



Article

A Protein Co-Conservation Network Model Characterizes Mutation Effects on SARS-CoV-2 Spike Protein

Lianjie Zeng^{1,2}, Yitan Lu³, Wenying Yan^{3,4,*}  and Yang Yang^{1,2,*} 

¹ School of Computer Science & Technology, Soochow University, Suzhou 215000, China

² Collaborative Innovation Center of Novel Software Technology and Industrialization, Nanjing 210000, China

³ Department of Bioinformatics, School of Biology and Basic Medical Sciences, Medical College of Soochow University, Suzhou 215123, China

⁴ Jiangsu Province Engineering Research Center of Precision Diagnostics and Therapeutics Development, Suzhou 215123, China

* Correspondence: wyyan@suda.edu.cn (W.Y.); yyang@suda.edu.cn (Y.Y.)

Abstract: The emergence of numerous variants of SARS-CoV-2 has presented challenges to the global efforts to control the COVID-19 pandemic. The major mutation is in the SARS-CoV-2 viral envelope spike protein that is responsible for virus attachment to the host, and is the main target for host antibodies. It is critically important to study the biological effects of the mutations to understand the mechanisms of how mutations alter viral functions. Here, we propose a protein co-conservation weighted network (PCCN) model only based on the protein sequence to characterize the mutation sites by topological features and to investigate the mutation effects on the spike protein from a network view. First, we found that the mutation sites on the spike protein had significantly larger centrality than the non-mutation sites. Second, the stability changes and binding free energy changes in the mutation sites were positively significantly correlated with their neighbors' degree and the shortest path length separately. The results indicate that our PCCN model provides new insights into mutations on spike proteins and reflects the mutation effects on protein function alternations.

Keywords: COVID-19; conservation; co-conservation; protein co-conservation weighted network; spike protein



Citation: Zeng, L.; Lu, Y.; Yan, W.; Yang, Y. A Protein Co-Conservation Network Model Characterizes Mutation Effects on SARS-CoV-2 Spike Protein. *Int. J. Mol. Sci.* **2023**, *24*, 3255. <https://doi.org/10.3390/ijms24043255>

Academic Editor: Yuriy L. Orlov

Received: 19 January 2023

Revised: 3 February 2023

Accepted: 4 February 2023

Published: 7 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has posed a serious global public-health emergency and was declared as a global pandemic by the World Health Organization (WHO) on 11 March 2020. The infection possesses a clustering onset and targets the respiratory framework of its host, causing influenza-like sickness with symptoms such as cough, fever, and then, in progressively serious cases, troubled breathing [1–3]. COVID-19 has spread rapidly over 200 countries and territories around the world, which has resulted in more than 661 million coronavirus cases and six million deaths (<https://www.worldometers.info/coronavirus/>, accessed on 24 December 2022).

SARS-CoV-2 has been confirmed to enter the host cells by contact between its receptor-binding domain (RBD) of the spike (S) protein and angiotensin converting enzyme 2 (ACE2) of the host cell [4,5]. Although substantial progress in clinical research has led to a better understanding of SARS-CoV-2 and the management of COVID-19, many countries have endured several waves of outbreaks due to the emergence of variants [6]. The S protein displays a high degree of genetic variability in the virus genome [7], and its variants make additional close contacts with ACE2, correlating with higher binding affinity and perhaps increased infectivity [8]. For example, the Omicron variant (B.1.1.529) can evade immunity from natural infection or vaccines and increase the risk of reinfections [9–11]. Acquiring a

comprehensive knowledge of the S protein variants is critical for the determination of a defensive approach against COVID-19.

The biological network approach, which is based on local and global topological features of the network, has been extensively used in various biological fields [12], and has also been applied in COVID-19 related research. Deisy constructed a network medicine framework to identify potentially effective drug candidates [13]. Luisa et al. applied protein contact networks and found that the allosteric modulation region modulates the binding of the spike protein with ACE2 in different conformational states of the spike protein [7,14]. One of the main network-based strategies is to convert a protein into a network to study the structure, function, and dynamics of proteins [15–17]. However, most of the network models for protein function study require the 3D structures of proteins that can be experimentally determined, but the process is still laborious and time-consuming. In particular, the emergence of new variants of COVID-19 brings challenges for structure-based network models if the structures from the variants are not determined.

In this study, we propose a protein co-conservation weighted network (PCCN) model simply based on protein sequence to investigate the mutation effects on the spike protein (Figure 1). We then compared the network topological features of the mutation and non-mutation sites at the residue and variant levels. Finally, we mined the correlation of topological features with the mutation effects on the spike protein stability changes and binding free energy changes.

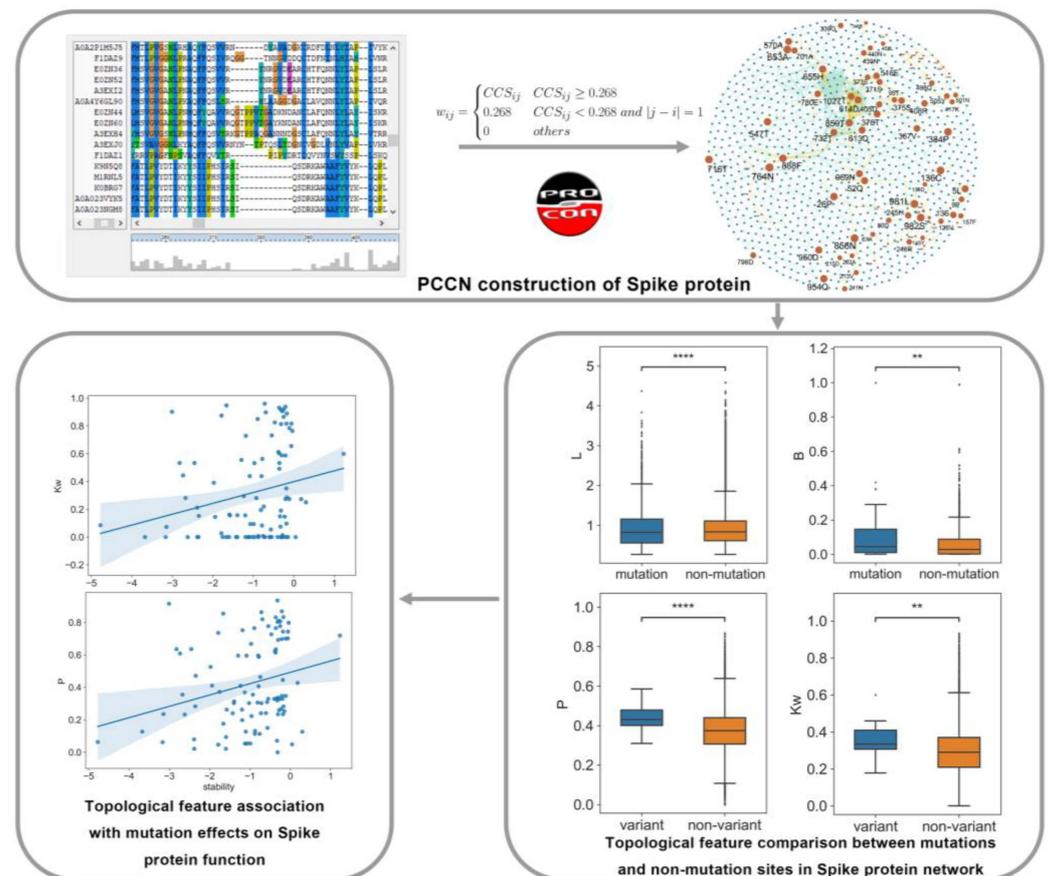


Figure 1. Overview of our method to study the mutation effects on the spike protein. ** $p < 0.01$; **** $p < 0.0001$.

2. Results

2.1. Conservation and Co-Conservation of the Spike Protein

To calculate the conservation of the spike protein, the protein family was initially searched in InterPro [18] for the Spike glycoprotein, beta-coronavirus (IPR042578), and 11,372 related sequences were found. Among these sequences, those within the same species had high similarity, so only one of them was retained, leaving 263 sequences that were aligned to calculate the conservation (CS) and co-conservation (CCS) using ProCon [19,20].

The distribution of CS is shown in Figure 2. The left panel shows the conservation of each residue, which shows a trend of normal distribution. Notably, the scores in the range of 0.9 to 1.0 were particularly high. On the right, the CCS among residues is shown, which followed an overall trend of normal distribution, except that the CCS was small. The results demonstrated that the rate of residues with high conservation residues and low co-conservation was relatively high, which indicated that many of the residues in the spike protein were conserved, but their co-conservation was relatively weak.

