Diagnostic Pathology: Open Access

Superantigens, Antigen Precision and Their Application in Coronavirus

Peijun Zuo[1]*, Longlong Zuo[2], Zhihong Li[1] and Liping Li[1]

[1]*Injury Prevention Research Center, Shantou University Medical College, Shantou, China*
[2]*Department of Medicine and Therapeutics, Chinese University of Hong Kong, Hong Kong, China*

Abstract

Aim: The aim of this study was to find a method to identify epidemic and sporadic pathogen strains. The method includes A. Detecting the antigenic amino acid sequence of the pathogen strain to determine the primary superantigen. B. If the primary superantigen contains tryptophan, the pathogen strain is an epidemic strain. C. If the primary superantigen does not contain tryptophan amino acid, it is a sporadic strain. D. Determine the outbreak time interval between the pathogen strain and the reference strain based on the increase in antigen precision between the pathogen strain and the reference strain. The method can predict whether a series of infectious diseases will become epidemic or scattered, as well as the date of outbreak, such as Severe Acute Respiratory Syndrome 1 (SARS I), Middle East Respiratory Syndrome (MERS) and Coronavirus Disease 2019 (COVID-19, SARS II).

Keywords: Superantigens; Antigen precision; SARS; Protein; In-population; Antibody

Introduction

Pathogens are the viruses, bacteria and parasites causing outbreaks around the world. Some infectious agents can cause global pandemics, while others spread only locally. For example, within the same coronavirus family, the prevalent strains causing global pandemics are the Severe Acute Respiratory Syndrome 1 (SARS I) strain and the coronavirus disease 2019 (COVID-19, SARS II) coronavirus strain, which causes Middle East Respiratory Syndrome (MERS) is only found in the Middle East. How to identify epidemic strains and sporadic strains is urgently needed and of great significance.

Existing technology can only identify epidemic strains of cholera groups O1 and O139 [1]. The existing technology can only be used for epidemic strains of *vibrio cholerae* and cannot identify other bacteria, nor can it identify epidemic strains and scattered strains of viruses. The main reason is the lack of understanding of the protein basis between epidemic and non-epidemic strains.

The purpose of this study was to provide a method for identifying epidemic and sporadic pathogen strains. Another aim of this study was to provide a method for determining the date of outbreak of the causative strain.

Materials and Methods

Method for determining superantigens

Using glycine as the node, the amino acid sequence of the antigen is cut with glycine to obtain the longest multiple amino acid peptide chain, which is the superantigen. The shortest amino acid peptide chain among superantigens is the primary superantigen, and the longest amino acid peptide chain is the last superantigen. Coronaviruses, the shortest amino acid peptide chain is the primary superantigen, the second shortest amino acid peptide chain is the secondary superantigen, the third shortest amino acid peptide chain is the tertiary superantigen, and the longest amino acid peptide chain is the quaternary superantigen.

Through global population studies, we identified 4 such candidate viral superantigens, and all were isolated from global populations. The dormant periods of these candidates, or the time intervals between global pandemics caused by them, were searched and analyzed among real viruses hosted and spread among all human populations around the world [2,3].

Calculation of antigen precision

Calculate the standard deviation of the molecular weight of amino acids in a superantigen to obtain its antigenic precision. Each superantigen fragment contains amino acids with different molecular weights. Excel was used to calculate the mean and Standard Deviation (SD) of amino acid molecular weights. The SD of amino acid molecular weight is an indicator of antigen precision.

Outbreak dates and protein sequences of corona virus.

We will study superantigens in all populations around the world. Coronavirus has been circulating in the entire population for 20 years, and some protein superantigens have mutated and broken out on different dates. This information provides the opportunity to study whether antigen precision correlates with the date of the outbreak to which it contributes. On the website https://www.ncbi.nlm.nih.gov/, select "Proteins in all databases". Enter access number to search for coronavirus protein sequences.

On 16 November, 2002, Severe Acute Respiratory Syndrome 1 (SARS I) was outbroken for epidemic led by primary super antigen SARS I D614: On 16 November, 2002, SARS I broke out in Guangdong Province, China [4,5]. The sequence of protein AAP72986 contains the spike protein in wild SARS I. It was the first coronavirus to break out. The first case is a 45-year-old man who has infected eight people. There are equal numbers of cases from the second and third generations. It quickly spread to Southeast Asia and Canada. Before the emergence of Middle East Respiratory Syndrome (MERS) and SARS II, it was an emerging, highly pathogenic coronavirus that infected more than 8,000

people and killed nearly 800 between November, 2002 and September, 2003 [4,5]. SARS I spread to more than 29 countries and regions, with a death rate of 9.6% [5].

On 1 September, 2012, MERS was emerged leading by primary super antigen MERS D614: Ten years later, on 1 September, 2012, middle east respiratory syndrome broke out. The MERS coronavirus was first recorded in the Kingdom of Saudi Arabia (KSA) in 2012 and later in 2014, 180 cases were detected in two years with a death rate of 43%. The sequence of protein QJX19948 contains the spike protein in wild MERS. This is the second coronavirus outbreak. Compared with the 8,000 cases of SARS I, MERS cases were scattered outbreaks in local areas.

