International Journal of Recent Engineering Research and Development (IJRERD)
ISSN: 2455-8761
www.ijrerd.com || Volume 10 – Issue 01 || Jan - Feb 2025 || PP. 01-05

Sequence Characteristics of Surface Glycoprotein Coding Gene in SARS-CoV-2 (Lineage: KP.3)

Gun Li, Ting Wei
Department of Biomedical Engineering, School of Electronic Information Engineering, Xi'An Technological University, Xi'An, Shaanxi, China

Abstract: At the end of 2019, a pneumonia outbreak caused by the SARS-CoV-2 had brought great harm to humans. The SARS-CoV-2 was highly pathogenic and spreads rapidly, triggering a serious global public health event. SARS-CoV-2 belongs to the family of corona viruses and has a high similarity to the SARS-CoV virus that emerged in 2003, with a sequence consistency of 79%. Recently, new variant of SARS-CoV-2 Lineage: KP.3 emerged in many areas. In order to study genetic characteristics of the surface glycoprotein coding gene in this new variant, codon usage pattern including the relative synonymous codon usage, base compositions, and the difference of them to humans were calculated. Further, the accessible surface area was calculated via its protein sequence. The results showed that there is obvious codon usage bias in the KP.3 variant of SARS-CoV-2, and that there are some accessible surface areas in the protein sequence.
Keywords: surface glycoprotein, codon usage pattern, SARS-CoV-2, KP.3 variant

1. Introduction

The genome of SARS-CoV-2 is about 30 kb and encodes about 9,860 amino acids. The surface glycoprotein, which has been identified as the most immunogenic protein of SARS-CoV-2, is expressed in the early phases of SARS-CoV-2 infection [1]. SARS-CoV-2 is more infectious than other corona viruses mainly because the S protein of the former has a stronger affinity for ACE2 on the surface of human cells. Surface protein is an antigen that mainly induces neutralizing antibodies of the novel corona virus, and the primary studies focused on the membrane protein or the surface protein [2]. More and more *in-silico* methods were used in this study, and the results showed that, from the perspective of secondary structure, surface glycoprotein remains the main target of many anti-corona virus drugs [3], and researches on its structure will further contribute to the development of more effective antiviral drugs. However, some new emerging lineages of SARS-CoV-2 are rapidly producing substitutes through point and recombination mutations that escape neutralizing antibodies [4]. Mapping of its genetic characteristics and studying its protein sequence characteristics will not only help us understand how the novel corona virus antigen evolves, but also evaluate whether immune escape occurs in new lineages. Some researches have already explored the characteristics of the KP.3 variant, e.g. it displayed increased immune evasion [5]. In the present study, genetic characteristics of the surface glycoprotein coding gene in the KP.3 variant, relative synonymous codon usage, base compositions, and the protein sequence characteristics were calculated and compared.

2. Methods

The gene sequences were searched by "SARS-CoV-2 KP complete" and the search results were further filtered by lineage number checked to ensure the gene sequences used in the present study are all of KP.3 variants (see Table.1).

Table 1: SARS-CoV-2 KP.3 surface glycoprotein coding sequences considered in the present study

Accession Number	Lineage	Accession Number	Lineage
PQ634678.1	KP.3.1.1	PQ667801.1	KP.3.1.1
PQ287374.1	KP.3.1.1	PQ667802.1	KP.3.1.1
PQ287400.1	KP.3	PQ766552.1	KP.3.1.1
PQ634677.1	KP.3.1.1	PQ766554.1	KP.3.1.1

The relative synonymous codon usage (RSCU) and the base compositions of SARS-CoV-2 KP.3 surface glycoprotein coding sequences were counted; the differences between the RSCU and the base composiotions of surface glycoprotein coding sequences and humans were further calculated and compared visually. The accessible surface area of surface glycoprotein (XLQ25762.1) linked from PQ766552.1 was calculated to explore the accessible characteristics.

3. Results

3.1 RSCU of the KP.3 surface glycoprotein coding sequences

The RSCU values of the SARS-CoV-2 KP.3 surface glycoprotein coding sequences (8 sequences shown in the Table 1) were calculated, and the result is shown in the Table 2.

