Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Peptides 24 (2003) 1837–1845
PEPTIDES
www.elsevier.com/locate/peptides

Prediction of amino acid pairs sensitive to mutations in the spike protein from SARS related coronavirus

Guang Wu*, Shaomin Yan

DreamSciTech Consulting Co. Ltd., 301, Building 12, Nanyou A-zone, Jainnan Road, CN-518054, Shenzhen, PR China

Received 4 October 2003; received in revised form 27 October 2003; accepted 28 October 2003

Abstract

In this study, we analyzed the amino acid pairs affected by mutations in two spike proteins from human coronavirus strains 229E and OC43 by means of random analysis in order to gain some insight into the possible mutations in the spike protein from SARS-CoV. The results demonstrate that the randomly unpredictable amino acid pairs are more sensitive to the mutations. The larger is the difference between actual and predicted frequencies, the higher is the chance of mutation occurring. The effect induced by mutations is to reduce the difference between actual and predicted frequencies. The amino acid pairs whose actual frequencies are larger than their predicted frequencies are more likely to be targeted by mutations, whereas the amino acid pairs whose actual frequencies are smaller than their predicted frequencies are more likely to be formed after mutations. These findings are identical to our several recent studies, i.e. the mutations represent a process of degeneration inducing human diseases.

Keywords: Amino acid pairs; Coronavirus; Mutations; SARS

1. Introduction

Although the severe acute respiratory syndrome (SARS) has gone after hitting the world for several months, everyone is intuitively expecting the possible return of SARS in the near future, and human logic seems to have such an assumption, i.e. if the SARS would return, it would be another mutated form following its first battle with humans. This is possible because the accumulating evidence shows that there are several mutations in SARS related coronavirus (SARS-CoV). So far 15 point mutations have been documented in SARS-CoV proteins: two in the 221 amino-acid-long membrane glycoprotein [2,15,17,19,24], three in the 1255 amino-acid-long spike glycoprotein [2,15,17,24] and ten in the 7073 amino-acid-long replicase polyprotein [2,8,11,15–17,24]. Naturally we would expect these three SARS-CoV proteins to have other forms of mutations rather than those documented, and the new mutations would lead to the difficulties in diagnosis and treatment of SARS.

An intriguing question is whether or not we can predict the new mutations of SARS-CoV. If so, it would be greatly helpful for identification of SARS-CoV, and a great advance in understanding of the evolutionary process in SARS-CoV. Also, it would be useful for studying the mutant patterns in other human coronavirus, which would give us some insight into the mutations in coronavirus.

Among encoded structural replicase, spike, envelope, membrane, nucleocapsid proteins from human SARS-CoV, the spike protein is incorporated into the viral envelope. The spike proteins of coronaviruses are large, type I membrane glycoproteins that are responsible for both binding to receptors on host cells and for membrane fusion. The spike proteins of some coronaviruses are cleaved into S1 and S2 subunits. S proteins also contain important virus-neutralizing epitopes, and amino acid changes in the spike proteins can dramatically affect the virulence and in vitro host cell tropism of the virus [6,7,22]. Still at present it is only the spike glycoprotein, in which a considerable amount of mutations has been documented. Using the Blastp program to align three spike glycoproteins from humans, we find little cue on the likelihood of which amino acid would mutate in SARS spike glycoprotein.

In the past three years, we have developed two models to analyze the primary structure in proteins [25] conducted a series of studies on mutations in different proteins [26–34]. Our studies show there is a clearly probabilistic pattern in the amino acids, which are subject to mutations. In this

* Corresponding author. Tel.: +86-755-2202-9353; fax: +86-755-2520-8256.
E-mail address: hongguanglishibahao@yahoo.com (G. Wu).

0196-9781/$ – see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/j.peptides.2003.10.008

study, we use our model to analyze two spike glycoproteins from human coronavirus in order to gain the insight on the prediction of amino acid pairs being sensitive to mutations in human SARS-CoV.

2. Materials and methods

The amino acid sequences of the spike glycoproteins were obtained from the Swiss-Protein data bank [1]. The access number is P59594 for human SARS-CoV with 3 point mutations [2,15,17,24], P15423 for human coronavirus strain 229E with 38 point mutations [3–5,10,12,20–23] and P36334 for human coronavirus strain OC43 with 80 point mutations [10,13,14,18]. In order to determine the amino acid pairs probabilistically sensitive to mutations, we conduct the following calculations [25], which is briefly described follows with the SARS-CoV spike protein as the example.