On 12 December, 2019, first wave of Corona Virus Disease 2019 (COVID-19) was caused pandemic leading by primary superantigen SARS II D614: On 12 December, 2019, 17 years after the SARS I outbreak, the first wave of COVID-19 was witnessed, caused by the SARS II D614 primary superantigen [6]. This strain was used as a reference strain. 12 December, 2019, is used as the reference date for the outbreak of the reference strain. SARS II D614 was used as the reference superantigen. The amino acid sequence of the spike protein QHD43416 is in wild SARS II D614, the third coronavirus outbreak.

COVID-19 is a devastating disease that, despite the availability of a vaccine, can still kill those infected. The disease is caused by the SARS II coronavirus, a RNA-positive virus that has been divided into four major pandemic waves (D614, D614G, Delta, and Omicron) [6].

The original strain contains 4 superantigens. They are the primary superantigen D614, 37 amino acids; the second superantigen N148, 38 amino acids; the tertiary superantigen I358, 41 amino acids; and the quaternary superantigen F718,43 amino acids [7].

On 10 March, 2020, second wave of COVID-19 was caused pandemic leading by sub-primary super antigen SubD614G: However, on 10 March, 2020, during the epidemic, after 2.9 months of stability, the D614 superantigen did mutate to D614G and became dominant [8], and caused the second wave of SARS II pandemic. The amino acid sequence of protein "7KDK_A" contains mutated SARS II D614G. This is the fourth coronavirus outbreak. The D614G mutation remains within the delta and omicron waves.

In October, 2020 third wave of COVID-19 was caused pandemic leading by secondary super antigen SARS II N148: In October, 2020 after 10 months of stability, the N148 secondary superantigen mutated into mutated N148 (E156del, F157del, R158G). The amino acid sequence of protein "7V8B_A" contains the SARS II Delta subtype, with complete mutations of N148 (E156del, F157del, R158G), T19R, G142D, L452R, T478K, D614G, P681R and D950N. This is the fifth coronavirus outbreak. These N148 mutations caused the SARS II Delta subtype pandemic, the third wave of the SARS II pandemic in the world [9].

On 23 November, 2021 fourth wave of COVID-19 caused pandemic leading by tertiary super antigen SARS II I358: On 23 November, 2021, after 23.4 months of stability, the I358 superantigen finally mutated into the SARS II Omicron fragment, triggering the fourth wave of the SARS II pandemic. The amino acid sequence of protein 7QO9_a contains Omicron isoforms. This is the sixth epidemic of coronavirus. Initial cases of SARS II Omicron subtype reported to begin pandemic in South Africa [10].

Plot the antigenic precision of super antigens and their stability

Excel was used to plot the correlation between the antigenic

precision of superantigens and their stability.

Plot the antigenic precision increased amounts and their stability, or the time interval of a global pandemic to the reference date

Calculate the antigenic precision of the superantigen responsible for the outbreak. Compared with the reference antigen precision of superantigen SARS II D614, if the antigen precision increase occurred later in time, the antigen precision was recorded as a positive increase, otherwise, it was recorded as a negative increase. Excel was used to plot the increase in antigenic precision for superantigens versus the time interval between their outbreak date and the reference strain's outbreak date.

Find correlation among super antigen stability y, antigenic precision increased amounts x1, and if contains tryptophan (W) x2

If the superantigen contains tryptophan (W), the value of x_2 is 1, otherwise, the value is 0.

Results

In the past 20 years, from the SARS I outbreak on 16 November, 2002, to the MERS outbreak on 1 September, 2012, to the SARS II D614 outbreak on 12 December, 2019, to the SARS II D614G outbreak on 10 March, 2020, by the outbreak of SARS II Delta in October, 2020 and by the outbreak of SARS II Omicron on 23 November, 2021, there have been a total of 6 outbreaks of human coronavirus infection in all mankind. Whether these 6 outbreaks are programmed by superantigens of the coronavirus requires analysis.

Antigenic precision is a characteristic of superantigens. If the superantigen enters the human body, responding antibodies will be produced. Some antibodies can neutralize its superantigens, protecting people from reinfection. Normally, you won't experience any unpleasant sensations during the neutralization process.

To identify the superantigen of the coronavirus, we will analyze the spike protein sequence of the Wuhan wild strain SARS II D614 (accession number, QHD43416). SARS I, MERS, SARS II D614, SARS II D614G, SARS II Delta, SARS II Omicron superantigens are also analyzed to calculate superantigen precision [5]. All coronavirus spike proteins contain four superantigens: D614, N148, I358, and F718. The results are shown in Table 1.

The first outbroken corona virus SARS I caused by primary super antigen SARS I D614, which was G-free peptide with tryptophan (W) and 37 amino acids in length

The primary superantigen in the original strain, SARS I D614, caused the SARS I epidemic. As shown in Table 1, the average molecular weight of the amino acids in the SARS I D614 superantigen is 126.8, and the standard deviation (SD) of the amino acid molecular weights is 26.98. Compared with the reference SD of 25.23 for SARS II D614, which appeared later in time, the SD of the later SARS I D614 decreased, resulting in increase of -1.75, as shown in Table 2.

This peptide contains the largest amino acid, tryptophan (W, molecular weight, 204.2262). This G-free fragment contains the D614 amino acid in the original strain, which triggered the epidemic in Guangdong, China on 16 November, 2002. Compared with the outbreak date of the reference strain SARS II D614 on 12 December, 2019, the difference between the two outbreak dates was 204.9 months, as shown in Table 2.

SARS I secondary superantigen SARS I N148 does not contain tryptophan, the largest amino acid. The precision of this SARS I N148 fragment is 23.17, as shown in Table 1.

