HUMAN BIOLOGY
2024, VOL. 94, ISSUE 2
ORIGINAL ARTICLE

OPEN ACCESS

Antigen Evolution from D614, to D614G, to Delta, and to Omicron Subtype Of SARS-Cov-2

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

ABSTRACT

Since 2019, the antigens from Severe Acute Respiratory Syndrome Corona Virus 2 (SARS-CoV-2) are keeping in evolution from initial D614, to D614G, to Delta, and to Omicron. The reasons of why the D614 subtype of SARS-CoV-2 was mutated to D614G first, to Delta second, and to Omicron third, are very importance in public health.

To determine if the "roughness" of the antigens is difference among the candidates leading to mutated to D614G, Delta, and Omicron sub-type of SARS-CoV-2. The hypothesis is that, the "roughest" antigen mutated first, the least "rough" antigen mutated last.

Seven of the peptides from SARS-CoV-2 were analyzed. The finding was that, the order for the evolution of its antigens was exactly from D614G "rough" status to Delta "precise" status, to Omicron "most precise" status.

In this paper, we would like to show the initial data how the antigens from initial D614, conducted evolution to D614G, to Delta, to Omicron of SARS-CoV-2. This finding can help the development of reagents for detecting both D614G "rough", Delta "precise", Omicron "most precise" antigens, or even help for the development of the vaccines against mutated SARS-CoV-2. And finally, to help the control of epidemic covid-19.

According to such order, the D614G "rough", Delta "precise", and Omicron "most precise" antigens of the SARS-CoV-2 can be designed in silicon, developed in laboratory, and confirmed in animals. The way should be very tough and long.

INTRODUCTION

Corona virus disease 2019 (Covid-19) has been broken out since 2019.

In December 2019, corona virus disease 2019 was first reported in China Zhou et al. (2020). The epidemic, which was first broken out in Wuhan, China, started on December 12, 2019, caused 2,794 laboratory-confirmed infections. These infections had 80 deaths until 26 January 2020.

Two of the whole sequences from strains of SARS-CoV-2, MN908947 Wu et al. (2020), MT007544 Caly et al. (2020), were reported on March 18, 2020 and June 25, 2020, respectively. Both are D614. That means, from the CDS 21563 to 25384 of the whole genome, it is gene "S". This "S" gene's product is "surface glycoprotein", and a "structural protein". In this protein, the 614th amino acid is "D". This D614 strain was the first generation of SARS-CoV-2. The protein ID is "QHD43416".

Based on the 6244 global cases of covid-19, there were only 9.9% of D614G before March 1, 2020, but it rapidly increased to 54% on March 10, 2020 in only 10 days, and became to the domain strain Korber et al. (2020).

On June 21, 2020, World Health Organization (WHO) Situation Report recorded over 8.7 million COVID-19 cases and 460,000 deaths. The numbers are increased daily Korber et al. (2020).

On August 20, 2020 Korber et al. (2020) and October 8, 2020, the detailed sequencing data of D614G strain of SARS-CoV-2 was reported Gobeil et al. (2020). The protein IDs are "7KDK_A", "7LX5_B" Pymm et al. (2021) and "7KEC_C". It is a trimer protein. The authors mentioned that, SARS-CoV-2 D614G is a 3 RBD down Spike Protein Trimer without the P986-P987 stabilizing mutations (S-GSAS-D614G). However, by analyzed in detail, the 7LX5-B is still in the non-mutated status, D614. This might not interfere the function of the other two units, already mutated to D614G, in this trimer.

[1]School of Public Health, Shantou University, Guangdong, China
[2]The Chinese University of Hong Kong, Hong Kong, China

Correspondence to: Dr. Peijun Zuo, School of Public Health, Shantou University, Guangdong, China.
E-mail: peijunzuo@hotmail.com.

Keywords: Antigen, Evolution, D614, D614G, SARS-CoV-2.

From the D614G domain strain epidemic day, March 10, 2020, seven months later, in October 2020, the Delta type became the domain strain over others.

Until January 2021, there were 78 million people had been infected, and more than 1.7 million patients were dead Gaebler et al. (2021). Nine folds more cases and 4 folds more deaths were increased in the 6 months duration.

It is still in pandemic status after two years.