2.1. Amino acid pairs in spike proteins

The spike protein from human SARS-CoV consists of 1255 amino acids. The first and second amino acids are considered as an amino acid pair, the second and third as another pair, the third and fourth, until the 1254th and 1255th, thus there are 1254 pairs. Because there are 20 types of amino acids, an amino acid pair can be composed from any of 20 types of amino acids so there are 400 types of theoretically possible amino acid pairs. Again there are 1254 pairs in the spike protein, which are more than 400 types of theoretically possible amino acid pairs, clearly some of 400 types should appear more than once. Meanwhile we may expect that some of 400 types are absent from the spike protein. Similarly there are 1172 and 1352 amino acid pairs in the spike proteins from strain 229E and OC43, respectively.

2.2. Actual frequency and randomly predicted frequency in amino acid pairs

The randomly predicted frequency is governed by the simple permutation principle [9]. For example, there are 39 arginines (R) and 96 serines (S) in SARS-CoV spike protein, the random frequency of amino acid pair "RS" would be 3 ($39/1255 \times 96/1254 \times 1254 = 2.983$). Actually we can find three "RS"s in the spike protein, so the actual frequency of "RS" is 3. Hence, we have three relationships between actual and predicted frequencies, i.e. the actual frequency is smaller than, equal to and larger than the predicted frequency, respectively.

2.3. Randomly predictable present amino acid pairs in SARS-CoV spike protein

As described in the last section, the predicted frequency of randomly present pair "RS" would be 3 and "RS" does appear 3 times in the spike protein, so the presence of "RS" is randomly predictable.

2.4. Randomly unpredictable present amino acid pairs in SARS-CoV spike protein

There are 84 alanines (A) in SARS-CoV spike protein, the frequency of random presence of "AA" would be 6 ($84/1255 \times 83/1254 \times 1254 = 5.555$), i.e. there would be 6 "AA"s in the spike protein. In fact, the "AA" appears 10 times in the spike protein, so the presence of "AA" is randomly unpredictable. This illustrates the case that the actual frequency of "AA" is larger than its predicted frequency. Another case is that the actual frequency is smaller than the predicted one. For example, there are 91 valines (V) in the spike protein and the predicted frequency of "AV" is 6 ($84/1255 \times 91/1254 \times 1254 = 6.091$), whereas the actual frequency is only three.

2.5. Randomly predictable absent amino acid pairs in SARS-CoV spike protein

There are 11 tryptophans (W) in SARS-CoV spike protein, the frequency of random presence of "RW" would be 0 ($39/1255 \times 11/1254 \times 1254 = 0.342$), i.e. the "RW" would not appear in the spike protein, which is true in the real situation. Thus the absence of "RW" is randomly predictable.

2.6. Randomly unpredictable absent amino acid pairs in SARS-CoV spike protein

There are 99 threonines (T) in SARS-CoV spike protein, the frequency of random presence of "RT" would be 3 ($39/1255 \times 99/1254 \times 1254 = 3.076$), i.e. there would be three "RT"s in the spike protein. However, no "RT" is found in this protein, therefore the absence of "RT" from the spike protein is randomly unpredictable.

2.7. Mutations in randomly predictable and unpredictable amino acid pairs

A point missense mutation results in two amino acid pairs being substituted by another two. As each pair has its actual and predicted frequencies, the difference between them represents a probabilistic measure for the comparison in substituted and substituting amino acid pairs before and after mutation. After calculating the predicted frequency and comparing with the actual frequency, we can classify the substituted amino acid pairs into the predictable/unpredictable amino acid pairs.

3. Results

Table 1 details the appearance of theoretically possible types of amino acids in three spike proteins, for example,

Table 1
Number of theoretical types of amino acid pairs in the spike proteins from different human coronaviruses

Appearance	SARS-CoV		Strain 229E		Strain OC43	
	Number	Percentage	Number	Percentage	Number	Percentage
0	59	14.75	86	21.5	66	16.5
1	76	19	78	19.5	61	15.25
2	61	15.25	60	15	56	14
3	51	12.75	45	11.25	55	13.75
4	46	11.5	43	10.75	41	10.25
5	38	9.5	18	4.5	35	8.75
6	22	5.5	19	4.75	30	7.5
7	15	3.75	14	3.5	17	4.25
8	15	3.75	12	3	11	2.75
9	7	1.75	8	2	11	2.75
10	4	1	7	1.75	6	1.5
11	4	1	5	1.25	6	1.5
12	1	0.25	3	0.75	4	1
13	1	0.25	0	0	0	0
14	0	0	0	0	0	0
15	0	0	0	0	1	0.25
16	0	0	1	0.25	0	0
17	0	0	0	0	0	0
18	0	0	0	0	0	0
19	0	0	0	0	0	0
20	0	0	1	0.25	0	0