Table 2: RSCU values of the SARS-CoV-2 KP.3 surface glycoprotein coding sequences

Amino Acid	Codon	Number	RSCU	Amino Acid	Codon	Number	RSCU
Phe	UUU	472	1.494	Tyr	UAU	344	1.564
	UUC	160	0.506		UAC	96	0.436
Leu	UUA	223	1.565	Ter	UAA	8	3
	UUG	152	1.067		UAG	0	0
	CUU	304	2.133	His	CAU	104	1.444
	CUC	88	0.618		CAC	40	0.556
	CUA	72	0.505	Gln	CAA	344	1.458
	CUG	16	0.112		CAG	128	0.542
Ile	AUU	344	1.675	Asn	AAU	392	1.181
	AUC	120	0.584		AAC	272	0.819
	AUA	152	0.74	Lys	AAA	359	1.301
Met	AUG	112	1		AAG	193	0.699
Val	GUU	376	2.022	Asp	GAU	352	1.419
	GUC	160	0.86		GAC	144	0.581
	GUA	112	0.602	Glu	GAA	272	1.447
	GUG	96	0.516		GAG	104	0.553
Ser	UCU	289	2.206	Cys	UGU	224	1.4
	UCC	80	0.611		UGC	96	0.6
	UCA	209	1.595	Ter	UGA	0	0
	UCG	24	0.183	Trp	UGG	104	1
Pro	CCU	208	1.891	Arg	CGU	80	1.5
	CCC	24	0.218		CGC	8	0.15
	CCA	208	1.891		CGA	8	0.15
	CCG	0	0		CGG	16	0.3
Thr	ACU	352	1.833	Ser	AGU	144	1.099
	ACC	72	0.375		AGC	40	0.305
	ACA	312	1.625	Arg	AGA	128	2.4
	ACG	32	0.167		AGG	80	1.5
Ala	GCU	328	2.13	Gly	GGU	360	2.195
	GCC	64	0.416		GGC	120	0.732
	GCA	208	1.351		GGA	144	0.878
	GCG	16	0.104		GGG	32	0.195

The results show that some codons, e.g. UGA, UAG and CCG, are not used in the sequences. Only the UAA is used as the terminal codons in these sequences. Further, the base compositions were counted (Figure 1) and the result shows the coding sequences are AU-biased.

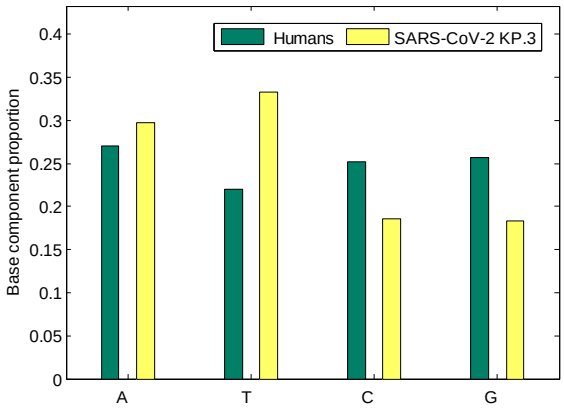

Figure 1: Base compositions of the SARS-CoV-2 KP.3 surface glycoprotein coding sequences

Further, the distance between the sequences and humans were calculated, and the results are shown in the Figure 2.

