American Journal of Infectious Diseases 1 (2): 116-123, 2005
ISSN 1553-6203

Searching of Main Cause Leading to Severe Influenza A Virus Mutations and Consequently to Influenza Pandemics/Epidemics

Guang Wu and Shaomin Yan
Computational Mutation Project, DreamSciTech Consulting, 301, Building 12, Nanyou A-zone Jiannan Road, Shenzhen, Guangdong Province, CN-518054, China

Abstract: The unpredictable mutations in the proteins from influenza A virus lead to the great difficulty in prevention of possible outbreak of bird flu and pandemic/epidemic of influenza. This unpredictability is due to the fact that we know little about the causes that lead to the mutations. In three of our recent studies on the hemagglutinins from influenza A virus, we unintentionally noticed the periodicity of mutations in hemagglutinins similar to the periodicity of sunspot. We calculated the amino-acid pair predictability and amino-acid distribution rank, which are developed by us over last several years and can numerically present the evolution of proteins in question, of 1217 full-length hemagglutinins from influenza A viruses. We then used the fast Fourier transform to determine the periodicity of mutations in the hemagglutinins. We compare the periodicities of mutations in influenza A virus hemagglutinins with those of solar and galactic cosmic rays and find a main periodicity of the mutations identical to that of sunspot and neutron rate (11 years/circle). Then we plot the sunspot number with respect to the historical pandemics/epidemics/non-pandemic new strains over last three centuries and compare the recorded sunspots with the historical pandemics before 1700. Both show a good agreement between sunspot activity and influenza related events. As the histories of Sun and galaxy are incomparably much longer than the history of influenza virus, the only logical deduction is that the hemagglutinin periodicities, which are identical to the periodicities of solar and galactic cosmic rays, are attribute to the solar and galactic activity. As the hemagglutinin is a sample of influenza A virus, we can logically deduce the role of migratory wild birds on the outbreak of bird flu and influenza, that is, cosmic rays are heading towards the polar regions, where more mutations occur in influenza A virus either within the wild birds or in their living environments and as the winter approaches, these waterfowl fly forwards warm south bringing back the new mutated influenza A virus leading to outbreak of bird flu or influenza.

Key words: Bird flu, cosmic ray, hemagglutinin, mutation, pandemic/epidemic of influenza, sunspot

INTRODUCTION

Never before the pandemic/epidemic of influenza threatens humans to such a degree, this is so because we are still unable to predict the mutations, which lead to the possible human to human transmission, in the proteins from influenza A virus, of which the surface proteins, hemagglutinin and neuraminidase, are particularly subject to the frequent mutations.

On the other hand, we could be in a better position in timing and dealing with the possible pandemic/epidemic of influenza if we would know the causes that lead to severe mutations. One clue that contributes to the spread of bird flu is the seasonal migration of wild birds. The question raised from here is where the wild birds get the new mutated influenza A viruses during their travelling.

In three of our recent studies on the proteins from influenza A virus[1-3], we unintentionally noticed one of periodicities of hemagglutinin mutations similar to the periodicity of sunspot. This is very suggestive because it is well known that radiation can induce mutations. Perhaps this would be one of causes of mutations in the proteins from influenza A virus. Along this line of thought, we conduct the current study.

MATERIALS AND METHODS

Hemagglutinin sequences: The complete amino-acid sequences of 1217 hemagglutinins from influenza A viruses isolated from 1918 to 2005 are obtained from the Medline Protein databank (Table 1).

Solar and galactic cosmic ray data: The sunspot numbers, neutron and proton monitor data are obtained from the Royal Observatory of Belgium harbours the Sunspot Index Data center[4], the Moscow Neutron Monitor in Moscow cosmic ray station[5] and CELIAS/SEM material[6], respectively.

Corresponding Author: Guang Wu, Computational Mutation Project, DreamSciTech Consulting, 301, Building 12, Nanyou A-zone, Jiannan Road, Shenzhen, Guangdong Province, CN-518054, China, Tel/Fax: +86 755 2528 8156

Table 1: Hemagglutinins of influenza A viruses analyzed in this study

Subtypes	Species	Number	Years
H1	Avian	14	1976–1999
	Human	55	1918–2001
	Swine	65	1930–2002
H2	Avian	19	1961–1998
	Human	22	1957–2005
H3	Avian	44	1969–2003
	Human	100	1968–2005
	Equine	46	1963–1994
	Seal	2	1992
	Swine	48	1977–2002
H4	Avian	33	1956–2000
	Seal	3	1982
	Swine	2	1999
H5	Avian	251	1959–2005
	Human	24	1997–2004
	Leopard	1	2004
	Swine	2	2003
	Tiger	1	2004
	unknown	3	2001
H6	Avian	75	1972–2004
H7	Avian	147	1934–2004
	Human	4	1934–2003
	Equine	22	1956–1977
	Seal	1	1980
H8	Avian	6	1968–1994
H9	Avian	141	1966–2003
	Human	7	1998–1999
	Swine	23	1998–2003
H10	Avian	3	1949
	Minks	3	1984
H11	Avian	3	1956
H12	Avian	2	1976
H13	Avian	8	1977–1984
	Whale	3	1984
H14	Avian	3	1982
H15	Avian	2	1979, 1983
H16	Avian	3	1999
Unidentified	Avian	5	1963–2000
	Equine	2	1971, 1976
	Human	14	1994–1998
	Swine	5	1986–1998

Computational mutation calculation: The rationale and detailed description of our approaches have been published in our previous publications (for details, see review articles[3,7,8]) and we describe our approaches with the example of the hemagglutinin of 1918 "Spanish" influenza A virus (accession number AF117241), which is the earliest complete sequence documented in the data bank[9,10].

