G16	0.00526	0.00526	0.00244	0.00526	0.00526	0.00526
G13 (meta) G16	0.00600	0.00600	0.00600	0.01460	0.00600	0.00600
G16 (meta) G17	0.05590	0.05590	0.05590	0.05590	0.04608	0.05590
G14 (meta) G16	0.13550	0.13550	0.13550	0.13550	0.13550	0.13550
G9 (meta) G16	0.03668	0.03668	0.03668	0.03668	0.03668	0.03668
G8 (meta) G10	0.10574	0.10292	0.10574	0.10574	0.10574	0.10574
G9 (meta) G10	0.15798	0.15516	0.15798	0.15798	0.15798	0.15798
G10 (meta) G14	0.01420	0.01702	0.01420	0.01420	0.01420	0.01420

G10 (meta) G11	0.14676	0.14394	0.14676	0.14676	0.14676	0.14676
G3 (meta) G10	0.19062	0.18780	0.19062	0.19062	0.19062	0.19062
G10 (meta) G13	0.11530	0.11248	0.11530	0.10670	0.11530	0.11530

2.1.4. Unsupervised Machine learning analysis of point mutations associated with SARS-CoV-2 Spike RBD (Chain E)

To analyze the impact of point mutations associated with SARS-CoV-2 Spike RBD (Chain E), we employed our previous novel approach combining graph theory and machine learning [61, 86, 87]. Our method utilized graph-theoretic molecular weighted invariants or descriptors, as detailed in Table 4, for the wildtype SARS-CoV-2 Spike RBD (Chain E) and its mild and severe mutations. We applied an unsupervised machine learning technique, specifically hierarchical clustering, to visualize the variations between each SARS-CoV-2 Spike RBD (Chain E) point mutation and the wildtype SARS-CoV-2 Spike RBD (Chain E). Hierarchical clustering provided insights into the structure and relationships within the datasets. The analysis was conducted using R statistical software, employing the single linkage function and Euclidean distance to generate a dendrogram for the SARS-CoV-2 Spike RBD mutation phenotypes. We opted for the single-linkage function to minimize potential biases in clustering the SARS-CoV-2 Spike RBD (Chain E)-related point mutations [104, 105]. This approach allowed us to create a visual representation of how virtual SARS-CoV-2 Spike RBD (Chain E)-related point mutations affected the entire SARS-CoV-2 Spike RBD (Chain E) structure in comparison to the wildtype. Through this approach we were able to distinguish how the different virtual mutations were differed from the wildtype, providing an insight into how the point mutations impacted the entire SARS-CoV-2 Spike RBD.

2.2. *Ab Initio* Modeling of Mutated Protein Sequences Using I-TASSER

To further elucidate the structural and functional implications of spike protein mutations, we implemented an *ab initio* modeling approach using the I-TASSER platform [93]. This state-of-the-art predictive modeling tool was employed to generate high-resolution, three-dimensional structural models of the spike protein variants [93, 94, 106]. The I-TASSER algorithm utilizes a hierarchical approach, combining threading, fragment assembly, and atomic-level structure refinement to predict protein structure and function. We input the mutated spike protein sequences into the I-TASSER server[95], which then produced detailed structural models. These models were analyzed to identify potential alterations in protein folding, binding site configurations, and overall conformational changes resulting from the point mutations. The visual representations derived from this process provided crucial insights into the molecular-level effects of the mutations, complementing our graph-theoretic modeling approach. By integrating these computational methodologies, we were able to establish a comprehensive framework for understanding the relationship between sequence-level mutations and their macromolecular consequences, offering a multifaceted view of the spike protein's structural and functional adaptations.

2.3. Molecular Dynamics Simulations of Wildtype and Mutated Proteins in Water

To complement our findings from graph-theoretic modeling, we explored whether molecular dynamics simulations can adequately replicate the effects of point mutations leading to diverse phenotypes among the COVID-19 mutations analyzed in this work. MD simulations were employed to investigate the binding interactions and stability of wildtype and mutated proteins in a dynamic state. Using the WebGRO for Macromolecular Simulations server[111] based on GROMACS [107], modeled structures were prepared in orthorhombic simulation boxes solvated with simple point

charge (SPC) water, counterions, and 0.15 M NaCl to mimic physiological conditions. Energy minimization was performed using the steepest descent integrator, followed by equilibration at 300 K and 1.1023 bar [112][108]. For robustness, triplicate 50 ns simulations were conducted with varied random sampling seed inputs. Similarly, wildtype protein structures from PDB and mutated variants generated via I-TASSER [93] were solvated using SPC/E water models, with the OPLS-AA force field describing protein interactions. Position restraints on protein atoms were applied during initial equilibration before full MD production runs. To further validate our findings, we conducted a confirmatory test using MD simulations. The simulation was performed over an extended period of 100 nanoseconds to ensure comprehensive sampling of the molecular behavior. This approach allowed us to observe the dynamic properties of the system over a longer timescale, providing additional insights into the stability and conformational changes of the molecular structures under study. Post-simulation analyses included root mean square deviation (RMSD) plots to assess trajectory stability and deviations in atomic positions. Minimum and maximum RMSD values, standard deviations, and standard errors were calculated to compare wildtype and mutated variants. Additionally, the Kolmogorov-Smirnov test was used to determine whether mutations altered the distribution of protein conformations relative to the wildtype. Visualization and analysis of trajectories were performed using VMD [113][109] and BIOVIA Discovery Studio (Dassault Systèmes BIOVIA. Discovery Studio Modeling Environment, Release 2017. Dassault Systèmes; San Diego, CA, USA. 2017). This comprehensive approach enabled a detailed comparison of protein-water interactions and stability across variants, providing insights into mutational effects on protein behavior.

3. Results

3.1. Graph-Theoretic Modeling Reveals Impact of Point Mutations on SARS-CoV-2 Spike RBD

We developed a three-level weighted hierarchical graph-theoretic model of the SARS-CoV-2 Spike RBD corresponding to chain E (shown in Figure 4). The model consisted of a foundation level with 20 vertex-weighted amino acids, a middle level with 19 vertex-weighted subdomain graphs, and a top level where subdomain graphs are consolidated into single weighted vertices. Key features of the model included vertex connections established using a 6-angstrom proximity threshold, subdomain graph vertex weights based on molar mass changes between adjacent vertices per average degree, and a virtual mutation process applied for five specific mutations (K417N, N440K, L452R, N501Y, E484A). Mutant-specific vertex-weighted graphs were created using I-TASSER[93] for *ab initio* modeling, and new graph-theoretic molecular descriptors were computed for mutated subdomains and applied to the top-level graph. This approach enabled observation of both local (subdomain) and global (entire RBD) effects of each point mutation on the SARS-CoV-2 Spike protein, facilitating understanding of how point mutations lead to different COVID-19 phenotypes.

Figure 5 presents a dendrogram using the graph-theoretic invariants/descriptors from Table 4 (see methods section), offering a visual representation of the impact of each virtual point mutation on the SARS-CoV-2 Spike RBD (Chain E) relative to the wildtype structure. This hierarchical clustering diagram illustrates the structural and functional relationships between the mutated variants and the original wildtype RBD (Chain E).

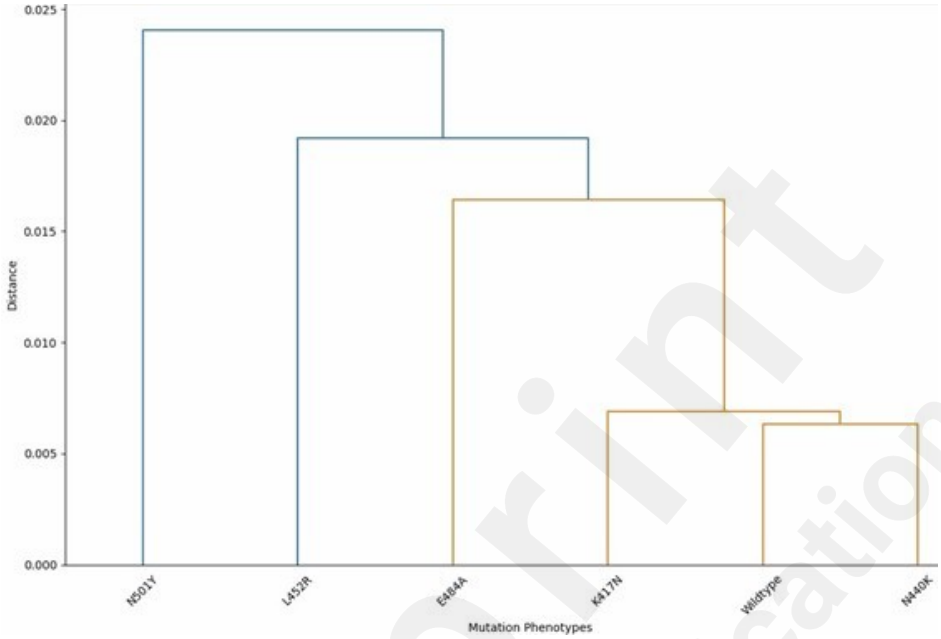


Figure 5: Clustering analysis of COVID-19 mutation-associated point mutations in SARS-CoV-2 RBD. This figure illustrates the hierarchical clustering of point mutations associated with different phenotypes of COVID-19 in the SARS-CoV-2 RBD. The analysis was performed using R statistical software, employing molecular descriptors of the top-level graph (as listed in **Table 4**) with the built-in functions 'hclust' and single linkage method. Euclidean distance was used as the similarity measure for clustering.

Dendrogram analysis (Figure 4) revealed varying degrees of divergence between the wildtype and the analyzed mutations based on Euclidean distance. The severe N501Y mutation, characterized by the substitution of asparagine with tyrosine at position 501, showed the greatest divergence from the wildtype, with an approximate Euclidean distance of 0.023. The L425R mutation, involving the replacement of leucine with a positively charged arginine at position 425, also exhibited a significant difference, with a Euclidean distance of approximately 0.019. Similarly, the E484A mutation, where glutamic acid is substituted with alanine at position 484, demonstrated a divergence of approximately 0.017 Euclidean distance from the wildtype. In contrast, the K147N mutation, which replaces lysine with asparagine at position 147 within the N-terminal domain, was closer to the wildtype but still distinct, with an approximate Euclidean distance of 0.007. These results underscore the structural and functional variability introduced by these mutations relative to the wildtype.

3.2 *Ab initio* Models Reveal Structural Alterations in Mutated SARS-CoV-2 Spike RBD

To gain deeper insights into the impact of COVID-19-specific mutations on the SARS-CoV-2 Spike RBD (Chain E) protein conformation, *ab initio* models were generated using I-TASSER[95] and visualized in Cytoscape [100]. These models provide a comparative view of selected point mutations (N440K, E484A, N501Y, K417N, and L452R) against the wildtype structure. Figures 6, 7, and 8 illustrate the structural changes induced by each mutation, revealing that even a seemingly mild mutation like E484A can significantly affect secondary structures and global protein folding. The models demonstrate alterations in protein folding patterns, changes in structural integrity, and modifications to local and global conformations. These visualizations highlight the profound impact of mutations on the SARS-CoV-2 Spike RBD, emphasizing how small changes can lead to substantial structural rearrangements. Such alterations may influence the virus's infectivity, immune evasion, and interaction with the ACE2 receptor. Notably, the N501Y mutation shows significant local and regional changes in the *ab initio* model, affecting protein folding and structural integrity. These modifications are more pronounced when compared to the wild-type structure, indicating that the N501Y mutation substantially impacts the RBD's conformation as shown by the Ab initio modeling results in Figure 7F. The E484A mutation, despite appearing mild, demonstrates detrimental effects on secondary structures and global protein folding (see Figure 6A). Other mutations (N440K, K417N, L452R) exhibit varying degrees of impact on the RBD structure (see Figure 7A, 8 and 6D), potentially influencing the spike protein's function and stability. These structural insights provide a foundation for understanding the molecular mechanisms behind the enhanced transmissibility and potential immune evasion of SARS-CoV-2 variants, contributing to our knowledge of the virus's evolution and informing future therapeutic strategies.

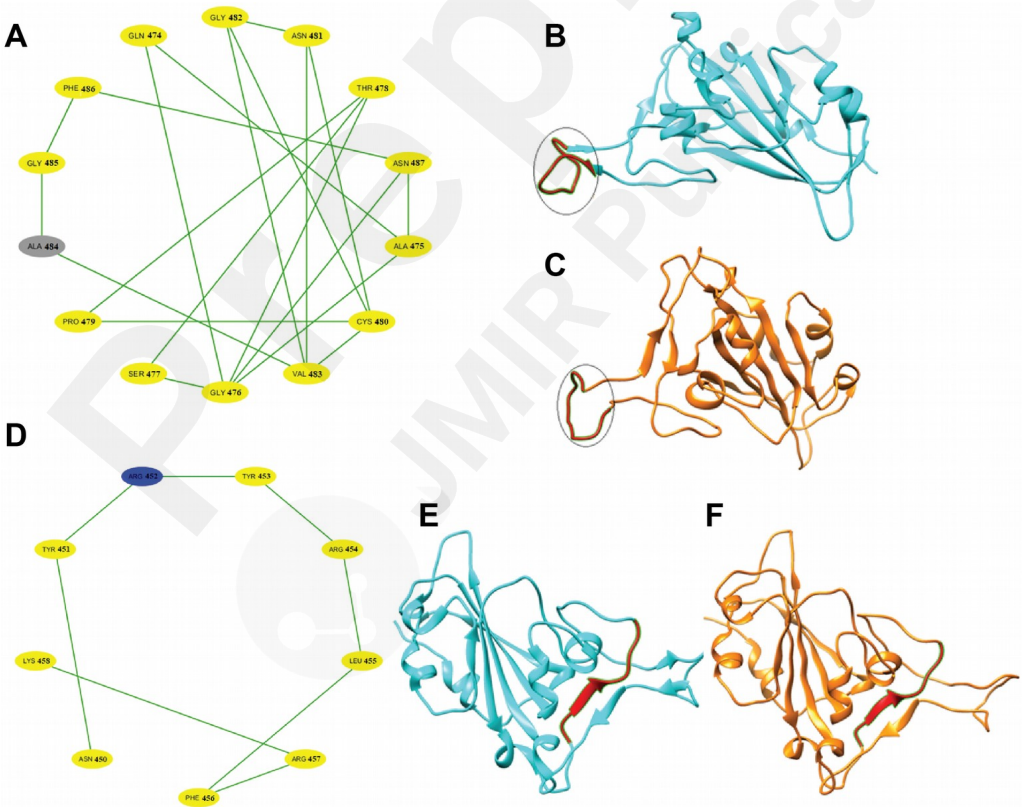


Figure 6: Local and global effects of mutations E484A and L452R on the SARS-CoV-2 Spike RBD (Chain E) compared to wildtype. (A) Mutated subdomain graph G15 illustrating the mild E484A mutation. (B) *Ab initio* model of wildtype SARS-CoV-2 Spike RBD (Chain E). (C) *Ab initio* model of SARS-CoV-2 Spike RBD (Chain E) following the mild E484A mutation. (D) Mutated subdomain graph G13 depicting the severe L452R mutation. (E) *Ab initio* model of wildtype SARS-CoV-2 Spike RBD (Chain E). (F) *Ab initio* model of SARS-CoV-2 Spike RBD (Chain E) after the severe L452R mutation, illustrating significant alterations in protein folding and structural integrity.