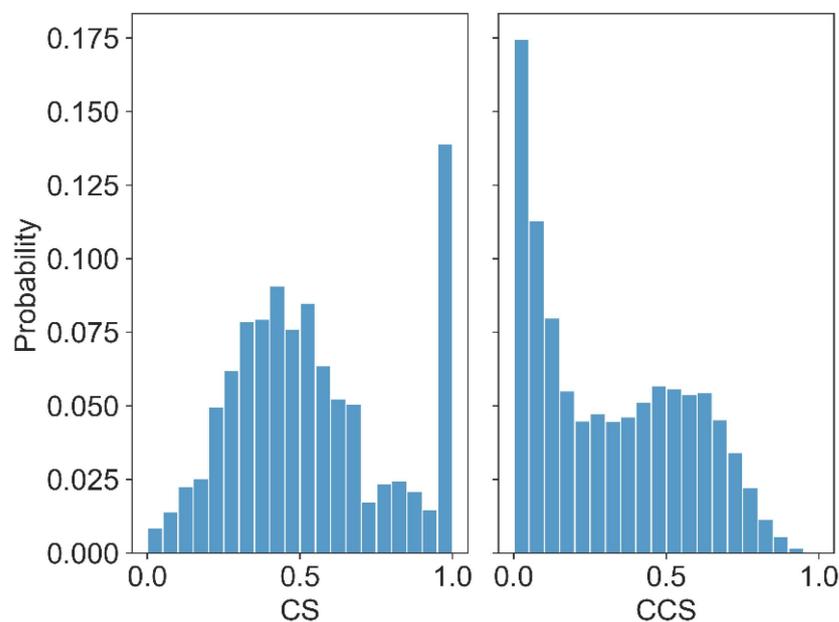


Figure 2. The distribution of the residue conservation scores in the spike protein.

Next, an investigation into the conservation and co-conservation of mutation sites in variants that indicate a significant impact on transmissible, severity, and immunity was conducted. As shown in Figure 3, most variants had substitutions of D614G, followed by substitutions at sites 484E, 681P and 501N, which are relatively common and considered to increase infectivity with clear evidence [21–30]. Regarding the co-conservation of mutation sites in the variants, most of the variants had strong co-conservation site pairs, which may indicate that the strong co-conservation sites may play an important role in infectiousness. However, we also found that although some sites (such as 484E, 681P, and 417K) appeared more frequently, they did not have strong co-conservation with other sites.

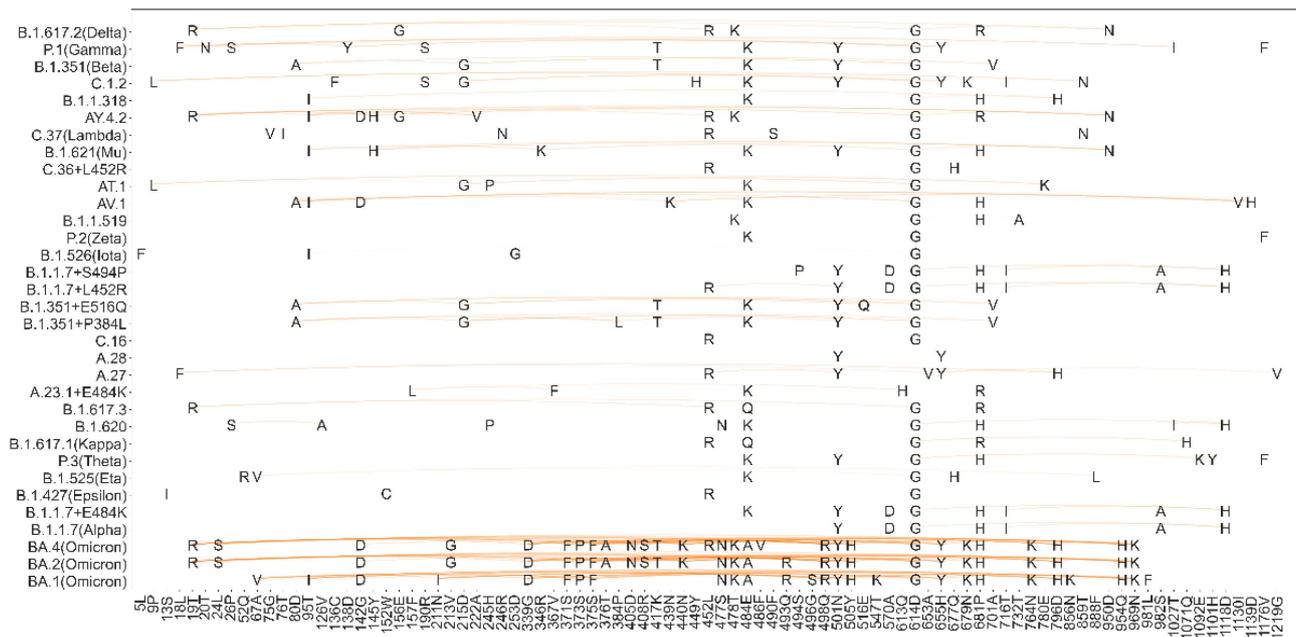


Figure 3. The co-conservation of mutation sites in popular variants.

2.2. Protein Co-Conservation Network Construction

A protein co-conservation weighted network of spike protein was constructed based on the CCS score that was calculated from the sequence of the protein. As shown in Figure 1, there were 1273 amino acids in the spike protein, which corresponded to 1273 nodes in the network. A total of 264,046 (16.3%) of all possible residue pairs had strong co-conservation as edges in the network, and the weights of the edges were CCS. Thus, the PCCN model could reveal the co-conservation information of residues from a network view.

2.3. Network Topological Features of PCCN of Spike Protein

Next, the topological features of PCCN were investigated and the correlation between these network features and the infectious or pathogenicity of SARS-CoV-2 was further explored. The network analysis was performed at two levels: residues and variants.

2.3.1. Mutation Sites Showed High Centrality in PCCN at the Residue Level

In the network, 89 nodes (residues) produced substitution mutations in the main variants. To compare the topological features of these 89 mutation nodes with the non-mutation nodes, we randomly selected 89 nodes from the non-mutation nodes and computed their average network parameters for 1000 times to present the network features of non-mutation nodes, which were then compared with the 89 mutation nodes. As shown in Figure 4, we calculated the average neighborhood weighted degree (Kw), page rank (P), degree centrality (D), betweenness (B), closeness (C), and shortest path length (L) and found that the mutation nodes had significantly larger Kw , P , D , B , but smaller L than the corresponding non-mutation nodes. The results indicate that the mutation nodes demonstrated more centrality than the non-mutation nodes and that they were topologically important in the PCCN. Moreover, the CS and CCS of the mutation nodes were significantly smaller than those of the non-mutation nodes, which showed that the conservation and co-conservation of the mutation nodes were relatively weak. Table S1 lists the detailed network feature information of the nodes.

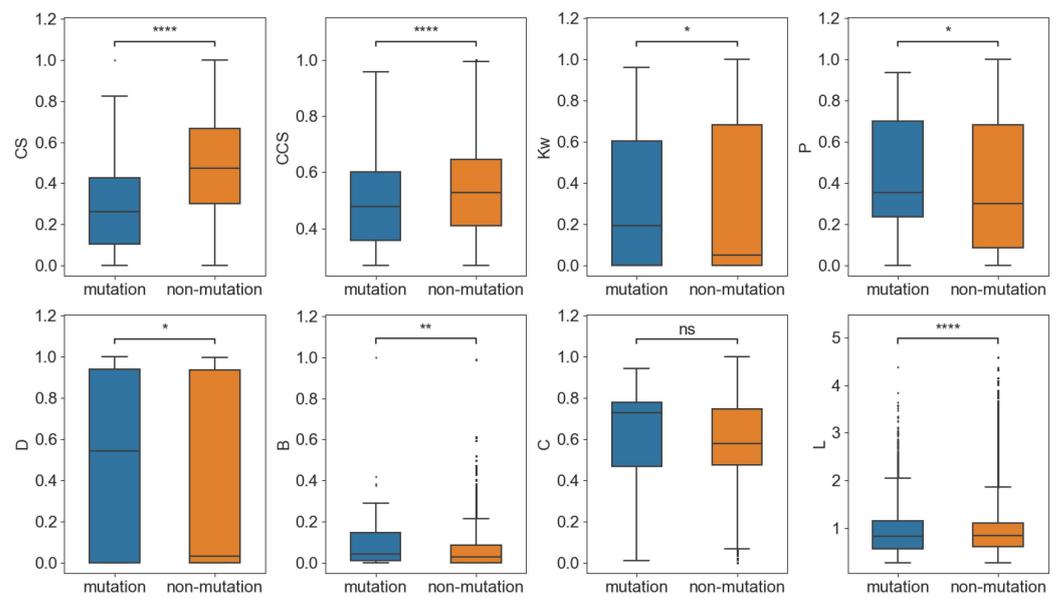


Figure 4. Comparison of the network features of the mutation and non-mutation sites. * $p < 0.05$; ** $p < 0.01$; **** $p < 0.0001$.