Amino-acid pair predictability: As we know that an amino-acid pair in a protein is composed of any 20 kinds of amino acids, so theoretically there are 400 possible types of amino-acid pairs. In terms of amino-acid pairs, distinguishing of protein differs in the number of possible types of amino-acid pairs and/or in the frequency of each type. The 1918 hemagglutinin is composed of 566 amino acids, thus there are 565 amino-acid pairs. Of 400 possible types, 137 are absent and 263 present: 111 types appear once, 71 twice, 42 three, 24 four, 8 five, 5 six, 1 nine and 1 eleven times, respectively.

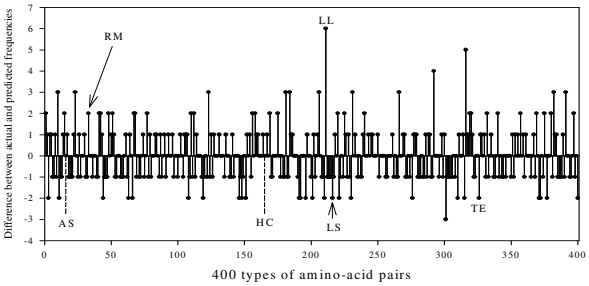

Fig. 1: Predictability of 400 types of amino-acid pairs from "AA", "AR", . . . to . . . "VY", "VV" in the hemagglutinin of 1918 "Spanish" influenza virus

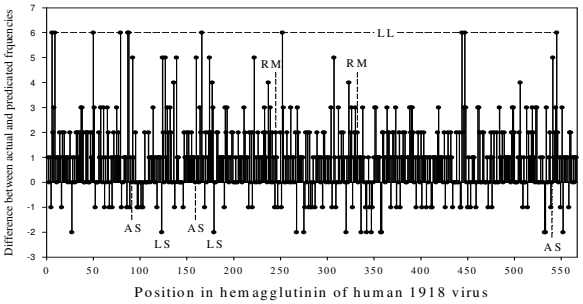

Fig. 2: Predictability of 565 amino-acid pairs in the hemagglutinin of 1918 "Spanish" influenza virus

Randomly predictable present type of amino-acid pair with predictable frequency: There are 37 alanines (A) and 49 serines (S) in the 1918 hemagglutinin, the frequency of random presence of amino-acid pair "AS" is 3 (37/566 × 49/565 × 565 = 3.203), that is, "AS" would appear three times in this hemagglutinin. Actually we do find three "AS"s in the hemagglutinin, so the actual frequency of "AS" is 3. In this case both the presence of the type "AS" and its frequency are predictable and the difference between its actual and predicted frequencies is 0, locating on the zero-line in both Fig. 1 and 2.

Randomly predictable present type of amino-acid pair with unpredictable frequency: There are 51 leucines (L) in the 1918 hemagglutinin, the frequency of random presence of amino-acid pair "LL" is 5 (51/566 × 50/565 × 565 = 4.505), i.e. there would be five "LL"s in this hemagglutinin. But actually "LL" appears eleven times, so the presence of "LL" is predictable, but its frequency is unpredictable and the difference between its actual and predicted frequencies is 6 (Fig. 1 and 2).

This is also the case that the actual frequency is larger than the predicted one. Another case is that the actual frequency is smaller than its predicted one. For instance, the predicted frequency of "LS" is 4 (51/566 × 49/565 × 565 = 4.415), while its actual frequency is 2 so the difference between its actual and predicted frequencies is -2 (Fig. 1 and 2).

Table 2: Possible distribution patterns regarding 11 "W"s in 11 imaging parts of the 1918 hemagglutinin