The second outbroken corona virus was MERS caused by primary superantigen MERS D614, which was G-free peptide without tryptophan (W) and with 29 amino acids in length

As shown in Table 1, MERS D614 does not contain tryptophan, the largest amino acid. The Standard Deviation of its amino acid molecular weight is 28.11 Da. Compared with the reference SD of SARS II D614

of 25.23 Da, the SD of the latecomer SARS II D614 has decreased compared with the former, recording an increase of -2.88, as shown in Table 2.

On 1 September, 2012, the original strain in the Middle East contained the superantigen MERS D614, leading to the second outbreak of the coronavirus anywhere in the world. Since it has not spread to other regions, it is not a global epidemic, but rather a local outbreak. Compared with the reference outbreak date of 12 December, 2019 for the reference superantigen SARS II D614, the two outbreak dates differ by 87.37 months, as shown in Table 2.

	2002 SARS I	2012 MERS	2019 SARS II
Super antigen	D614	D614（Potential）	D614
Number of amino acids	37	29	37
Mean of molecular weight	126.8	133.6	129.3
Standard Deviations (SD)	26.98	28.11	25.23
Contains tryptophan (W)	1 (Yes)	0 (No)	1 (Yes)
Super antigen	N148（Potential）	N148	N148
Number of amino acids	31	37	38
Mean of molecular weight	132.4	142.7	140.1
Standard Deviation ns (SD)	23.17	25.2	26.92
Contains tryptophan (W)	0 (No)	1 (Yes)	1 (Yes)
Super antigen	I358	I358（Potential）	I358
Number of amino acids	41	70	41
Mean of molecular weight	136	129.6	134.2
Standard Deviations (SD)	29.08	23.89	31.66
Contains tryptophan (W)	1 (Yes)	0 (No)	1 (Yes)
Potential super antigen	F718	F718	F718
Number of amino acids	43	33	43
Mean of molecular weight	124.7	133.1	125.3
Standard Deviations (SD	21.66	23.89	20.69
Contains tryptophan (W)	0 (No)	0 (No)	0 (No)

Table 1: The molecular weight and Standard Deviations (SD) or antigenic precision of super antigens D614, N148, I358 and F718.

Super antigens	Antigen precision	Antigen precision increased	Stability (Ms)
SARS I D614	26.98	-1.75	204.9 (SARS I)
SARS II D614	25.23	SARS II D614 is a referenced super antigen	-
SARS II D614G	26.74	D614G was broken in two parts by G	-
SubD614G	27.11	1.88	2.9 (D614G)
SARS II N148	26.92	1.69	10 (Delta)
SARS II I358	31.66	6.43	23.4 (Omicron)
MERS	28.11	-2.88	87.37

Table 2: The super antigen precision increased amounts and its stability (Compared with reference SARS II D614).

The third outbreak corona virus was reference SARS II D614, caused by primary super antigen SARS II D614, which was G-free peptide with tryptophan (W) and 37 amino acids in length

As shown in Table 1, the average molecular weight of the superantigen SARS II D614 is 129.3 Da, and its amino acid molecular weight SD is 25.23 Da. This superantigen contains the largest amino acid tryptophan. The amino acid sequence of SARS II D614 is TNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYST, with a total length of 37 amino acids. On 12 December, 2019, the SARS II D614 superantigen in the original strain of the virus in Wuhan, China led the first wave of the SARS II epidemic and the third outbreak of the coronavirus. The precision of this superantigen is 25.23 Da, as shown in Table 1.

The fourth outbreak corona virus SARS II D614G, in Italian strain, led second wave of SARS II pandemic

SARS II D614 has been confirmed to have undergone a subsequent mutation. It mutated into SARS II D614G on 10 March, 2020, first appearing in Italy and leading the second wave of the SARS II pandemic. Since the first outbreak of the reference strain on 12 December, 2019, the conservative months for SARS II D614 have been 2.9 months. Therefore, the stability of SARS II D614 is 2.9 months. Due to the mutation of SARS II D614 superantigen to SARS II D614G, the original superantigen was cleaved into two parts by glycine G. The remaining 24 amino acids containing W were immediately used as a new superantigen SubD614G. The precision of the new superantigen has been increased to 27.11. Compared with the reference SD 25.23 of SARS II D614, the SD of SubD614G increased and was recorded as an increase of 1.88, as shown in Table 2. The SARS II D614G mutation was preserved in the SARS II Delta and SARS II Omicron subtypes.

The fifth outbreak corona virus SARS II N148, in Indian strain, led third wave of SARS II, SARS II Delta pandemic

Another superantigen is the second superantigen SARS II N148, which has a length of 38 amino acids and is in the second position after the first superantigen SARS II D614 in the superantigen chain. The detailed amino acid sequence is VYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE. SARS II N148 contains "W". The superantigen has also been confirmed as a mutated antigen in the SARS II delta variant. SARS II mutated in India in October, 2020 and triggered the third wave of SARS II epidemic. Therefore, the stability of SARS II N148 is 10 months from the initial outbreak date of the reference strain on 12 December, 2019. The precision of this superantigen before mutation was 26.92, higher than the precision of SARS II D614, as shown in Table 1. Compared with the reference SD 25.23 of SARS II D614, the SD of SARS II N148 increased by 1.69, as recorded in Table 2.