Normally, once a microbe infected to a person, in around 14 days, the generated IgG could protect the infected person. The gained immunity could against the infected microbe, and the outcome should be recovery. In the other end, the microbe could not be easily transmitted from one person to another. So, the epidemic status should also be controlled in 14 days.

Why the corona virus disease 2019 is already keeping epidemic for 2 years? The major reason should, at least, be the generated antibody, IgG, could not protect the host to get infect the new mutant strains of SARS-CoV-2, with brand new antigens.

The SARS-CoV-2, the pathogen of the corona virus disease 2019 is not stable for the antigens in its spike. That means, it is keeping changing. In these two years period, the antigens of the SARS-CoV-2 are keeping shifting from D614, to D614G, to Delta, and finally to Omicron.

The epidemic data shows the D614G is the next generation of D614. As the consequence, the D614G had a higher infectivity Zhang et al. (2020) and a higher toxicity to human. It gained the stronger capability to survival from the human immunity.

There are too many papers mentioned such evolution of the antigens, from D614, to D614G, to Delta, and to Omicron. But limited number papers demonstrated the reason why the D614 went to D614G evolution, and the D614G to Delta, to Omicron.

Once an antigen invades to human body, the immune system will generate a specific antibody to against such antigen. When such antigen wants to invade human body again in the second time, it must do some mutation to perform evolution, otherwise will fail to enter the human body by the defense of established immune system specific to against the non-mutant strain.

Normally, one antigen matches one antibody. That is true, there exist many cross reactions between antigens and antibodies, but the efficiency might not as good as the original antigen-antibody pair. The Antibody-Dependent Enhancement (ADE) might happen due to such cross reactions. The matching of antigen and antibody is very similar to the pairing of lock and key.

In 1990s, at Austria, one student owed one key to open the lock of his or her self's and did not be able to open any other locks. But their professor held one key to open all locks of his or her students. The professor's key is very similar to any of the students' keys, but the difference should be a little bit "smaller" than the student's keys. Why it was "smaller", but not "bigger" than the students' keys? The smaller key could enter the space of the original students' keys, but the bigger key could not.

Same rules might fit the evolution of the antigens. The evolution antigen needs to be smaller and similar compared by its non-mutant original antigen. It is true at least in the case of D614 to D614G, D614G to Delta, to Omicron. Because "G" is the smallest amino acid among all 20 amino acids, and for sure, "smaller" than "D". D614G is smaller than D614. Delta type is also smaller than D614G, same as Omicron.

This paper will try to find a regulation or a rule why the mutant was happened in the position of D614, but not in other position. Or if the other mutation happened, why the order was first from D614 to D614G, second from D614G to Delta, third to Omicron.

MATERIALS AND METHODS

The amino acid sequences of protein "QHD43416" contained D614, "7KDK_A", "7LX5_B", "7KEC_C" contained D614G were searched from NCBI.

The amino acid sequences of SARS-CoV-2, the D614 and the D614G mutant were compared by the online software of Clustal Omega.

In the alignment result of Clustal Omega, all "G" amino acids were marked by the yellow color (Figure 1, 2, 3, 4).

The total amino acids and their molecular weights were listed in Table 1.

Seven candidates for potential mutant peptides were searched. The criterion is the author's hypothesis, the mutant is happened in the "bigger" amino acids. Compared with "G", all other 19 amino acids are "bigger" amino acids. That means, the non- "G" amino acids are potential part for mutation or evolution.

Thus, seven non- "G" fragments were selected.

From the longest, F718, 43 amino acids, to the D614 fragment, 37 amino acids, there are 3 candidates.

For compare with these longer candidates, 3 shorter candidates, one 36 amino acids, F58, two 35 amino acids, F58, Y1138 were selected.

For compare the different between the D614 and D614G, D614G was selected to analyze even it is a "G" contained fragment.

Such candidates were analyzed by their molecular weight of each amino acids. Their standard deviations (SD) were calculated by Excel.

Table 1: Total amino acids and their molecular weights.