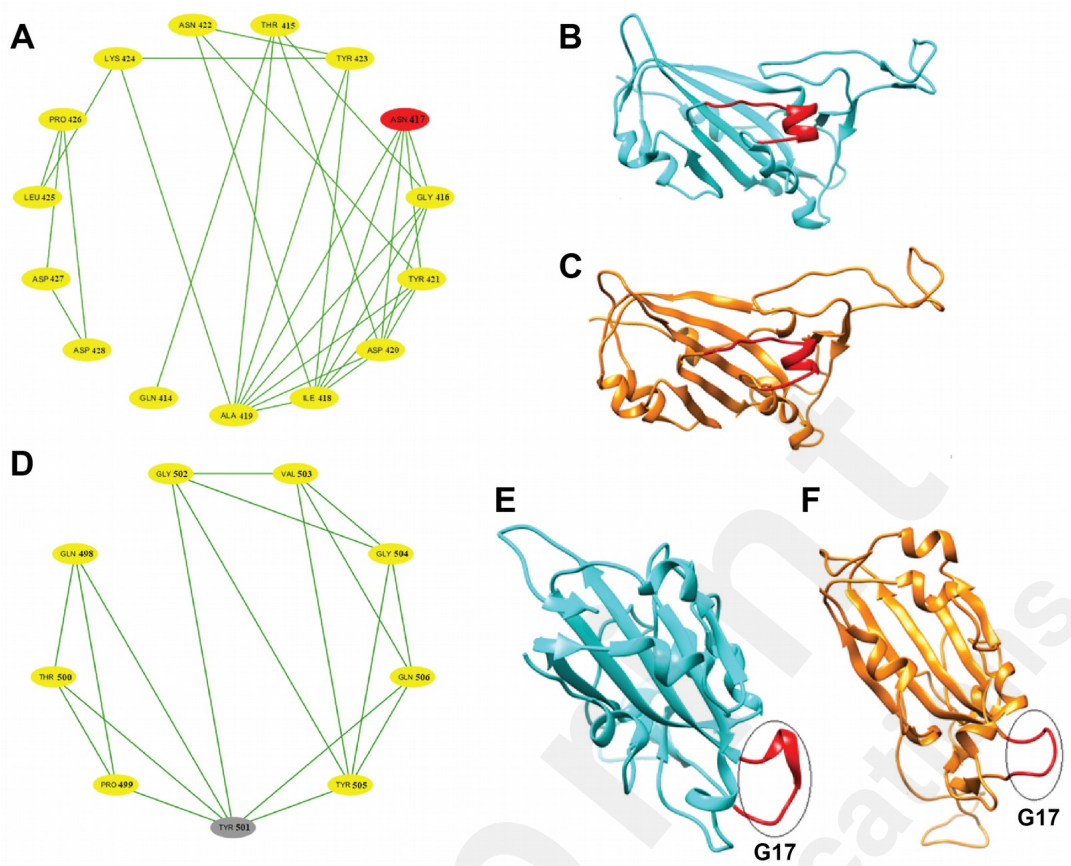


Figure 7: Local and global effects of mutations K417N and N501Y on the SARS-CoV-2 Spike RBD (Chain E) compared to wildtype. (A) Mutated subdomain graph G10 illustrating the mild K417N mutation. (B) *Ab initio* model of wildtype SARS-CoV-2 Spike RBD (Chain E). (C) *Ab initio* model of SARS-CoV-2 Spike RBD (Chain E) following the mild K417N mutation. (D) Mutated subdomain graph G17 depicting the severe N501Y mutation. (E) *Ab initio* model of wildtype SARS-CoV-2 Spike RBD (Chain E). (F) *Ab initio* model of SARS-CoV-2 Spike RBD (Chain E) after the N501Y mutation, illustrating significant alterations in protein folding and structural integrity

.

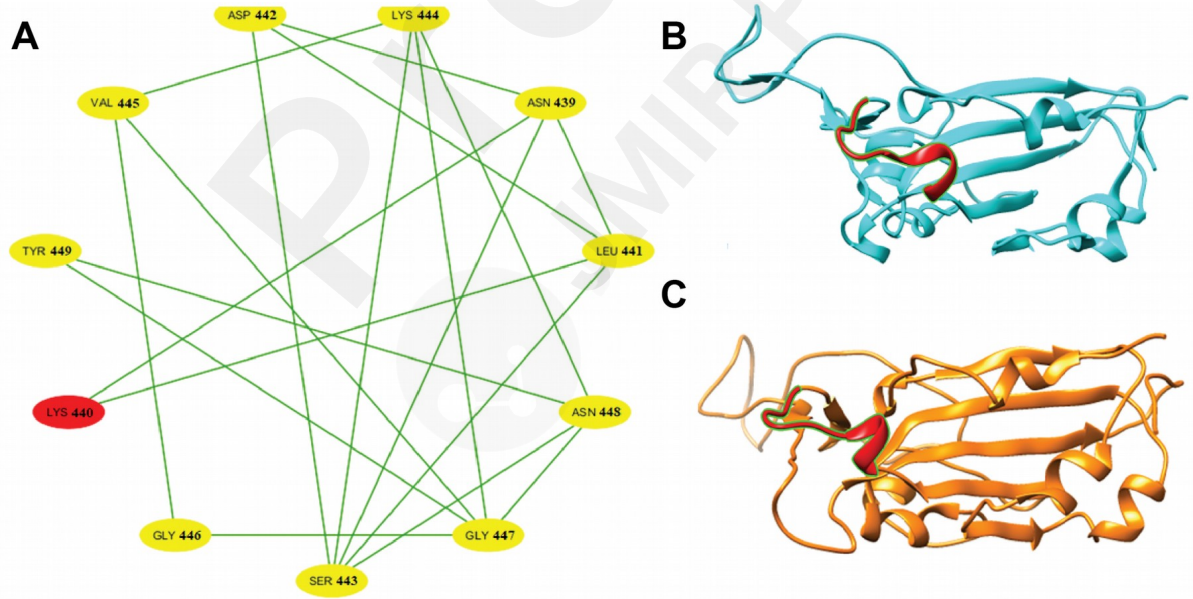


Figure 8: Local and global effects of mutation N440K on the SARS-CoV-2 Spike RBD (Chain E) compared to wildtype. (A) Mutated subdomain graph G12 illustrating the mild N440K mutation. (B) *Ab initio* model of wildtype SARS-CoV-2 Spike RBD (Chain E). (C) *Ab initio* model of SARS-CoV-2 Spike RBD (Chain E) following the mild N440K mutation.

3.3. Molecular Dynamic Simulation Analysis Unveil Differential Structural Stability Between Wildtype and Mutated SARS-Cov-2 Spike Proteins

To complement our graph-theoretic and *ab initio* modeling findings, we conducted MD simulations to investigate the effects of point mutations on COVID-19 variants. These simulations provided insights into the binding interactions and stability of wildtype and mutated SARS-CoV-2 spike proteins in dynamic states. Our approach aligns with recent studies that have employed MD simulations to examine the conformational behavior of SARS-CoV-2 spike protein variants.

We utilized the WebGRO server[111] based on GROMACS[112], to prepare modeled structures in orthorhombic simulation boxes. The proteins were solvated with simple point charge (SPC) water, counterions, and 0.15 M NaCl to mimic physiological conditions. Energy minimization was performed using the steepest descent integrator, followed by equilibration at 300 K and 1.1023 bar. For robustness, we conducted triplicate 50 ns simulations with varied random sampling seed inputs. Both wildtype protein structures from PDB and mutated variants generated via I-TASSER[95, 102] were simulated using the OPLS-AA force field and SPC/E water models. The results of our confirmatory test, which involved molecular simulation and dynamics studies conducted over a 100-nanosecond period, are presented in the Supplementary section, specifically in Figure 14. This extended simulation provides additional validation and insights into the molecular behavior observed in our primary analysis.

Post-simulation analyses focused on RMSD plots to assess trajectory stability and atomic position deviations. Visualization and analysis of trajectories were performed using VMD and BIOVIA Discovery Studio (Dassault Systèmes BIOVIA. Discovery Studio Modeling Environment, Release 2017. Dassault Systèmes; San Diego, CA, USA. 2017). This comprehensive approach enabled detailed comparisons of protein-water interactions and stability across variants, providing valuable insights into the effects of mutations on protein behavior in dynamic environments. As illustrated in Figure 9, RMSD analysis revealed differential structural stability between wildtype and mutated SARS-CoV-2 spike proteins.

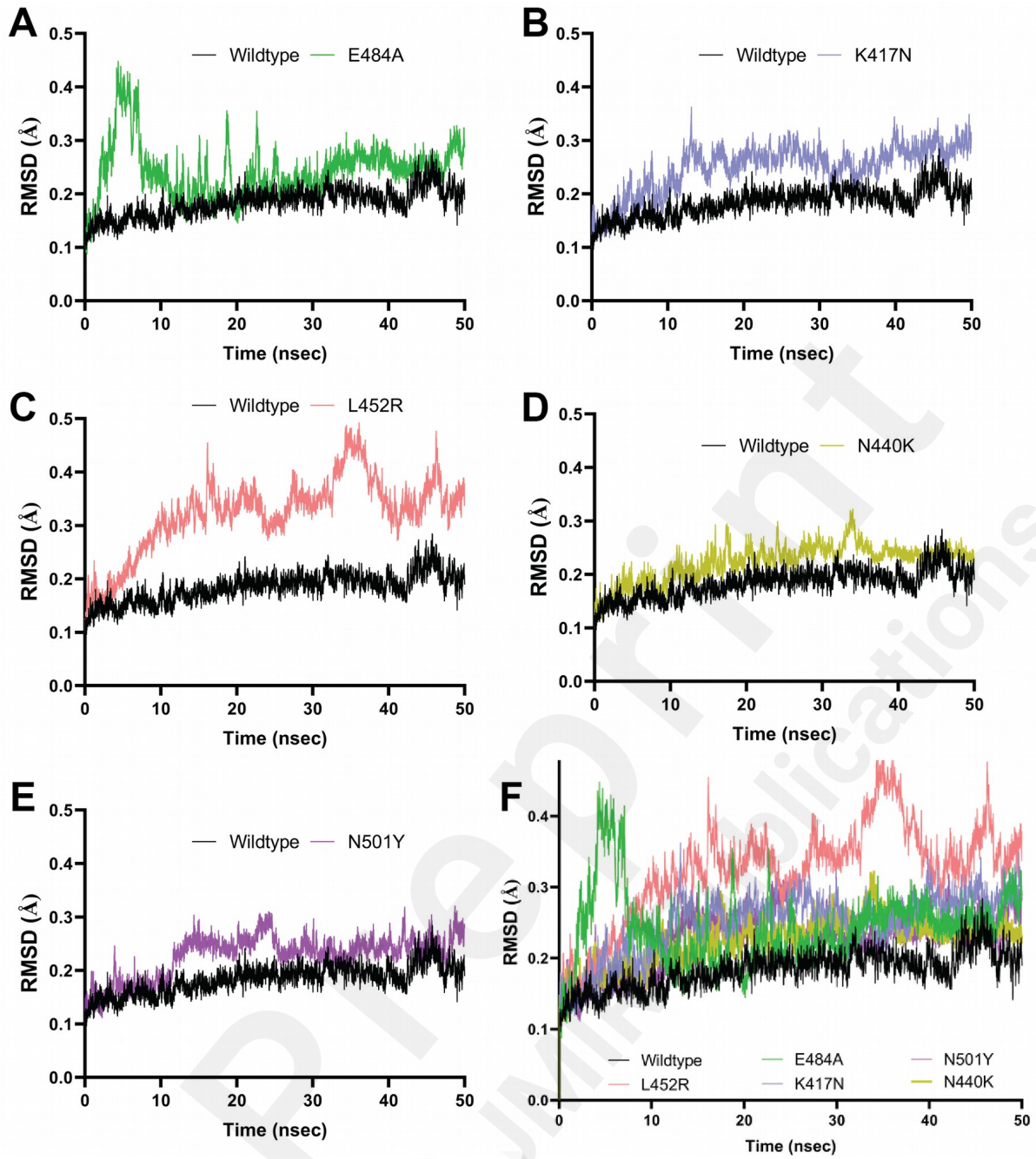


Figure 9: Molecular dynamics simulations reveal differential structural stability of wildtype and mutated SARS-CoV-2 spike proteins. Root Mean Square Deviation (RMSD) analysis was performed on wildtype and mutated spike proteins. The wildtype protein (black) exhibited high stability (RMSD range: 3.00E-07 to 0.284101; mean ± SEM: 0.1838 ± 0.000381). Mutated proteins showed varying degrees of instability: E484A (green; range: 4.00E-07 to 0.4477065; mean ± SEM: 0.2435 ± 0.000727), L452R (pink; mean ± SEM: 0.3224 ± 0.000984), K417N (lavender; range: 5.00E-07 to 0.3620131; mean ± SEM: 0.2492 ± 0.000624), N440K (olive; range: 5.00E-07 to 0.3220472; mean ± SEM: 0.2230 ± 0.00048), and N501Y (indigo; range: 4.00E-07 to 0.320323; mean ± SEM: 0.2279 ± 0.000559). Two-sample Kolmogorov-Smirnov tests confirmed significant conformational deviations for all mutations compared to wildtype (P-value < 2.2e-16, $\alpha = 0.05$).

As evident from Figure 9, the wildtype protein exhibited high stability with an RMSD range of 3.00E-07 to 0.284101 and a mean ± SEM of 0.1838 ± 0.000381. In contrast, mutated proteins demonstrated varying degrees of instability, with E484A showing an RMSD range of 4.00E-07 to 0.4477065 and a mean ± SEM of 0.2435 ± 0.000727, and L452R displaying a mean ± SEM of 0.3224 ± 0.000984. Other mutations, including K417N, N440K, and N501Y, also exhibited increased instability compared to the wildtype as depicted in Figure 9.

Two-sample Kolmogorov-Smirnov tests[114] confirmed significant conformational deviations for all

mutations compared to wildtype ($P < 2.2e-16$, $\alpha = 0.05$, see Figure 9A-F). To further validate our initial findings, we conducted an extended MD simulation for 100 nanoseconds. This confirmatory test revealed significant conformational deviations for all mutant structures compared to the wildtype. The results of these extended MD simulations provide additional evidence supporting the structural impacts of the mutations we investigated. Detailed trajectory analysis, including RMSD plots, are provided in the representative structural snapshots, are presented in Figure 14 in the supplementary section.

Discussion

Principal Results

The dendrogram analysis (Figure 5) highlights significant structural and functional differences between several mutations and the wildtype SARS-CoV-2 spike protein, particularly the N501Y, L452R, E484A, and K147N mutations. These differences, quantified by Euclidean distances from the wildtype, reflect the profound alterations induced by these mutations. The N501Y mutation (Asparagine to Tyrosine at position 501) and L452R mutation (Leucine to Arginine at position 452) exhibit substantial deviations from the wildtype, with Euclidean distances of approximately 0.023 and 0.019, respectively. These mutations significantly impact the Spike RBD, enhancing infectivity and immune evasion. As illustrated in Figure 4, these mutations are distinctly clustered within the dendrogram, highlighting their significant divergence from the wildtype. The notable Euclidean distances (N501Y: 0.023 and L452R: 0.019) indicate substantial structural alterations, which correspond to their clinical relevance in enhancing viral transmission and facilitating immune escape. Specifically, the N501Y mutation increases binding affinity to the human ACE2 receptor, enhancing viral transmissibility and infectivity. It has been linked to variants of concern such as Alpha, Beta, and Gamma and enables infection across a broader range of hosts, including mice. This mutation represents a critical adaptive change in SARS-CoV-2 evolution, underscoring its role in the pandemic [30, 104, 105]. Similarly, the L452R mutation introduces a positively charged arginine in place of leucine within a hydrophobic region of the RBD. This disrupts local hydrophobic interactions and destabilizes the protein structure, contributing to immune evasion and enhanced infectivity. The dendrogram clustering (Figure 4) underscores its significant impact, consistent with clinical observations of resistance to neutralizing antibodies and increased viral transmission [106, 107].

The E484A mutation (Glutamic Acid to Alanine at position 484) and the K147N mutation (Lysine to Asparagine at position 147) also exhibit deviations from the wildtype, with Euclidean distances of approximately 0.017 and 0.0075, respectively. Although these mutations show less pronounced structural alterations compared to N501Y and L452R, as indicated by their Euclidean distances in Figure 5, they still represent significant changes in protein structure and function, as illustrated in Figure 5. Research has demonstrated that the E484A mutation impairs antibody recognition, enhancing immune evasion [108]. Similarly, the K147N mutation, located in the N-terminal domain, reduces neutralization by antibodies [109, 110]. Although their effects are milder compared to N501Y and L452R, these findings highlight that even seemingly minor mutations can induce important structural changes with functional consequences.

Our study utilized I-TASSER [93] models and Cytoscape visualizations to investigate structural changes in mutated Receptor Binding Domains (RBDs) of the SARS-CoV-2 Spike protein. The analysis revealed significant alterations in the mutated RBDs compared to the wildtype, both at the local mutation sites and in the overall RBD conformation. These findings suggest potential long-range effects on protein dynamics. Our visualizations highlighted localized changes in protein folding near mutation sites, and potential impacts on the overall stability and flexibility of the RBD.