The sixth outbreak corona virus SARS II I358, in South African strain, led fourth wave of SARS II, SARS II Omicron pandemic

Another identified superantigen is SARS II I358, with a length of 41 amino acids. The amino acid sequence of the antigen is EVFNATRFASVYAWNRKRISNCVADYNSASFSTFKCY. The superantigen has been confirmed as the Omicron variant of SARS II. The superantigen I358 contains "W". The tertiary superantigen SARS II I358 mutations have been confirmed by the Omicron subtype.

The tertiary superantigen mutated in South Africa on 23 November, 2021, leading to the fourth SARS II outbreak, the SARS II Omicron pandemic. The stability of SARS II I358 is 23.4 months. Since the first outbreak of the reference strain on 12 December, 2019, it has maintained conservatism for 23.4 months. The precision of the antigen before mutation was 31.66, which is higher than the precision of the reference superantigen SARS II D614, as shown in Table 1. Compared with the reference SD 25.23 of SARS II D614, the SD of SARS II I358 increased by 6.43, as recorded in Table 2.

There is a dose-dependent relationship between the super antigenic stability and antigenic precision in SARS II D614, SARS II Delta, and SARS II Omicron

As shown in Figure 1, the antigen standard deviation and antigen precision of the reference superantigen SARS II D614 are 25.23 and 25.23, respectively. The conservative time for SARS II D614 mutation to SARS II D614G is 2.9 months, and the stability is 2.9 months. The antigen standard deviation of the superantigen SARS II N148 is 26.92, and the antigen precision is 26.92. The conservative time for the wild-type superantigen SARS IIN148 mutation to Delta type SARS IIN18 (E156del, F157del, R158G) is 10 months, and the stability is 10 months.

Similarly, the antigen SD and antigen precision of the tertiary superantigen SARS II I358 were 31.66 and 31.66, respectively. If x represents the antigen precision of the superantigen and y represents the stability of the superantigen before its mutation, the regression equation is $y=88.467 \ln(x) -282.05$. The correlation coefficient R is 0.998. After statistical testing, the probability of Class I errors occurring is $P=0.03$ and <0.05. There is a dose-dependent relationship between the antigen precision and stability of superantigens, which is statistically significant.

The actual occurrence date of the SARS II D614G mutation is 2.9 months after the outbreak of the reference strain, and the predicted mutation date is 3.5 months later. Similarly, the actual date of the SARS II Delta variant is 10 months later, while the predicted date of the variant is 9.3 months later. The actual date of SARS type II Omicron mutation is 23.4 months later, and the predicted mutation date is 23.6 months later. The maximum difference between a true mutation and a predicted mutation is 3 weeks. That is to say, the stability of superantigens can be calculated by their amino acid sequences.

There is a dose-dependent relationship between the stability or outbroken time interval between the pathogen strain and the reference strain, y, and antigenic precision increased amounts between the pathogen strain and the reference strain, x, in corona virus.

As shown in Table 2 and Figure 2, the antigen SD of the superantigen SARS I D614 is 26.98. The precision of the reference superantigen is 25.23. The precision of the antigen decreased from SARS I D614 to SARS II D614. The increase in antigen precision relative to the reference strain is -1.75. The conservative time for SARS I D614 is 204.9 months.

The conservative time for SARS II D614 mutation to SARS II D614G is 2.9 months, and the stability is 2.9 months. The superantigen SARS II D614 is broken down into two parts by glycine. The superantigen is immediately converted to sub D614G. It is sub D614G, leading the SARS II D614G pandemic.

The antigen standard deviation of sub D614G increased to 27.11. Compared with the reference SD 25.23 of SARS II D614, the SD of sub D614G increased and was recorded as an increase of 1.88, as shown in Table 2.
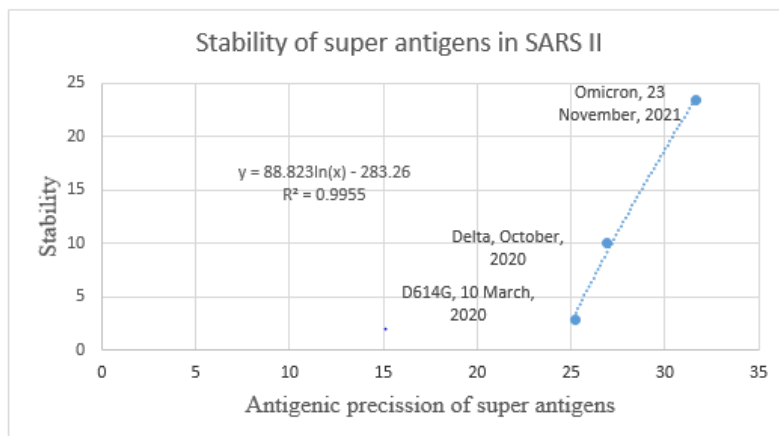
Figure 1: Dose-dependent relationship between stability of super antigen and antigen antigenic precision in SARS II. There is a dose-dependent relationship between the stability of superantigens and the precision of antigens. The probability of Type I errors occurring is P=0.03, which is less than 0.05 and has statistical significance. The predicted mutation sequence is consistent with actual history.
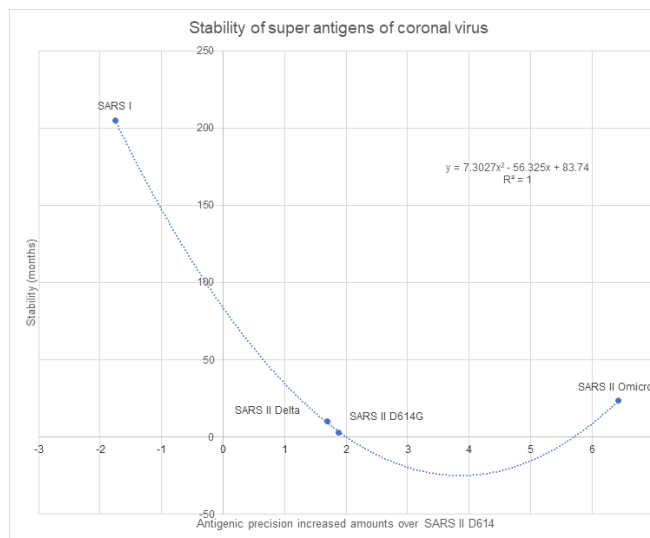
Figure 2: Dose-dependent relationship between stability of super antigen and antigenic precision increased amounts. There is a dose-dependent relationship between the stability of superantigens and the increase in antigen precision. The probability of class I errors occurring is P<0.03, which is statistically significant. The predicted mutation sequence and time are consistent with the actual history. The function of superantigens is transmitted from the primary to the secondary, and then to the tertiary. From SARS I to MERS, to SARS II D614, to SARS II D614G, to SARS II N147 or SARS II Delta, and then to SARS II I358 or SARS II Omicron, this programmatic propagation process lasted for 20 years, highly consistent with mathematical models.