Amino acid	Abbreviations	Chemical formula	Molecular weight, g/mol
Alanine	A	C3H7NO2	89.0935
Arginine	R	C6H14N4O2	174.2017
Asparagine	N	C4H8N2O3	132.1184
Aspartate	D	C4H7NO4	133.1032
Cysteine	C	C3H7NO2S	121.159
Glutamate	E	C5H9NO4	147.1299
Glutamine	Q	C5H10N2O3	146.1451
Glycine	G	C2H5NO2	75.0669
Histidine	H	C6H9N3O2	155.1552
Isoleucine	I	C6H13NO2	131.1736
Leucine	L	C6H13NO2	131.1736
Lysine	K	C6H14N2O2	146.1882
Methionine	M	C5H11NO2S	149.2124
Phenylalanine	F	C9H11NO2	165.19
Proline	P	C5H9NO2	115.131
Serine	S	C3H7NO3	105.093
Threonine	T	C4H9NO3	119.1197
Tryptophan	W	C11H12N2O2	204.2262
Tyrosine	Y	C9H11NO3	181.1894
Valine	V	C5H11NO2	117.1469

In all 20 amino acids, the smallest amino acid is Glycine, "G", the molecular weight is 75.0669. The biggest amino acid is Tryptophan, "W", the molecular weight is 204.2262.

RESULTS

Confirmed mutated peptide D614G, Delta, Omicron, and other candidates of potential mutant peptides

As showed in Figure 1, the candidates of mutant peptide are the arrowed fragments. One of them, is D614, TNTSNQVAVLYQDVNCTEVPVAIHADQLTPTW RVYST, 37 amino acids.

This non- "G" fragment contains the D614 amino acid. This D614 is already confirmed to be mutated to D614G. This potential mutated fragment is already become a confirmed mutated peptide.

The other longer non- "G" peptide is, N148, 38 amino acids, VYYHKNNKSWMESEFRVYSSANNCTFEYVSQPF LMDLE, showed in Figure 3.

Figure 1: The amino acids of I358 and D614.

```
QHD43416     YLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQPTESIVR     328
7KDK_A       ylqprtfllkynengtitdavdcaldplsetkctlksftvekgiyqtsnfrvqptesivr     328
7KEC_C       ylqprtfllkynengtitdavdcaldplsetkctlksftvekgiyqtsnfrvqptesivr     328
7LX5_B       ylqprtfllkynengtitdavdcaldplsetkctlksftvekgiyqtsnfrvqptesivr     420
             ************************************************************

                I358→→→→→→→→→→→→→→→→→→→→
QHD43416     FPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLN     388
7KDK_A       fpnitnlcpfgevfnatrfasvyawnrkrisncvadysvlynsasfstfkcygvsptkln     388
7KEC_C       fpnitnlcpfgevfnatrfasvyawnrkrisncvadysvlynsasfstfkcygvsptkln     388
7LX5_B       fpnitnlcpfgevfnatrfasvyawnrkrisncvadysvlynsasfstfkcygvsptkln     480
             ************************************************************

QHD43416     DLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGN     448
7KDK_A       dlcftnvyadsfvirgdevrqiapgqtgkiadynyklpddftgcviawnsnnldskvggn     448
7KEC_C       dlcftnvyadsfvirgdevrqiapgqtgkiadynyklpddftgcviawnsnnldskvggn     448
7LX5_B       dlcftnvyadsfvirgdevrqiapgqtgkiadynyklpddftgcviawnsnnldskvggn     540
             ************************************************************

QHD43416     YNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPY     508
7KDK_A       ynylyrlfrksnlkpferdisteiyqagstpcngvegfncyfplqsygfqptngvgyqpy     508
7KEC_C       ynylyrlfrksnlkpferdisteiyqagstpcngvegfncyfplqsygfqptngvgyqpy     508
7LX5_B       ynylyrlfrksnlkpferdisteiyqagstpcngvegfncyfplqsygfqptngvgyqpy     600
             ************************************************************

QHD43416     RVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRD     568
7KDK_A       rvvvlsfellhapatvcgpkkstnlvknkcvnfnfngltgtgvltesnkkflpfqqfgrd     568
7KEC_C       rvvvlsfellhapatvcgpkkstnlvknkcvnfnfngltgtgvltesnkkflpfqqfgrd     568
7LX5_B       rvvvlsfellhapatvcgpkkstnlvknkcvnfnfngltgtgvltesnkkflpfqqfgrd     660
             ************************************************************

                 D614→→→→→→→→→→→→→→
QHD43416     IADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQ     628
7KDK_A       iadttdavrdpqtleilditpcsfggvsvitpgtntsnqvavlyqgvnctevpvaihadq     628
7KEC_C       iadttdavrdpqtleilditpcsfggvsvitpgtntsnqvavlyqgvnctevpvaihadq     628
7LX5_B       iadttdavrdpqtleilditpcsfggvsvitpgtntsnqvavlyqdvnctevpvaihadq     720
             *********************************************.**************

                          D614→→→→→
QHD43416     LTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSPRRARSVA     688
7KDK_A       ltptwrvystgsnvfqtragcligaehvnnsyecdipigagicasyqtqtnspgsassva     688
7KEC_C       ltptwrvystgsnvfqtragcligaehvnnsyecdipigagicasyqtqtnspgsassva     688
7LX5_B       ltptwrvystgsnvfqtragcligaehvnnsyecdipigagicasyqtqtnspgsassva     780
             *****************************************************  * ***
```