the third row shows how many types do not appear. From the viewpoint of amino acid pairs, no matter the length of a protein is, the number of its theoretically possible types cannot be more than 400, and therefore the difference between proteins is either how many types of theoretically possible amino acid pairs appear or how many times a theoretically possible type of amino acid pair repeats or both. Table 1 shows 59, 86 and 66 types are absent from the spike protein of SARS-CoV, strain 229E and strain OC43 (third row in the Table), respectively. Still Table 1 shows that 76, 78 and 61 types appear once in the spike protein of SARS-CoV, strain 229E and strain OC43 (fourth row in the Table), respectively, and so on. The absent types include 17 randomly predictable and 42 randomly unpredictable with regard to SARS-CoV spike protein, 37 randomly predictable and 49 randomly unpredictable with regard to strain 229E spike protein, and 12 randomly predictable and 54 randomly unpredictable with regard to strain OC43 spike protein.

Still we can classify the present amino acid pairs as randomly predictable and unpredictable with respect to theoretically possible types and pairs, because some theoretically possible types appear many times (from row 5 to row 23 in Table 1). The columns 3, 4, 5 and 6 in Table 2 show how many predictable and unpredictable types and pairs in human spike proteins. When corresponding the position of each mutation to predictable pairs and unpredictable pairs, we find that a vast majority of mutations occurs at the unpredictable pairs (columns 7, 8, 9 and 10 in Table 2).

Fig. 1 shows the ratios of frequency difference (AF−PF) versus mutation number per each type of substituted amino acid pairs in spike proteins. It can be seen that there is a general tendency in the ratios, i.e. the larger the difference, the higher the chance of mutation occurring. Therefore, the difference between actual and predicted frequencies indicates the potential chance of mutation occurring in amino acid pairs.

Table 2
Occurrence of mutations with respect to randomly predictable and unpredictable amino acid pairs in the spike proteins from different human coronaviruses

Spike protein	Amino acid pairs	Types		Pairs		Mutations		Ratio	
		Number	Percentage	Number	Percentage	Number	Percentage	Mutations/types	Mutations/pairs
SARS-CoV	Predictable	86	25.22	226	18.02	0	0	0/86 = 0	0/226 = 0
	Unpredictable	255	74.78	1028	81.98	3	100	3/255 = 0.012	3/1028 = 0.003
Strain 229E	Predictable	81	25.8	206	17.58	1	2.63	1/81 = 0.012	1/206 = 0.005
	Unpredictable	233	74.2	966	82.42	37	97.37	37/233 = 0.159	37/966 = 0.038
Strain OC43	Predictable	97	29.04	286	21.15	4	5	4/97 = 0.041	4/286 = 0.014
	Unpredictable	237	70.96	1066	78.85	76	95	76/237 = 0.321	76/1066 = 0.071

