

Prediction of mutation position, mutated amino acid and timing in hemagglutinins from North America H1 influenza A virus

Shao-Min Yan¹, Guang Wu^{2*}

¹Guangxi Academy of Sciences, 98 Daling Road, Nanning, Province Guangxi 530007, China. ²DreamSciTech Consulting, 301, Building 12, Nanyou A-zone, Jiannan Road, Shenzhen, Province Guangdong 518054, China. Correspondence should be addressed to Guang Wu (hongguanglishibahao@yahoo.com)

Received Jan. 8th, 2009; revised Feb. 9th, 2009; accepted Feb. 10th, 2009

ABSTRACT

This study was trying to predict the mutations in H1 hemagglutinins of influenza A virus from North America including the predictions of mutation position, the predictions of would-be-mutated amino acids and the predictions of time of occurrence of mutations. The results paved a possible way for accurate, precise and reliable prediction of mutation in proteins from influenza A virus.

Keywords: Hemagglutinin, Influenza, Mutation, Neural Network, Prediction

1. INTRODUCTION

Mathematical modelling provides a promising hope to predict the mutation in proteins from influenza A virus, not only because the history shows that accurate, precise and reliable predictions are mainly based on mathematical modelling, but also the prediction of mutations at protein can be classified as prediction of mutation position, prediction of mutated amino acid and timing of mutation [1]. All these require the use of more sophisticated mathematical tools.

Perhaps, the best way to predict the mutation is to find its cause, thus a mutation could occur if the same cause appears again. However, the causes, which led to historical mutations, might not leave any sign to trace, and the evolved proteins from influenza A virus may no longer be sensitive to the causes, which led to mutations in the past. All these mean that the mutation causes would be poor predictors for prediction of mutations, while the preparedness for possible pandemic/epidemic of influenza would lag behind the appearance of influenza without prediction. On the other hand, no matter what mutation cause is, any cause would leave signs in a protein, otherwise, no mutation would be recorded. These signs can be arguably used for prediction. This is the basic consideration for prediction of mutations using modelling.

Generally, the amino acids in a protein is represented as alphabet, thus a number of models use amino-acid symbols as operating units, for example, sequence alignment, phylogenetics, and multi-sequence comparison, by which the history of proteins of interests can be traced [2,3]. Unfortunately, these symbol-based approaches cannot accurately and precisely answer the predictions proposed, because they cannot operate in sophisticated mathematical models, whose operating units are values.

In this view, the protein science actually is at the historical phase of searching for the ways to represent a protein sequence as a numeric sequence, and it is hoped that the numeric sequence is sensitive to mutations, positions of amino acids in protein sequence, composition of protein sequence, length of protein sequence, neighbouring amino acids.

In fact, currently there are several ways to transfer a protein sequence into a numeric sequence, and the most profound one would be the use of the physicochemical property to represent a protein sequence [4] as well as related approaches [5-10].

On the other hand, other approaches are also developed, for example, the approaches based on random mechanism to quantify each amino acid in a protein as well as a protein in whole [1].

This study was designed to predict the mutation positions, the mutated amino acids and the time of occurrence of mutations in the hemagglutinins from North America H1 influenza A viruses using neural network, because the hemagglutinin is the major surface antigen of influenza viruses, against which neutralizing antibodies are elicited during virus infection and vaccination [11-15]. Among various types, the H1 influenza virus is the cause for several historical disasters, such as 1918 Spanish flu, 1977 Russian flu, 1950 and 1988 epidemics [16-18].

2. MATERIALS AND METHODS

The amino acid sequences and corresponding RNA sequences of 494 hemagglutinins from North America influenza A/H1 viruses isolated from 1918 to 2008 were obtained from the influenza virus resources [19]. Forty-

six identical hemagglutinins were excluded, thus the remaining 448 hemagglutinins were used in this study.