The amino acid sequence of protein "QHD43416" contained D614, "7KDK_A", "7LX5_B", "7KEC_C" contained D614G are searched from NCBI. The amino acid sequences of SARS-CoV-2, the D614 and the D614G mutant were compared with the online software of Clustal Omega. In the alignment result of Clustal Omega, all "G" amino acids are marked by the yellow color. I358 and D614 are indicated by the arrows. D614G is in the same position of D614 but mutated in the fragments of "7KDK_A" and "7KEC_C". I358 is already mutated in Omicron.

This fragment is also a confirmed peptide already mutated to Delta type, exactly on the base of D614G mutant.

The other similar peptide is, I358, 41 amino acids EVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCY, showed in Figure 1. This fragment is also a confirmed peptide already mutated to Omicron type, exactly on the base of D614G mutant.

The longest non- "G" peptide is F718, 43 amino acids AENSVAYSNNSLAIPTNFTISVTTEILPVSMTKTSVDCTMYIC, showed in Figure 2.

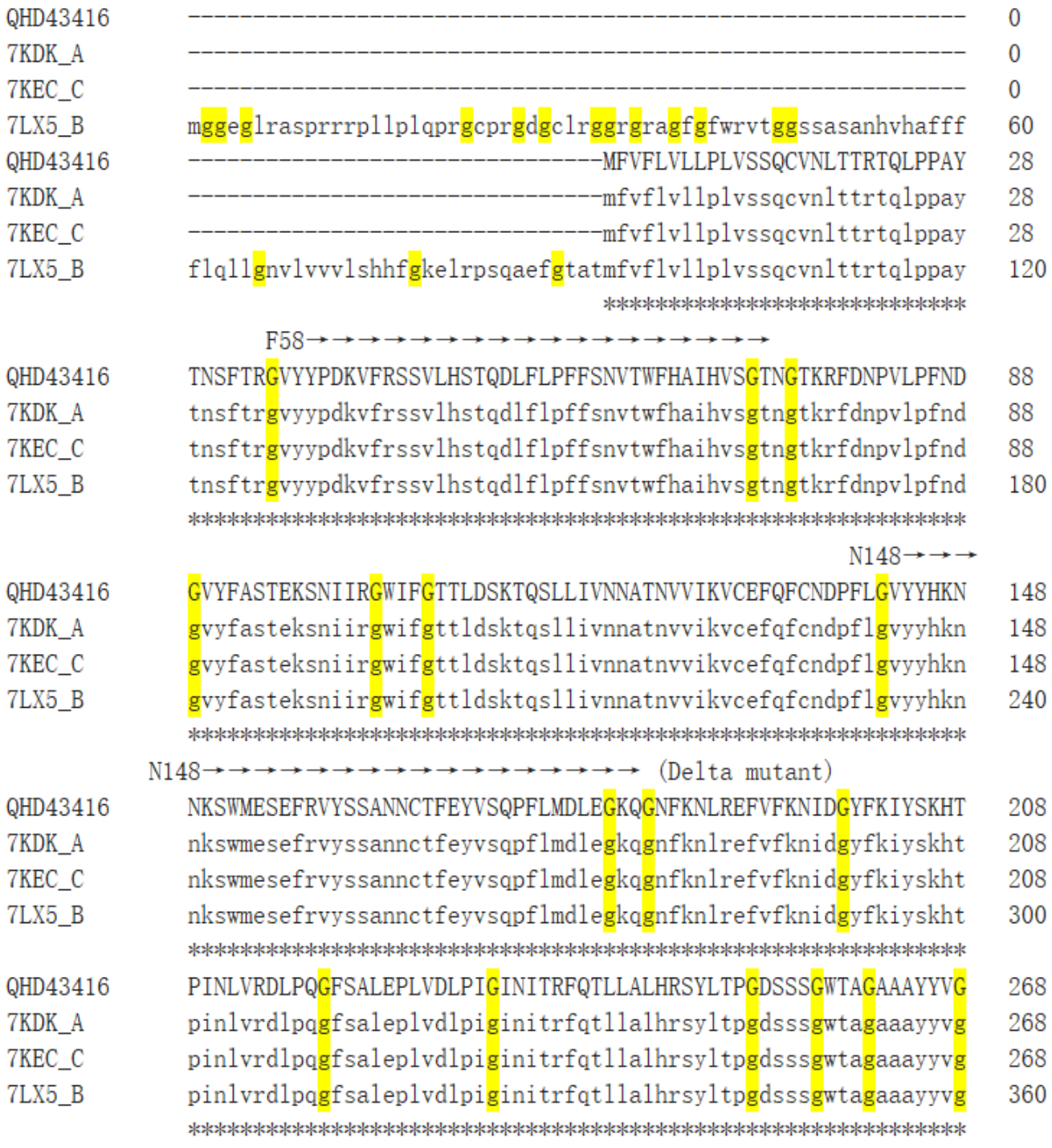
Figure 2: The amino acids of F718 and I1018.

```
                          F718→→→→→→→→→→→→→→→→→→→→→→→→→→→→→→
QHD43416     SQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDSTE   748
7KDK_A       sqsiiaytmslgaensvaysnnsiaiptnftisvtteilpvsmtktsvdctmyicgdste   748
7KEC_C       sqsiiaytmslgaensvaysnnsiaiptnftisvtteilpvsmtktsvdctmyicgdste   748
7LX5_B       sqsiiaytmslgaensvaysnnsiaiptnftisvtteilpvsmtktsvdctmyicgdste   840
             ************************************************************

QHD43416     CSNLLLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPD   808
7KDK_A       csnlllqygsfctqlnraltgiaveqdkntqevfaqvkqiyktppikdfggfnfsqilpd   808
7KEC_C       csnlllqygsfctqlnraltgiaveqdkntqevfaqvkqiyktppikdfggfnfsqilpd   808
7LX5_B       csnlllqygsfctqlnraltgiaveqdkntqevfaqvkqiyktppikdfggfnfsqilpd   900
             ************************************************************

QHD43416     PSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPLLTDE   868