Finally, the top three mutation sites and co-mutation sites with the highest network parameters are listed in Table 1. Several substitutions had more than one high parameter such as 498Q, 440N, 516E, 1027T, and 614D. Then, we investigated these five substitution site relations with their allosteric modulation region (AMR) and RBD and found that 614D and 1027T had more interactions with both AMR and RBD (Figure 5). The results indicate that there was more of a co-conservation tendency among 614D, AMR, and RBD, which may be the reason for the mutation of 614D and will allow for a modulated structural rearrangement for membrane fusion [31] and increase the infectivity and stability of virions [25].

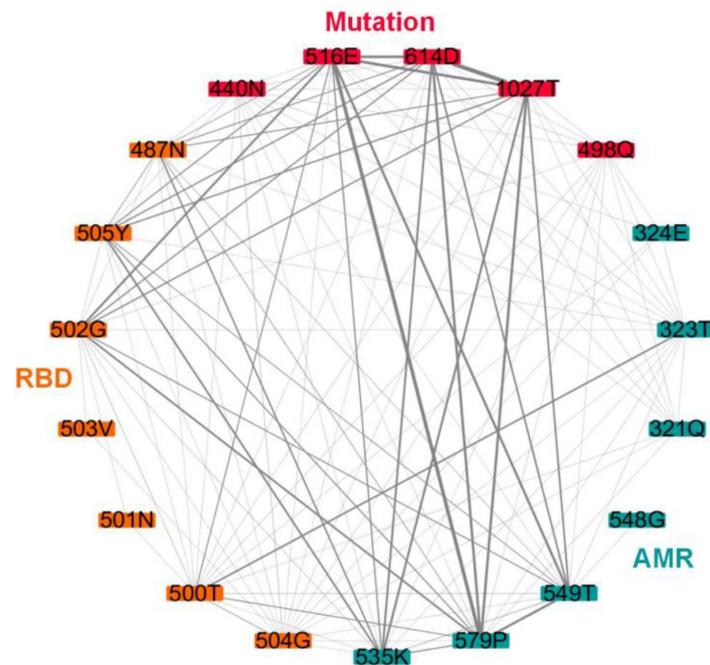


Figure 5. Interactions among top mutation sites with AMR and RBD. Red nodes: mutation sites; green nodes: AMR; orange nodes: RBD.

Table 1. The top three substitutions and co-substitutions with the highest network characteristics.

Parameters	Top Three
<i>Kw</i>	516E, 376T, 1027T
<i>CS</i>	764N, 981L, 856N
<i>P</i>	969N, 859T, 1027T
<i>D</i>	95T, 516E, 215D
<i>B</i>	405D, 498Q, 440N
<i>C</i>	440N, 498Q, 1071Q
<i>CCS</i>	614D-1027T, 614D-859T, 376T-408R
<i>L</i>	190R-679N, 5L-253D, 190R-484E

2.3.2. Mutation Sites Showed High Centrality in PCCN at the Variant Level

In this element of the study, 33 special variants were collected that have caused great harm to human society and are considered as highly infectious or virulent. To study the variants from a network view, we compared the network features of variants with those of the non-variants. For the variant group, the average parameter of mutation sites in each variant was used to present the feature for the variant; therefore, there were 33 values in the variant group for each parameter. For the non-variant group, random selection was applied to achieve the same number of non-mutation sites as each variant and computed the average parameter 1000 times. For example, there were five substitutions in the Beta variant, and five nodes were taken from the non-mutation sites 1000 times. The distribution of network characteristics for the same number of nodes is shown in Supplementary Figures S1–S8.

As a result, the variant group also showed significantly higher *Kw*, *P*, *D*, and *B* than the non-variant group (Figure 6). The results indicated that the mutation sites that occurred on the same variants have high centrality and may play an important role in their conservation. In addition, the *CS* and *CCS* of mutation nodes in the variants were significantly smaller than the non-mutation nodes in the non-variant group, which showed that the conservation and co-conservation of mutation nodes were relatively weak at the variant level.

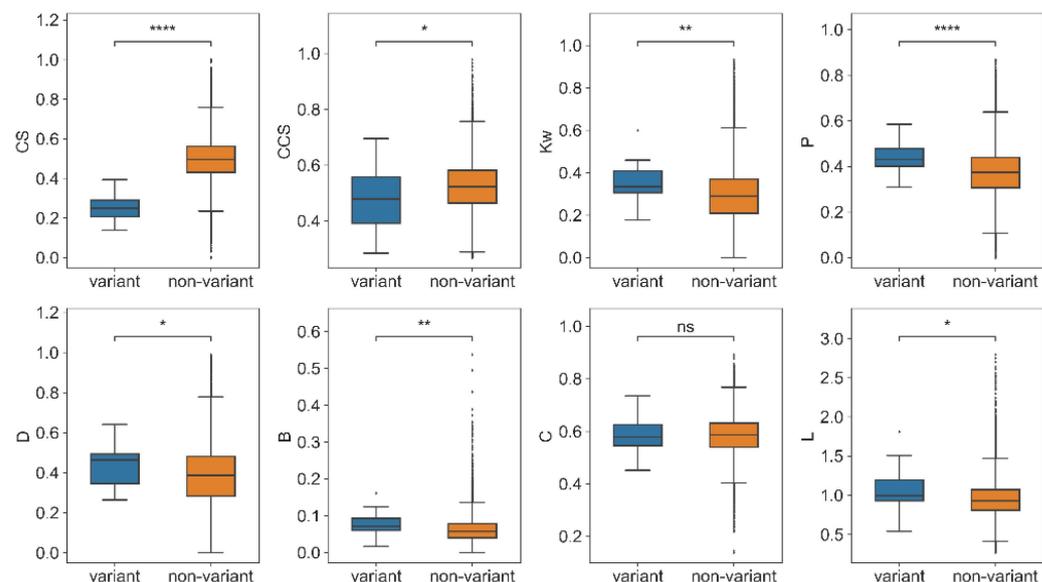


Figure 6. Comparison of the network features of the variant and non-variant groups. * $p < 0.05$; ** $p < 0.01$; **** $p < 0.0001$.

Finally, Omicron BA.4 was taken, which contained the most mutation sites among the variants, as an example to perform the network analysis. As shown in Figure 7, similar to previous results, the CS and CCS of the mutation sites in BA.4 were significantly smaller than the non-variant group, which means that both the conservation and co-conservation of the mutation sites in the variant were lower. For the network parameters, only the shortest path length in Omicron and the non-variant group showed significant differences.

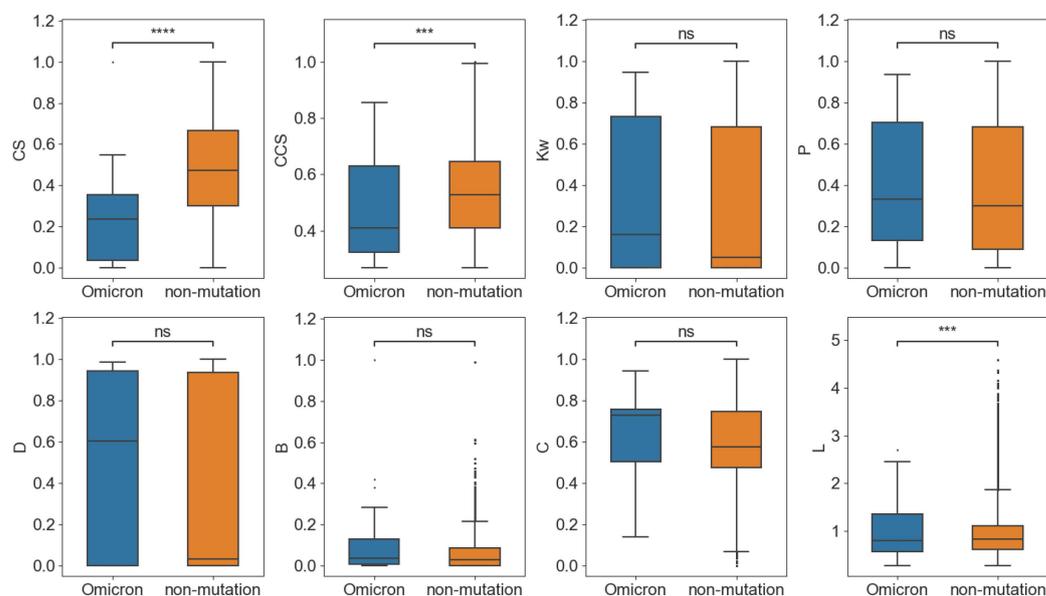


Figure 7. Comparison of the network features of Omicron BA.4 and the non-variant group. *** $p < 0.001$; **** $p < 0.0001$.