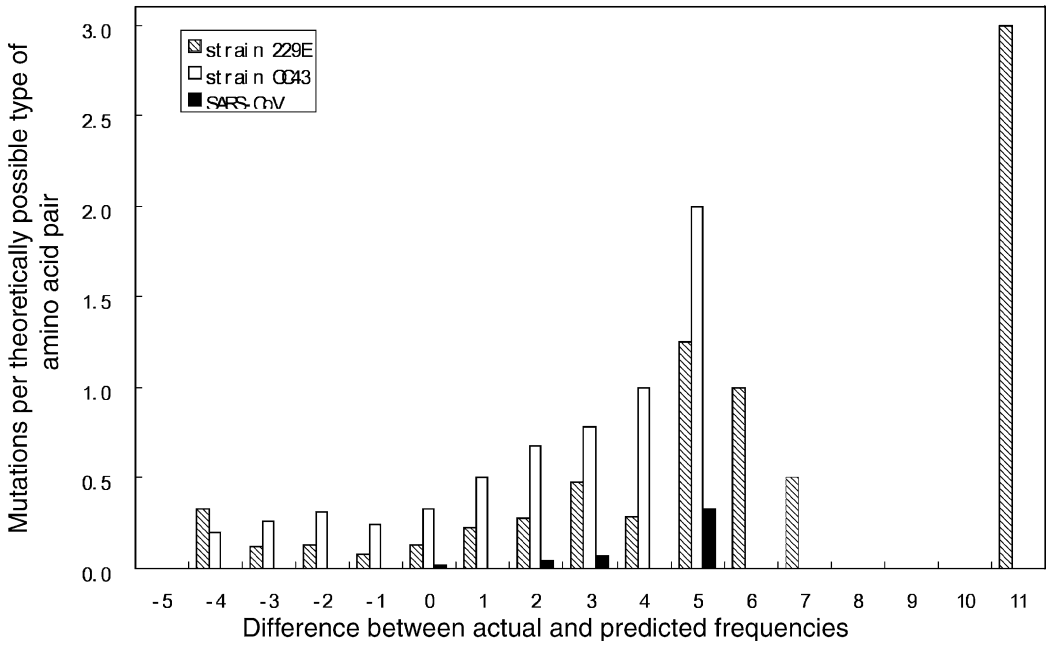

Fig. 1. Ratios of difference between actual and predicted frequencies versus mutations per theoretically possible type of amino acid pair in the spike proteins from different human coronaviruses.

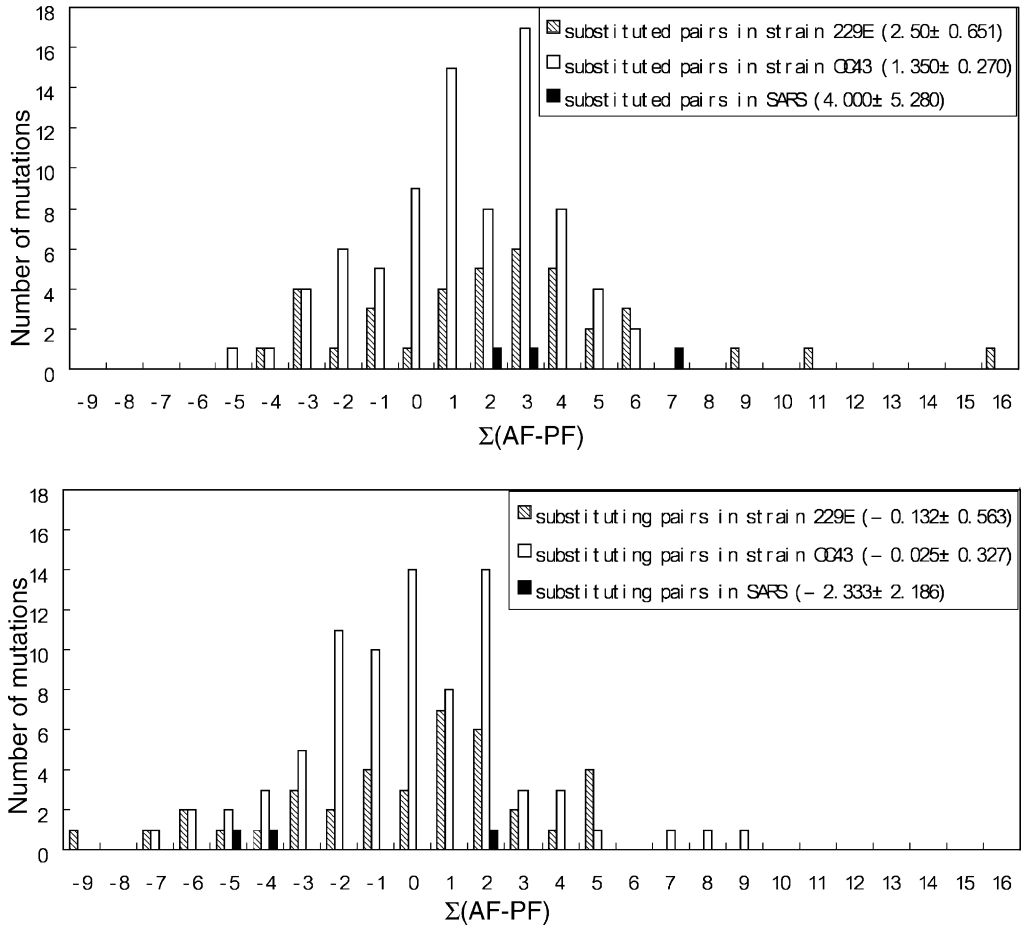

Fig. 2. Sum of differences between actual and predicted frequencies in substituted and substituting amino acid pairs in the spike proteins from different human coronaviruses (the data are presented as mean ± S.E.).