7KDK_A       pskpskrsfiedllfnkvtladagfikqygdclgdiaardlicaqkfngltvlpplltde   868
7KEC_C       pskpskrsfiedllfnkvtladagfikqygdclgdiaardlicaqkfngltvlpplltde   868
7LX5_B       pskpskrspiedllfnkvtladagfikqygdclgdiaardlicaqkfngltvlpplltde   960
             ******** ***************************************************

QHD43416     MIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFN   928
7KDK_A       miaqytsallagtitsgwtfgagaalqipfamqmayrfngigvtqnvlyenqklianqfn   928
7KEC_C       miaqytsallagtitsgwtfgagaalqipfamqmayrfngigvtqnvlyenqklianqfn   928
7LX5_B       miaqytsallagtitsgwtfgagpalqipfpmqmayrfngigvtqnvlyenqklianqfn   1020
             *********************** ****** *****************************

QHD43416     SAIGKIQDSLSSTASALGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLDKVE   988
7KDK_A       saigkiqdslsstasalgklqdvvnqnaqalntlvkqlssnfgaissvlndilsrldkve   988
7KEC_C       saigkiqdslsstasalgklqdvvnqnaqalntlvkqlssnfgaissvlndilsrldkve   988
7LX5_B       saigkiqdslsstpsalgklqdvvnqnaqalntlvkqlssnfgaissvlndilsrldppe   1080
             ************* *******************************************  *
                            I1018→→→→→→→→→→→→→→→→→→→→→→→→→→
QHD43416     AEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRVDFCGKGYH   1048
7KDK_A       aevqidrlitgrlqslqtyvtqqliraaeirasanlaatkmsecvlgqskrvdfcgkgyh   1048
7KEC_C       aevqidrlitgrlqslqtyvtqqliraaeirasanlaatkmsecvlgqskrvdfcgkgyh   1048
7LX5_B       aevqidrlitgrlqslqtyvtqqliraaeirasanlaatkmsecvlgqskrvdfcgkgyh   1140
             ************************************************************

QHD43416     LMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDGKAHFPREGVFVSNGTHWFVTQRN   1108
7KDK_A       lmsfpqsaphgvvflhvtyvpaqeknfttapaichdgkahfpregvfvsngthwfvtqrn   1108
7KEC_C       lmsfpqsaphgvvflhvtyvpaqeknfttapaichdgkahfpregvfvsngthwfvtqrn   1108
7LX5_B       lmsfpqsaphgvvflhvtyvpaqeknfttapaichdgkahfpregvfvsngthwfvtqrn   1200
             ************************************************************
```