2.4. Stability Changes of Spike Protein upon Mutations Significantly Correlated with the Node Page Rank and Neighbors' Degree

To study the mutation effects for the spike protein function from a conservation network view, we performed the correlation analysis between the topological features of the mutation sites and mutation effects on the protein stability changes.

DeepDDG [32] was used to predict the changes in stability for each point mutation and each mutation pair based on the spike protein structure. As shown in Figure 8a, most of the mutations were destabilizing mutations, for example, G142D, E156G, and C136F were the top three mutations that obtained the smallest stability scores, but D138Y obtained the largest positive stability score, which means that the mutation increased the stability. Then, we conducted a correlation analysis between the topological features and the stability changes upon mutations and found that three topological parameters including page rank, average neighborhood weighted degree, and degree centrality were significantly correlated with stability changes (Table S2). In particular, there was a positive correlation between the page rank (P) and stability changes with $r = 0.271$ and $p = 6.07 \times 10^{-3}$ (Figure 8b). A similar correlation pattern was also found between the average neighborhood weighted degree (Kw) and stability changes, as shown in Figure 8c ($r = 0.253$ and $p = 1.10 \times 10^{-2}$). The results indicate that if the nodes themselves had a high degree of centrality, high page rank, or their neighbors had a high degree, their mutations tended to destabilize the spike protein.

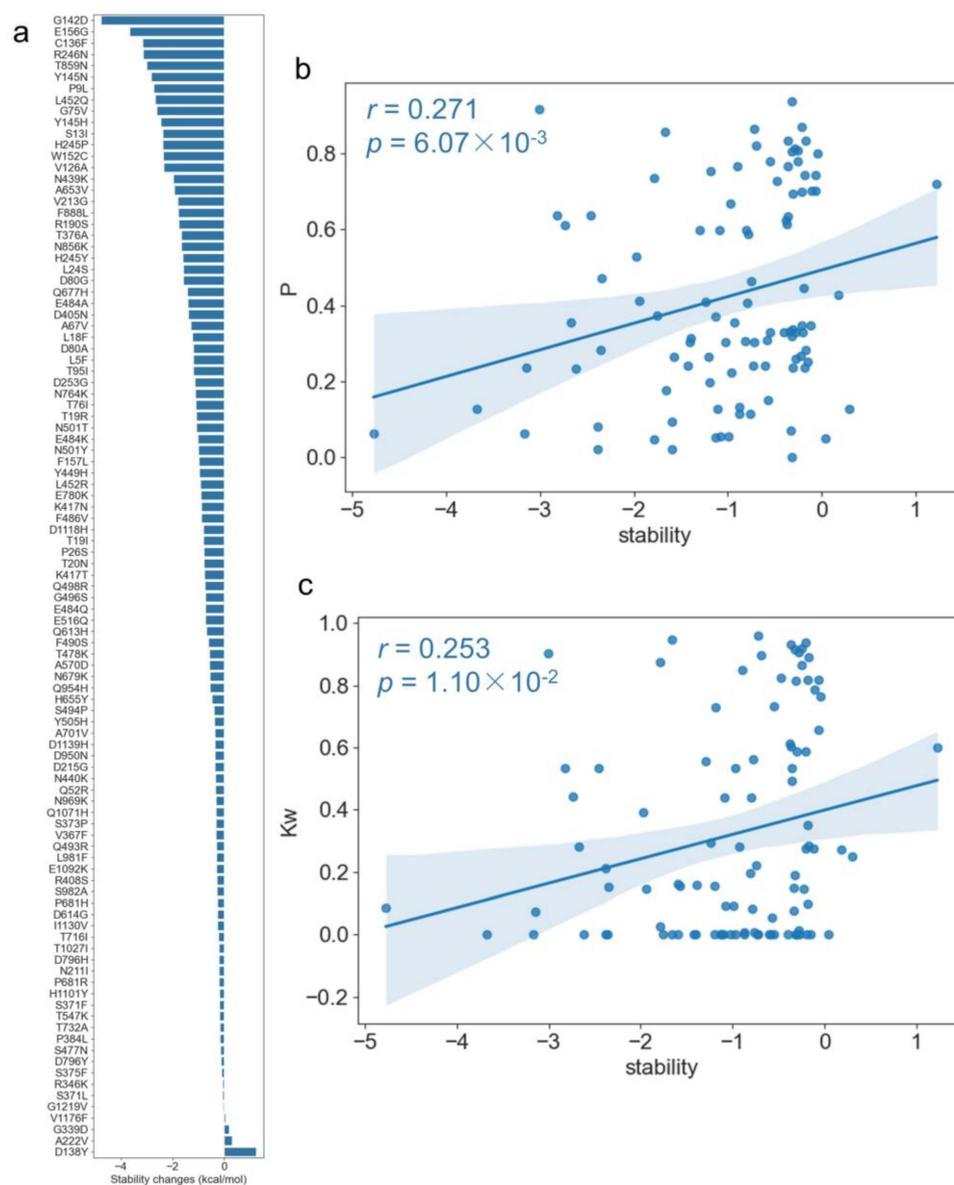


Figure 8. The correlation between the stability changes of the spike protein upon the mutations and PCCN topological features. (a). Stability changes in the spike protein upon the point mutations. (b). Page rank was positively significantly correlated with stability changes. (c). The average neighborhood weighted degree was positively significantly correlated with the stability changes.

2.5. Binding Free Energy Changes of Spike Protein upon Mutations Significantly Correlated with CCS and Average Shortest Path Length

The binding of the spike protein with the human receptor ACE2 plays a central role in the molecular machinery of the SARS-CoV-2 infection of human cells. Mutations in the spike protein affect the binding affinity with ACE2 and have been established to be proportional to the infectivity of different viral variants in the host cells [4,33–36]. Therefore, we analyzed the relationship between the PCCN topological features and binding free energy (BFE) changes upon mutation to investigate the variant effects on protein binding from a network view based only on the sequence information.

TopNetTree [37] was employed to estimate the BFE changes upon the mutations listed in Table 2 of the SARS-CoV-2 spike protein with ACE2 (PDB: 7A98 [38]). As shown in Figure 9a, all of the mutations increased the BFE between the spike glycoprotein and ACE2. The changes in BFE were significantly negatively correlated with the closeness centrality with the p -value of 0.045 and the coefficient of -0.207 . Next, we analyzed the co-mutation effects on the spike

protein. The correlation of the shortest path length and CCS between two mutations with the BFE changes of the two mutations were calculated, respectively. As shown in Figure 9b,c, the BFE changes were significantly positively correlated with CCS ($r = 0.233$ and $p = 1.00 \times 10^{-3}$) and shortest path length ($r = 0.179$ and $p = 6.15 \times 10^{-8}$) from the two mutation sites. The results demonstrate that if the two mutation sites have the tendency of high co-conservation or larger distance, they might induce larger BFE changes. Moreover, the top three co-mutation pairs with the largest BFE were D215G–D614G, S371F–D614G, and D614G–T732A. Notably, D614G was involved in all of them, which has been reported to have an important impact on infectivity [21–27]; although it is not in the RBD, it has an important impact on BFE. The results indicate that our PCCN model based on sequence information could also reflect the mutation effects on the binding affinity changes between the spike protein and ACE2. The detailed correlation parameters are listed in Table S3.

Table 2. Summary of the 33 variants of the spike protein.