2.1 Amino-Acid Pair Predictability

According to the permutation [1, 20, 21], for example, there are 47 asparagines “N” and 37 valines “V” in the hemagglutinin, strain A/swine/Ontario/53518/03(H1N1), accession number DQ280219, the frequency of amino-acid pair NV is 3 ($47/566 \times 37/565 = 3.072$), that is, NV would appear three times in this hemagglutinin. Actually 3 NVs can be found in this hemagglutinin, so NV is predictable and the difference between its predicted and actual frequency is 0. Again, there are 48 leucines “L” in DQ280219 hemagglutinin, and the frequency of random presence of LL is 4 ($48/566 \times 47/565 = 3.986$), i.e. there would be four LLs in the hemagglutinin. But LL appears nine times in reality, so the difference between its predicted and actual frequency is -5. After such calculations [22], each amino-acid pair had its difference between predicted and actual frequency. As a point mutation is relevant to a single amino acid, which connects with two neighbouring amino acids except for the terminal one and constructs two amino-acid pairs, so each amino acid has the sum of difference between predicted and actual frequency in two neighbouring amino-acid pairs, which is the first quantification for each amino acid in a hemagglutinin. Nevertheless, any hemagglutinin must have a certain amount of predictable amino-acid pairs, by which the percentage of how many amino-acid pairs predictable can be found. This predictable portion is the quantification for a whole hemagglutinin.

2.2 Amino-Acid Distribution Probability

According to the occupancy of subpopulations and partitions, the positions of any type of amino acids in hemagglutinin can be viewed as a certain distribution [10, 23-31], whose probability is $\frac{r!}{q_0! \times q_1! \times \dots \times q_n!} \times \frac{r!}{r_1! \times r_2! \times \dots \times r_n!} \times n^{-r}$

[32], where r is the number of amino acids, n is the number of partitions, r_n is the number of amino acids in the n -th partition, q_n is the number of partitions with the same number of amino acids, and ! is the factorial function. For instance, there are 36 lysines “K” in DQ280219

hemagglutinin. Their predicted and actual distribution probabilities are 0.0419 and 0.0020 [33], so the ratio of predicted versus actual distribution probabilities is 20.95, whose natural logarithm is 3.0421, which is the second quantification for each amino acid in a hemagglutinin.

2.3 Future Composition of Amino Acids

The relationship between 64 RNA codons and translated amino acids is governed by translation probability [1, 34-36], based on which the amino acid mutating probability can be determined. For example, alanine “A” has the 12/36 chance of mutating to “A”, but cysteine “C” has no chance of mutating to “A”, then both aspartic acid “D” and glutamic acid “E” have the 2/18 chance of mutating to “A”, and so on. In total, the future composition of amino acid “A” is 6.1271% in DQ280219 hemagglutinin, whereas its current composition is only 5.1146% (29/567), and the ratio is 1.1980 (6.1271%/5.1146%), thus the future composition of amino acids is got [1], and assigned the ratio of predicted versus actual compositions to each amino acid [37], which is the third quantification for each amino acid in hemagglutinin [1].

Although there are countless mutation causes impacting a parent protein, these causes should leave their traces in the protein, which should be measured out using these three quantifications, which in fact represent the countless mutation causes.

2.4 Prediction of Mutation Position

Any mutation cause can lead to occurrence or non-occurrence of mutation, which can be classified as unity and zero after comparing a parent protein with its daughter protein. In this way, the occurrence or non-occurrence of mutation in a parent protein becomes a binary sequence. Thus, two datasets can be got, the mutation cause dataset, which are three quantifications, and the mutation consequence dataset, which is a binary sequence. Moreover, these two datasets have the position-to-position relationship (Table 1), which is the cause-mutation relationship. Mathematically this relationship is the problem of classification, which can be solved either using the logistic regression in statistics or neural network. The feed forward backpropagation neural network

Table 1. Inputs and target of DQ280219 hemagglutinin sequence.

Position	Amino acid	Quantified hemagglutinin sequence			Mutation sequence
		I	II	III	
1	M	-2	0.0000	1.2569	0
...
276	R	-2	1.2809	1.9392	0
277	G	-1	2.3790	0.7887	0
278	H	0	0.0000	1.2396	1
279	G	0	2.3790	0.7887	1
280	S	1	4.0008	1.1081	0
...
566	I	-2	1.1285	0.9590	0

would be applied to this relationship to predict the mutation position.

2.5 Prediction of would-be-mutated Amino Acid

The prediction was made using the amino-acid mutating probability, which was based on the relationship between RNA codons and translated amino acids [1].

2.6 Timing of Mutation

As each hemagglutinin is different one from another due to mutation, each hemagglutinin would be quantified differently one from another. Along the time axis, all hemagglutinins would construct their evolutionary process, and the timing of the mutation would be possible by detailed analysis of this evolutionary process.