The antigen precision of wild-type SARS IIN148 is 26.92, which is 1.69 higher than the antigen precision of the reference strain SARS IID614. The conservative period for wild-type SARS II N148 mutations to Delta type SARS II N148 (E156del, F157del, R158G) is 10 months, and the stability is 10 months.

Similarly, the antigen SD of the superantigen SARS II I358 is 31.66, and the antigen precision is 31.66. The increase in antigen precision compared to the reference strain SARS II D614 is 6.43. The conservative time for wild-type SARS II I358 mutation to Omicron SARS II I358 is 23.4 months, and the stability is 23.4 months.

If x represents the increase in antigen precision of the superantigen, and y represents the stability of the superantigen, that is, its mutation to the reference strain or the required number of months from the reference strain, then the regression equation is $y=7.3027x^2-56.325x+83.74$, with a correlation coefficient of R=1. After statistical testing, the probability of Class I errors occurring is P<0.02. There is a dose-dependent relationship between the improvement of antigen precision and the stability of superantigens, which is statistically significant.

Correlation among super antigen stability y, antigenic precision increased amounts x1, and if contains tryptophan (W) x2. If it contains tryptophan (W), the value of x2 is 1, otherwise it is 0.

If x1 represents the increase in antigen precision of the superantigen, x2 represents whether the superantigen contains tryptophan, and y represents the stability of the superantigen when mutated to or from the reference strain, the regression equation is $y=7.3027x1^2-56.325x1+222.9195x2-139.1754$, with a correlation coefficient of R=1. After statistical testing, the probability of Class I errors occurring is P<0.02. The increase in antigen precision of superantigens (whether or not they contain tryptophan W), as well as the presence of tryptophan in superantigens, has a dose-dependent relationship with the stability of superantigens, which is statistically significant.

Discussion

The first outbreak corona virus SARS I caused by super antigen SARS I D614, which was G-free peptide with tryptophan (W) and 37 amino acids in length

The superantigen chain in coronavirus is composed of D614, N148, I358, and F718. SARS I D614 was the primary trigger for the outbreak of SARS I. Only one superantigen, SARS I D614, can sustain the epidemic for 9 months. For infectious diseases with longer epidemics, it is necessary to transfer the role of the primary superantigen to the secondary trigger superantigen SARS I N148. However, SARS I N148 lacks tryptophan (W). If there were no tryptophan, the SARS I N148 superantigen would not have been triggered, and the epidemic would have stopped. Although the tertiary superantigen SARS I I358 is complete, without the secondary superantigen, the primary SARS I D614 cannot be directly transmitted to the tertiary superantigen. So, SARS I can only be prevalent for 9 months.

The antigen precision of the primary superantigen SARS I D614 is 26.98. Compared with SARS II D614, the precision increase is x1=-1.75, and the tryptophan state is x2=1. The regression equation for the stability of the superantigen SARS I D614 is $y=7.3027x1^2-56.325x1+222.9195x2-139.1754=204.9$ months from SARS II D614. On the same day, SARS I broke out on 16 November, 2002. Both x1 and x2 are internal genetic factors that caused the outbreak of SARS I D614.

The precision of the superantigen of SARS I N148 is 23.17. Compared with SARS II D614, the increase in precision is x1=2.06, and the tryptophan state is x2=0. The regression equation for the stability of the superantigen SARS I N148 is $y=7.3027x1^2-56.325x1+222.9195x2-139.1754=-224.2$ months. The time of the negative month does not exist. Therefore, the SARS I N148 superantigen was not stimulated.

The second outbreak corona virus MERS caused by primary super antigen MERS D614, which was G-free peptide without tryptophan (W) and 29 amino acids in length.

The superantigen chain in coronavirus is composed of D614, N148, I358, and F718. D614 was the primary trigger for the global outbreak of the pandemic. MERS, its MEARS D614 does not contain tryptophan (W). If there was no tryptophan, the primary superantigen of MERS D614 would not have been triggered and stopped circulating. This is a local outbreak of infectious disease that has not spread worldwide. Although MEARS N148 is complete, it cannot be triggered directly without the launch of the primary superantigen MERS D614.