These structural changes provide a molecular basis for understanding the observed impacts on binding affinity and dynamics of the Spike protein. A prime example of these effects is the N501Y mutation in the SARS-CoV-2 Spike RBD, known to enhance ACE2 receptor binding. Our study showed that this mutation induces significant local and regional changes in the ab initio model of Chain E. Specifically, N501Y leads to substantial alterations in protein folding and notable changes in structural integrity. These modifications are more pronounced when compared to the wild-type structure, indicating that the N501Y mutation substantially impacts the RBD's conformation. Such structural changes suggest that this mutation may have considerable implications for the RBD's function, potentially affecting viral behavior and interactions with host cells.

Additionally, our research, as demonstrated in Figure 6, further illustrates that even small local perturbations can significantly affect the overall structure of the spike protein's RBD. Both L452R and N501Y mutations disrupt protein folding and secondary structures of the SARS-CoV-2 Spike protein RBD, as evident from our results in Figures 6, 7, and 8. Consistent with research findings, the L452R mutation, at a local level, introduces a positively charged residue that alters hydrophobic interactions and stability. On a global scale, it enhances spike protein stability, promotes viral fusion with host membranes, and strengthens ACE2 receptor binding [110]. Similarly, the N501Y mutation is known to form additional hydrogen bonds and π-π interactions with ACE2 locally, while globally shifting the spike protein into an "open" prefusion conformation that facilitates receptor engagement. Collectively, these mutations increase ACE2 binding affinity, enhance viral infectivity and transmissibility, and contribute to immune evasion by reducing neutralizing antibody recognition.

Conclusively, our ab initio models offer a valuable structural framework for interpreting experimental data on these mutations. They also provide hypotheses for future investigations into their functional consequences, such as altered receptor binding or antibody recognition. This research contributes significantly to our understanding of how SARS-CoV-2 mutations affect the virus's structure and function, potentially informing future strategies for treatment and prevention.

To gain further insights into the structural stability and dynamic behavior of both wildtype and mutated spike proteins, we conducted two sets of MD simulations. Initially, we performed preliminary simulations for 50 nanoseconds, with results presented in Figure 9 of the main text, providing a foundational understanding of short-term dynamics and structural changes. To validate and extend these observations, we carried out longer, more comprehensive MD simulations for 100 nanoseconds, with results detailed in Figure 14 of the supplementary section. Both sets of simulations offered valuable additional insights, with the extended 100ns simulations allowing us to observe longer-term conformational changes and stability patterns, thus providing a more robust confirmation of our initial findings. These MD studies complemented our static structural analyses, offering a dynamic perspective on the effects of mutations on the spike protein's behavior over time. RMSD analysis revealed that the wildtype protein maintains a stable structure with minimal fluctuations (range RMSD: 3.00E-07 to 0.284101; mean ± SEM RMSD: 0.1838 ±0.000381). In contrast, mutated proteins exhibited varying degrees of instability. Our results suggested that the E484A mutation showed increased structural flexibility (range RMSD: 4.00E-07to 0.4477065; mean ± SEM RMSD: 0.2435 ± 0.000727), with statistically significant deviations from the wildtype. The L452R mutation caused the greatest instability among analyzed mutations with a mean RMSD of 0.3224 (standard error: 0.000984), reflecting its disruptive effect on protein conformation. The K417N mutation induced moderate instability (range RMSD: 5.00E-07 to 0.3620131; mean ± SEM RMSD: 0.2492±0.000624), while the N440K mutation remains relatively stable (range RMSD: 5.00E-07 to 0.3220472; mean ± SEM RMSD: 0.2230± 0.00048) but still deviates significantly from the wildtype. The N501Y mutation introduced structural changes with a mean ± SEM RMSD of 0.2279 ± 0.000559 and range RMSD: 4.00E-07 to 0.320323, with its global shape remains relatively

altered compared to the wildtype.

The two-sample Kolmogorov-Smirnov statistical test [114] confirmed that both mild mutations (E484A, N440K, K417N) and severe mutations (L452R and N501Y) resulted in significant conformational deviations in terms of their distributions from the wildtype for 50ns (see Figure 9A-F), with all P-values < 2.2e-16 at an alpha level of 0.05. A 100 ns MD simulation of wildtype and mutated SARS-CoV-2 spike proteins revealed significant differences in structural stability (see Figure 14, supplementary section). The wildtype protein exhibited moderate stability with an RMSD range of 3.00E-07 to 0.3198064 and a mean of 0.235059314 ± 0.000494983. Mutated proteins showed varying degrees of stability: E484A (range: 4.00E-07 to 0.4236228; mean ± SEM: 0.310775795 ± 0.00089635), K417N (range: 4.00E-07 to 0.4897364; mean ± SEM: 0.336187979 ± 0.000896928), L452R (range: 5.00E-07 to 0.561134; mean ± SEM: 0.415135078 ± 0.000956638), N440K (range: 1.10E-06 to 0.3744204; mean ± SEM: 0.212346294 ± 0.000861528), and N501Y (range: 3.00E-07 to 0.3636647; mean ± SEM: 0.252887616 ± 0.000484278). E484A, K417N, and L452R mutations displayed higher RMSD values, indicating increased instability compared to the wildtype. Notably, the L452R mutation demonstrated the highest instability among severe mutations. In contrast, N440K mutation showed lower RMSD means, suggesting milder effects on protein stability. These findings indicate that mutations can substantially impact the structural dynamics of the spike protein, potentially affecting its function and interactions with host receptors. The observed changes in structural stability across different mutations correlate with clinical observations, suggesting a mechanistic link between altered protein dynamics and enhanced viral properties[109, 110, 112, 113]. Specifically, these structural changes may contribute to the increased infectivity, immune evasion capabilities, and resistance to therapeutic antibodies observed in variants carrying these mutations. This alignment between MD simulations and clinical data underscores the importance of studying protein structural changes in understanding and predicting the behavior of SARS-CoV-2 variants [109, 110, 112, 113].

Comparison with Prior Work

A few studies have applied graph-theoretic models to investigate the effects of point mutations on protein structures and their associated disease phenotypes [86, 87, 112]. This research on SARS-CoV-2 spike protein mutations, particularly N501Y, L452R, E484A, and K147N, provides a detailed computational molecular analysis of the structural and functional alterations induced by these mutations. Specifically, it highlights the impact of mutations on the RBD of the spike protein, quantified using Euclidean distances and MD simulations. For instance, the N501Y and L452R mutations significantly disrupt protein folding, enhance ACE2 binding affinity, and contribute to immune evasion [22, 30, 31, 43]. These findings align with previous studies like Knisley et al.'s graph-theoretic analysis of cystic fibrosis mutations in NBD1 of CFTR proteins, which quantified local and global structural changes caused by mutations [112]. However, unlike prior studies that relied solely on graph-theoretic metrics to model structural perturbations in CFTR proteins [23, 86, 87, 112-114], this research integrated MD simulations to compute RMSD values for mutated spike proteins in comparison to the wildtype. This approach reveals varying degrees of instability among the analyzed mutations, providing insights into mutation-induced conformational dynamics absent in earlier works. Similarly, while Kakraba et al. focused on CFTR mutations in NBD2 [61, 86], and Netsey et al. examined point mutations like Glu6Val in sickle cell hemoglobin [87] this study uniquely addresses structural changes in SARS-CoV-2 spike protein RBD mutations. By combining quantitative approaches with molecular-level dynamics, this research offers additional strengths over previous methodologies that primarily used graph-theoretic models to analyze mutation effects [86, 87, 112]. Together, these complementary studies underscore the diverse applications of computational methods in unraveling mutation-driven phenomena across biological systems. While previous research has addressed distinct biological contexts—such as cystic fibrosis, sickle cell

disease, our current research emphasizes the importance of integrating multifaceted approach such as *ab initio* modeling and MD simulations and to achieve a deeper understanding of mutation-induced structural and functional changes.

Limitations and Future Directions

Despite these valuable insights from graph-theoretic analysis and MD simulations, certain limitations remain with this current study. Specifically, our graph-based model struggled to fully capture the impact of mutations like N440K—associated with increased infectivity and immune evasion—as it may require more comprehensive molecular descriptors [112, 113]. To address this concern, future research should aim to refine these models for better representation of such mutations. Additionally, other mutations like S477N and T478K might be included in future studies. Also, future studies can use our graph-theoretic modeling approach to predict point mutations with potentially devastating consequences by analyzing changes in vertex weights and combinatorial descriptors in protein structure graphs, thereby providing a valuable tool for disease surveillance and early intervention strategies.

Conclusions

This study provides a detailed computational analysis of key SARS-CoV-2 spike protein mutations, including N501Y, L452R, E484A, and K147N, and their structural and functional impacts. Using dendrogram clustering, Euclidean distance measurements, and MD simulations, the research highlights how these mutations disrupt protein stability and alter the RBD. Mutations such as N501Y and L452R significantly enhance ACE2 binding affinity, viral transmissibility, and immune evasion, while even milder mutations like E484A and K147N contribute to structural perturbations and reduced antibody recognition. RMSD analysis revealed varying degrees of instability among mutated proteins, with L452R causing the greatest disruption. These findings align with clinical observations of increased infectivity and immune resistance associated with these mutations. While the study underscores the utility of computational models in understanding mutation-driven phenomena, it also highlights areas for future research, such as refining models to better capture the effects of additional mutations like S477N and T478K.

Overall, this research advances our understanding of SARS-CoV-2 evolution and provides critical insights for monitoring viral mutations and developing effective therapeutic strategies.

Acknowledgements

The authors would like to thank the Office of the President, the Provost, and the Dean of the Celia Scott Weatherhead School of Public Health and Tropical Medicine for their outstanding support provided to SK for the article processing fee, which was made possible through start-up funds.

Author Contributions

Samuel Kakraba (S.K.) led the entire study design. Edem Kofi Netsey (E.K.N.) conducted the graph-theoretic modeling under S.K.'s supervision. Joseph Asante Jnr. (J.A.J.) performed molecular simulation and dynamics under S.K.'s supervision. S.K. analyzed and interpreted all results in this study with inputs from J.A.J., Jeffrey G. Shaffer (J.G.S.), and Sudesh K. Srivastav (S.K.S.). The manuscript was written by S.K., E.K.N., Samuel M. Naandam (S.M.N.), Kuukua Egyinba Abraham (K.E.A.), Aayire Clement Yadem (A.C.Y.), Gabriel Owusu (G.O.), J.G.S., Desmond Yemeh (D.Y.), Ellis Owusu-Dabo (E.O.D.), Seydou Doumbia (S.D.), Stephen Manortey (S.M.), Chris E. Morkle (C.E.M.), Steve Szabo (S.S.), Ernest Yankson (E.Y.), Mamadou Sangare (M.S.), and S.K.S. All authors reviewed and approved the final manuscript version.

Conflicts of Interest

The authors declare that there are no conflicts of interest or competing interests that could be perceived to bias this work. This includes, but is not limited to, financial relationships, personal relationships, professional affiliations, or intellectual property interests that might have influenced the research, its interpretation, or its presentation.

Abbreviations

RBD: Receptor Binding Domain
N501Y: Asparagine replaced with Tyrosine at position 501
L452R: Leucine replaced by Arginine at position 452
K417N: Lysine replaced with Asparagine at position 417.
E484A: Glutamic Acid replaced with Alanine at position 484
COVID-19: Coronavirus Disease 2019
SARS-CoV-2: Severe Acute Respiratory Syndrome Coronavirus 2
Graph-theoretic model: A model using graph theory for structural analysis
3CL: 3C-like protease, also known as main protease
N440K: Asparagine replaced with Lysine at position 440
S477N: Serine replaced with Asparagine at position 477
ACE2 receptor: Angiotensin-Converting Enzyme 2 receptor
T478K: Threonine replaced by Lysine at position 478
E484A: Amino acid glutamic acid (E) at position 484 is replaced by alanine (A)
VOCs: Variants of Concern
B.1.1.7: Alpha variant, first identified in the United Kingdom
UK: United Kingdom
P.1: Gamma variant, first identified in Brazil
P2: A lineage of the Gamma variant, identified in Brazil
B.1.351: Beta variant, first identified in South Africa
E484K: Glutamic Acid replaced with Lysine at position 484
PPI: Protein-Protein Interaction
CF: Cystic Fibrosis
SCD: Sickle Cell Disease
Y1219G: Tyrosine replaced by Glycine at position 1219
G1271E: Glycine to Glutamic Acid mutation at position 1271
1A3N: PDB ID for the structure of the hemoglobin protein
E6V: Glutamic Acid is replaced by Valine at position 6mutation
V23I Valine replaced by Isoleucine at position 23
K82N Lysine replaced by Asparagine at position 82
ACE2 Angiotensin-Converting Enzyme 2
6M0J PDB ID for the structure of the SARS-CoV-2 spike protein bound to ACE2
UCSF Chimera A visualization software for interactive molecular visualization
RBM Receptor-Binding Motif
I-TASSER Iterative Threading Assembly Refinement, a method for predicting protein structure and function
CRISPR-Cas9 Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-associated protein 9, a technology for editing genomes
RSMD: Root Mean Square Deviation
SEM: Standard Error of the Mean
CFTR: Cystic Fibrosis Transmembrane Conductance Regulator

CF: Cystic Fibrosis
NBD1: Nucleotide Binding Domain 1
NBD2: Nucleotide Binding Domain 2

References

1. Shrestha, N., et al., *The impact of COVID-19 on globalization.* One Health, 2020. **11**: p. 100180.
2. Msemburi, W., et al., *The WHO estimates of excess mortality associated with the COVID-19 pandemic.* Nature, 2023. **613**(7942): p. 130-137.
3. Cascella, M., et al., *Features, Evaluation, and Treatment of Coronavirus (COVID-19)*, in *StatPearls*. 2024, StatPearls Publishing

4. Pollard, C.A., M.P. Morran, and A.L. Nestor-Kalinoski, *The COVID-19 pandemic: a global health crisis.* Physiol Genomics, 2020. **52**(11): p. 549-557.
5. Huang, Y., et al., *Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19.* Acta Pharmacologica Sinica, 2020. **41**(9): p. 1141-1149.
6. Ayyadevara, S., et al., *Aspirin-Mediated Acetylation Protects Against Multiple Neurodegenerative Pathologies by Impeding Protein Aggregation.* Antioxid Redox Signal, 2017. **27**(17): p. 1383-1396.
7. Kakraba, S., et al., *A Novel Microtubule-Binding Drug Attenuates and Reverses Protein Aggregation in Animal Models of Alzheimer's Disease.* Front Mol Neurosci, 2019. **12**: p. 310.
8. Kakraba, S., et al., *Thiadiazolidinone (TDZD) Analogs Inhibit Aggregation-Mediated Pathology in Diverse Neurodegeneration Models, and Extend Life- and Healthspan.* Pharmaceuticals, 2023. **16**(10).
9. Balasubramaniam, M., et al., *Aggregate Interactome Based on Protein Cross-linking Interfaces Predicts Drug Targets to Limit Aggregation in Neurodegenerative Diseases.* iScience, 2019. **20**: p. 248-264.
10. Bowroju, S.K., et al., *Design and Synthesis of Novel Hybrid 8-Hydroxy Quinoline-Indole Derivatives as Inhibitors of Aβ Self-Aggregation and Metal Chelation-Induced Aβ Aggregation.* Molecules, 2020. **25**(16).
11. Shi, Y., et al., *Exploration of the Shared Molecular Mechanisms between COVID-19 and Neurodegenerative Diseases through Bioinformatic Analysis.* International Journal of Molecular Sciences, 2023. **24**(5): p. 4839.
12. Huang, P., et al., *Links between COVID-19 and Parkinson's disease/Alzheimer's disease: reciprocal impacts, medical care strategies and underlying mechanisms.* Transl Neurodegener, 2023. **12**(1): p. 5.
13. Li, C., et al., *COVID-19 and risk of neurodegenerative disorders: A Mendelian randomization study.* Translational Psychiatry, 2022. **12**(1): p. 283.
14. Idrees, D. and V. Kumar, *SARS-CoV-2 spike protein interactions with amyloidogenic proteins: Potential clues to neurodegeneration.* Biochemical and Biophysical Research Communications, 2021. **554**: p. 94-98.
15. Bhardwaj, T., et al., *Amyloidogenic proteins in the SARS-CoV and SARS-CoV-2 proteomes.* Nat Commun, 2023. **14**(1): p. 945.
16. Idrees, D. and V. Kumar, *SARS-CoV-2 spike protein interactions with amyloidogenic proteins: Potential clues to neurodegeneration.* Biochem Biophys Res Commun, 2021. **554**: p. 94-98.