I	II	III	IV	V	VI	VII	VIII	IX	X	XI	Probability	Rank
W	W	W	W	W	W	W	W	W	W	W	1.3991e-4	30
	W	W	W	W	W	W	W	W	W	WW	7.6948e-3	14
		W	***W***	***W***	***W***	***W***	***W***	***W***	***WW***	***WW***	***0.0693***	***6***
			W	W	W	W	W	WW	WW	WW	0.1616	2
				W	W	W	WW	WW	WW	WW	0.1010	4
					W	WW	WW	WW	WW	WW	0.0121	11
		W	W	W	W	W	W	W	W	WWW	0.0115	12
			W	W	W	W	W	W	WW	WWW	0.1077	3
				W	W	W	W	WW	WW	WWW	0.2020	1
					W	W	WW	WW	WW	WWW	0.0808	5
						WW	WW	WW	WW	WWW	3.3665e-3	17
				W	W	W	W	W	WWW	WWW	0.0269	9
					W	W	W	WW	WWW	WWW	0.0539	7
						W	WW	WW	WWW	WWW	0.0135	10
						W	W	WWW	WWW	WWW	2.9924e-3	18
							WW	WWW	WWW	WWW	4.2749e-4	28
			W	W	W	W	W	W	W	WWWW	7.6948e-3	14
				W	W	W	W	W	WW	WWWW	0.0404	8
					W	W	W	WW	WW	WWWW	0.0404	8
						W	WW	WW	WW	WWWW	6.7330e-3	15
					W	W	W	W	WWW	WWWW	0.0135	10
						W	W	WW	WWW	WWWW	0.0135	10
							WW	WW	WWW	WWWW	9.6185e-4	22
							W	WWW	WWW	WWWW	6.4124e-4	25
						W	W	W	WWWW	WWWW	1.1222e-3	21
							W	WW	WWWW	WWWW	4.8093e-4	27
								WWW	WWWW	WWWW	2.0039e-5	35
				W	W	W	W	W	W	WWWWW	2.6932e-3	19
					W	W	W	W	W	WWWWW	8.0796e-3	13
						W	W	WW	WW	WWWWW	4.0398e-3	16
							WW	WW	WW	WWWWW	1.9237e-4	29
						W	W	W	WWW	WWWWW	1.7955e-3	20
							W	WW	WWW	WWWWW	7.6948e-4	24
								WWW	WWW	WWWWW	1.6031e-5	36
							W	W	WWWW	WWWWW	1.9237e-4	29
								WW	WWWW	WWWWW	2.4046e-5	34
								W	WWWWW	WWWWW	4.8093e-6	38
					W	W	W	W	W	WWWWWW	5.3864e-4	26
						W	W	W	WW	WWWWWW	8.9773e-4	23
							W	WW	WW	WWWWWW	1.9237e-4	29
							W	W	WWW	WWWWWW	1.2825e-4	31
								WW	WWW	WWWWWW	1.6031e-5	36
								W	WWWW	WWWWWW	8.0154e-6	37
									WWWWW	WWWWWW	1.7812e-7	43
						W	W	W	W	WWWWWWW	6.4124e-5	32
							W	W	WW	WWWWWWW	5.4963e-5	33
								WW	WW	WWWWWWW	3.4352e-6	40
								W	WWW	WWWWWWW	4.5803e-6	39
									WWWW	WWWWWWW	1.2723e-7	44
							W	W	W	WWWWWWWW	4.5803e-6	39
								W	WW	WWWWWWWW	1.7176e-6	41
									WWW	WWWWWWWW	6.3615e-8	45
								W	W	WWWWWWWWW	1.9084e-7	42
									WW	WWWWWWWWW	2.1205e-8	46
									W	WWWWWWWWWW	4.2410e-9	47
										WWWWWWWWWW	3.8554e-11	48

Bold italic font indicates the real distribution pattern in the 1918 hemagglutinin.

Randomly unpredictable present type of amino-acid pair: There are 20 arginines (R) and 8 methionines (M) in the 1918 hemagglutinin and the predicted frequency of "RM" is 0 ($20/566 \times 8/565 \times 565 = 0.283$), so "RM" would not appear in this hemagglutinin. However it appears twice in the reality, thus the presence of "RM" is unpredictable. Naturally its frequency is unpredictable too and the difference between its actual and predicted frequencies is 2 (Fig. 1 and 2).

Randomly predictable absent type of amino-acid pair: There are 13 histidines (H) and 16 cysteines (C) in the 1918 hemagglutinin. The predicted frequency of "HC" is 0 ($13/566 \times 16/565 \times 565 = 0.367$), i.e. "HC" would not appear in this hemagglutinin, which is true in the real situation. Thus the absence of "HC" with its frequency is predictable and the difference between its actual and predicted frequencies is 0, which appears in Fig. 1 but not in Fig. 2.

Randomly unpredictable absent type of amino-acid pair: There are 37 threonines (T) and 35 glutamic acids (E) in the 1918 hemagglutinin. The predicted frequency of "TE" is 2 ($37/566 \times 35/565 \times 565 = 2.288$), i.e. there would be two "TE"s in this hemagglutinin. However no "TE" appears in the actuality (Fig. 2), therefore the absence of "TE" from the 1918 hemagglutinin is unpredictable. Naturally its frequency is unpredictable too and the difference between its actual and predicted frequencies is -2 (Fig. 1).

Predictable and unpredictable portions of amino-acid pairs: After the calculations described above, the amino-acid pairs in a protein can be classified as predictable and unpredictable portions with respect to type and frequency and the sum of both predictable and unpredictable portions is 100%. Of the amino-acid pairs in the 1918 hemagglutinin, the unpredictable type and frequency are 68.75 and 76.81%, respectively. Either predictable portion or unpredictable portion can serve as a quantitative measure to present a protein.