The antigen precision of MERS D614 is 28.11, and compared with SARS II D614, the precision increase is x1=-2.88, and the tryptophan

state is x2=0. The regression equation for the stability of super antigen MERS D614 is $y=7.3027x1^2-56.325x1+222.9195x2-139.1754=87.37$ months. On the same day, MERS D614 broke out on 1 September, 2012. X1 and x2 are both internal genetic factors that program MERS D614 to erupt on a specific date.

The third outbreak corona virus SARS II D614, caused by super antigen SARS II D614, which was G-free peptide with tryptophan (W) and 37 amino acid in length.

As shown in Table 1, the antigen precision of SARS I D614 is 26.98, and the antigen precision of MERS D614 is 28.11, both higher than SARS II D614, 25.23. Among the fragments containing "W", the SD of SARS II D614 and 25.23 were the smallest among the superantigens D614, N148, and I358. This means that the state is "the roughest" and this location is the most likely place for mutations to occur. The site has been confirmed to have undergone a D614G mutation. The SARS II coronavirus with low antigen precision cannot spread in a short period of time after the previous epidemic, such as SARS I D614, which cannot even after a sporadic virus with high antigen precision, MERS D614. It must wait for the immune level of the entire population to gradually disappear.

If x represents the increase in antigen precision of the superantigen and y represents the stability of the superantigen before mutation, the regression equation is $y=7.3027x2-56.325x+83.74$, with a correlation coefficient of R=1.

From the equation, it can be seen that for mutated viruses with low antigen precision , such as the reference strain SARS II D614 with the lowest antigen precision, the minimum value of x is close to 0, indicating that the antibody level of the population is close to the antigen precision of the reference strain. Only then can the reference strain possibly evade immune monitoring in the population, which is the time of the reference strain outbreak. The minimum waiting time is 83.74 months. The waiting time is approximately 6.978 years.

In the SARS II D614 case, it waited for 17 years after the high antigen precision SARS I outbreak on 16 November, 2002, and 7.281 years after the high antigen precision MERS outbreak. Low antigen precision viruses require a minimum of 6.978 years to wait for immune decline in individuals who have been infected with similar viruses with high precision.

On 12 December, 2019, COVID-19 broke out in Wuhan, China. The original strain was SARS II D614. Three longest superantigens containing W were found in the SARS II D614 or QHD43416 sequences, with lengths greater than or equal to 37 amino acids. They are SARS II D614, SARS II N148, and SARS II I358, respectively.

For SARS II D614, the predicted outbreak date is 12 December, 2019. The predicted outbreak date of SARS II D614G is 10 March, 2020; For SARS II N148 or SARS II Delta, the predicted outbreak date is 12 October, 2020. For SARS II I358 or SARS II Omicron type, the predicted outbreak date is 23 November, 2021.

Even 17 years later, the predicted outbreak date will be the same as the actual outbreak date. Before the variation of COVID-19, all the superantigens SARS II D614, SARS II N148 and SARS II I358 existed independently. The prevalence and outbreak of coronavirus are driven by superantigen chains, with D614 as the primary antigen, N148 as the secondary antigen, I358 as the tertiary antigen, and F718 as the quaternary antigen.

The fourth outbreak corona virus SARS II D614G, in Italian strain, led second wave of SARS II pandemic

The second wave of SARS II was caused by a mutation from SARS II D614 to SARS II D614G. The SD of SARS II D614 is 25.23, which is the smallest among the four longest superantigens. It first mutated on 10 March, 2020, with the mutated strain being SARS II D614G. The relationship between antigens and antibodies is very similar to that of locks and keys. A high-precision key can open any lock with low precision and equal precision. In the same way, high-precision antigens generate high-precision antibodies, which can recognize and capture any low or equally precise antigen. The teeth of a high precision key are slightly smaller than those of a low precision key, otherwise it cannot enter the space of a low precision lock. For the same reason, the mutated high precision SARS II D614G is slightly smaller than SARS II D614, partly because G is the smallest amino acid, and partly because the human immune system may only recognize the amino acid between the two Gs as a superantigen. If D mutates to G, the human immune system needs to re recognize two smaller superantigens that are divided into two segments, such as SubD614G.

Why does the mutation occur in SARS II D614 instead of longer fragments, SARS II N148 or SARS II I358. The virus hopes to stay in the human body for as long as possible. It will not mutate to the highest precision state, as the human body can produce the highest precision antibodies to capture and neutralize it. If this situation really occurs, the virus can only survive in a person's body for 14 days. If combined with the time of transmission in the population, the virus may only stay in the population for 3 to 9 months, for example, the antigen precision of SARS I is very high, only spreading for 9 months.

The best way for a virus is to mutate to the next level closest to its precision. The SARS II D614G mutation confirms this, with an increase in precision from 25.23 in SARS II D614 to 27.11 in the subpremary superantigen SubD614G.

Therefore, the structure of SARS II D614G is a result of evolution and a more "precise" or complex state. The human body must produce more precise antibodies to recognize it. After the SARS II D614G mutation, the primary superantigen SARS II D614 is broken down into two parts by G, and the action of the superantigen is transmitted to the sub-primary superantigen SubD614G. The average duration of the epidemic is 9 months, with the shortest being 2.9 months, such as SARS II D614.