In the alignment result of Clustal Omega, all "G" amino acids are marked by the yellow color. F718 and I1018 are indicated by the arrows.

In the other hand, some of non- "G" peptide are shorter than 37 amino acids. F58, 36 amino acids, VYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHVS, showed in Figure 3.

Figure 3: The amino acids of F58 and N148.



In the alignment result of Clustal Omega, all "G" amino acids are marked by the yellow color. F58 and N148 are indicated by the arrows. Delta mutant, E156del, F157del, R158G is already happened in N148, after the D614G mutant.

There are many other peptides without "G" amino acid, such as I1018, 35 amino acids, RLQSLQTYVTQQLIRAAEIRASANLAATKMSECVL, showed in Figure 2; Y1138, 35 amino acids, IVNNTVYDPLQPELDSFKEELDKYFKNHTSPDVDL, showed in Figure 4.

OPEN ACCESS

Figure 4: The amino acids of Y1138.

```
                                          Y1138→→→→→→→→→→→→→→→→→→
QHD43416      FYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGD  1168
7KDK_A        fyepqiittdntfvsgncdvvigivnntvydplqpeldsfkeeldkyfknhtspdvdlgd  1168
7KEC_C        fyepqiittdntfvsgncdvvigivnntvydplqpeldsfkeeldkyfknhtspdvdlgd  1168
7LX5_B        fyepqiittdntfvsgncdvvigivnntvydplqpeldsfkeeldkyfknhtspdvdlgd  1260
              ************************************************************

QHD43416      ISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKW----------------  1212
7KDK_A        isginasvvniqkeidrlnevaknlneslidlqelgkyeqgsgyipeaprdgqayvrkdg  1228
7KEC_C        isginasvvniqkeidrlnevaknlneslidlqelgkyeqgsgyipeaprdgqayvrkdg  1228
7LX5_B        isginasvvniqkeidrlnevaknlneslidlqelgkyeqgsgyipeaprdgqayvrkdg  1320
              ****************************************   :

QHD43416      PWYIWLGFIAGLIAIVMVTIMLCCMTSC------CSCLKGCCSCGSCCKFDEDDSEPVLK  1266
7KDK_A        ewvllstflgrslevlfqgpghhhhhhhhsawshpqfekgggsggggs-ggsawshpqfe  1287
7KEC_C        ewvllstflgrslevlfqgpghhhhhhhhsawshpqfekgggsggggs-ggsawshpqfe  1287
7LX5_B        ewvllstflgrslevlfqgpghhhhhhhhsawshpqfekgggsggggs-ggsawshpqfe  1379
               * :   *:.  : :::                  .  **  * *. .  ..  *.* ::

QHD43416      GVKLHYT                                                       1273
7KDK_A        k------                                                       1288
7KEC_C        k------                                                       1288
7LX5_B        k------                                                       1380
```

In the alignment result of Clustal Omega, all "G" amino acids are marked by the yellow color. Y1138 is indicated by the arrow.

Because the shorter peptides owed lower possibility to be mutated, only 3 of them were analyzed. F58, I1018, Y1138, these three are examples shorter than the fragment of D614, 37 amino acids.

The molecular weight of confirmed mutated peptide D614, D614G, Delta, Omicron and other candidates of potential mutant peptides

As showed in Table 2, the D614, peptide with 37 amino acids, has a mean molecular weight of 129.3, the standard deviation (SD) for the individual molecular weight of its amino acids is 25.23. This peptide does contain the biggest amino acid, Tryptophan (W, molecular weight, 204.2262).