Type mutation and frequency mutation: As there are 400 types of theoretically possible amino-acid pairs and we use the 100% to classify them as predicable and unpredictable types, thus 0.25% represents one of 400 types, so a 0.25% change indicates that one of 400 types mutates to an unpredictable type from a predictable one or vice versa. This is the type mutation. However, the situation related to the frequency of amino-acid pairs is dependent on the length of amino-acid sequence. Hence, about 0.18% (1/565) change can be regarded as a modification in an amino-acid pair in the 1918 hemagglutinin, as there are 565 amino-acid pairs in the protein. This is the frequency mutation.

Amino-acid distribution probability/rank: The position of any 20 kinds of amino acids in a protein can be determined by experimental approach, so each kind of amino acids has a certain distribution pattern with respect to its position in the protein. Furthermore a certain distribution pattern can be associated with a certain probability, which can be calculated according to the occupancy problems of subpopulations and partitions[11]. For a certain distribution of a kind of amino acid in a protein, its distribution probability is equal to $r!/(q_0! \times q_1! \times ... \times q_n!) \times r!/(r_1! \times r_2! \times ... \times r_n!) \times n^{-r}$, where ! is the factorial function, i.e. $n! = n \times (n - 1) \times (n - 2) \times ... \times 1$, $0! = 1$ by definition, r is the number of a kind of amino acid, q is the number of parts with the same number of amino acids and n is the number of grouped parts in the protein for a kind of amino acid.

For example, there are 11 tryptophans (W) in the 1918 hemagglutinin, how do these 11 "W"s distribute among 566 amino acids in this hemagglutinin? We can imagine to group it into 11 parts and each one contains about 51 amino acids (566/11 = 51.45). Table 2 lists all 56 possible distribution patterns regarding 11 "W"s in 11 imaging parts and their distribution probabilities and ranks. The first 11 columns in Table 2 show that the 1918 hemagglutinin is grouped into 11 imaging parts and the first 11 cells in each row represent a possible distribution pattern of "W"s and the last two columns display the corresponding distribution probability and rank.

As different distribution patterns can have the same distribution probability, we rank the distribution probabilities according to a descending order, thus the largest distribution probability is ranked as one. Again in our example, there are 56 possible distributions for 11 "W"s in 11 parts in Table 2, while there are only 48 distribution ranks. In general, the smaller the distribution rank is, the larger the distribution probability is. Although there are many possible distributions for a kind of amino acids in a hemagglutinin (such as 56 possible distributions for 11 "W"s), the hemagglutinin in question possesses only one distribution pattern, therefore there is only one distribution probability/rank for each kind of amino acids and a maximum of 20 distribution probabilities or ranks in a protein.

Similarly, we imagine to group 8 parts for 8 metheonines (M), 13 parts for 13 histidines (H), 16 parts for 16 cysteines (C), 17 parts for 17 glutamines (Q), 19 parts for 19 phenylalanines (F) and prolines (P), 20 parts for 20 arginines (R), 25 parts for 25 aspartic acids (D), 26 parts for 26 tyrosines (Y), 32 parts for 32 isoleucines (I) and valines (V), 33 parts for 33 lysines (K), 35 parts for 35 glutamic acids (E), 37 parts for 37 alanines and threonines (T), 42 parts for 42 asparagines (N), 44 parts for 44 glycines (G), 49 parts for 49 serines (S), 51 parts for 51 leucines (L) in the hemagglutinin of 1918 "Spanish" influenza virus and conduct the similar calculations.

As different kinds of amino acids have different contributions to a protein, we standardize them by means of the distribution rank per amino acid, which is calculated by dividing the rank of each kind of amino acids by the number of corresponding amino acids. In the 1918 hemagglutinin, the distribution rank of "W"s is 6/11 = 0.545, because these 11 "W"s distribute in the hemagglutinin with the distribution probability (0.0693). Accordingly, the sum of ranks for all 20 kinds of amino acids is 22.8698, thus the distribution rank in this hemagglutinin is 22.8698/20 = 1.1435. Naturally, the distribution rank can serve as a quantitative measure to present a protein.

Fast Fourier transform: One of important applications of Fourier analysis is to determine the periodicity in a chaotic fluctuating dataset, so we use it to analyze the

potential periodicity of hemagglutinin mutations over time.

RESULTS AND DISCUSSION

One advantage of our approaches is that we can use either amino-acid pair predictability or amino-acid distribution rank as a single value to numerically present a hemagglutinin. In this way, we can present each hemagglutinin with an unambiguous value seeing that we can view each hemagglutinin as a sample from its evolution and a sample from the evolution of influenza A virus. Also, both amino-acid pair predictability and amino-acid distribution rank are very sensitive to mutations in hemagglutinins[3,8,12], in particular, our first approach reveals the mutation in individual amino-acid pair and the second one indicates the sensitive position. We can therefore compare different hemagglutinins with respect to species, subtypes, etc. Moreover, we can use the sophisticated mathematical methods to analyze hemagglutinins[1,2,13].