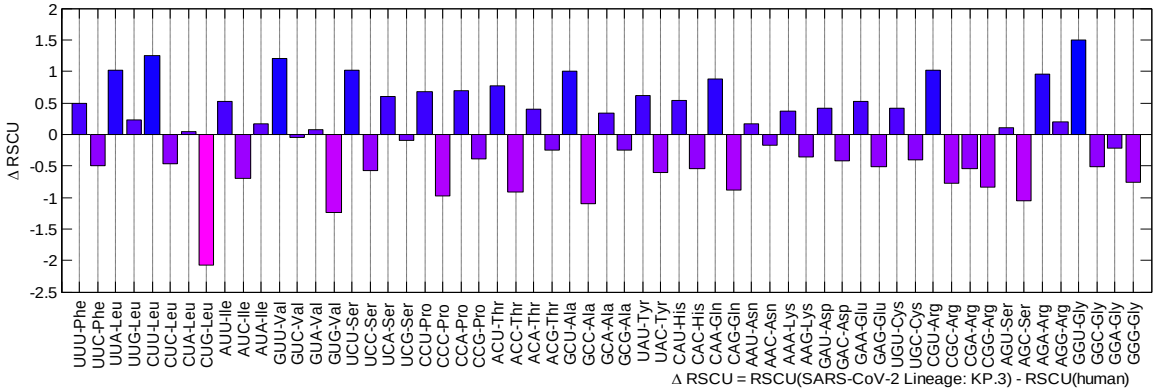

Figure 2: Comparison of the RSCU values of surface glycoprotein coding sequences with humans

3.2 Amino acids used in the surface glycoprotein sequences

Amino acids used in the surface glycoprotein sequences were compared and the results are shown in the Figure 3.

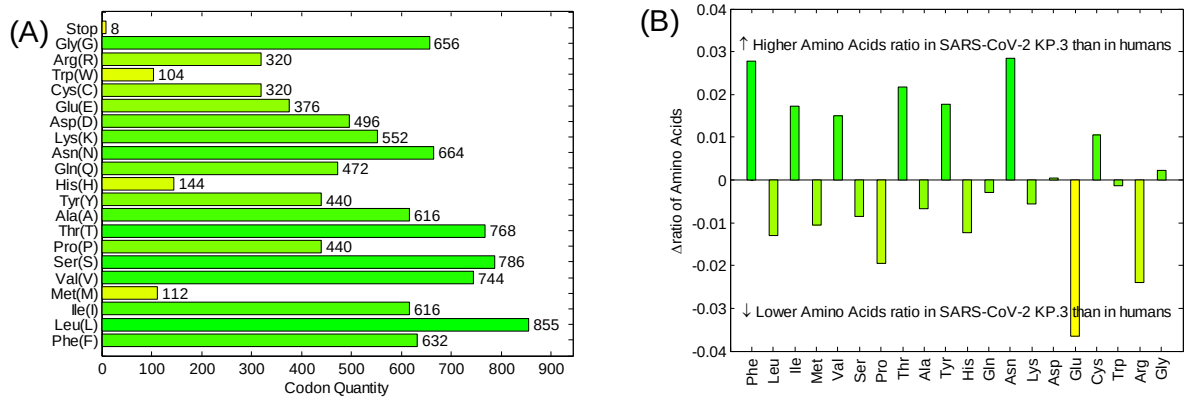

Figure 3: Amino acids used in the surface glycoprotein sequences. (A) The number of amino acids used in the 8 SARS-CoV-2 KP.3 surface glycoprotein coding sequences, (B) difference of the Amino acids used between surface glycoprotein coding genes and humans.

3.3 Accessible surface area of the S protein sequence

The accessible surface areas of surface glycoprotein, take the sequence XLQ25762.1 linked from PQ766552.1 as an example, was calculated to explore the accessible characteristics. The accessible surface areas are denoted by the green region, see the Figure 4.