17. Biotechnology Information, N.C., *PubChem Patent Summary for US-2023125667-A1*. https://pubchem.ncbi.nlm.nih.gov/patent/US-2023125667-A1.
18. Eberle, R.J., et al., *Tau protein aggregation associated with SARS-CoV-2 main protease.* PLoS One, 2023. **18**(8): p. e0288138.
19. Baranova, A., H. Cao, and F. Zhang, *Causal effect of COVID-19 on Alzheimer's disease: A Mendelian randomization study.* J Med Virol, 2023. **95**(1): p. e28107.
20. Rai, S.N., et al., *Exploring the Paradox of COVID-19 in Neurological Complications with Emphasis on Parkinson's and Alzheimer's Disease.* Oxid Med Cell Longev, 2022. **2022**: p. 3012778.
21. Baazaoui, N. and K. Iqbal, *COVID-19 and Neurodegenerative Diseases: Prion-Like Spread and Long-Term Consequences.* Journal of Alzheimer's Disease, 2022. **88**(2): p. 399-416.
22. Andre, M., et al., *From Alpha to Omicron: How Different Variants of Concern of the SARS-Coronavirus-2 Impacted the World.* Biology (Basel), 2023. **12**(9).
23. Thakur, S., et al., *SARS-CoV-2 Mutations and Their Impact on Diagnostics, Therapeutics and Vaccines.* Front Med (Lausanne), 2022. **9**: p. 815389.
24. Pondé, R.A.A., *Physicochemical effect of the N501Y, E484K/Q, K417N/T, L452R and T478K mutations on the SARS-CoV-2 spike protein RBD and its influence on agent fitness and on attributes developed by emerging variants of concern.* Virology, 2022. **572**: p. 44-54.
25. Willett, J.D.S., et al., *SARS-CoV-2 rapidly evolves lineage-specific phenotypic differences when passaged repeatedly in immune-naïve mice.* Communications Biology, 2024. **7**(1): p. 191.
26. Hattab, D., et al., *SARS-CoV-2 journey: from alpha variant to omicron and its sub-variants.* Infection, 2024. **52**(3): p. 767-786.
27. Huo, J., et al., *A delicate balance between antibody evasion and ACE2 affinity for Omicron BA.2.75.* Cell Reports, 2023. **42**(1): p. 111903.
28. Chatterjee, S., et al., *A Detailed Overview of SARS-CoV-2 Omicron: Its Sub-Variants, Mutations and Pathophysiology, Clinical Characteristics, Immunological Landscape, Immune Escape, and Therapies.* Viruses, 2023. **15**(1).
29. Yang, W.T., et al., *SARS-CoV-2 E484K Mutation Narrative Review: Epidemiology, Immune Escape, Clinical Implications, and Future Considerations.* Infect Drug Resist, 2022. **15**: p. 373-385.
30. Yang, W.-T., et al., *SARS-CoV-2 E484K Mutation Narrative Review: Epidemiology, Immune Escape, Clinical Implications, and Future Considerations.* Infection and Drug Resistance, 2022. **15**(null): p. 373-385.
31. Tian, F., et al., *N501Y mutation of spike protein in SARS-CoV-2 strengthens its binding to receptor ACE2.* Elife, 2021. **10**.
32. Liu, Y., et al., *The N501Y spike substitution enhances SARS-CoV-2 transmission.* bioRxiv, 2021.
33. Khandia, R., et al., *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.* Environmental Research, 2022. **209**: p. 112816.
34. Carabelli, A.M., et al., *SARS-CoV-2 variant biology: immune escape, transmission and fitness.* Nat Rev Microbiol, 2023. **21**(3): p. 162-177.
35. Pickering, B., et al., *Divergent SARS-CoV-2 variant emerges in white-tailed deer with deer-to-human transmission.* Nature Microbiology, 2022. **7**(12): p. 2011-2024.
36. Hatmal, M.M., et al., *Comprehensive Structural and Molecular Comparison of Spike Proteins of SARS-CoV-2, SARS-CoV and MERS-CoV, and Their Interactions with ACE2.* Cells, 2020. **9**(12).
37. Shang, J., et al., *Structural basis of receptor recognition by SARS-CoV-2.* Nature, 2020.

581(7807): p. 221-224.
38. Zhang, Y., R. Yan, and Q. Zhou, *ACE2, B0AT1, and SARS-CoV-2 spike protein: Structural and functional implications.* Current Opinion in Structural Biology, 2022. **74**: p. 102388.
39. Shang, J., et al., *Cell entry mechanisms of SARS-CoV-2.* Proceedings of the National Academy of Sciences, 2020. **117**(21): p. 11727-11734.
40. Wang, Q., et al., *Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2.* Cell, 2020. **181**(4): p. 894-904.e9.
41. Cantón, R., et al., *New variants of SARS-CoV-2.* Rev Esp Quimioter, 2021. **34**(5): p. 419-428.
42. Wang, P., et al., *Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7.* Nature, 2021. **593**(7857): p. 130-135.
43. Law, J.L.M., et al., *SARS-COV-2 recombinant Receptor-Binding-Domain (RBD) induces neutralizing antibodies against variant strains of SARS-CoV-2 and SARS-CoV-1.* Vaccine, 2021. **39**(40): p. 5769-5779.
44. Garcia-Beltran, W.F., et al., *Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity.* Cell, 2021. **184**(9): p. 2372-2383.e9.
45. Khan, M.Z.I., et al., *An overview of viral mutagenesis and the impact on pathogenesis of SARS-CoV-2 variants.* Front Immunol, 2022. **13**: p. 1034444.
46. Huuskonen, S., et al., *The comprehensive SARS-CoV-2 'hijackome' knowledge base.* Cell Discovery, 2024. **10**(1): p. 125.
47. Mostefai, F., et al., *Refining SARS-CoV-2 intra-host variation by leveraging large-scale sequencing data.* NAR Genomics and Bioinformatics, 2024. **6**(4): p. lqae145.
48. Loguercio, S., et al., *Understanding the host-pathogen evolutionary balance through Gaussian process modeling of SARS-CoV-2.* Patterns, 2023. **4**(8): p. 100800.
49. Chen, Y., et al., *Structure analysis of the receptor binding of 2019-nCoV.* Biochemical and Biophysical Research Communications, 2020. **525**(1): p. 135-140.
50. Xia, X., *Domains and Functions of Spike Protein in Sars-Cov-2 in the Context of Vaccine Design.* Viruses, 2021. **13**(1).
51. Yang, Z., et al., *Optimizing Parkinson's Disease Prediction: A Comparative Analysis of Data Aggregation Methods Using Multiple Voice Recordings via an Automated Artificial Intelligence Pipeline*, in *Preprints*. 2024, Preprints.
52. Kakraba, S., et al., *AI-enhanced multi-algorithm R shiny app for Predictive Modeling and analytics- A case study of Alzheimer's disease diagnostics (preprint).* JMIR Preprints, 2024.
53. Abraham, K.E., et al., *Parkinson-s-disease-prediction-code-and-data-repository-main*. 2025, Unpublished.
54. Shaffer, J.G., et al., *Basic SMART pred R shiny web application for Machine learning and Deep learning tasks*. 2024, Unpublished.
55. Dash, M., H. Liu, and J. Yao. *Dimensionality reduction of unsupervised data*. in *Proceedings Ninth IEEE International Conference on Tools with Artificial Intelligence*. 1997.
56. Forina, M., C. Armanino, and V. Raggio, *Clustering with dendrograms on interpretation variables.* Analytica Chimica Acta, 2002. **454**(1): p. 13-19.
57. Negi, S.S., C.H. Schein, and W. Braun, *Regional and temporal coordinated mutation patterns in SARS-CoV-2 spike protein revealed by a clustering and network analysis.* Scientific Reports, 2022. **12**(1): p. 1128.
58. Huang, Q., et al., *A New Way to Trace SARS-CoV-2 Variants Through Weighted Network Analysis of Frequency Trajectories of Mutations.* Front Microbiol, 2022. **13**: p. 859241.
59. Sinha, A., S. Sangeet, and S. Roy, *Evolution of Sequence and Structure of SARS-CoV-2 Spike Protein: A Dynamic Perspective.* ACS Omega, 2023. **8**(26): p. 23283-23304.
60. Becker, M., et al., *Immune response to SARS-CoV-2 variants of concern in vaccinated*

individuals. Nature Communications, 2021. **12**(1): p. 3109.
61. Kumari, M., et al., *A critical overview of current progress for COVID-19: development of vaccines, antiviral drugs, and therapeutic antibodies.* Journal of Biomedical Science, 2022. **29**(1): p. 68.
62. Kakraba, S. and D. Knisley, *A graph-theoretic model of single point mutations in the cystic fibrosis transmembrane conductance regulator.* JOURNAL OF ADVANCES IN BIOTECHNOLOGY, 2016. **6**(1): p. 780-786.
63. Netsey, E.K., et al., *A Mathematical Graph-Theoretic Model of Single Point Mutations Associated with Sickle Cell Anemia Disease.* JOURNAL OF ADVANCES IN BIOTECHNOLOGY, 2021.
64. Das, S., et al., *A new graph-theoretic approach to determine the similarity of genome sequences based on nucleotide triplets.* Genomics, 2020. **112**(6): p. 4701-4714.
65. Buphamalai, P., et al., *Network analysis reveals rare disease signatures across multiple levels of biological organization.* Nature Communications, 2021. **12**(1): p. 6306.
66. Galindez, G., et al., *Network-based approaches for modeling disease regulation and progression.* Computational and Structural Biotechnology Journal, 2023. **21**: p. 780-795.
67. Cheng, T.M., et al., *Prediction by graph theoretic measures of structural effects in proteins arising from non-synonymous single nucleotide polymorphisms.* PLoS Comput Biol, 2008. **4**(7): p. e1000135.
68. Zitnik, M., et al., *Current and future directions in network biology.* Bioinformatics Advances, 2024. **4**(1): p. vbae099.
69. Goh, K.-I., et al., *The human disease network.* Proceedings of the National Academy of Sciences, 2007. **104**(21): p. 8685-8690.
70. Nicholson, D.N. and C.S. Greene, *Constructing knowledge graphs and their biomedical applications.* Comput Struct Biotechnol J, 2020. **18**: p. 1414-1428.
71. Valous, N.A., et al., *Graph machine learning for integrated multi-omics analysis.* Br J Cancer, 2024. **131**(2): p. 205-211.
72. Meng, Z., et al., *Heterogeneous biomedical entity representation learning for gene-disease association prediction.* Brief Bioinform, 2024. **25**(5).
73. Ratajczak, F., et al., *Speos: an ensemble graph representation learning framework to predict core gene candidates for complex diseases.* Nature Communications, 2023. **14**(1): p. 7206.
74. Li, H., et al., *CGMega: explainable graph neural network framework with attention mechanisms for cancer gene module dissection.* Nature Communications, 2024. **15**(1): p. 5997.
75. Kwon, J.J., et al., *On knowing a gene: A distributional hypothesis of gene function.* Cell Systems, 2024. **15**(6): p. 488-496.
76. Gualdi, F., B. Oliva, and J. Piñero, *Predicting gene disease associations with knowledge graph embeddings for diseases with curtailed information.* NAR Genomics and Bioinformatics, 2024. **6**(2): p. lqae049.
77. Callahan, T.J., et al., *An open source knowledge graph ecosystem for the life sciences.* Scientific Data, 2024. **11**(1): p. 363.
78. Manipur, I., et al., *Community Detection in Protein-Protein Interaction Networks and Applications.* IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2023. **20**(1): p. 217-237.
79. Zhang, W., et al., *Network-based machine learning and graph theory algorithms for precision oncology.* npj Precision Oncology, 2017. **1**(1): p. 25.
80. Shen, R., *Graph mining and module detection in protein-protein interaction networks*. 2014, State University of New York at Albany: United States -- New York. p. 97.

81. Knisley, D. and J. Knisley, *Predicting protein-protein interactions using graph invariants and a neural network.* Computational biology and chemistry, 2011. **35**(2): p. 108-113.
82. Manzoni, C., P.A. Lewis, and R. Ferrari, *Network Analysis for Complex Neurodegenerative Diseases.* Current Genetic Medicine Reports, 2020. **8**(1): p. 17-25.
83. Razi, A., et al., *Network-Based Enriched Gene Subnetwork Identification: A Game-Theoretic Approach.* Biomed Eng Comput Biol, 2016. **7**(Suppl 2): p. 1-14.
84. Rout, T., et al., *Essential proteins in cancer networks: a graph-based perspective using Dijkstra's algorithm.* Network Modeling Analysis in Health Informatics and Bioinformatics, 2024. **13**(1): p. 42.
85. Wang, J., et al., *Dynamic protein interaction network construction and applications.* Proteomics, 2014. **14**(4-5): p. 338-52.
86. Keskin, O., N. Tuncbag, and A. Gursoy, *Predicting Protein–Protein Interactions from the Molecular to the Proteome Level.* Chemical Reviews, 2016. **116**(8): p. 4884-4909.
87. Kakraba, S., *A Hierarchical Graph for Nucleotide Binding Domain 2*. 2015, East Tennessee State University: United States -- Tennessee. p. 61.
88. Netsey, E.K., et al., *A Mathematical Graph-Theoretic Model of Single Point Mutations Associated with Sickle Cell Anemia Disease.* JOURNAL OF ADVANCES IN BIOTECHNOLOGY, 2021. **9**: p. 1-14.
89. Brunk, E. and U. Rothlisberger, *Mixed Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulations of Biological Systems in Ground and Electronically Excited States.* Chemical Reviews, 2015. **115**(12): p. 6217-6263.
90. Salo-Ahen, O.M.H., et al., *Molecular Dynamics Simulations in Drug Discovery and Pharmaceutical Development.* Processes, 2021. **9**(1): p. 71.
91. Tuckerman, M.E., et al., *Ab Initio Molecular Dynamics Simulations.* The Journal of Physical Chemistry, 1996. **100**(31): p. 12878-12887.
92. Lappala, A., *The next revolution in computational simulations: Harnessing AI and quantum computing in molecular dynamics.* Current Opinion in Structural Biology, 2024. **89**: p. 102919.
93. Yao, N., et al., *Applying Classical, Ab Initio, and Machine-Learning Molecular Dynamics Simulations to the Liquid Electrolyte for Rechargeable Batteries.* Chemical Reviews, 2022. **122**(12): p. 10970-11021.
94. Schlick, T., et al., *Biomolecular modeling and simulation: a field coming of age.* Quarterly Reviews of Biophysics, 2011. **44**(2): p. 191-228.
95. Zhang, Y., *I-TASSER server for protein 3D structure prediction.* BMC Bioinformatics, 2008. **9**(1): p. 40.
96. Kakraba, S.,

A Hierarchical Graph for Nucleotide Binding Domain 2, in *Mathematical Sciences*. 2015, East Tennessee State University.