Figure 3 shows the evolutionary process of influenza A virus hemagglutinins over almost a century in terms of amino-acid pair predictability and amino-acid distribution rank in comparison with solar and galactic cosmic rays. We include the galactic cosmic rays for comparison because the solar cosmic rays are more concentrated towards lower energies while the galactic cosmic rays are more concentrated towards higher energies[14] although we only noticed the relationship between one of periodicities of hemagglutinin mutations and that of sunspot number. In Fig. 3, the historical mutations are recorded as the fluctuations in unpredictable type, frequency and rank. The general trend can be seen through the regressed line and trend channel, but the fluctuations imply that the mutations occur all the time. In comparison with solar and galactic cosmic activity, the fluctuations of unpredictable type, frequency and distribution rank are irregular. This nevertheless imposes the big difficulty in prediction of influenza A virus mutations.

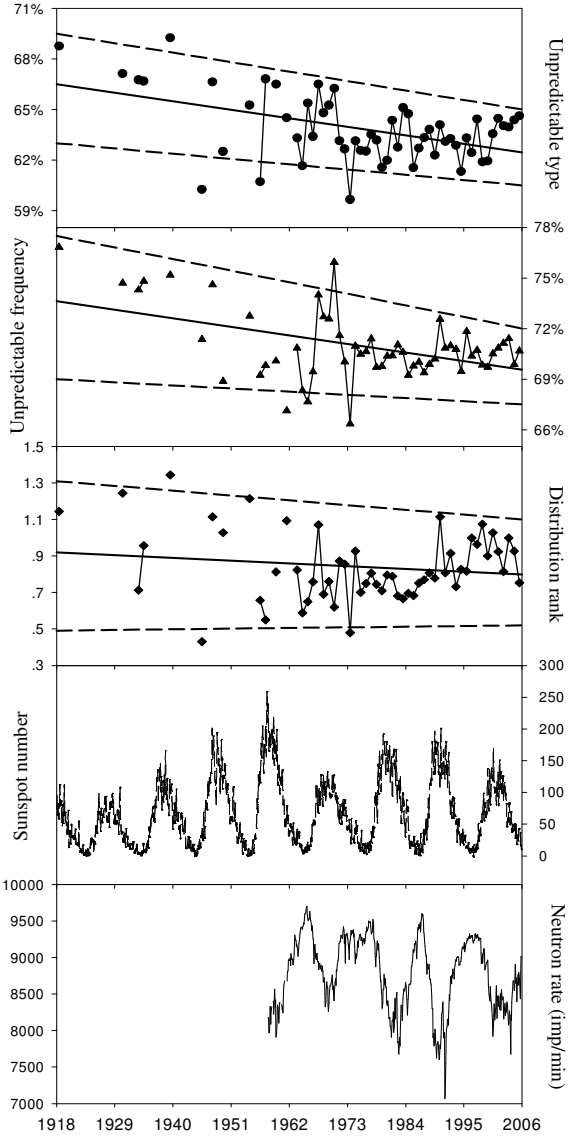

Fig. 3: Unpredictable type of amino-acid pairs, unpredictable frequency of amino-acid pairs, amino-acid distribution rank, monthly average sunspot number and monthly average neutron rate along the time course. Each symbol in top three panels presents the mean value of unpredictable type, frequency and rank in given year. The solid and dashed lines in top three panels are regressed lines and trend channel lines

Figure 4 demonstrates that irregular fluctuations in unpredictable type, frequency and distribution rank have different periodicities after fast Fourier transform. It can be found in Fig. 4 that the unpredictable type, frequency and distribution rank do have the periodicities in consistent with the periodicities of sunspot numbers and neutron monitor data including the best-known 11-year periodicity of solar activity. At the periodicity of 11 years, the unpredictable type and frequency provide different interpretations for hemagglutinin mutations. The unpredictable type has 1.35% type mutations, that is, about 5.4 (1.35/0.25%) types of amino-acid pairs would experience mutation at the interval of 11 years, meanwhile the unpredictable frequency has 4.54 frequency mutations, say, about 25 (4.54%/0.18) amino-acid pairs would experience mutation at the interval of 11 years. Another important periodicity is about 7 years in unpredictable type, frequency, distribution rank and neutron rate. In fact, we used 7 years as a periodicity to timing the mutation trend in hemagglutinins in our previous studies[1,2] and this periodicity is identical to the interval found in the analysis on influenza history[15]. Still, small periodicities such those between 2 to 3 years are

identical to the observation on the migrating wild aquatic birds in North America[16].

The periodicities, which are not in consistent with the periodicities of solar and galactic rays, can be attribute to various causes, of which our approaches can provide a piece of explanation.

Table 3: Recorded sunspot activity and influenza pandemics or epidemics before 1700

Recorded sunspot year	Pandemics/epidemics
	412 BC
1171	1173
1511	1510 (probably)
1556	1557 (possibly)
1590*	1580 (agreed)

*, deduced 1579 was the maximum.