WHO Label	Lineage + Additional Mutations	Spike Mutations of Interest	Category	No. of Substitutions
Omicron	BA.1	A67V T95I G142D N211I G339D S371L S373P S375F S477N T478K E484A Q493R G496S Q498R N501Y Y505H T547K D614G H655Y N679K P681H N764K D796Y N856K Q954H N969K L981F	VOC	27
Omicron	BA.2	T19I L24S G142D V213G G339D S371F S373P S375F T376A D405N R408S K417N N440K S477N T478K E484A Q493R Q498R N501Y Y505H D614G H655Y N679K P681H N764K D796Y Q954H N969K	VOC	28
Omicron	BA.4	T19I L24S G142D V213G G339D S371F S373P S375F T376A D405N R408S K417N N440K L452R S477N T478K E484A F486V Q498R N501Y Y505H D614G H655Y N679K P681H N764K D796Y Q954H N969K	VOC	29
Alpha	B.1.1.7	N501Y A570D D614G P681H T716I S982A D1118H	DEV	7
n/a	B.1.1.7 + E484K	N501Y A570D D614G P681H T716I S982A D1118H E484K	DEV	8
Epsilon	B.1.427	S13I W152C L452R D614G	DEV	4
Eta	B.1.525	Q52R A67V E484K D614G Q677H F888L	DEV	6
Theta	P.3	E484K N501Y D614G P681H E1092K H1101Y V1176F	DEV	7
Kappa	B.1.617.1	L452R E484Q D614G P681R Q1071H	DEV	5
n/a	B.1.620	P26S V126A H245Y S477N E484K D614G P681H T1027I D1118H	DEV	9
n/a	B.1.617.3	T19R L452R E484Q D614G P681R	DEV	5
n/a	A.23.1+E484K	F157L V367F Q613H P681R E484K	DEV	5
n/a	A.27	L18F L452R N501Y A653V H655Y D796Y G1219V	DEV	7
n/a	A.28	N501T H655Y	DEV	2
n/a	C.16	L452R D614G	DEV	2
n/a	B.1.351+P384L	D80A D215G K417N E484K N501Y D614G A701V P384L	DEV	8
n/a	B.1.351+E516Q	D80A D215G K417N E484K N501Y D614G A701V E516Q	DEV	8
n/a	B.1.1.7+L452R	N501Y A570D D614G P681H T716I S982A D1118H L452R	DEV	8
n/a	B.1.1.7+S494P	N501Y A570D D614G P681H T716I S982A D1118H S494P	DEV	8
Iota	B.1.526	L5F T95I D253G D614G	DEV	4
Zeta	P.2	E484K D614G V1176F	DEV	3
n/a	B.1.1.519	T478K D614G P681H T732A	DEV	4
n/a	AV.1	D80G T95I G142D N439K E484K D614G P681H I1130V D1139H	DEV	9
n/a	AT.1	P9L D215G H245P E484K D614G E780K	DEV	6
n/a	C.36+L452R	D614G Q677H L452R	DEV	3
Mu	B.1.621	T95I Y145N R346K E484K N501Y D614G P681H D950N	DEV	8
Lambda	C.37	G75V T76I R246N L452Q F490S D614G T859N	DEV	7
n/a	AY.4.2	T19R T95I G142D Y145H E156G A222V L452R T478K D614G P681R D950N	DEV	11
n/a	B.1.1.318	T95I E484K D614G P681H D796H	DEV	5
n/a	C.1.2	P9L C136F R190S D215G Y449H E484K N501Y D614G H655Y N679K T716I T859N	DEV	12
Beta	B.1.351	D80A D215G K417N E484K N501Y D614G A701V	DEV	7
Gamma	P.1	L18F T20N P26S D138Y R190S K417T E484K N501Y D614G H655Y T1027I V1176F	DEV	12
Delta	B.1.617.2	T19R E156G L452R T478K D614G P681R D950N	DEV	7

VOC: variants of concern; DEV: de-escalated variants.

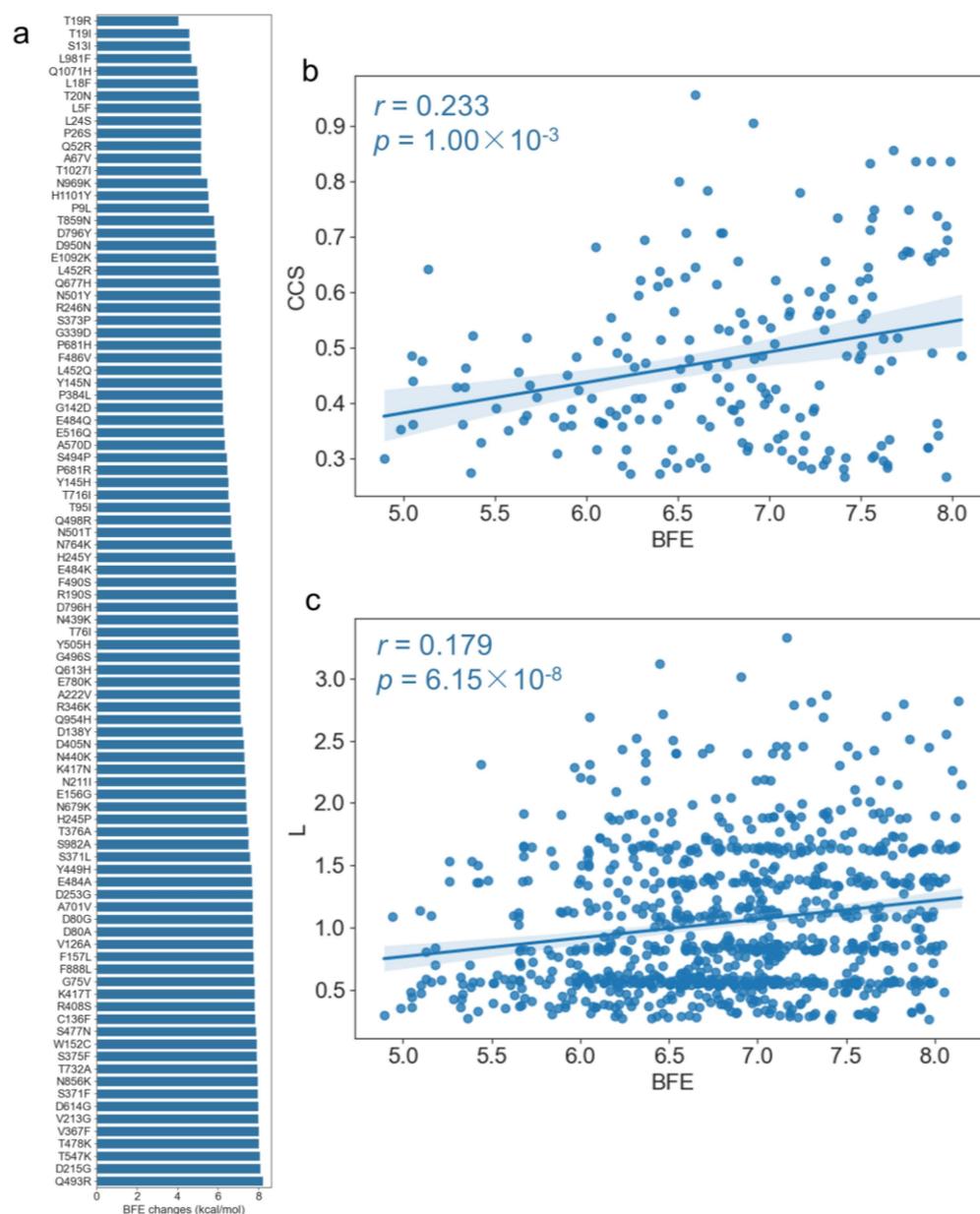


Figure 9. Correlation between the binding free energy changes of the spike protein upon mutations and the PCCN topological features. (a) Stability changes of the spike protein upon point mutations. (b) Page rank was positively significantly correlated with the stability changes. (c) BEF changes were positive significantly correlated with the average shortest path length L .

3. Discussion

Over the past years, the constant emergence of new variants of SARS-CoV-2 has presented challenges in the prevention of the spread and treatment of COVID-19. Studying the mutation effects on the spike protein from SARS-CoV-2 remains important. In this study, we proposed a sequence-based network model called PCCN that contains conservation and co-conservation information to study the mutation effects on the spike protein from a network view. We found that the mutation sites in the PCCN not only demonstrated lower conservation and co-conservation tendencies than the non-mutation sites, but also showed significant topological differences such as the average neighborhood weighted degree and page rank, both from non-mutation sites at the residue level and variant level. The results mean that the PCCN model could capture the topological features of mutation sites and thus has the potential to predict the mutation sites from non-mutation sites.