97. Lan, J., et al., *Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor.* Nature, 2020. **581**(7807): p. 215-220.
98. Berman, H.M., et al., *The Protein Data Bank.* Nucleic Acids Research, 2000. **28**(1): p. 235-242.
99. Pettersen, E.F., et al., *UCSF ChimeraX: Structure visualization for researchers, educators, and developers.* Protein Sci, 2021. **30**(1): p. 70-82.
100. Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks.* Genome Res, 2003. **13**(11): p. 2498-504.
101. Sternberg, A. and C. Naujokat, *Structural features of coronavirus SARS-CoV-2 spike protein: Targets for vaccination.* Life Sci, 2020. **257**: p. 118056.
102. Yang, J., et al., *The I-TASSER Suite: protein structure and function prediction.* Nat Methods,

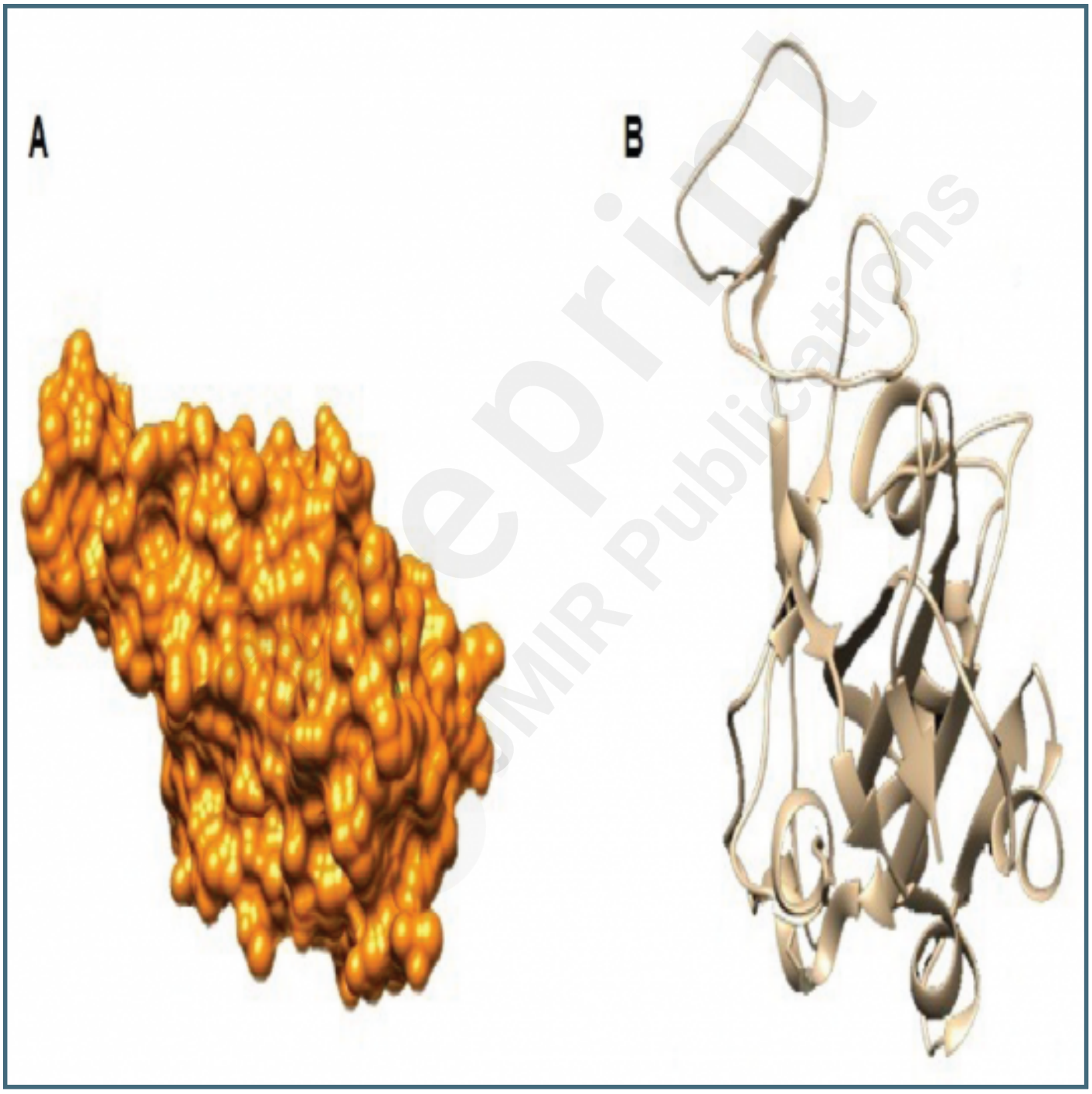
2015. **12**(1): p. 7-8.
103. Sanyal, D., et al., *An exploration of the SARS-CoV-2 spike receptor binding domain (RBD) – a complex palette of evolutionary and structural features.* bioRxiv, 2020.
104. Tian, F., et al., *N501Y mutation of spike protein in SARS-CoV-2 strengthens its binding to receptor ACE2.* eLife, 2021. **10**: p. e69091.
105. Starr, T.N., et al., *Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding.* Cell, 2020. **182**(5): p. 1295-1310.e20.
106. Aleem, A., A.B. Akbar Samad, and S. Vaqar, *Emerging Variants of SARS-CoV-2 and Novel Therapeutics Against Coronavirus (COVID-19)*, in *StatPearls*. 2024, StatPearls Publishing

107. Ahmad, A., M.A.M. Fawaz, and A. Aisha, *A comparative overview of SARS-CoV-2 and its variants of concern.* Infez Med, 2022. **30**(3): p. 328-343.
108. Poudel, S., et al., *Highly mutated SARS-CoV-2 Omicron variant sparks significant concern among global experts - What is known so far?* Travel Med Infect Dis, 2022. **45**: p. 102234.
109. Alba, J.M.G., et al., *Emergence of New SARS-CoV2 Omicron Variants after the Change of Surveillance and Control Strategy.* Microorganisms, 2022. **10**(10).
110. Anand, U., et al., *The spread of the omicron variant: Identification of knowledge gaps, virus diffusion modelling, and future research needs.* Environmental Research, 2023. **225**: p. 115612.
111. *WebGRO for Macromolecular Simulations* Available from: https://simlab.uams.edu/.
112. Abraham, M.J., et al., *GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers.* SoftwareX, 2015. **1-2**: p. 19-25.
113. Humphrey, W., A. Dalke, and K. Schulten, *VMD: Visual molecular dynamics.* Journal of Molecular Graphics, 1996. **14**(1): p. 33-38.
114. Berger, V.W. and Y. Zhou, *Kolmogorov-Smirnov Test: Overview*, in *Wiley StatsRef: Statistics Reference Online*. 2014, John Wiley & Sons, Ltd: Chichester, UK.

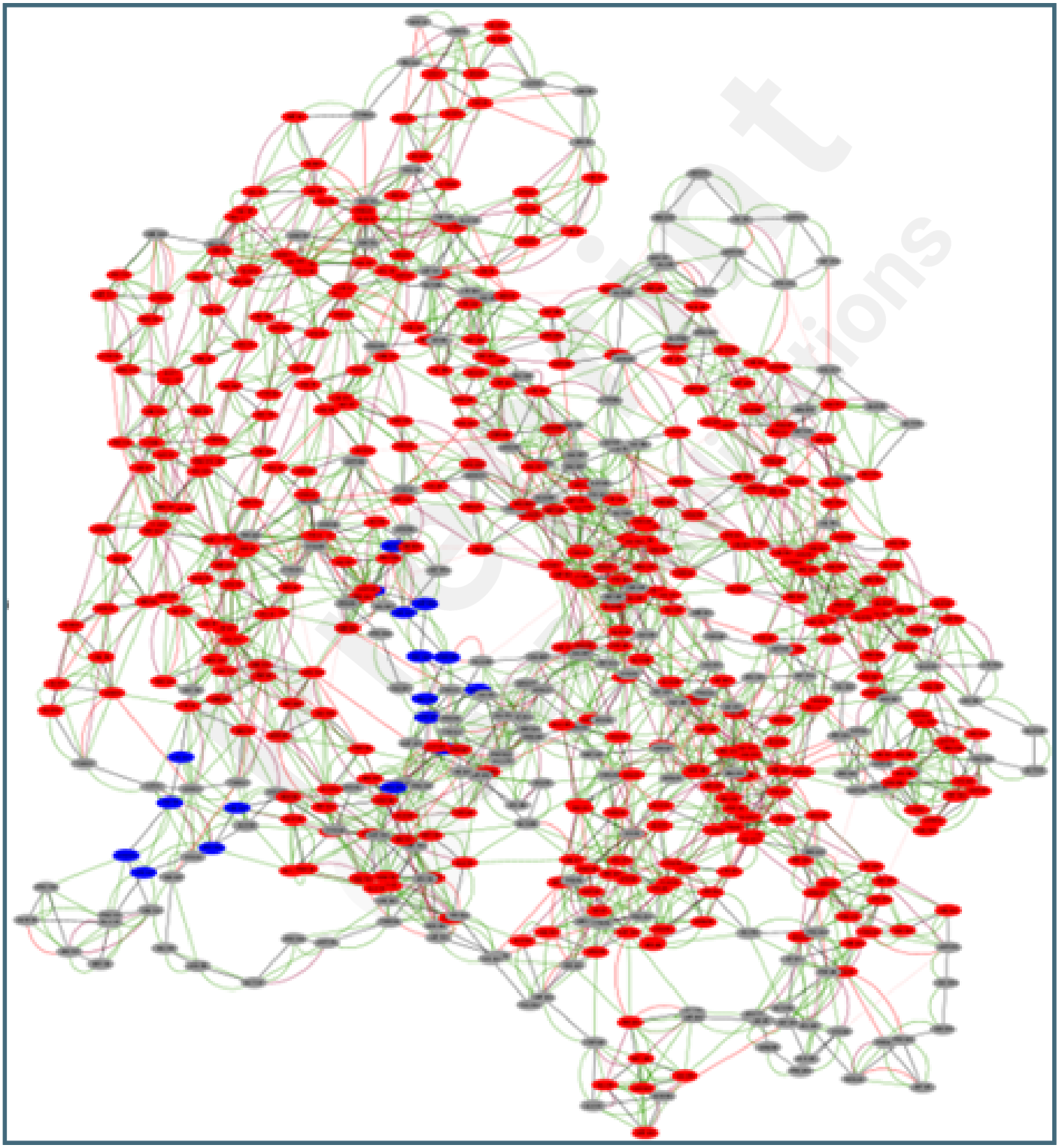
Supplementary Files

Figures

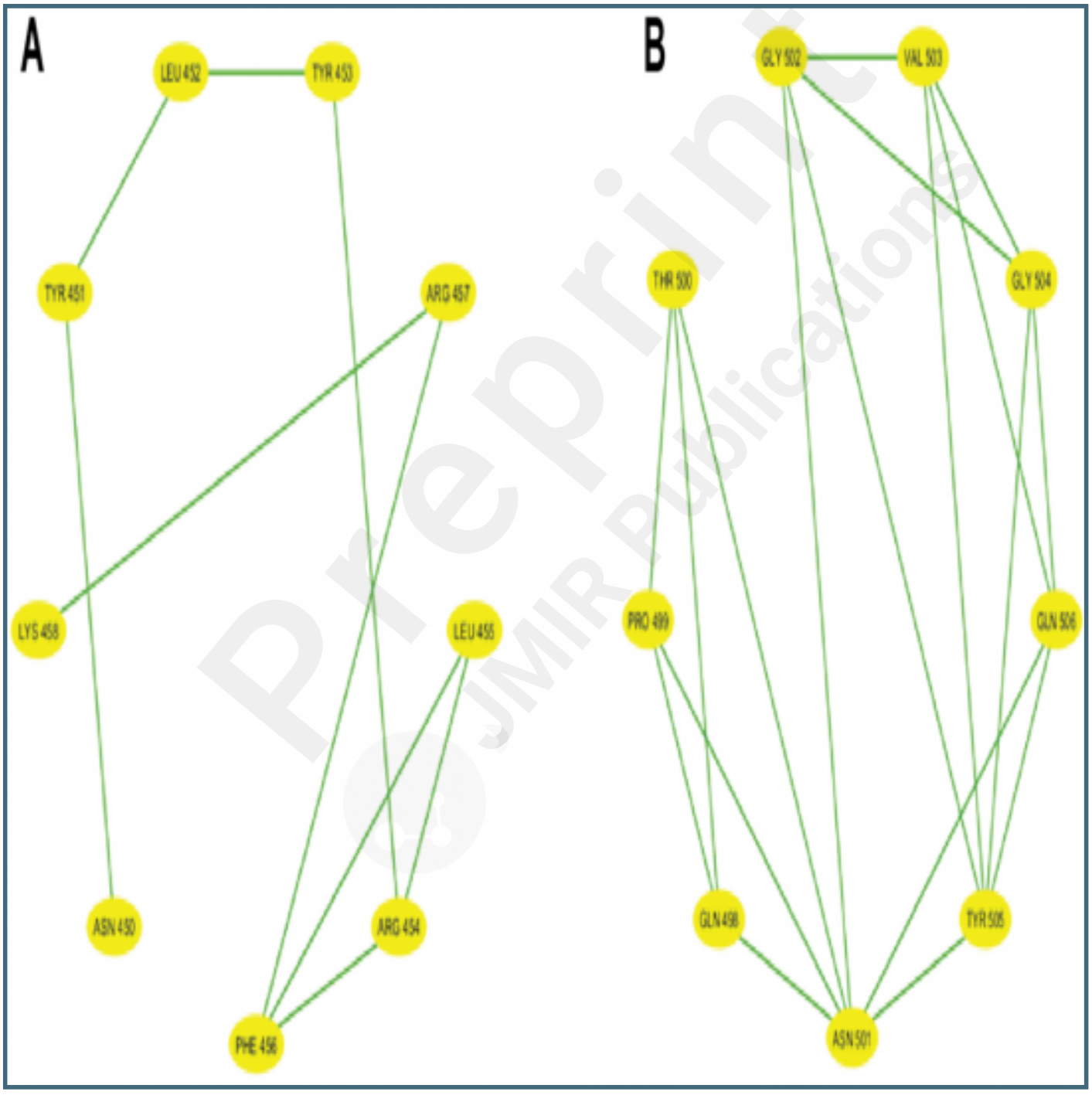
Structure of SARS-CoV-2 spike receptor-binding domain (RBD). A) Quaternary structure of the SARS-CoV-2 spike RBD (Chain E) in complex with ACE2. The RBD forms a concave outer surface that accommodates the N-terminal helix of ACE2, illustrating the key interaction interface for viral entry. The RBD is shown in one color, while ACE2 is depicted in a contrasting color to highlight the binding interface. B) Secondary structure of the SARS-CoV-2 spike RBD. The RBD features a twisted five-stranded antiparallel ? sheet (?1, ?2, ?3, ?4, and ?7) with short connecting helices and loops. Four disulfide bonds are present: three in the core (Cys336–Cys361, Cys379–Cys432, and Cys391–Cys525) stabilizing the ? sheet structure, and one (Cys480–Cys488) connecting loops at the distal end of the receptor-binding motif (RBM).