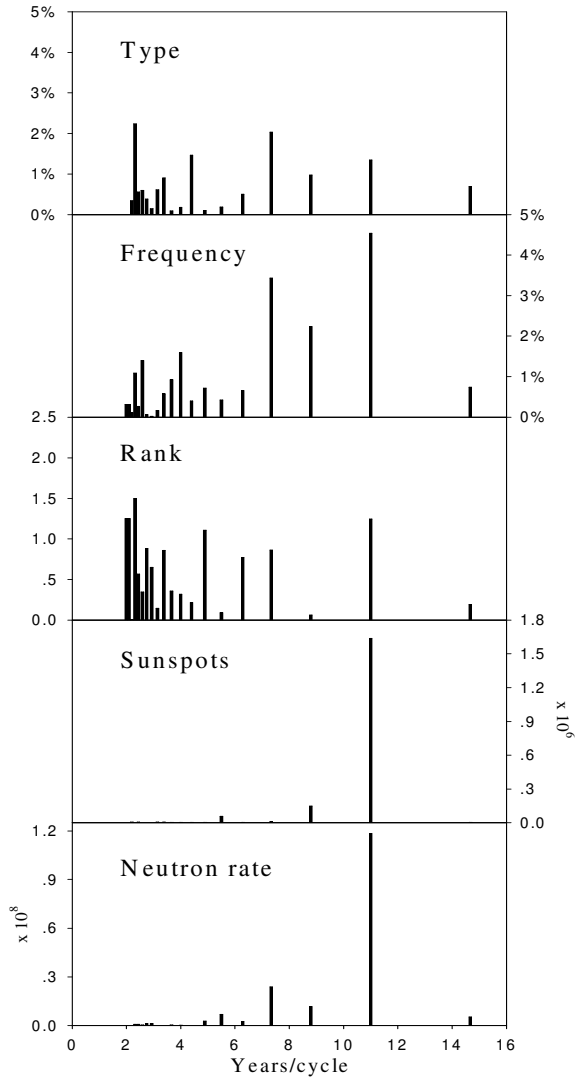

Fig. 4: Periodicity of unpredictable type of amino-acid pairs, unpredictable frequency of amino-acid pairs, amino-acid distribution rank, sunspot number and neutron rate over the last 43 years. The periodicities in top three panels are calculated from 1163 influenza A virus hemagglutinins

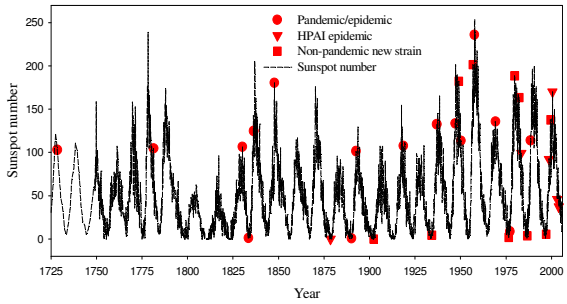

Fig. 5: Historical pandemics/epidemics, high pathogenic avian influenza (HPAI) epidemics and non-pandemic new strains with respect to the sunspot number (yearly sunspot number before 1749 and thereafter monthly sunspot number)

Nature should deliberately construct the amino-acid pairs whose actual frequency differs from the predicted one. The functional amino-acid pairs should be deliberately evolved, so a protein would keep only absolutely necessary unpredictable amino-acid pairs. During the evolutionary process, nature tries to minimize the unpredictable portion through mutations, which can be evidenced by seeing the decreased regressed lines of unpredictable type and frequency in Fig. 3. However, this process may bring about new unpredictable amino-acid pairs, so the evolutionary process is continuing. The construction of a protein with large predictable portions of amino-acid pairs and with small amino-acid distribution rank is certainly a way to adapt to the fast-changes in surroundings and environments, as the construction speed of a protein might be crucial for its survival, this would require the least time and energy.

Other explanations can include the mutations due to the pressure of the antibody[18] and due to the selective pressure for the appearance of host cell variant with altered receptor binding specificities[19].

Therefore, the superposition of various periodicities and random factors results in the irregular fluctuations in Fig. 3. On the other hand, there are more type mutations with shorter periodicity, implicating that at first the mutation leads to the appearance or disappearance of a type of amino-acid pair and then the mutation continues to modify the present type of amino-acid pairs in long term.

Figure 5 gives the clear evidence that there is a strong link between historical pandemics/epidemics and solar activity. Apparently, the historical pandemics or epidemics occurred at either the maximum or the minimum of sunspots. This means that the solar activity does induce the mutations in influenza A virus, however the occurrence of pandemics/epidemics would be dependent on the virulence of mutated virus as we can see that the non-pandemic new strains based on available data appeared at either the maximum or the minimum of sunspots, too. A piece of evidence, which can explain the fact that more influenza related events