Table 3
Sum of difference between actual and predicted frequencies with respect to amino acid pairs being substituted and substituting by mutations

SARS-CoV	Σ(AF–PF)		Strain 229E	Σ(AF–PF)		Strain OC43	Σ(AF–PF)	
mutation position	Substituted pairs	Substituting pairs	Mutation position	Substituted pairs	Substituting pairs	Mutation position	Substituted pairs	Substituting pairs
77	3	−4	98	6	4	29	3	0
244	2	−5	120	2	−2	29	3	1
577	7	2	176	3	−1	40	−2	−7
			210	3	0	62	−5	−2
			223	−3	−6	63	−2	0
			230	2	−3	115	3	−2
			230	2	0	115	3	0
			248	16	−1	116	3	1
			270	6	2	152	−2	2
			295	−3	−3	161	−1	1
			300	1	1	167	0	0
			307	5	5	173	3	−2
			336	4	2	173	3	−2
			401	2	2	190	1	−1
			414	−4	3	222	2	3
			424	2	5	248	1	3
			430	4	1	252	2	−1
			441	−1	−9	272	5	−4
			444	−2	5	283	3	0
			462	4	0	288	2	−3
			481	−1	−6	291	3	2
			488	4	2	303	1	−3
			530	9	2	308	4	2
			577	3	−7	329	3	1
			578	−3	−1	334	−2	−1
			590	11	1	451	4	−4
			642	3	5	454	−3	2
			681	1	1	467	0	2
			700	−3	1	488	1	−3
			711	4	1	496	4	−1
			714	3	−1	544	1	−1
			765	5	−2	557	1	7
			775	3	1	566	5	−5
			846	0	−4	570	−1	1
			871	1	−3	579	1	0
			937	6	−5	587	6	0
			971	−1	2	603	1	−1
			1005	1	3	612	0	−5
						630	3	4
						641	−1	2
						665	5	−3
						694	−1	5
						700	0	−2
						728	−3	−2
						758	−4	−1
						783	0	0
						802	1	0
						817	3	0
						824	2	−2
						833	3	2
						884	4	9
						896	3	4
						912	1	−2
						915	3	−2
						933	2	2
						944	−2	2
						955	1	8
						955	1	−1
						969	−3	−1
						975	0	1

Table 3 (*Continued*)

SARS-CoV	Σ(AF–PF)		Strain 229E	Σ(AF–PF)		Strain OC43	Σ(AF–PF)	
mutation position	Substituted pairs	Substituting pairs	Mutation position	Substituted pairs	Substituting pairs	Mutation position	Substituted pairs	Substituting pairs
						993	4	4
						1012	2	−6
						1016	1	0
						1039	4	−1
						1058	1	−2
						1059	−2	−4
						1074	4	1
						1089	4	0
						1160	0	2
						1189	1	2
						1193	6	2
						1197	−1	1
						1202	−3	3
						1211	3	0
						1220	2	2
						1231	0	2
						1246	5	−6
						1265	0	−3
						1331	2	0
						1342	3	−2

AF: actual frequency; PF: predicted frequency.

As the point missense mutations substitute one type of amino acid to another one, we can gain some insight into the mutation tendency after comparing the difference between actual and predicted frequencies in substituted and substituting amino acid pairs. For the numerical analysis, we calculate the difference between actual frequency (AF) and predicted frequency (PF) in amino acid pairs before and after mutation, i.e. Σ(AF−PF). For instance, a mutation at position 244 substitutes "I" to "T" which results in two amino acid pairs "DI" and "IW" changing to "DT" and "TW", because the amino acid is "D" at position 243 and "W" at position 245. The actual frequency and predicted frequency are 7 and 5 for "DI", 1 and 1 for "IW", 2 and 6 for "DT", and 0 and 1 for "TW", respectively. Thus, the difference between actual frequency and predicted frequency is 2 with regard to the substituted amino acid pairs, i.e. $(7-5)+(1-1)=2$, and -5 with regard to the substituting amino acid pairs, i.e. $(2-6)+(0-1)=-5$. In this way, we can compare the frequency difference in the amino acid pairs affected by mutations. Fig. 2 shows the difference between actual and predicted frequencies in both substituted and substituting amino acid pairs in spike proteins. It can be seen that the substituting pairs distribute more centrally and symmetrically than the substituted pairs do. The sum of differences between actual and predicted frequencies is statistically smaller in substituting amino acid pairs than in substituted ones in Table 3 (the Student's *t*-test, $P < 0.05$). These statistical differences suggest that the mutations lead to the deduction of difference between actual and predicted frequencies. From a probabilistic viewpoint, this means that the mutations are more likely to occur, and these findings are similar to the results in our recent studies [26–34].