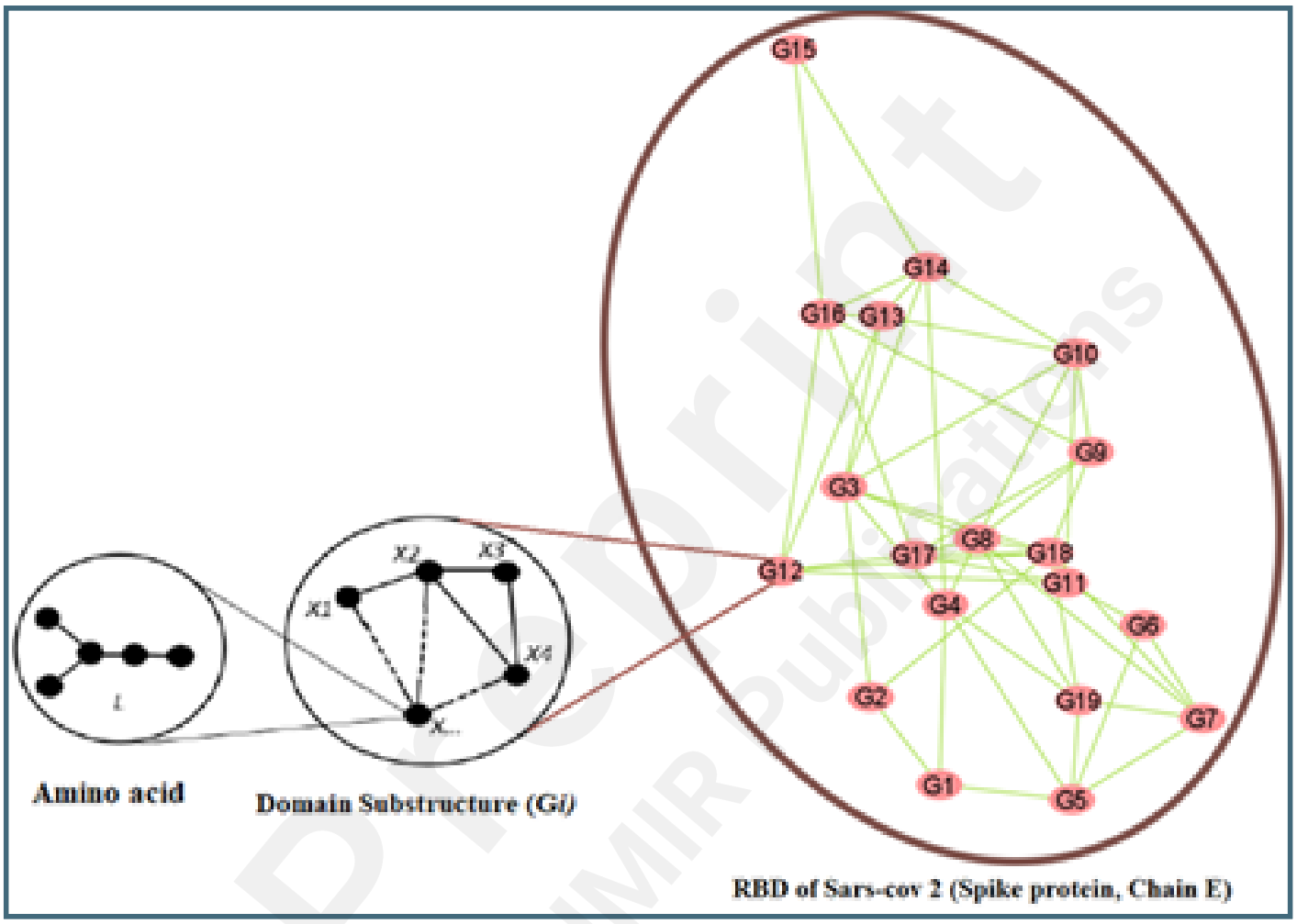
Network Graph of SARS-CoV-2 Spike Receptor-Binding Domain (RBD). This figure illustrates a network graph representation of the SARS-CoV-2 Spike protein's RBD, derived from the crystal structure 6M0J obtained from the Protein Data Bank. Each vertex in the graph represents an individual amino acid residue of the RBD. Edges connecting these vertices were established using a 6-angstrom proximity threshold, where two residues are connected if any of their atoms are within 6Å of each other. This representation provides a comprehensive view of the RBD structure, from individual amino acids to their network interactions.



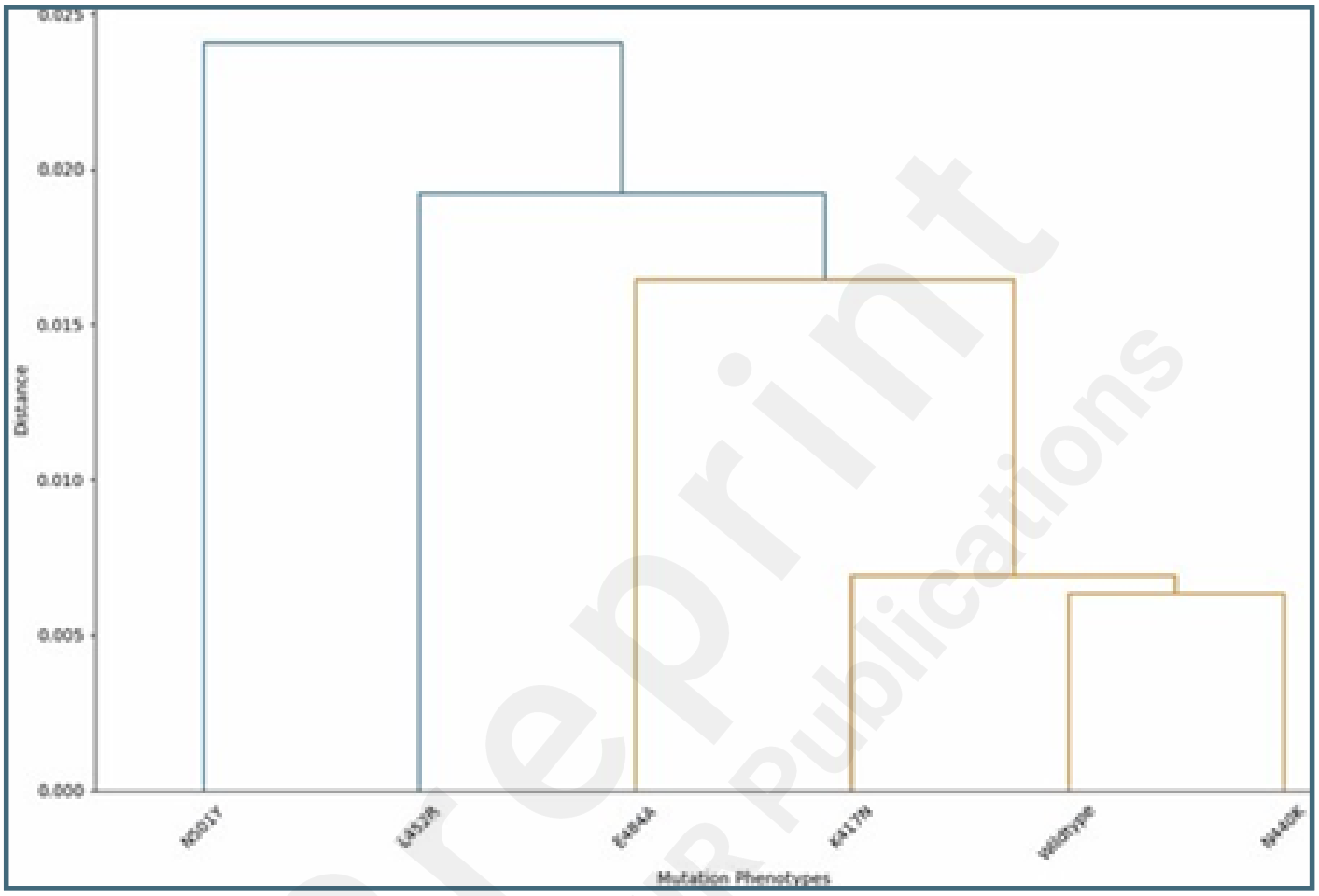
Subdomain graphs of SARS-CoV-2 Spike protein RBD (Chain E) where N501Y and L452R mutations occur. A) Subdomain graph for G13 containing N501Y mutation. B) Subdomain graph for G17 containing L452R mutation. Ab initio models were generated for each subsequence, employing a proximity threshold of 6 angstroms and determining endpoints based on each amino acid residue's center of mass. These graphs illustrate the local structural context surrounding the mutation sites, with nodes representing amino acid residues and edges indicating spatial proximity. The central nodes (N501Y in A and L452R in B) highlight the locations of the severe mutations, while surrounding nodes depict neighboring residues that may influence or be affected by these mutations. This representation aids in visualizing potential structural and functional impacts of these critical mutations on the Spike protein's RBD.



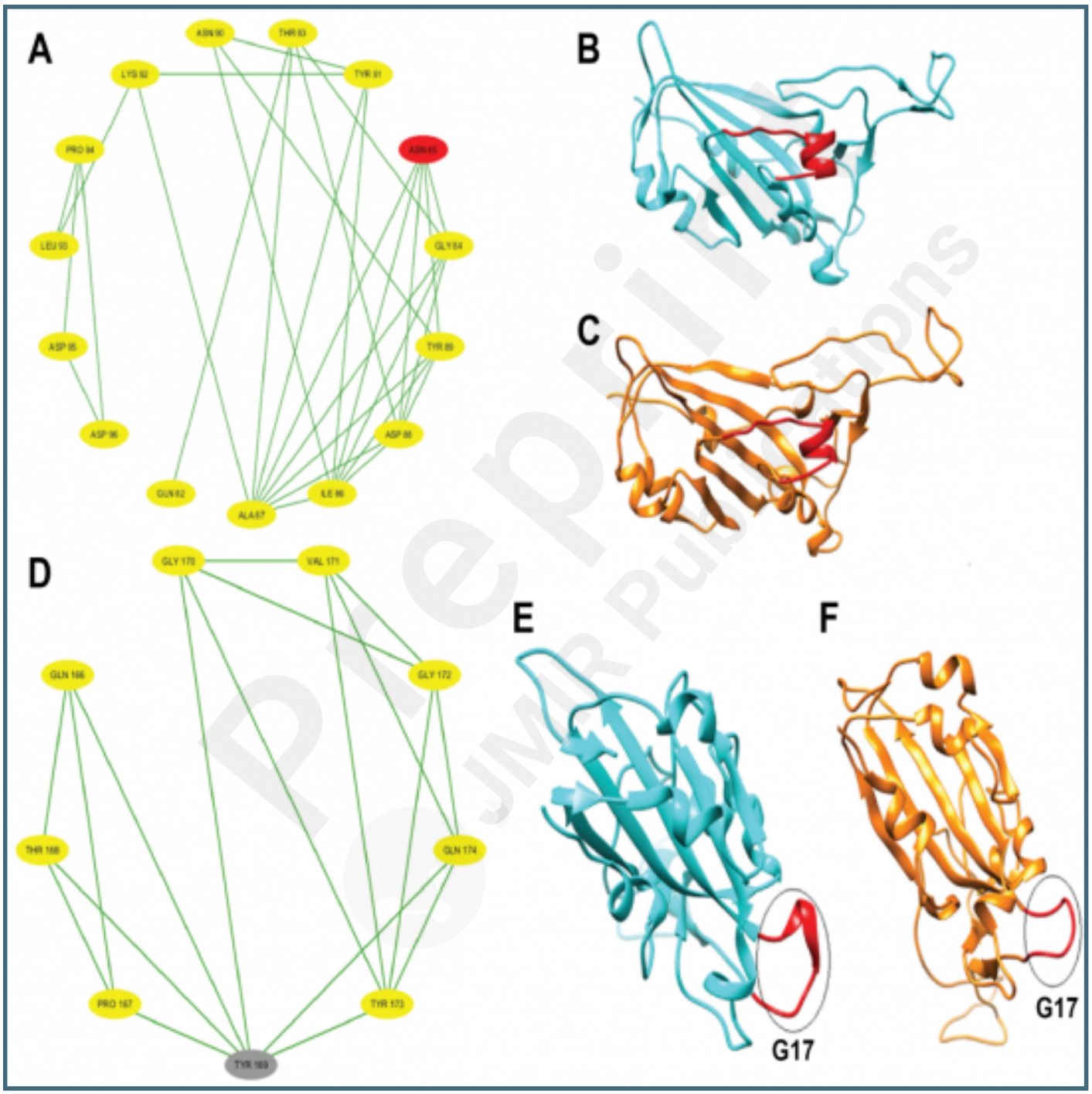
Hierarchical Graph Model of SARS-CoV-2 Spike Receptor-Binding Domain (RBD) Chain E. This figure depicts a three-level hierarchical graph model of the SARS-CoV-2 Spike RBD (Chain E), illustrating the protein's structure and interactions at different scales. The lower level shows 20 vertex-weighted amino acids, the middle level presents 19 vertex-weighted subdomain graphs, and the top level displays a condensed view with single weighted vertices representing each subdomain. Edges between vertices are established using a 6-angstrom proximity threshold, providing a comprehensive view from individual amino acids to subdomain interactions.



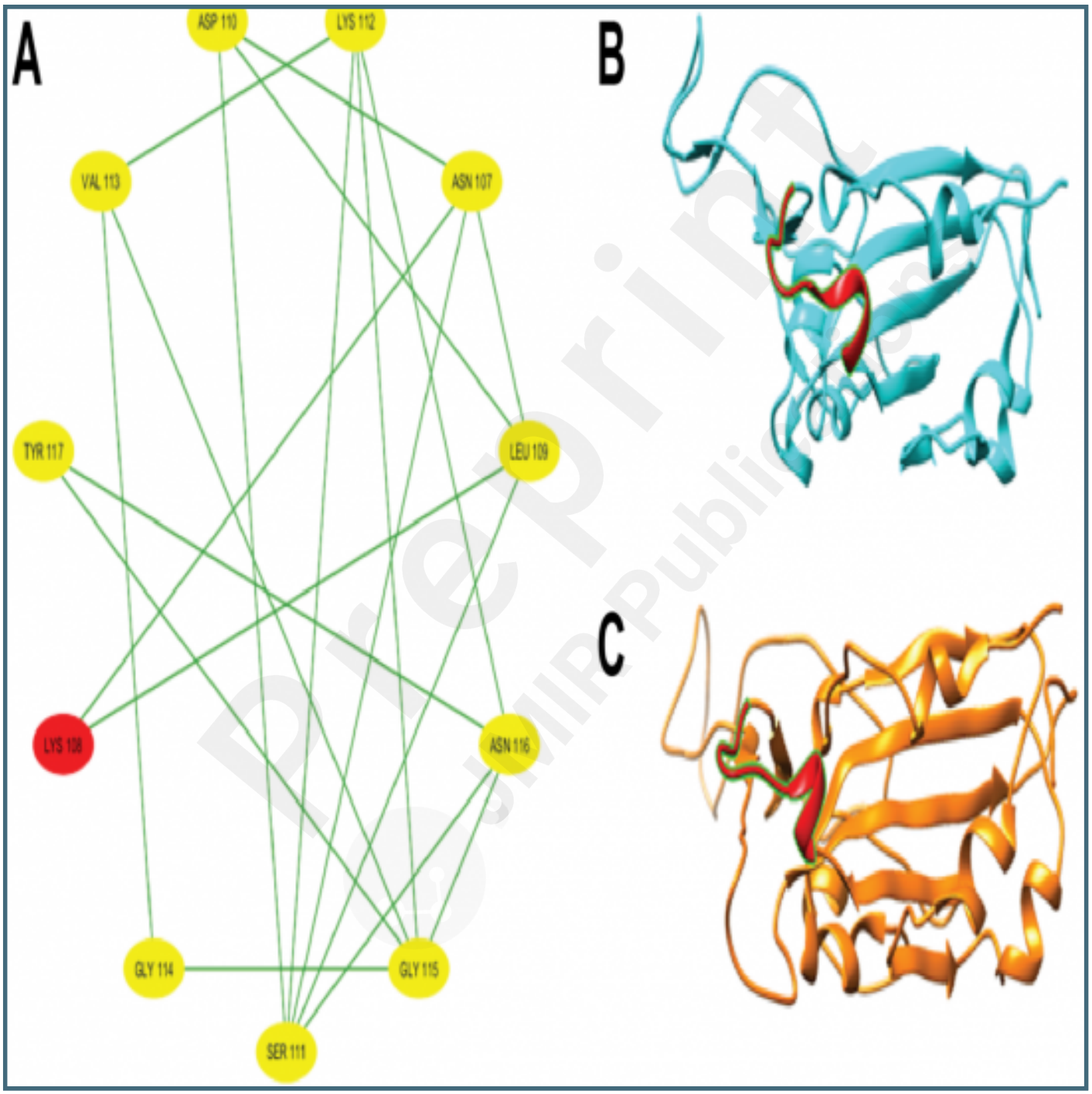
Clustering analysis of COVID-19 mutation-associated point mutations in SARS-CoV-2 RBD. This figure illustrates the hierarchical clustering of point mutations associated with different phenotypes of COVID-19 in the SARS-CoV-2 RBD. The analysis was performed using R statistical software, employing molecular descriptors of the top-level graph (as listed in Table 4) with the built-in functions ‘hclust’ and single linkage method. Euclidean distance was used as the similarity measure for clustering.