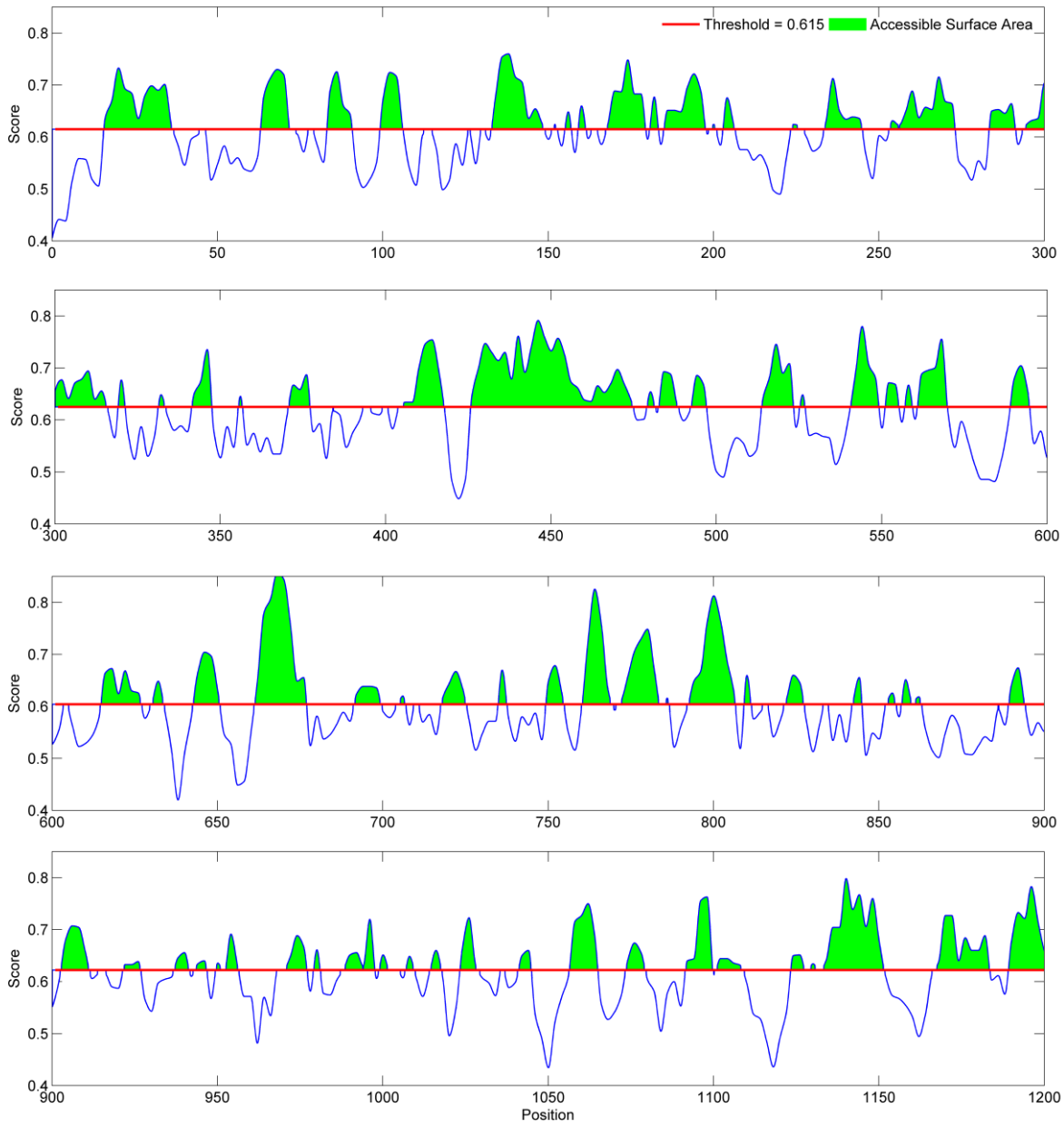

Figure 4: Accessible surface area of the surface protein in SARS-CoV-2 KP.3 variant

4. Discussion

The study of coding gene stability is helpful to understand the effect of virus variation on vaccine-induced immune response, so as to guide the optimization and improvement of vaccines [6-8]. If the coding gene of in SARS-CoV-2 KP.3 variant is stable, the vaccine designed based on the virus strain may effective; and if the gene is unstable, the vaccine needs to be adjusted according to the mutation of the virus. Previous studies explored many genetic characterizations of respiratory syncytial virus surface glycoproteins [9] and the SARS-CoV-2 surface glycoprotein alignments [10], etc [11, 12].

Because of the important negative function of SARS-CoV-2 surface glycoprotein in human body [13], the stability of glycoprotein genes is closely related to the transmissibility and pathogenicity of viruses. In the present study, the results show the codon usage of surface glycoprotein in SARS-CoV-2 KP.3 variant is different from that of humans obviously, and that amino acid usage is also different from that of humans. The research results of the surface glycoprotein encoding gene of the novel corona virus can be used to evaluate the transmissibility of the virus. The variation of some key sites may lead to changes in the structure of

glycoprotein, which in turn affect the ability of the virus to bind to the host cell, invasion efficiency and transmission characteristics. By studying the genetic stability, the transmission risk and pathogenic potential of the mutated strains can be better evaluated.