As mentioned in Section 2, the actual frequency can be equal to, larger than or smaller than the predicted frequency. Accordingly we can look at these relationships with respect to the substituted (Table 4) and substituting (Table 5) pairs. Table 4 reveals that more than 75% of mutations occur at the pairs, whose actual frequency is larger than their pre-

Table 4
Classification of substituted amino acid pairs with respect to mutations in the spike proteins from different human coronaviruses

Spike protein	Amino acid pairs		Mutations in SARS-CoV		Mutations in strain 229E		Mutations in strain OC43	
	I	II	Number	Percentage	Number	Percentage	Number	Percentage
Predictable	AF = PF	AF = PF	0	0	1	2.63	4	5
Unpredictable	AF > PF	AF > PF	1	33.33	12	31.58	19	23.75
	AF > PF	AF = PF	2	66.67	10	26.32	20	25
	AF > PF	AF < PF	0	0	10	26.32	23	28.75
	AF < PF	AF = PF	0	0	3	7.89	9	11.25
	AF < PF	AF < PF	0	0	2	5.26	5	6.25

AF: actual frequency; PF: predicted frequency.

Table 5
Classification of substituting amino acid pairs with respect to mutations in the spike proteins from different human coronaviruses

Amino acid pairs		Mutations in SARS-CoV		Mutations in strain 229E		Mutations in strain OC43	
I	II	Number	Percentage	Number	Percentage	Number	Percentage
AF = 0, PF > 0	AF = 0, PF > 0	0[a]	0	0[a]	0	2[a]	2.5
AF = 0, PF > 0	AF = PF = 0	0[a]	0	0[a]	0	1[a]	1.25
AF = 0, PF > 0	AF = PF > 0	0[a]	0	0[a]	0	0[a]	0
AF = 0, PF > 0	AF < PF, AF ≠ 0	1[a]	33.33	0[a]	0	4[a]	5
AF = 0, PF > 0	AF > PF	0[a]	0	1[a]	2.63	7[a]	8.75
AF = PF = 0	AF = PF = 0	0	0	0	0	0	0
AF = PF = 0	AF = PF > 0	0	0	0	0	0	0
AF = PF = 0	AF < PF, AF ≠ 0	0[a]	0	0[a]	0	0[a]	0
AF = PF = 0	AF > PF	1	0	1	2.63	0	0
AF < PF, AF ≠ 0	AF < PF, AF ≠ 0	1[a]	33.33	9[a]	23.68	6[a]	7.5
AF < PF, AF ≠ 0	AF = PF > 0	0[a]	0	2[a]	5.26	11[a]	13.75
AF < PF, AF ≠ 0	AF > PF	0[a]	0	11[a]	28.95	26[a]	32.5
AF = PF > 0	AF = PF > 0	0	0	0	0	3	3.75
AF > PF	AF > PF	0	0	8	21.05	9	11.25
AF = PF > 0	AF > PF	1	33.33	6	15.79	11	13.75

[a] It indicates one or both substituting amino acid pairs with their actual frequency smaller than predicted frequency. These amino acid pairs are 66.67, 60.52 and 69% of total amino acid pairs in SARS-CoV, strain 229E and strain OC43, respectively.

dicted frequency in one or both substituted pairs (the first three rows in unpredictable pairs). Comparing the first three rows with the last two rows in unpredictable pairs, we can see that the mutations are more likely to target the pairs whose actual frequencies are larger than predicted frequencies. Therefore, Table 4 suggests which type of amino acid pairs are more likely to be substituted, i.e. the different sensitivities of amino acid pairs to mutations in spike proteins.

Table 5 shows in which types of amino acid pairs the mutations are likely or unlikely to form in spike proteins. We can find that more than 60% of mutations result in one or both substituting pairs whose actual frequencies are smaller than their predicted frequencies. Taking the results in both Tables 3 and 4 into account, the mutations are likely to attack the pairs whose actual frequencies are larger than their predicted ones and the consequences of mutations are likely to form the pairs whose actual frequencies are smaller than their predicted ones. In such a manner, the mutations reduce the difference between actual and predicted frequencies (Fig. 2).

4. Discussion

In this study, we have analyzed the amino acid pairs affected by mutations in three spike proteins in order to gain some insight into the possible mutations from SARS-CoV. Firstly, the present results demonstrate that the randomly unpredictable amino acid pairs are more sensitive to the mutations (Table 2), although these 3 spike proteins are constructed by different types of amino acid pairs which repeat different times (Table 1). Furthermore, the larger the difference between actual and predicted frequencies is, the higher the chance of mutation occurring is (Fig. 1). The effect induced by mutations is to reduce the difference between actual and predicted frequencies (Fig. 2). Finally, the amino acid pairs whose actual frequencies are larger than their predicted frequencies are more likely to be targeted by mutations (Table 4), whereas the amino acid pairs whose actual frequencies are smaller than their predicted frequencies are more likely to be formed after mutations (Table 5). These findings are identical to our recently publications [26–34].