The stability and binding free energy of the spike protein were correlated with SARS-CoV-2 infectivity [4,33–36]. Thus, we analyzed the correlation between the topological features of PCCN and changes in the stability and binding free energy of spike protein upon mutations. The stability changes were significantly positively correlated with the node page rank and neighbors' degree, while BFE changes were positively correlated with the shortest path length, and correlated negatively with the closeness centrality. The results indicate that our PCCN model could reflect the changes in the physico-chemical characteristics of the spike protein upon mutation.

However, this study had two potential limitations. First, the limited number of variants may result in bias for the topological feature comparison analysis between the variants and non-variants. Second, the relationship between the topological features and spike protein function needs further validation in large datasets or experimental verification. In future study, we will continue to collect the variants to explore and enhance the application of PCCN for the SARS-CoV-2 study.

Taken together, the sequence-based PCCN presented here not only suggests topological differences in mutation sites from non-mutations, but also provides new insights into the mutation effects on the spike protein physico-chemical characteristics without any protein structure information. Moreover, the PCCN was only based on the sequence information, which means that it can be quickly migrated to study the effects of continuously emerging variants of SARS-CoV-2. This is an adaptive way to face the emergence of new strains or viruses in the future.

4. Materials and Methods

4.1. Dataset

Since the outbreak of SARS-CoV-2, many variants have emerged on the spike protein that are highly infectious and have caused great destruction to the world such as Omicron [39,40], Beta [41–44], Gamma [44–46], and Delta [47]. In this paper, we focused on the variants provided by the ECDC. A total of 48 variants were collected from the ECDC website (<https://www.ecdc.europa.eu/en/covid-19/variants-concern>, accessed on 9 July 2022). The variants were divided into four types: variant of concern (VOC), variant of interest (VOI), variant under monitoring (VUM), and de-escalated variants (DEV). Then, 33 variants with currently available evidence of increasing transmissibility, immunity, or infection severity were kept for further analysis (Table 2). The substitutions of variants were selected from [48]. A total of 101 substitutions were observed that corresponded to 89 distinct mutation sites.

4.2. Conservation Calculation

The conservation score (CS) of each residue and the co-conservation score (CCS) between residues in the spike protein were calculated by ProCon [19,20], which is based on multiple sequence alignment (MSA) and information theory. The sequences for MSA were obtained from the InterPro database [18] by looking for its protein family to find similar sequences in different species. Then, sequences were aligned by ClustalX2 [49] and used as ProCon input to compute the spike protein conservation information.

The raw CS that was calculated by ProCon is between 0 and 3, and the CCS between residues is between 0 and 350. The scores were normalized by a min–max normalization method as shown in Equation (1), and they were mapped between 0 and 1:

$$X = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

where x_{max} and x_{min} are the maximum and minimum scores, respectively.

Co-conservation scores for all possible residue pairs can be calculated by ProCon, and the larger the score, the stronger the co-conservation tendency. However, most of the residues pairs had weak co-conservation. We tried to keep more residue pairs with relatively strong co-conservation to construct the network. As shown in Figure 2, when the

co-conservation score was smaller than 0.268, the number of residue pairs increased sharply, which means that they had less of a tendency of co-conservation. Therefore, residue pairs with CCS larger than 0.268 were selected as strong co-conservation.

In order to simplify the representation of the score, the CS and CCS denote the normalized conservation score and co-conservation score, respectively, in the later section.

4.3. Protein Co-Conservation Network Construction

The spike protein was represented as a network consisting of a set of nodes and edges. Each amino acid residue is presented as a node. The node set V is defined as follows:

$$V = \{v_i | 1 < i < N\} \quad (2)$$

where i is the residue id in the protein; v_i is the corresponding residue (node) for i ; and N is the number of residues in the protein.

An edge was placed between two residues if the CCS between them was larger than 0.268. Thus, the adjacent matrix W of the PCCN was defined as follows:

$$w_{ij} = \begin{cases} CCS_{ij} & CCS_{ij} \geq 0.268 \\ 0.268 & CCS_{ij} < 0.268 \text{ and } |j - i| = 1 \\ 0 & \text{others} \end{cases} \quad (3)$$

When two nodes had strong co-conservation, that is, the co-conservation score was greater than or equal to the threshold of 0.268, there was an edge between the nodes and the weight of the edge was the CCS. The links between adjacent residues along the backbone were also kept and their weights were set at the threshold 0.268 if the CCS between them was less than the threshold. Thus, the PCCN is a network model that contains co-conservation information among residues.

4.4. Network Analysis

After constructing the PCCN, we analyzed the network topological characteristics of the residues in the network. Six network characteristics were used in our work: average neighborhood weighted degree (Kw), page rank (P) [50], closeness centrality (C), betweenness centrality (B), degree centrality (D), and shortest path length (L). For node v_i , the parameters Kw , D , B , and C are defined as follows:

$$Kw = \frac{\sum_{j \in \{j | w_{ij} \neq 0\}} w_{ij}}{n_i} \quad (4)$$

$$D = \frac{n_i}{N - 1} \quad (5)$$

$$B = \frac{2}{(N - 1)(N - 2)} \sum_{s, t \in V} \frac{\sigma(s, t | v_i)}{\sigma(s, t)} \quad (6)$$

$$C = \frac{n_i - 1}{N - 1} \frac{n_i - 1}{\sum_{j=1}^{n_i-1} \sigma(n_i, n_j)} \quad (7)$$

where n_i is the of adjacent nodes number of v_i , and $\sigma(s, t)$ is the shortest path length from node s to node t .

Page rank was originally designed to measure the importance of the website pages. Here, it was used to measure the importance of a node [50]. While it is an algorithm of directed graphs, each edge of the conservative interaction network will be converted to two directed edges. The shortest path length is the length of the path with the minimum sum of edge weights. Since all nodes are connected, there is an average shortest path between any pair. These network parameters were implemented by NetworkX [51].

4.5. Estimations of Binding Free Energy (BFE) Change

The change in binding free energy ($\Delta\Delta G$) upon a single mutation from variants was estimated using TopNetTree [37]. The effects of two substitutions on the BFE changes were evaluated as follows:

$$\Delta\Delta G_{ij} = \frac{\Delta\Delta G_i + \Delta\Delta G_j}{2} \quad (8)$$

where $\Delta\Delta G_i$ and $\Delta\Delta G_j$ are the binding free energy changes following mutation i and j , respectively; and $\Delta\Delta G_{ij}$ is the combined changes of both mutations i and j .

4.6. Estimations of Stability Changes

The effect of the amino acid substitutions on stability was calculated using DeepDDG [32]. Similar to the binding free energy changes of two substitutions, the effects of two mutations on the stability changes were defined as follows:

$$s_{ij} = \frac{s_i + s_j}{2} \quad (9)$$

where S_i and S_j are the stability changes following mutation i and j and S_{ij} is the combined changes of stability following the mutation i and j . The structure of the spike protein was obtained from the website of Zhang's laboratory (<https://zhanggroup.org/COVID-19/index.html#table1>, accessed on 30 October 2022) with the identification code QHD43416 that was generated by the D-I-TASSER/C-I-TASSER pipeline [52].

4.7. Statistics

The Spearman's correlation coefficient was used to analyze the correlation between the protein properties and network characteristics. The Wilcoxon rank-sum test was applied to analyze the statistically significant differences and was implemented by SciPy [53].

5. Conclusions

In conclusion, we proposed a network model known as PCCN to study the mutation effects on the spike protein only from the sequence information. The topological features of PCCN showed a significant difference at both the residue and variant levels. In particular, the page rank and average neighborhood weighted degree could reflect the mutation effects on the spike protein stability, while the co-conservation score and shortest path length could indicate the mutation effects on the binding free energy changes. The PCCN model offers a new approach for studying variants that may cause future pandemics based on the sequence and topology information.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms24043255/s1>.

Author Contributions: Conceptualization, W.Y.; Methodology, Y.Y.; Software, L.Z.; Data curation, L.Z. and Y.L.; Writing the original draft preparation, W.Y. and L.Z.; Writing the review and editing, Y.Y.; Visualization, L.Z. and Y.L.; Supervision, W.Y. and Y.Y.; Project administration, W.Y.; Funding acquisition, Y.Y. and W.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Key Project of the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (20KJA520010); the National Natural Science Foundation of China (62272335); the Collaborative Innovation Center of Novel Software Technology and Industrialization; the Key Research and Development Program of Jiangsu Province (BE2020656); the Jiangsu Students Innovation and Entrepreneurship Training Program (202210285081Z); and the Priority Academic Program Development of Jiangsu Higher Education Institutions.