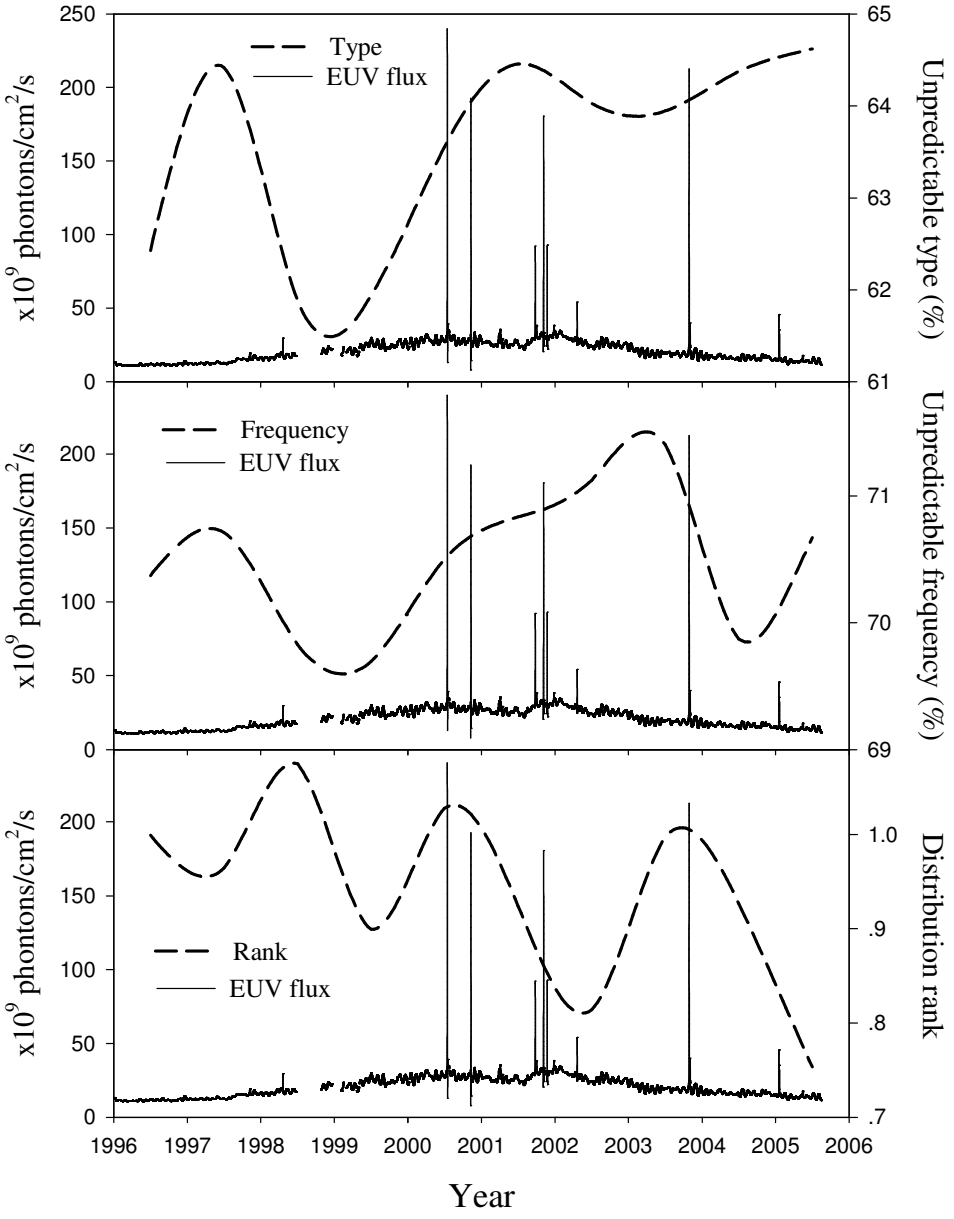

Fig. 6: Effects of extreme ultraviolet (EUV) flux (26–34 nm) on unpredictable type and frequency of amino-acid pairs and on amino-acid distribution rank. The data are calculated from 713 influenza A virus hemagglutinins from 1996 to 2005

are recorded over past 50 years in Fig. 5, is that four chaotic solar cycles have been recorded with the 300% higher-than-normal intensity of ultraviolet radiation[20].

Table 3 lists the outbreaks of pandemics/epidemics and sunspots before the systematic observation of sunspot activity[21]. There is a good agreement between recorded sunspots and identified pandemics or epidemics[22,23]. Although no recorded sunspot is documented in 412 BC, there was a lunar eclipse in 413 BC[24].

On the other hand, it is interesting that the pandemics/epidemics can occur at the minimal level of sunspot number, which is against the concept that strong solar activity induces the mutations. However, more galactic cosmic rays reach the Earth during the solar minimums because the Sun emits plasma and magnetic fields expelling some cosmic rays, i.e., the galactic cosmic activity counterbalances solar activity[14].

Intriguingly, it seems that more high pathogenic avian influenza epidemics were recorded since 2000 in Fig. 5. These phenomena can be attribute to the other type of galactic cosmic rays because five largest particle flares were recorded on 14 July 2000, 8

November 2000, 4 November 2001, 28 October 2003 and 20 January 2005. Figure 6 provides an explanation for the recently frequent outbreaks of bird flu, which may partially due to the extreme ultraviolet (EUV) flux, as can be seen the spikes of EUV flux interrupted the patterns of unpredictable type and frequency.