The length of the sub-primary superantigen SubD614G is only 24 amino acids. According to the selection rules of the secondary superantigen, it should be F58 with 36 amino acids, as shown in Table 3. The precision of F58 antigen is 27.29, which is 2.06 higher than the precision of reference antigen of 25.23. Substitute x=2.06 into the formula $y=7.3027x^2-56.325x+83.74= -1.3000$ to obtain the result with a negative time interval. Due to the negative y value of -1.3000 months between the predicted outbreak date and the reference strain outbreak date, which is a negative interval month, this is theoretically impossible and cannot be achieved in practice. Therefore, F58 cannot be used as a secondary superantigen. The virus can only hand over the task of the secondary superantigen to the originally planned N148. As a result, a third wave of the epidemic broke out, and the fact that it actually occurred, the principle of selecting super antigens in theory, and the quantitative calculation of antigen precision were all mutually confirmed with the mathematical models they fitted, which could make the correlation coefficient of the mathematical models reach 1. Any mistake made by either party will reduce the correlation coefficient, reaching 1, which can also prove the correctness of the mathematical model description. They further demonstrate that the methods and mathematical models determined in this work are accurate and reliable.

Peptides	Amino acids	Mean of molecular weight	SD	Contains "W"
F718	43	125.3	20.69	N
I358	41	134.2	31.66	Y
N148	38	140.1	26.92	Y
D614	37	129.3	25.23	Y
D614G	37	127.7	26.74	Y
F58	36	136.6	27.29	Y
I1018	35	128.2	26.58	N
Y1138	35	135.3	18.77	N

Table 3: Molecular weight and standard deviation of D614, N148, I358 and other potential mutant antigens.

The fifth outbreak corona virus SARS II N148, in Indian strain, led third wave of SARS II Delta pandemic

The third wave of SARS II was caused by a mutation of SARS II D614G to SARS II N148 or SARS II Delta subtype. Which superantigens will undergo mutations in the next step?.

The SD of SARS II N148 is 26.92, while the SD of SARS II I358 is 31.66. The length of SARS II N148 is 38 amino acids, while the length of SARS II I358 is 41 amino acids. The virus hopes to fire first and later mutate the next secondary superantigen, SARS II N148, with the closest but slightly more precise and a slightly longer length.

The third wave of SARS II was caused by the SARS II N148 or SARS II Delta subtype that occurred in India in October, 2020. SARS II N148 did indeed activate and mutate into a new superantigen at the correct time and in the correct order, with three site mutations: E156del, F157del, and R158G. The final antigen precision increased from 26.92 to 28.65. The higher precision superantigens SARS II D614, SARS II D614G, and SARS II N148 are factors that escape the previously lower precision immune system of the entire population, all of which increase the infectivity of the virus.

The predicted outbreak date for SARS II Delta is 12 October, 2020, and literature reports indicate an outbreak in October, 2020. Like the primary superantigen, the secondary superantigen SARS II N148 is also cleaved into two parts by the R158G site (E156del, F157del, R158G) in SARS II N118. The role of superantigens is transformed into SARS II I358. Unlike the primary superantigen SARS II D614, SARS II D614 only has one amino acid mutation, while SARS II N148 has at least 10 amino acid mutations. Figure 2, can also be seen that as the precision of the antigen increases, the stability of the superantigen first decreases and then increases, with a segment below 0 indicating that it will not actually infect humans. This may be due to the human immune system being particularly sensitive to monitoring this segment, just like the human eye is only sensitive to the wavelength of visible light and not sensitive to ultraviolet and infrared radiation. So, although the antigen precision of SubD614G is higher than that of SARS II N147, it is still detected by the immune system at the edge of the sensitive area, and ultimately must be ignited before SARS II N147 and transfer the responsibility for leading epidemic to SARS II N147 later.

The sixth outbreak corona virus SARS II I358, in South African strain, led fourth wave of SARS II Omicron pandemic

The fourth wave of the epidemic caused by the activation and mutation of the SARS II superantigen is the transfer of transmission responsibility from the secondary superantigen SARS II N148 or SARS II Delta to the tertiary superantigens SARS II I358 or SARS II Omicron.

The antigen precision of SARS II N148 (Delta) has increased from 26.92 to 28.65. Which superantigen will be activated and mutated in the next step?.

The antigen precision of SARS II I358 is 31.66, with a length of 41 amino acids. This superantigen is the final choice for the virus to activate and then mutate. It did indeed mutate into the SARS II Omicron strain on 23 November, 2021, triggering the fourth wave of SARS II in South Africa.

Antigen precision and latency or stability are shown to be dose-dependent. It has statistical significance, indicating that the evolutionary sequence is correct and consistent with the actual epidemic history. The SARS II D614G mutation occurs first, followed by the SARS II N148 (Delta) subtype, and the SARS II I358 (Omicron) subtype.

Although the precision of antigens did not predict the prevalence of alpha, beta, and gamma subtypes, the failure of predictions did prove that this prediction was correct. Alpha, beta, and gamma subtypes have not formed the major waves in the world. These three subtypes all appeared one month before the Delta type, in September 2020. If it is the main subtype, it is difficult to cause a global outbreak of another subtype within a month. There are no cases of these three subtypes in China. This antigen precision can only predict major infectious disease epidemics and outbreaks.

Superantigens SARS II D614, SARS II D614G, SARS II N148, and SARS II I358 are factors that contribute to immune system escape and increase viral infectivity. The predicted outbreak date for SARS II I358 or SARS Omicron is 23 November, 2021, the same day as the actual outbreak.