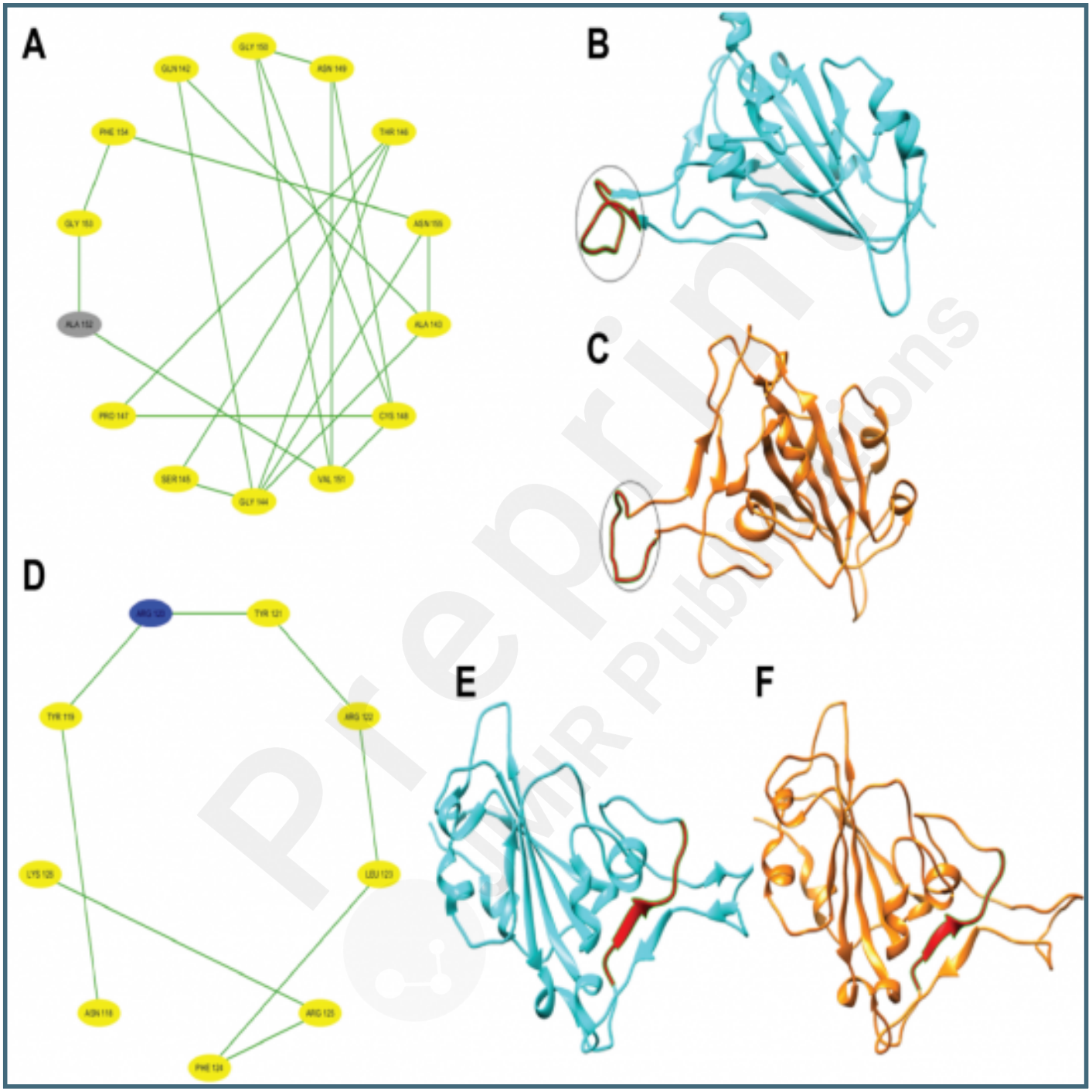
Local and global effects of mutations E484A and L452R on the SARS-CoV-2 Spike RBD (Chain E) compared to wildtype. (A) Mutated subdomain graph G15 illustrating the mild E484A mutation. (B) Ab initio model of wildtype SARS-CoV-2 Spike RBD (Chain E). (C) Ab initio model of SARS-CoV-2 Spike RBD (Chain E) following the mild E484A mutation. (D) Mutated subdomain graph G13 depicting the severe L452R mutation. (E) Ab initio model of wildtype SARS-CoV-2 Spike RBD (Chain E). (F) Ab initio model of SARS-CoV-2 Spike RBD (Chain E) after the severe L452R mutation, illustrating significant alterations in protein folding and structural integrity.



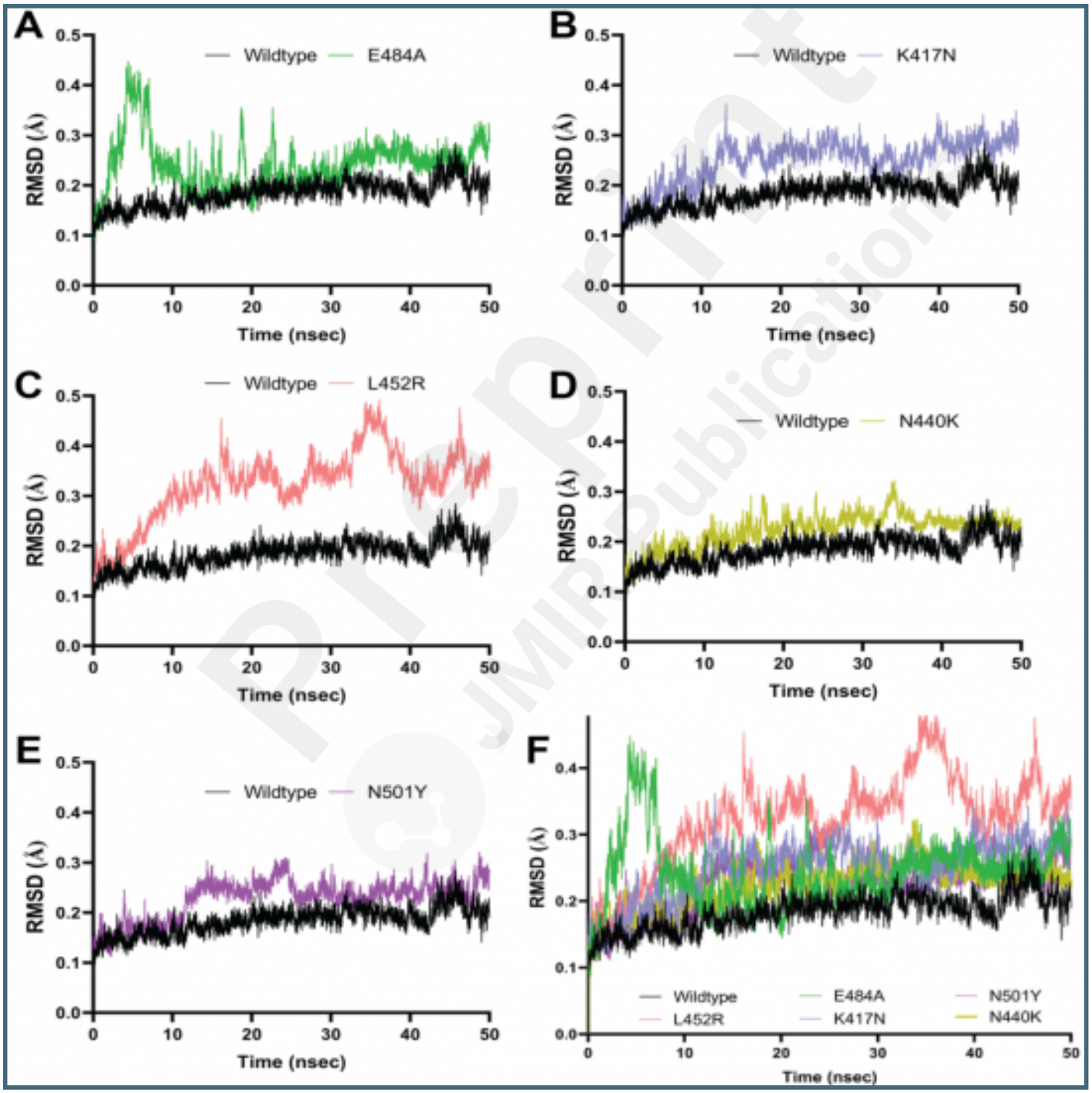
Local and global effects of mutations K417N and N501Y on the SARS-CoV-2 Spike RBD (Chain E) compared to wildtype. (A) Mutated subdomain graph G10 illustrating the mild K417N mutation. (B) Ab initio model of wildtype SARS-CoV-2 Spike RBD (Chain E). (C) Ab initio model of SARS-CoV-2 Spike RBD (Chain E) following the mild K417N mutation. (D) Mutated subdomain graph G17 depicting the severe N501Y mutation. (E) Ab initio model of wildtype SARS-CoV-2 Spike RBD (Chain E). (F) Ab initio model of SARS-CoV-2 Spike RBD (Chain E) after the N501Y mutation, illustrating significant alterations in protein folding and structural integrity.



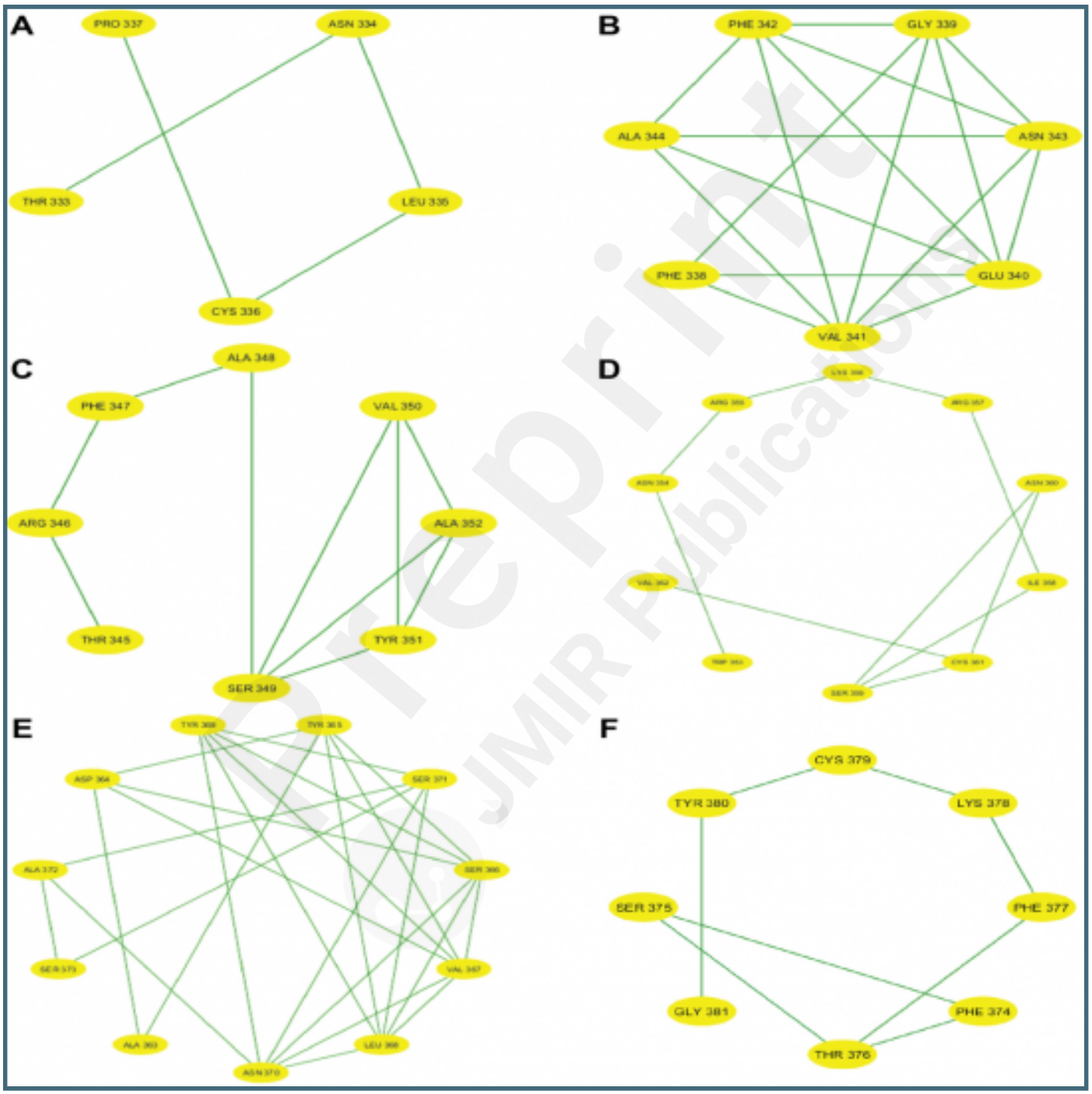
Local and global effects of mutation N440K on the SARS-CoV-2 Spike RBD (Chain E) compared to wildtype. (A) Mutated subdomain graph G12 illustrating the mild N440K mutation. (B) Ab initio model of wildtype SARS-CoV-2 Spike RBD (Chain E). (C) Ab initio model of SARS-CoV-2 Spike RBD (Chain E) following the mild N440K mutation.



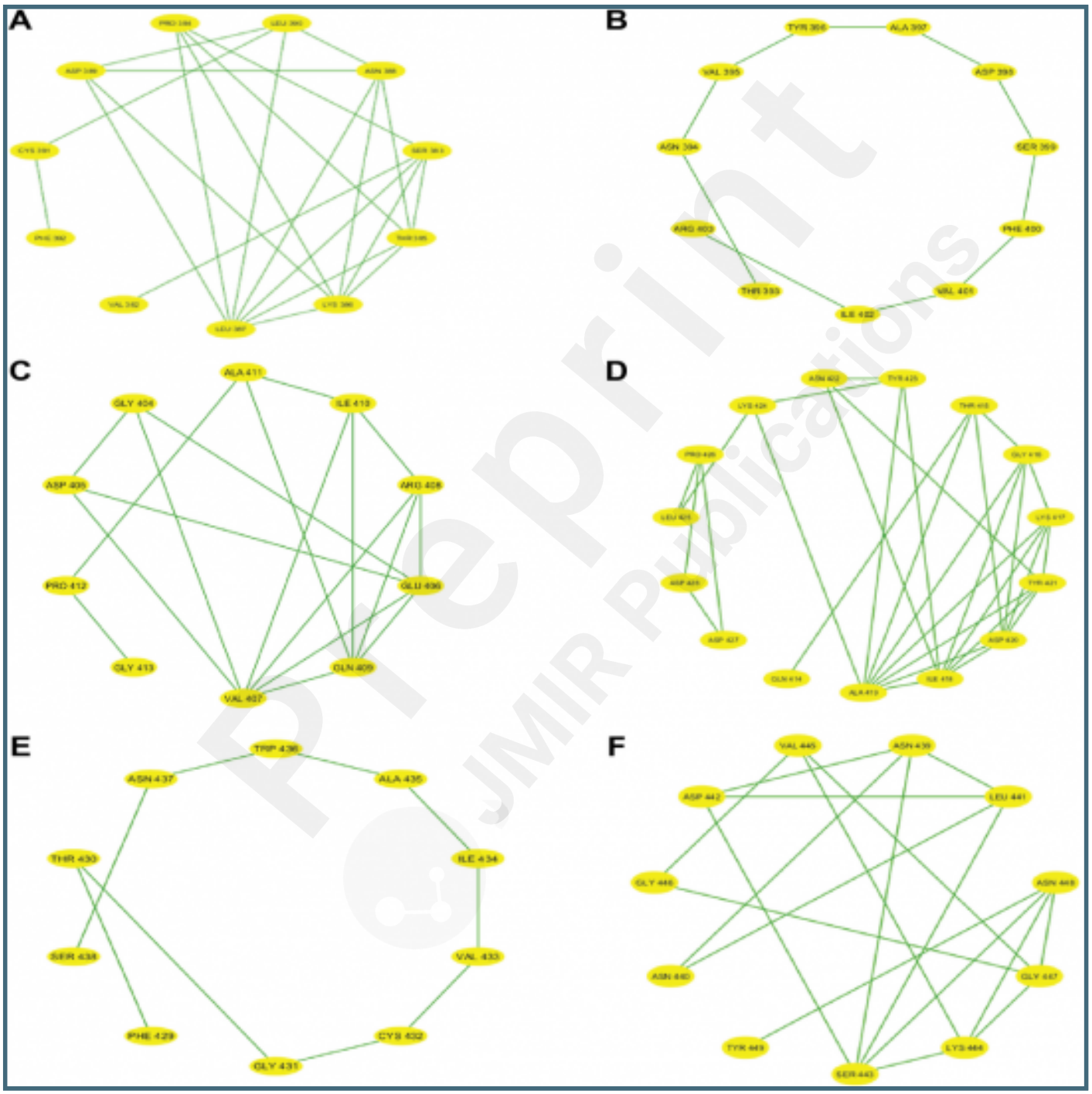
Molecular dynamics simulations reveal differential structural stability of wildtype and mutated SARS-CoV-2 spike proteins. Root Mean Square Deviation (RMSD) analysis was performed on wildtype and mutated spike proteins. The wildtype protein (black) exhibited high stability (RMSD range: 3.00E-07 to 0.284101; mean ± SEM: 0.1838 ± 0.000381). Mutated proteins showed varying degrees of instability: E484A (green; range: 4.00E-07 to 0.4477065; mean ± SEM: 0.2435 ± 0.000727), L452R (pink; mean ± SEM: 0.3224 ± 0.000984), K417N (lavender; range: 5.00E-07 to 0.3620131; mean ± SEM: 0.2492 ± 0.000624), N440K (olive; range: 5.00E-07 to 0.3220472; mean ± SEM: 0.2230 ± 0.00048), and N501Y (indigo; range: 4.00E-07 to 0.320323; mean ± SEM: 0.2279 ± 0.000559). Two-sample Kolmogorov-Smirnov tests confirmed significant conformational deviations for all mutations compared to wildtype (P-value < 2.2e-16, ? = 0.05).