Informed Consent Statement: Not applicable.

Data Availability Statement: The source code and data used in analysis are available at <https://github.com/XDcat/PCCN>, accessed on 16 January 2023.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chen, N.; Zhou, M.; Dong, X.; Qu, J.; Gong, F.; Han, Y.; Qiu, Y.; Wang, J.; Liu, Y.; Wei, Y.; et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study. *Lancet* **2020**, *395*, 507–513. [[CrossRef](#)] [[PubMed](#)]
2. Hossain, M.U.; Ahammad, I.; Bhattacharjee, A.; Chowdhury, Z.M.; Hossain Emon, M.T.; Chandra Das, K.; Keya, C.A.; Salimullah, M. Whole genome sequencing for revealing the point mutations of SARS-CoV-2 genome in Bangladeshi isolates and their structural effects on viral proteins. *RSC Adv.* **2021**, *11*, 38868–38879. [[CrossRef](#)]
3. Menni, C.; Sudre, C.H.; Steves, C.J.; Ourselin, S.; Spector, T.D. Quantifying additional COVID-19 symptoms will save lives. *Lancet* **2020**, *395*, e107–e108. [[CrossRef](#)] [[PubMed](#)]
4. Hoffmann, M.; Kleine-Weber, H.; Schroeder, S.; Krüger, N.; Herrler, T.; Erichsen, S.; Schiergens, T.S.; Herrler, G.; Wu, N.-H.; Nitsche, A.; et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **2020**, *181*, 271–280. [[CrossRef](#)]
5. Zhou, P.; Yang, X.-L.; Wang, X.-G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H.-R.; Zhu, Y.; Li, B.; Huang, C.-L.; et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **2020**, *579*, 270–273. [[CrossRef](#)]
6. Cascella, M.; Rajnik, M.; Aleem, A.; Dulebohn, S.C.; Di Napoli, R. Features, Evaluation, and Treatment of Coronavirus (COVID-19). In *StatPearls*; StatPearls Publishing: Treasure Island, FL, USA, 2022.
7. Di Paola, L.; Hadi-Alijanvand, H.; Song, X.; Hu, G.; Giuliani, A. The Discovery of a Putative Allosteric Site in the SARS-CoV-2 Spike Protein Using an Integrated Structural/Dynamic Approach. *J. Proteome Res.* **2020**, *19*, 4576–4586. [[CrossRef](#)] [[PubMed](#)]
8. Harrison, A.G.; Lin, T.; Wang, P. Mechanisms of SARS-CoV-2 Transmission and Pathogenesis. *Trends Immunol.* **2020**, *41*, 1100–1115. [[CrossRef](#)]
9. Cele, S.; Jackson, L.; Khoury, D.S.; Khan, K.; Moyo-Gwete, T.; Tegally, H.; San, J.E.; Cromer, D.; Scheepers, C.; Amoako, D.G.; et al. Omicron extensively but incompletely escapes Pfizer BNT162b2 neutralization. *Nature* **2022**, *602*, 654–656. [[CrossRef](#)] [[PubMed](#)]
10. Pulliam, J.R.C.; Schalkwyk, C.V.; Govender, N.; Gottberg, A.V.; Cohen, C.; Groome, M.J.; Dushoff, J.; Mlisana, K.; Moultrie, H. Increased risk of SARS-CoV-2 reinfection associated with emergence of Omicron in South Africa. *Science* **2022**, *376*, eabn4947. [[CrossRef](#)]
11. Hui, K.P.Y.; Ho, J.C.W.; Cheung, M.-c.; Ng, K.-c.; Ching, R.H.H.; Lai, K.-I.; Kam, T.T.; Gu, H.; Sit, K.-Y.; Hsin, M.K.Y.; et al. SARS-CoV-2 Omicron variant replication in human bronchus and lung ex vivo. *Nature* **2022**, *603*, 715–720. [[CrossRef](#)]
12. Yan, W.; Zhang, D.; Shen, C.; Liang, Z.; Hu, G. Recent Advances on the Network Models in Target-based Drug Discovery. *Curr. Top. Med. Chem.* **2018**, *18*, 1031–1043. [[CrossRef](#)] [[PubMed](#)]
13. Morselli Gysi, D.; do Valle, Í.; Zitnik, M.; Ameli, A.; Gan, X.; Varol, O.; Ghiassian, S.D.; Patten, J.J.; Davey, R.A.; Loscalzo, J.; et al. Network medicine framework for identifying drug-repurposing opportunities for COVID-19. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2025581118. [[CrossRef](#)]
14. Hadi-Alijanvand, H.; Di Paola, L.; Hu, G.; Leitner, D.M.; Verkhivker, G.M.; Sun, P.; Poudel, H.; Giuliani, A. Biophysical Insight into the SARS-CoV2 Spike-ACE2 Interaction and Its Modulation by Hepcidin through a Multifaceted Computational Approach. *ACS Omega* **2022**, *7*, 17024–17042. [[CrossRef](#)] [[PubMed](#)]
15. Yan, W.; Zhou, J.; Sun, M.; Chen, J.; Hu, G.; Shen, B. The construction of an amino acid network for understanding protein structure and function. *Amino Acids* **2014**, *46*, 1419–1439. [[CrossRef](#)] [[PubMed](#)]
16. Liang, Z.; Verkhivker, G.M.; Hu, G. Integration of network models and evolutionary analysis into high-throughput modeling of protein dynamics and allosteric regulation: Theory, tools and applications. *Brief Bioinform.* **2020**, *21*, 815–835. [[CrossRef](#)]
17. Yan, W.; Hu, G.; Liang, Z.; Zhou, J.; Yang, Y.; Chen, J.; Shen, B. Node-weighted amino acid network strategy for characterization and identification of protein functional residues. *J. Chem. Inf. Model.* **2018**, *58*, 2024–2032. [[CrossRef](#)]
18. Blum, M.; Chang, H.-Y.; Chuguransky, S.; Grego, T.; Kandasamy, S.; Mitchell, A.; Nuka, G.; Paysan-Lafosse, T.; Qureshi, M.; Raj, S.; et al. The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* **2020**, *49*, D344–D354. [[CrossRef](#)]
19. Yang, Y. Bioinformatics Research on Disease-related Amino Acid Variations. Ph.D. Thesis, Soochow University, Suzhou, China, 2015.
20. Shen, B.; Vihinen, M. Conservation and covariance in PH domain sequences: Physicochemical profile and information theoretical analysis of XLA-causing mutations in the Btk PH domain. *Protein. Eng. Des. Sel.* **2004**, *17*, 267–276. [[CrossRef](#)]
21. Li, Q.; Wu, J.; Nie, J.; Zhang, L.; Hao, H.; Liu, S.; Zhao, C.; Zhang, Q.; Liu, H.; Nie, L.; et al. The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. *Cell* **2020**, *182*, 1284–1294.e1289. [[CrossRef](#)]
22. Socher, E.; Conrad, M.; Heger, L.; Paulsen, F.; Sticht, H.; Zunke, F.; Arnold, P. Mutations in the B.1.1.7 SARS-CoV-2 Spike Protein Reduce Receptor-Binding Affinity and Induce a Flexible Link to the Fusion Peptide. *Biomedicines* **2021**, *9*, 525. [[CrossRef](#)]