The quaternary super antigen

The SD of the super antigen SARS II F718 is 20.69, even less than 25.23. Why hasn't this "rough" antigen mutated? The possible reason may be the distribution of amino acids, as it does not contain any of the largest amino acids, tryptophan (W). If any superantigen does not contain "W", it may not show a trend of first order mutations.

Excluding potential epidemic candidate strains without "W" amino acids in superantigen is a statistical bias issue, as the absence of the amino acid with the highest molecular weight may interfere with SD calculations. Tryptophan does have its own biochemical functions, one of which is that it can be translated from the "stop codon". According to reports, Thymine Guanine Adenine (TGA termination codon) in spirochetes is the termination codon for tryptophan, while in other species, the termination signal [11].

Tryptophan may play an important role in infectivity. This requires further research. At least four superantigens, D614, N148, I358, and F718, have been discovered from the coronavirus. The successive collaboration of these four superantigens has driven the spread of this disease through a precise program. The average prevalence time guided by one superantigen is 9 months, and the total prevalence time guided by four superantigens is 36 months. This is in line with the Chinese saying that the epidemic should not exceed 3 years.

In summary, just like how SARS II D614 mutated into SARS II D614G, then again mutated into SARS II N148 (Delta) subtype, and finally mutated into SARS II I358 (Omicron), it started with a "rough" state and ended with a "precise" state. The goal of evolution is a precise state. For SARS II, the correlation coefficient R between antigen precision and stability is 0.998.

If SARS I and MERS are added to the mathematical model, the correlation coefficient R between antigen precision and stability reaches 1. Some scientists may be skeptical about this based on empirical evidence that the correlation is close to 1 but cannot reach 1. If we ignore cases of epidemic diseases with a correlation coefficient of 1, we may miss a way to discover the truth. We have no reason to ignore it. For the sake of caution, we calculated more than 3 times, with a decimal point accurate to 7, and found that R is really 1. The R of SARS II has reached 0.998. After adding the data of SARS I, the R reaches 1, and the fitting is getting better and better. Logically, it can be inferred that this is correct.

Conclusion

There are at least four superantigens acting as chain reactions in the epidemic population. In the cases of SARS I, MERS, and SARS

II, they are D614, N148, I358, and F718, respectively. The complete D614 is the primary superantigen, the complete N148 is the secondary superantigen, and the complete I358 is the tertiary superantigen. The leading role of superantigens is transmitted from the primary to the secondary, and then to the tertiary. From SARS I to MERS, to SARS II D614, to SARS II D614G, to SARS II N147 or SARS II Delta, and then to SARS II I358 or SARS II Omicron, this program-controlled infection process has lasted for 20 years and conforms to the mathematical models. The increase in precision of superantigens and the presence or absence of tryptophan are determining factors for the outbreak date of coronavirus induced diseases.

Author Contributions

Peijun Zuo searched for the information, performed the analysis, and wrote the paper. Professor Liping Li provided the key advice. Longlong Zuo and Zhihong Li did interpretation of data.

Research Ethics Approval

This is not a traditional epidemiological study, although to find the relation between protein molecular and the epidemic disease caused by the protein mutation. It did not contain any samples from human and animals. A research ethics approval is not applicable.

Acknowledgments

References

1. Xu SH, Li YX, Li ST, Wu Q, Sun FQ, et al. (2010) Epidemic condition and biological characteristics of non-O1/non-O139 *vibrio cholerae* in Haizhu district of Guangzhou. Chin J Prev Med 44:1087-90.

2. Cotten M, Watson SJ, Zumla AI, Makhdoom HQ, Palser AL, et al. (2014) Spread, circulation and evolution of the Middle East respiratory syndrome coronavirus. mBio 5:e01062-13.

3. Zuo P, Zuo L, Li Z, Liu C, Li L (2022) Antigen evolution from D614 to G614 to Delta and to Omicron subtype of SARS-CoV-2.

4. Rabenau HF, Cinatl J, Morgenstern B, Bauer G, Preiser W, et al. (2005) Stability and inactivation of SARS coronavirus. Med Microbiol Immunol 194:1-6.

5. Zhang D, Lu J (2021) Epidemiology of infectious diseases. Epidemic practice workbook. Sun Yat-sen-university press.

6. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, et al. (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579:270-273.

7. Zuo P, Zuo L, Li Z, Li L (2023) Virulence and infectivity were associated with different fragments in the delta subtype of SARS-CoV-2 new COVID-19 variants. Intech Open 10:25-32

8. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, et al. (2020) Tracking changes in SARS-CoV-2 Spike: Evidence that D614G increases infectivity of the COVID-19 virus. Cell 182:812-827.

9. Pascarella S, Ciccozzi M, Zella D, Bianchi M, Benedetti F, et al. (2021) SARS-CoV-2 B.1.617 Indian variants: Are electrostatic potential changes responsible for a higher transmission rate?. J Med Virol 93:6551-6556.

10. Aleem A, Ab AS, Slenker AK (2022) emerging variants of sars-cov-2 and novel therapeutics against coronavirus (COVID-19). Stat Pearls 8:10-15.

11. Meng Q, Gu W, Bi K, Ji H, Wang W (2010) Spiralin-like protein SLP31 from Spiroplasma eriocheiris as a potential antigen for immunodiagnostics of tremor disease in Chinese mitten crab Eriocheir sinensis. Folia Microbiol 55:245-50.