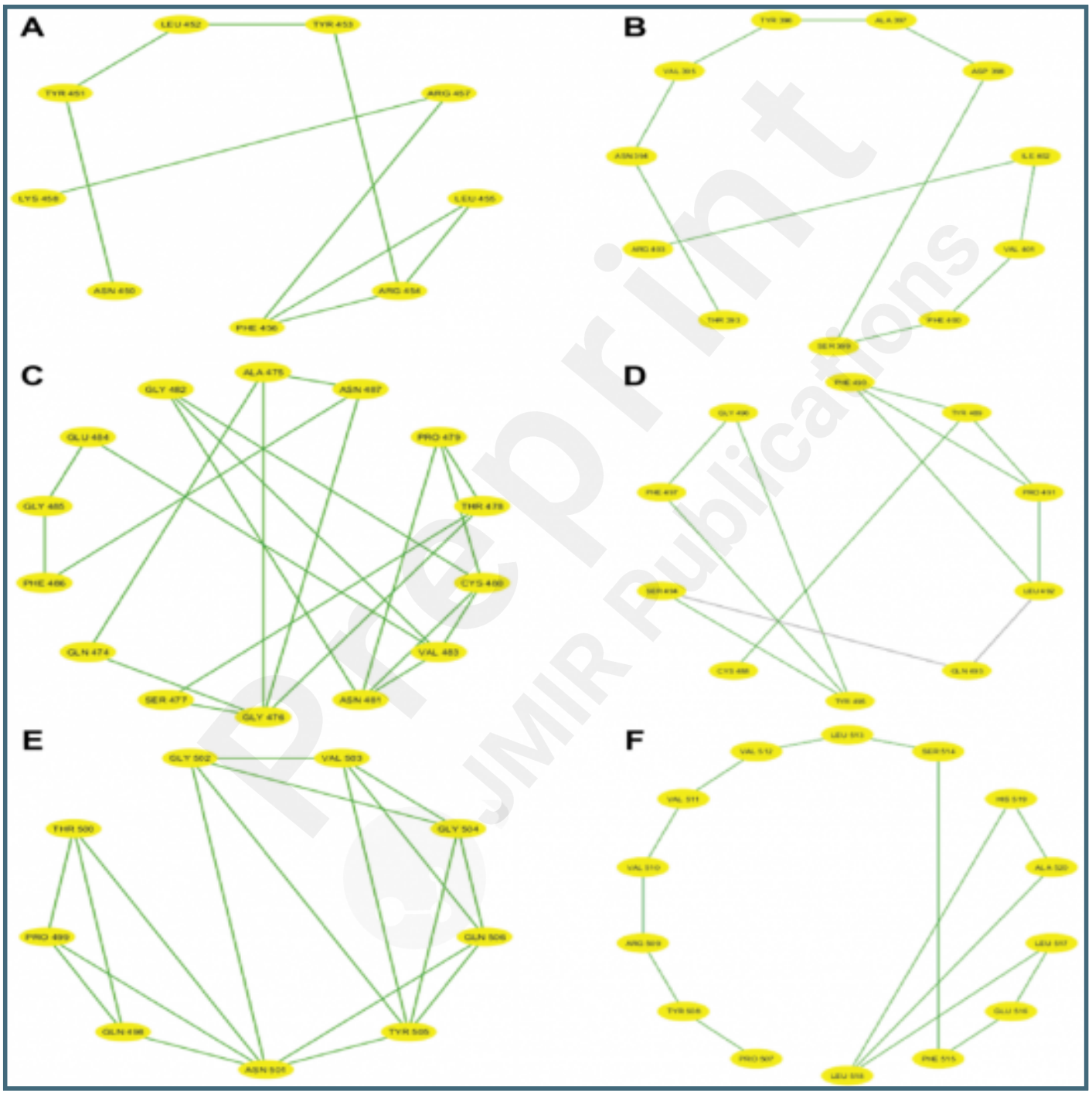
Nested interaction graphs (G1-G6) derived from ab initio modeling of RBD subsequences using I-TASSER. Graphs represent local structural context of amino acids, with nodes as residues and edges indicating spatial proximity (?6Å between residue centers of mass). A) Subdomain graph G1 corresponding to subsequence S1. B) Subdomain graph G2 corresponding to subsequence S2. C) Subdomain graph G3 corresponding to subsequence S3. D) Subdomain graph G4 corresponding to subsequence S4. E) Subdomain graph G5 corresponding to subsequence S5. F) Subdomain graph G6 corresponding to subsequence S6.



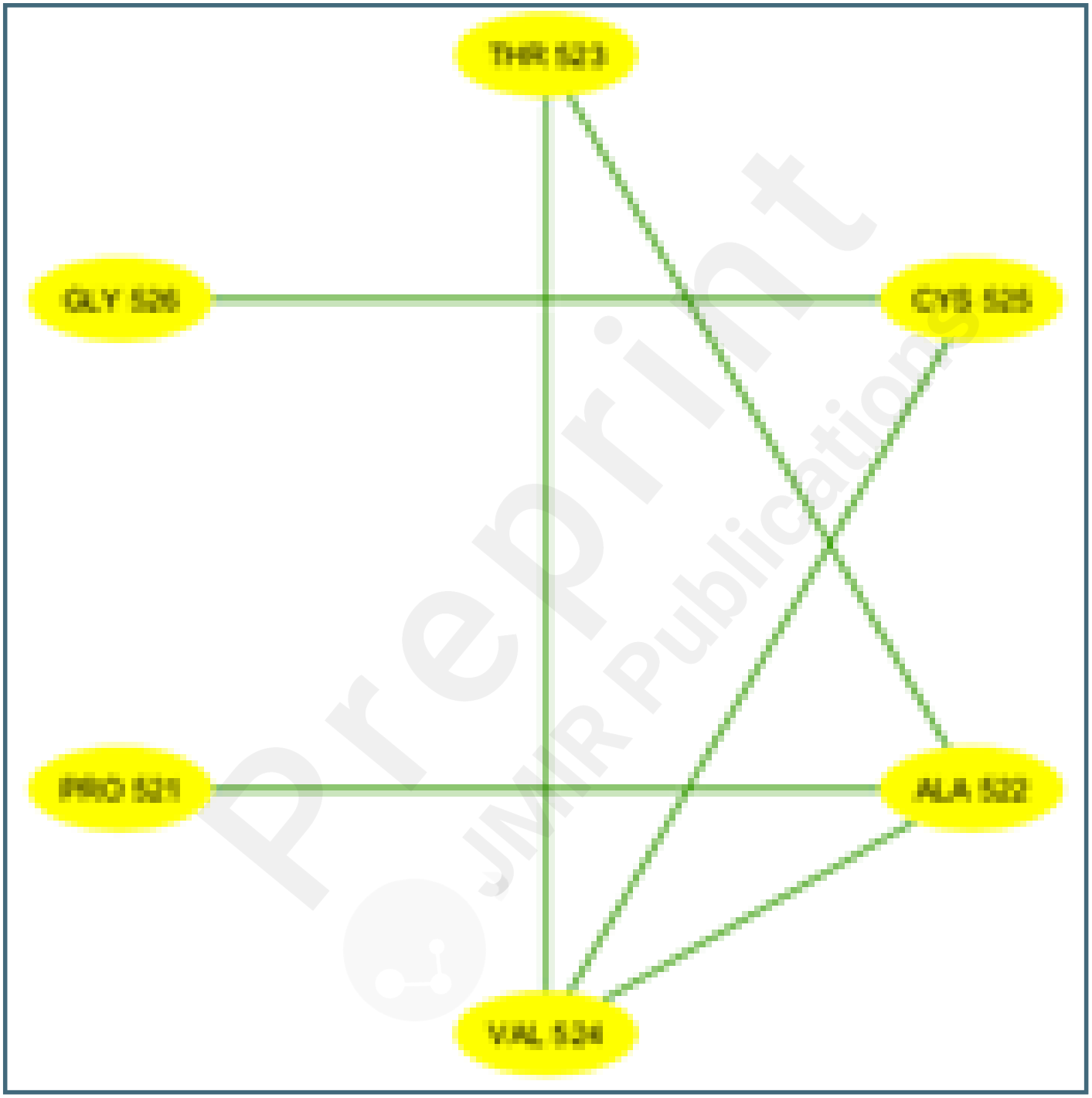
Nested interaction graphs (G7-G12) derived from ab initio modeling of RBD subsequences using I-TASSER. Graphs represent local structural context of amino acids, with nodes as residues and edges indicating spatial proximity (?6Å between residue centers of mass). A) Subdomain graph G7corresponding to subsequence S7. B) Subdomain graph G8 corresponding to subsequence S8. C) Subdomain graph G9 corresponding to subsequence S9. D) Subdomain graph G10 corresponding to subsequence S10. E) Subdomain graph G11 corresponding to subsequence S11. F) Subdomain graph G12 corresponding to subsequence S12.



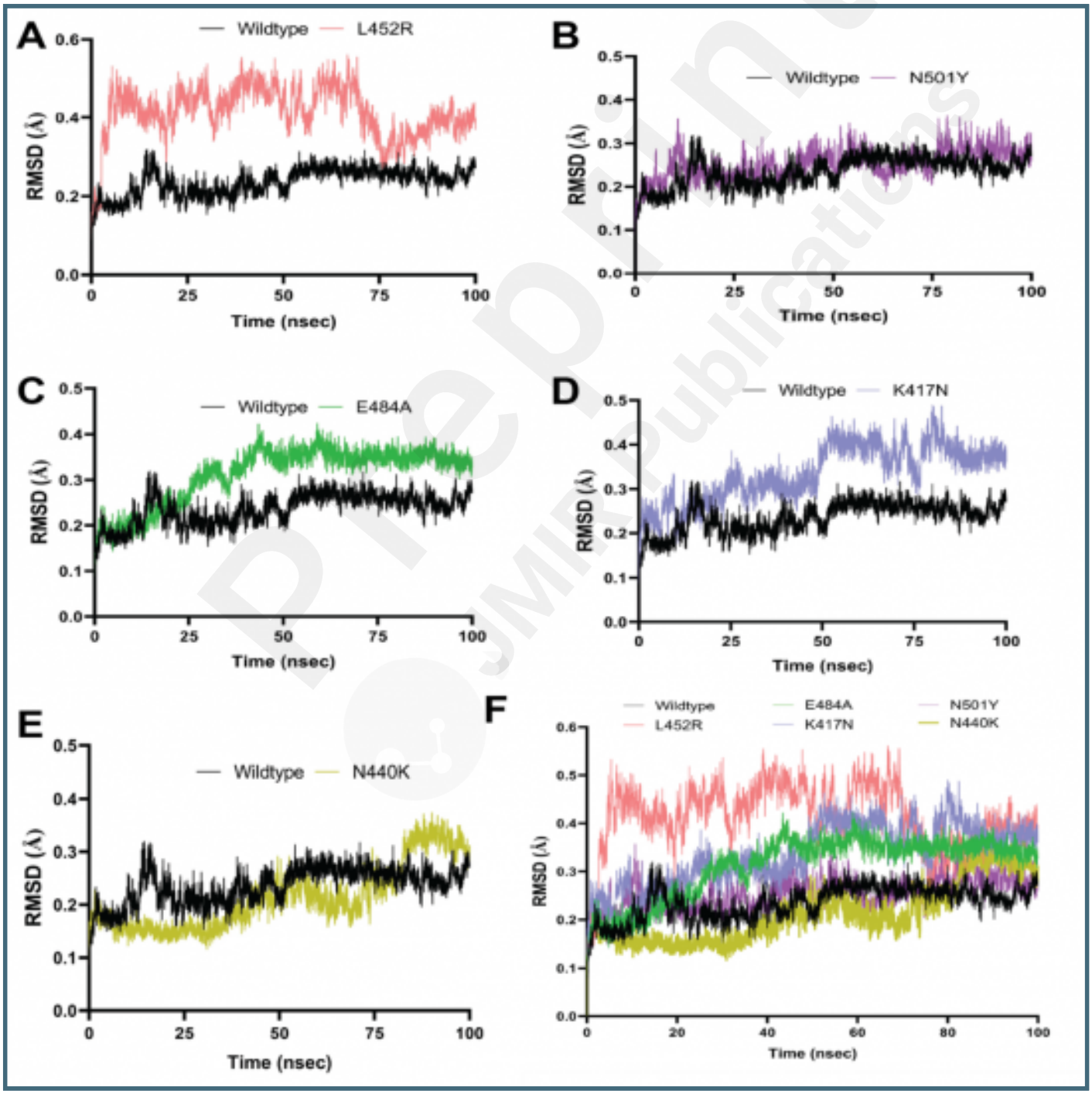
Nested interaction graphs (G13-G18) derived from ab initio modeling of RBD subsequences using I-TASSER. Graphs represent local structural context of amino acids, with nodes as residues and edges indicating spatial proximity (?6Å between residue centers of mass). A) Subdomain graph G13 corresponding to subsequence S13. B) Subdomain graph G14 corresponding to subsequence S14. C) Subdomain graph G15 corresponding to subsequence S15. D) Subdomain graph G16 corresponding to subsequence S16. E) Subdomain graph G17 corresponding to subsequence S17. F) Subdomain graph G18 corresponding to subsequence S18.



Nested interaction graphs(G19) derived from ab initio modeling of RBD subsequences using I-TASSER. Graphs represent local structural context of amino acids, with nodes as residues and edges indicating spatial proximity (?6Å between residue centers of mass). Subdomain graph G19 corresponding to subsequence S19.



MD simulations reveal differential structural stability of wildtype and mutated SARS-CoV-2 spike proteins. Root Mean Square Deviation (RMSD) analysis was performed on wildtype and mutated spike proteins. The wildtype protein exhibited moderate stability (RMSD range: 3.00E-07 to 0.3198064; mean ± SEM: 0.235059314 ± 0.000494983). Mutated proteins showed varying degrees of stability: E484A (range: 4.00E-07 to 0.4236228; mean ± SEM: 0.310775795 ± 0.00089635), K417N (range: 4.00E-07 to 0.4897364; mean ± SEM: 0.336187979 ± 0.000896928), L452R (range: 5.00E-07 to 0.561134; mean ± SEM: 0.415135078 ± 0.000956638), N440K (range: 1.10E-06 to 0.3744204; mean ± SEM: 0.212346294 ± 0.000861528), and N501Y (range: 3.00E-07 to 0.3636647; mean ± SEM: 0.252887616 ± 0.000484278). Two-sample Kolmogorov-Smirnov tests confirmed significant conformational deviations for all mutations compared to wildtype (P-value =0.0000, ? = 0.05), except for N440K (P-value = 1.43E-232) and N501Y (P-value = 1.18E-64). For all severe mutations, the L452R mutation showed the highest instability, while N440K exhibited the lowest RMSD mean compared to other mild mutations.



Multimedia Appendixes

Supplementary Information: Structural Analysis and Molecular Dynamics of SARS-CoV-2 Spike RBD Mutations.
URL: http://asset.jmir.pub/assets/5bed99cc404c135206525404d8c5f62f.docx