Combining the results with our previous studies, our model suggests that the mutations go along a pathway, which is probabilistically more likely to occur. As such a pathway is less energy- and time-consuming, in fact, the mutations represent a process of degeneration inducing human diseases. Although this study shows that the mutations in the spike proteins from strains 229E and OC43 go along the direction of degeneration, i.e. the mutations go along a probabilistically easy pathway, the documented evidence in literature still cannot suggest whether or not the mutations belong to degeneration in the spike protein from human SARS-CoV.

If the potential mutations in the spike protein from SARS-CoV would go along a probabilistically easy pathway, according to the results obtained from our analysis, we should pay more attention to the amino acid pairs with the following characteristics for potential mutations, i.e. the amino acid pairs with large difference between actual and predicted frequencies and their actual frequencies larger than predicted frequencies. Table 6 lists the amino acid pairs with the frequency difference being larger than 3 in spike protein from SARS-CoV, as these amino acid pairs seem to be more vulnerable to mutations (Fig. 1). With these sensitive amino acid pairs in mind, we can easily determine their positions. For example, the “NF”s are located at positions 129–130, 178–179, 230–231, 304–305, 526–527, 528–529, 699–700, 783–784, 951–952, 1056–1057 and 1090–1091 in spike protein from SARS-CoV. Moreover we notice that

Table 6
Amino acid pairs being more likely to be targeted by mutation in the spike protein from human SARS-CoV

Difference between actual and predicted frequencies	Amino acid pair	Actual frequency	Predicted frequency
6	NF	11	5
6	DV	11	5
6	FN	11	5
5	HT	6	1
5	TS	13	8
5	VV	12	7
4	AA	10	6
4	QI	7	3
4	GI	9	5
4	IA	9	5
4	PF	8	4
4	TQ	8	4

the positions of "FN"s overlap with "NF"s at positions from 526 to 530, at which the highly possible mutations would be more likely to occur. This hypothesis can be supported by the mutations found in other proteins, such as human collagen α5(IV) chain precursor [29], p53 protein [34] and so on. In such a manner, we could predict the potential mutations in the spike protein from human SARS-CoV with possible amino acid pairs and positions.

Acknowledgments

The authors wish to thank the anonymous referees for their insightful comments, which sharpen up the points presented in this study. Also the special thanks go to the first referee for correcting the English in the previous version of manuscript.