Acknowledgments

The authors thank the Science and Technology Department of Shaanxi Province for funding - key research and development project from the Science and Technology Department of Shaanxi Province (project code: 2023-YBGY-261).

References

[1]. Daichi Yamasoba, K. Uriu, Arnon Plianchaisuk, Y. Kosugi, Lin Pan, J. Zahradník, et al. Virological characteristics of the SARS-CoV-2 KP.3, LB.1, and KP.2.3 variants. The Lancet Infectious Diseases, 2024, 24(8): e482-e483.

[2]. Weizhu Yan, Yanhui Zheng, Xiaotao Zeng, Bin He, Wei Cheng. Structural biology of SARS-CoV-2: open the door for novel therapies", Signal Transduction and Targeted Therapy, 2022, 7: 26.

[3]. Sarmad Frogh Arshada, Rehana Rehanab, Muhammad Asif Saleemc, Imran Ahmad Khanh, M. Ali Hayssami, Muhammad Anwarj, et al. Multi-epitopes vaccine design for surface glycoprotein against SARS-CoV-2 using immunoinformatic approach. Heliyon, 2024, 10(2): e24186.

[4]. Ashley L Taylor, Tyler N Starr. Deep mutational scanning of SARS-CoV-2 Omicron BA.2.86 and epistatic emergence of the KP.3 variant. Virus Evolution, 2024, 10(1): veae067.

[5]. Kun Xu, Yaling An, Xueyuan Liu, Haitang Xie, Dedong Li, Ting Yang, et al. Neutralization of SARS-CoV-2 KP.1, KP.1.1, KP.2 and KP.3 by human and murine sera. npj Vaccines, 2024, 9: 215.

[6]. Prerna Arora, Christine Happle, Amy Kempf, Inga Nehlmeier, Metodi V Stankov, Alexandra Dopfer-Jablonka, et al. Impact of JN.1 booster vaccination on neutralisation of SARS-CoV-2 variants KP.3.1.1 and XEC. The Lancet Infectious Diseases, 2024, 24(12): e732-e733.

[7]. Diego Gomes Teixeira, João Firmino Rodrigues-Neto, Dayse Caroline Severiano da Cunha, Selma Maria Bezerra Jeronimo. Understanding SARS-CoV-2 spike glycoprotein clusters and their impact on immunity of the population from Rio Grande do Norte, Brazil. Infection, Genetics and Evolution, 2024, 118: 105556.

[8]. Regina C. Ambe, Shubhang Bhalla, Alejandra Alvarado, Jose Barragan, Jorge Cervantes. Bacille-Calmette-Guerin modulates human macrophage and dendritic cell response to SARS-CoV-2 S-glycoprotein. Infectious Medicine, 2023, 2(3): 241-245.

[9]. Yu Ping Fang, Chun Chin Chang, De Wei Lai, Chun Yi Lee. Genetic characterization of respiratory syncytial virus surface glycoproteins F and G in Taiwan, 2017–2021. Journal of Microbiology, Immunology and Infection, 2024, 57(4): 564-572.

[10]. Done Stojanov. Data on multiple SARS-CoV-2 surface glycoprotein alignments. Data in Brief, 2021, 38: 107414.

[11]. Samuel L. Freeman, A. Sofia F. Oliveira, Andrea E. Gallio, Annachiara Rosa, Maria K. Simitakou, Christopher J. Arthur, et al. Heme binding to the SARS-CoV-2 spike glycoprotein. Journal of Biological Chemistry, 2023, 299(8): 105014

[12]. Evgenii V. Matveev, Gennady V. Ponomarev, Marat D. Kazanov. Genome-wide bioinformatics analysis of human protease capacity for proteolytic cleavage of the SARS-CoV-2 spike glycoprotein. Microbiology Spectrum, 2024, 12(2): 3.

[13]. Vittoria Monaco, Ilaria Iacobucci, Luisa Canè, Irene Cipollone, Veronica Ferrucci, Pasqualino de Antonellis, et al. SARS-CoV-2 uses Spike glycoprotein to control the host's anaerobic metabolism by inhibiting LDHB. International Journal of Biological Macromolecules, 2024, 278: 134638.