N148 has 38 amino acids, mean molecular weight is 140.1, SD is 26.92, with "W". Its SD is bigger than the SD of D614, 25.23, but smaller than any other SDs of "W" contained fragments. Delta subtype is already showed their mutation in this peptide, with the second order after D614G mutation.

I358 has 41 amino acids, mean molecular weight is 134.2, SD is 31.66, with "W". Its SD is bigger than the SD of D614, 25.23, and also bigger than the SD of Delta subtype N148, 26.92. Omicron subtype is already showed their mutation in this peptide, with the third order after D614G, Delta mutation.

The other peptide's mean molecular weights, their SDs, and the status of contain "W" or not are also indicated in Table 2.

The details are as following.

F718 has 43 amino acids, mean molecular weight is 125.3, SD is 20.69, without "W".

D614G has 37 amino acids, mean molecular weight is 127.7, SD is 26.74, with "W". Its SD is bigger than the SD of D614, 25.23.

F58 has 36 amino acids, mean molecular weight is 136.6, SD is 27.29, with "W".

I1018 has 35 amino acids, mean molecular weight is 128.2, SD is 26.58, without "W".

Y1138 has 35 amino acids, mean molecular weight is135.3, SD is 18.77, without "W".

Although F58, I1018 and Y1138 contain shorter peptide than D614, the 37 amino acids peptide. They are selected as references and analyzed for comparing with their longer buddies.

DISCUSSION

Like the pairs of keys and locks, the more precise the key, the more fineness for their outlook structures. For proteins or antigens, the SD of molecular weight of their amino acids can work as an indicator of their "fineness". The bigger the SD, the more the "fineness" or "precise". In the other way, the smaller the SD, the more the "roughness" or "rough".

D614 already mutated to D614G. It's SD of the molecular weight is 25.23. The molecular weight for Aspartate (D) is 133.1032. For the smallest amino acid, Glycine (G), the molecular weight is 75.0669. D614 is already mutated or evolution to D614G, and its size is already gotten smaller. The SD for D614 is 25.23, it changed to 26.74 for D614G. Thus, the structure of D614G is the result of an evolution, a more "precise" or complex status.

This evolution is essential for the virus. The first generation of the virus, usually was in "rough" status, stimulated human host to generate "rough" antibody to against the virus invaders. The war will be end if the antibody killed the virus. The virus would not to accept such failing. To survival for its species, if possible, the virus has to evolution to a more "precise" status, like from D614 to mutate to D614G. The first generation of antibody would not be able to recognize the more "precise" or more complex antigens in the secondary generation of the virus.

Why the mutation was happed first in the fragment of D614, but not in the longer fragments, F718, I 358, N148? And why N148 is the second mutated in Delta type after D614G mutation? And I 348 is the third mutated in Omicron type after Delta mutation?

In these 3 longer fragments, the SD for I358 is 31.66, for N148, the SD is 26.92. Compared with the D614, both of their SD are bigger than 25.23. Such "precise" fragments would not have the reasons to mutate. And after the D614 mutated to D614G, the N148, the second "rough" fragment will promote to the "roughest" fragment, this cause the second mutation, from D614G (with non-mutated N148), to the Delta type with mutated N148 (E156del, F157del, R158G). Same reason, after N148 mutated to Delta, the I358, the third "rough" fragment will promote to the "roughest" fragment, this cause the third mutation, to the Omicron type.

The F718, with a SD of 20.69, even smaller than 25.23. Why this "rough" fragment did not mutate before D614?

The possible reason might be the distribution of the amino acids has some bias. It does not contain any of the biggest amino acid, Tryptophan (W). If any peptide did not contain "W", it might not trend to be mutated in the very beginning.

To exclude the non- "W" fragments as a potential candidate for the first order evolution, is a concern of statistic bias, interfere the SD calculation.

The tryptophan does have its own biochemical functions, one of them is that, it can be translated from a "stop codon" in mRNA. Without stop codon, the protein chain would be extended. It was reported that TGA codon in Spiroplasma is for tryptophan, instead of a stop signal in other species Meng et al. (2010).

In our knowledge, Spiroplasma species can generate non-specific super allergic reaction in the human body. For this reason, tryptophan may play some roles for the virus evolution, to help it escape from the watching of the immune system.

Tryptophan its self is also an inflammatory mediator. The potential evolution might not be happened first in the non- "W" fragments.