References

[1] Bairoch A, Apweiler R. The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 2000. Nucleic Acids Res 2000;28:45–8.
[2] Bellini WJ, Campagnoli RP, Icenogle JP, Monroe SS, Nix WA, Oberste MS, et al. The EMBL GenBank DDBJ databases, 2003. Submitted APR-2003 to the EMBL GenBank DDBJ databases.
[3] Bonavia A, Holmes KV, Viral and cellular changes in a human cell line persistently infected with human coronavirus HCoV-229E. The EMBL GenBank DDBJ databases, 2001. Submitted FEB-2001 to the EMBL GenBank DDBJ databases.
[4] Bonavia A, Zelus BD, Wentworth DE, Talbot PJ, Holmes KV. Identification of a receptor-binding domain of the spike glycoprotein of human coronavirus HCoV-229E. J Virol 2003;77:2530–8.
[5] Breslin JJ, Mork I, Smith MK, Vogel LK, Hemmila EM, Bonavia A, et al, Human coronavirus 229E: receptor binding domain and neutralization by soluble receptor at 37 °C . J Virol 2003;77:4435–8.
[6] Chen LL, Ou HY, Zhang R, Zhang CT. ZCURVE_CoV: a new system to recognize protein coding genes in coronavirus genomes, and its applications in analyzing SARS-CoV genomes. Biochem Biophys Res Commun 2003;307:382–8.
[7] Chou KC, Wei DQ, Zhong WZ. Binding mechanism of coronavirus main proteinase with ligands and its implication to drug design against SARS. Biochem Biophys Res Commun 2003;308:148–51.
[8] Emery S, Erdman D, Peret T, Ksiazek T. The EMBL GenBank DDBJ databases, 2003. Submitted APR-2003 to the EMBL GenBank DDBJ databases.
[9] Feller W. An introduction to probability theory and its applications, 3rd ed., vol. I. New York: Wiley; 1968.
[10] Gallagher TM, Buchmeier MJ. Coronavirus spike proteins in viral entry and pathogenesis. Virology 2001;279:371–4.
[11] Gu XM, Lai MD. Genome analysis of the SARS-associated virus. Zhejiang Da Xue Xue Bao Yi Xue Ban 2003;32:362–8.
[12] Hays JP, Myint SH. PCR sequencing of the spike genes of geographically and chronologically distinct human coronaviruses 229E. J Virol Methods 1998;75:179–93.
[13] Kuenkel F, Herrler G. Structural and functional analysis of the surface protein of Human coronavirus OC43. Virology 1993;195:195–202.
[14] Kuenkel F, Herrler G. Structural and functional analysis of the S proteins of two human coronavirus OC43 strains adapted to growth in different cells. Arch Virol 1996;1411:1123–31.
[15] Leung FC, Zeng F, Chan CWM, Chan CMY, Chen J, Chow KYC, et al. The EMBL GenBank DDBJ databases, 2003. Submitted APR-2003 to the EMBL GenBank DDBJ databases.
[16] Lin J-H, Chiu S-C, Yang J-Y, Wang S-F, Chen H-Y. Detection of a novel human coronavirus in a severe acute respiratory syndrome patient in Taiwan. The EMBL GenBank DDBJ databases, 2003. Submitted for publication.
[17] Marra M, Jones SJM, Holt R. The complete genome of the SARS associated coronavirus. The EMBL GenBank DDBJ databases, 2003. Submitted APR-2003 to the EMBL GenBank DDBJ databases.
[18] Mounir S, Talbot PJ. Molecular characterization of the S protein gene of human coronavirus OC43. J Gen Virol 1993;74:1981–7.
[19] Qin E, Zhu Q, Yu M, Fan B, Chang G, Si B, et al. SARS coronavirus BJ03 isolate genome sequence. The EMBL GenBank DDBJ databases, 2003. Submitted APR-2003 to the EMBL GenBank DDBJ databases.
[20] Raabe T, Schelle-Prinz B, Siddell SG. Nucleotide sequence of the gene encoding the spike glycoprotein of human coronavirus HCV 229E. J Gen Virol 1990;71:1065–73.
[21] Raabe T, Siddell S. Nucleotide sequence of the human coronavirus HCV 229E mRNA 4 and mRNA 5 unique regions. Nucleic Acids Res 1989;17:6387.
[22] Rota PA, Oberste MS, Monroe SS, Nix WA, Campagnoli R, Icenogle JP, et al. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science 2003;30:1394–9.
[23] Thiel V, Herold J, Schelle B, Siddell SG. Infectious RNA transcribed in vitro from a cDNA copy of the human coronavirus genome cloned in vaccinia virus. J Gen Virol 2001;82:1273–81.
[24] Tsui SKW, Lo DYM, Tam JS, Fung KP, Chim SSC, Au CC, et al. DNA sequence of a human coronavirus (CUHK-W1) from a patient with severe acute respiratory syndrome (SARS) in Hong Kong. The EMBL GenBank DDBJ databases, 2003. Submitted APR-2003 to the EMBL GenBank DDBJ databases.
[25] Wu G, Yan SM. Randomness in the primary structure of protein: methods and implications. Mol Biol Today 2002;3:55–69.
[26] Wu G, Yan SM. Prediction of presence and absence of two- and three-amino-acid sequence of human monoamine oxidase B from its amino acid composition according to the random mechanism. Biomol Eng 2001;18:23–7.
[27] Wu G, Yan S. Determination of amino acid pairs sensitive to variants in human low-density lipoprotein receptor precursor by means of a random approach. J Biochem Mol Biol Biophys 2002;6:401–6.
[28] Wu G, Yan SM. Estimation of amino acid pairs sensitive to variants in human phenylalanine hydroxylase protein by means of a random approach. Peptides 2002;23:2085–90.
[29] Wu G, Yan S. Analysis of amino acid pairs sensitive to variants in human collagen α5(IV) chain precursor by means of a random approach. Peptides 2003;24:347–52.
[30] Wu G, Yan S. Determination of amino acid pairs sensitive to variants in human β-glucocerebrosidase by means of a random approach. Protein Eng 2003;16:195–9.

[31] Wu G, Yan SM. Determination of amino acid pairs in human haemoglobulin-chain sensitive to variants by means of a random approach. Comp Clin Pathol 2003;12:21–5.
[32] Wu G, Yan S. Determination of amino acid pairs sensitive to variants in human Bruton's tyrosine kinase by means of a random approach. Mol Simul 2003;29:249–54.
[33] Wu G, Yan S. Determination of amino acid pairs sensitive to variants in human coagulation factor IX precursor by means of a random approach. J Biomed Sci 2003;10:451–4.
[34] Wu G, Yan S. Determination of amino acid pairs in human p53 protein sensitive to mutations/variants by means of a random approach. J Mol Mod 2003;9:337–41.