23. Zhang, L.; Jackson, C.B.; Mou, H.; Ojha, A.; Peng, H.; Quinlan, B.D.; Rangarajan, E.S.; Pan, A.; Vanderheiden, A.; Suthar, M.S.; et al. SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. *Nat. Commun.* **2020**, *11*, 6013. [CrossRef] [PubMed]
24. Becerra-Flores, M.; Cardozo, T. SARS-CoV-2 viral spike G614 mutation exhibits higher case fatality rate. *Int. J. Clin. Pract.* **2020**, *74*, e13525. [CrossRef] [PubMed]
25. Plante, J.A.; Liu, Y.; Liu, J.; Xia, H.; Johnson, B.A.; Lokugamage, K.G.; Zhang, X.; Muruato, A.E.; Zou, J.; Fontes-Garfias, C.R.; et al. Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* **2021**, *592*, 116–121. [CrossRef] [PubMed]
26. Yurkovetskiy, L.; Wang, X.; Pascal, K.E.; Tomkins-Tinch, C.; Nyalile, T.P.; Wang, Y.; Baum, A.; Diehl, W.E.; Dauphin, A.; Carbone, C.; et al. Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. *Cell* **2020**, *183*, 739–751.e738. [CrossRef]
27. Korber, B.; Fischer, W.M.; Gnanakaran, S.; Yoon, H.; Theiler, J.; Abfalterer, W.; Hengartner, N.; Giorgi, E.E.; Bhattacharya, T.; Foley, B.; et al. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* **2020**, *182*, 812–827.e819. [CrossRef]
28. Wang, Z.; Schmidt, F.; Weisblum, Y.; Muecksch, F.; Barnes, C.O.; Finkin, S.; Schaefer-Babajew, D.; Cipolla, M.; Gaebler, C.; Lieberman, J.A.; et al. mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *Nature* **2021**, *592*, 616–622. [CrossRef]
29. Gu, H.; Chen, Q.; Yang, G.; He, L.; Fan, H.; Deng, Y.Q.; Wang, Y.; Teng, Y.; Zhao, Z.; Cui, Y.; et al. Adaptation of SARS-CoV-2 in BALB/c mice for testing vaccine efficacy. *Science* **2020**, *369*, 1603–1607. [CrossRef]
30. Leung, K.; Shum, M.H.; Leung, G.M.; Lam, T.T.; Wu, J.T. Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. *Eurosurveillance* **2021**, *26*, 2002106. [CrossRef]
31. Zhang, J.; Cai, Y.; Xiao, T.; Lu, J.; Peng, H.; Sterling, S.M.; Walsh, R.M., Jr.; Rits-Volloch, S.; Zhu, H.; Woosley, A.N.; et al. Structural impact on SARS-CoV-2 spike protein by D614G substitution. *Science* **2021**, *372*, 525–530. [CrossRef]
32. Cao, H.; Wang, J.; He, L.; Qi, Y.; Zhang, J.Z. DeepDDG: Predicting the Stability Change of Protein Point Mutations Using Neural Networks. *J. Chem. Inf. Model.* **2019**, *59*, 1508–1514. [CrossRef]
33. Walls, A.C.; Park, Y.-J.; Tortorici, M.A.; Wall, A.; McGuire, A.T.; Velesler, D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* **2020**, *181*, 281–292.e286. [CrossRef]
34. Song, H.-D.; Tu, C.-C.; Zhang, G.-W.; Wang, S.-Y.; Zheng, K.; Lei, L.-C.; Chen, Q.-X.; Gao, Y.-W.; Zhou, H.-Q.; Xiang, H. Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 2430–2435. [CrossRef] [PubMed]
35. Qu, X.-X.; Hao, P.; Song, X.-J.; Jiang, S.-M.; Liu, Y.-X.; Wang, P.-G.; Rao, X.; Song, H.-D.; Wang, S.-Y.; Zuo, Y. Identification of two critical amino acid residues of the severe acute respiratory syndrome coronavirus spike protein for its variation in zoonotic tropism transition via a double substitution strategy. *J. Biol. Chem.* **2005**, *280*, 29588–29595. [CrossRef] [PubMed]
36. Li, W.; Shi, Z.; Yu, M.; Ren, W.; Smith, C.; Epstein, J.H.; Wang, H.; Cramer, G.; Hu, Z.; Zhang, H. Bats are natural reservoirs of SARS-like coronaviruses. *Science* **2005**, *310*, 676–679. [CrossRef]
37. Wang, M.; Cang, Z.; Wei, G.W. A topology-based network tree for the prediction of protein-protein binding affinity changes following mutation. *Nat. Mach. Intell.* **2020**, *2*, 116–123. [CrossRef]
38. Benton, D.J.; Wrobel, A.G.; Xu, P.; Rouston, C.; Martin, S.R.; Rosenthal, P.B.; Skehel, J.J.; Gamblin, S.J. Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. *Nature* **2020**, *588*, 327–330. [CrossRef]
39. England, P.H. SARS-CoV-2 variants of concern and variants under investigation in England. *Technical Briefing 12*. 2021. Available online: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1050999/Technical-Briefing-35-28January2022.pdf (accessed on 9 July 2022).
40. Pulliam, J.R.C.; van Schalkwyk, C.; Govender, N.; von Gottberg, A.; Cohen, C.; Groome, M.J.; Dushoff, J.; Mlisana, K.; Moultrie, H. Increased risk of SARS-CoV-2 reinfection associated with emergence of the Omicron variant in South Africa. *medRxiv* **2021**. [CrossRef]
41. Tegally, H.; Wilkinson, E.; Giovanetti, M.; Iranzadeh, A.; Fonseca, V.; Giandhari, J.; Doolabh, D.; Pillay, S.; San, E.J.; Msomi, N.; et al. Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature* **2021**, *592*, 438–443. [CrossRef]
42. Cele, S.; Gazy, I.; Jackson, L.; Hwa, S.-H.; Tegally, H.; Lustig, G.; Giandhari, J.; Pillay, S.; Wilkinson, E.; Naidoo, Y.; et al. Escape of SARS-CoV-2 501Y.V2 from neutralization by convalescent plasma. *Nature* **2021**, *593*, 142–146. [CrossRef]
43. Madhi, S.A.; Baillie, V.; Cutland, C.L.; Voysey, M.; Koen, A.L.; Fairlie, L.; Padayachee, S.D.; Dheda, K.; Barnabas, S.L.; Borhat, Q.E. Efficacy of the ChAdOx1 nCoV-19 Covid-19 vaccine against the B. 1.351 variant. *N. Engl. J. Med.* **2021**, *384*, 1885–1898. [CrossRef]
44. Funk, T.; Pharris, A.; Spiteri, G.; Bundle, N.; Melidou, A.; Carr, M.; Gonzalez, G.; Garcia-Leon, A.; Crispie, F.; O'Connor, L. Characteristics of SARS-CoV-2 variants of concern B. 1.1, B. 1.351 or P. 1: Data from seven EU/EEA countries, weeks 38/2020 to 10/2021. *Eurosurveillance* **2021**, *26*, 2100348. [CrossRef] [PubMed]
45. Faria, N.R.; Mellan, T.A.; Whittaker, C.; Claro, I.M.; Candido, D.d.S.; Mishra, S.; Crispim, M.A.; Sales, F.C.; Hawryluk, I.; McCrone, J.T. Genomics and epidemiology of the P. 1 SARS-CoV-2 lineage in Manaus, Brazil. *Science* **2021**, *372*, 815–821. [CrossRef] [PubMed]
46. Dejnirattisai, W.; Zhou, D.; Supasa, P.; Liu, C.; Mentzer, A.J.; Ginn, H.M.; Zhao, Y.; Duyvesteyn, H.M.; Tuekprakhon, A.; Nutalai, R. Antibody evasion by the P. 1 strain of SARS-CoV-2. *Cell* **2021**, *184*, 2939–2954.e2939. [CrossRef] [PubMed]

47. Sheikh, A.; McMenamin, J.; Taylor, B.; Robertson, C. SARS-CoV-2 Delta VOC in Scotland: Demographics, risk of hospital admission, and vaccine effectiveness. *Lancet* **2021**, *397*, 2461–2462. [[CrossRef](#)] [[PubMed](#)]
48. O'Toole, Á.; Hill, V.; Pybus, O.G.; Watts, A.; Bogoch, I.I.; Khan, K.; Messina, J.P.; Tegally, H.; Lessells, R.R.; Giandhari, J.; et al. Tracking the international spread of SARS-CoV-2 lineages B.1.1.7 and B.1.351/501Y-V2 with grinch. *Wellcome Open Res.* **2021**, *6*, 121. [[CrossRef](#)]
49. Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R.; et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948. [[CrossRef](#)]
50. Page, L.; Brin, S.; Motwani, R.; Winograd, T. *The PageRank Citation Ranking: Bringing Order to the Web*; Stanford InfoLab: Stanford, CA, USA, 1999.
51. Hagberg, A.; Swart, P.; S Chult, D. *Exploring Network Structure, Dynamics, and Function Using NetworkX*; Los Alamos National Lab.(LANL): Los Alamos, NM, USA, 2008.
52. Zheng, W.; Zhang, C.; Li, Y.; Pearce, R.; Bell, E.W.; Zhang, Y. Folding non-homologous proteins by coupling deep-learning contact maps with I-TASSER assembly simulations. *Cell Rep. Methods* **2021**, *1*, 100014. [[CrossRef](#)]
53. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **2020**, *17*, 261–272. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.