Sum up, like D614 mutated to D614G, D614G mutated to Delta, to Omicron, the SARS-CoV-2 might start conduct its evolution in the non- "G", fragments, contains "W" amino acid, with the "roughest" status. The end point for the evolution should be the "precise" status. The effect of evolution is dose dependent, with the dose of "roughness". The first evolution was happened from D614 to D614G, consisted with the smallest SD, 25.23, with "W". The second evolution was happed from D614G to Delta type, also consisted with the second smallest SD, 26.92, with "W". The Delta mutation is on the base of D614G mutation.

The mutation in the Omicron types of SARS-CoV-2 is also follow such rule. We would like to discuss that in other papers.

This finding can help the development of reagents for detecting both "rough" and "precise" antigens, or even help for the development of the vaccines against SARS-CoV-2. And finally, to help the control of epidemic covid-19.

According to such rule or route, the "common precise" antigens of the SARS-CoV-2 can be designed in silicon, developed in wet laboratory, and confirmed in animals. The way should be very tough and long.

KEY POINTS

Question: Why the D614 subtype of SARS-CoV-2 was mutated to D614G first, to Delta second, and to Omicron third?

Findings: In this analyzing, the peptide contain D614 has smallest SD of their molecular weight of its amino acids, 25.23, the Delta contained peptide has second smallest SD, 26.92, the Omicron contained peptide has third smallest SD, 31.66, a significant difference.

Meaning: In a SARS-CoV-2 pandemic, SDs of their molecular weight of their amino acids, may be used to forecast the coming mutations and the orders.

ABBREVIATIONS

A, Alanine, ADE, Antibody-Dependent Enhancement ,C, Cysteine,COVID-19, Corona virus disease 2019,D, Aspartate, Glutamate, F, Phenylalanine ,G, Glycine,H, Histidine, I, Isoleucine, K, Lysine, L, Leucine, M, Methionine, N, Asparagine, P, Proline, Q, Glutamine, R, Arginine, S, Serine,SARS-CoV-2, Severe acute respiratory syndrome corona virus 2,SD, standard deviation, T, Threonine, W, Tryptophan, WHO, World Health Organization, V, Valine, Y, Tyrosine.

DECLARATIONS

Author Contributions

Dr. Peijun Zuo searched for the information, performed the analysis and wrote the paper. Professor Dr. Liping Li provided the key advices.

Acknowledgments

This work was supported by project NTF21021, Research Starting Funding from Shantou University.

Conflicts of Interest

The authors declare no conflicts of interest.

Consent statement/Ethical approval

Not required.

REFERENCES

1.Zhou P, Yang X-L, Wang X-G, et al. 2020 Mar. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 579(7798):270-273.

2.Wu F, Zhao S, Yu B, et al. 2020 Mar. A new coronavirus associated with human respiratory disease in China. Nature. 579(7798):265-269.

3.Caly L, Druce J, Robers J, et al. 2020 Jun. Isolation and rapid sharing of the 2019 novel coronavirus (SARS-CoV-2) from the first patient diagnosed with COVID-19 in Australia. Med J Aust. 212(10):459-462.

4.Korber B, Fischer WM, Gnanakaran S, et al. 2020 Aug 20. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. Cell. 182(4):812-827.e19.

5.Gobeil S, Janowska k, McDowell S, et al. 2020 Oct 12. D614G mutation alters SARS-CoV-2 spike conformational dynamics and protease cleavage susceptibility at the S1/S2 junction. bioRxiv. 10.11.335299.

6.Pymm P, Adair A, Chan L-J, et al. 2021 May 11. Nanobody cocktails potently neutralize SARS-CoV-2 D614G N501Y variant and protect mice. Proc Natl Acad Sci U S A. 118(19): e2101918118.

7.Gaebler C, Wang Z, Lorenzi JCC, et al. 2021 Jan 4. Evolution of antibody immunity to SARS-CoV-2. bioRxiv. 11.03.367391.

8.Zhang L, Jackson CB, Mou H, et al. 2020 Jun 12. The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. bioRxiv. 06.12.148726.

9.Meng Q, Gu W, Bi K, et al. 2010 May. Spiralin-like protein SLP31 from Spiroplasma eriocheiris as a potential antigen for immunodiagnostics of tremor disease in Chinese mitten crab Eriocheir sinensis. Folia Microbiol (Praha). 55(3):245-50.