Without the numerical presentation of the hemagglutinins by our approaches and their fast Fourier transform, it would be relatively difficult to find various periodicities of hemagglutinin mutations. Again, the analysis on the history of influenza shows that the interval between pandemics has not significantly increased or decreased with the passage of time, suggesting that increased population and travel are not determining factors[15].

Therefore, our analysis can establish the relationship between severe mutations and effects of cosmic activity. We can logically deduce the role of waterfowl migration on the outbreak of bird flu and influenza. The Earth magnetic field protects the Earth from most cosmic rays and only the highest energy cosmic rays will penetrate the magnetic field and the atmosphere to hit the ground at the equator. Many cosmic rays penetrate the magnetic field, but are guided along the Earth's magnetic field lines towards the polar regions[14], where more mutations occur in influenza A virus either within the migratory wild birds or in their living environments. As the winter approaches, these waterfowl fly forwards warm south bringing back the new mutated influenza A virus leading to outbreak of bird flu or influenza.

ACKNOWLEDGMENTS

We thank the Royal Observatory of Belgium, the Moscow cosmic ray station and the CELIAS/SEM experiment on the Solar Heliospheric Observatory (SOHO) spacecraft for their free downloadable data used in this study.

REFERENCES

1. Wu, G. and S. Yan, 2005. Timing of mutation in hemagglutinins from influenza A virus by means of unpredictable portion of amino-acid pair and fast Fourier transform. Biochem. Biophys. Res. Commun., 333: 70-8.
2. Wu, G. and S. Yan, 2006. Timing of mutation in hemagglutinins from influenza A virus by means of amino-acid distribution rank and fast Fourier transform. Protein Pept. Lett., 13: 143-8.
3. Wu, G. and S. Yan. Mutation trend of hemagglutinin of influenza A virus: a review from computational mutation viewpoint. Acta Pharmacol. Sin., (in press)
4. http://www.sidc.oma.be/
5. http://helios.izmiran.rssi.ru/cosray/main.htm
6. http://www.usc.edu/dept/space_science/semdata.htm
7. Wu, G. and S.M. Yan, 2002. Randomness in the primary structure of protein: methods and implications. Mol. Biol. Today, 3: 55-69.
8. Wu, G. and S. Yan, 2006. Fate of influenza A virus proteins. Protein Pept. Lett. 13: 377-84.
9. Taubenberger, J.K., A.H. Reid, A.E. Krafft, K.E. Bijwaard and T.G. Fanning, 1997. Initial genetic characterization of the 1918 "Spanish" influenza virus. Science, 275: 1793-6.
10. Reid, A.H., T.G. Fanning, J.V. Hultin and J.K. Taubenberger, 1999. Origin and evolution of the 1918 "Spanish" influenza virus hemagglutinin gene. Proc. Natl. Acad. Sci., USA, 96: 1651-6.
11. Feller, W., 1968. An Introduction to Probability Theory and Its Applications. 3rd Edn. Vol, I. New York: Wiley.
12. Wu, G. and S. Yan, 2004. Fate of 130 hemagglutinins from different influenza A viruses. Biochem. Biophys. Res. Commun., 317: 917-24.
13. Wu, G. and S. Yan, 2005. Prediction of mutation trend in hemagglutinins and neuraminidases from influenza A viruses by means of cross-impact analysis. Biochem. Biophys. Res. Commun., 326: 475-82.
14. http://neutronm.bartol.udel.edu//listen/main.html
15. Potter, C.W., 2001. A history of influenza. J. Appl. Microbiol., 91: 572-9.
16. Krauss, S., D. Walker, S.P. Pryor, L. Niles, L. Chenghong, V.S. Hinshaw and R.G. Webster, 2004. Influenza A viruses of migrating wild aquatic birds in North America. Vector Borne Zoonotic. Dis, 4: 177-89.
17. No authors listed, 2004. Special issue: Space Radiation Biology. Biol. Sci. Space, 18: 200-60.
18. Donis, R.O., W.J. Bean, Y. Kawaoka and R.G. Webster, 1989. Distinct lineages of influenza virus H4 hemagglutinin genes in different regions of the world. Virology, 169: 408-17.
19. Hoffmann, E., J. Stech, I. Leneva, S. Krauss, C. Scholtissek, P.S. Chin, M. Peiris, K.F. Shorteidge and R.g. Webster, 2000. Characterization of the influenza A virus gene pool in avian species in southern China: was H6N1 a derivative or a precursor of H5N1? Virology, 74: 6309-15.
20. Davis, G.E.Jr. and W.E. Lowell, 2004. Chaotic solar cycles modulate the incidence and severity of mental illness. Med. Hypotheses, 62: 207-14.
21. ftp://ftp.ngdc.noaa.gov/ STP/SOLAR_ DATA/SUNSPOT_ NUMBERS/ANCIENT_ DATA/Early_Reports.
22. Hirsch, A., 1883. Handbook of Geographical and Historical Pathology. London: New Sydenham Society.
23. Pyle, G.F., 1986. The Diffusion of Influenza: Patterns and Paradigms. New Jersey: Rowan & Littlefield.
24. http://www.hooper-home.net/CHRONO/