2.7 Software and Statistics

The MatLab software [38] was used for prediction. The prediction sensitivity, specificity and total correct rate were calculated according to the published method [39].

3. RESULTS AND DISCUSSION

The performance of modelling was measured using the prediction sensitivity ($42.9\% \pm 31.4\%$), specificity ($99.5\% \pm 0.4\%$) and total correct rate ($99.0\% \pm 0.4\%$), because the predicted mutation positions can be classified as the positives, false positives, negatives and false negatives when comparing the predicted with the actual mutation positions. As seen, the prediction sensitivity was still low although the prediction specificity and total correct rate were quite high.

After the prediction of possible mutation positions using the neural network, the would-be-mutated amino ac-

ids at predicted positions can be predicted using the amino-acid mutating probability [1]. **Figure 1** illustrates the prediction of possible mutation positions and mutated amino acids at the predicted positions, where the solid line in the lower panel is the predicted mutation probability by the neural network with respect to each amino acid in ABX58602 hemagglutinin and the dotted line is the cut-off mutation probability of 0.5, that is, the amino acid whose mutation probability is larger than 0.5 risks mutation. For this hemagglutinin, there were two positions (98 and 507) whose mutation probability was larger than 0.5, so the amino acid E at these positions would have a larger chance of mutation. Meanwhile, the would-be-mutated amino acid can be determined using the amino-acid mutating probability (upper panel), where the amino acid "D" has the largest chance to appear. So the lower panel indicates the possible mutation positions with probability, and the upper panel displays the would-be-mutated amino acids with probability.

Figure 2 displays the evolution of North America H1 hemagglutinins. This predictable portion fluctuated over time, which represented the mutation process. With fast Fourier transform, which is suitable to find the periodicity in chaotic dataset, the mutation periodicity can be found from **Figure 2**, where (i) the evolutionary process of influenza A virus hemagglutinins from 1978 to 2008 contained many periodicities; (ii) each periodicity suggested different number of mutations along the time course; (iii) the periodicity with the biggest number of mutations was about 5.6 years, thus the time when mutations would occur in future can be estimated; and (iv) the hemagglutinin periodicity provides the chance to trace the possible mutation cause in nature, because each periodicity may correspond to a natural phenomenon.

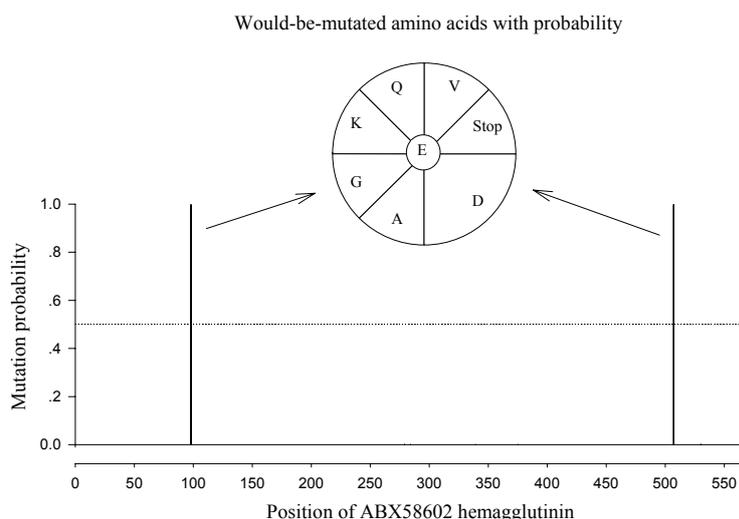


Figure 1. Prediction of mutation positions and would-be-mutated amino acids. On the lower panel: the x-axis represents the position of ABX58602 hemagglutinin from 1 to 565, because ABX58602 hemagglutinin is composed of 565 amino acids; the y-axis represents the mutation probability predicted using neural network model, where there are two probabilities larger than 0.5 at positions 98 and 507. On the upper panel, the centre of pie is labelled as "E" glutamic acid, which is the amino acid at positions 98 and 507 of ABX58602 hemagglutinin. The other letters represent the would-be-mutated amino acids, and the area occupied by letter represents the probability to mutate to this amino acid based on the amino-acid mutating probability, for example, "E" has the largest chance to mutate to "D".

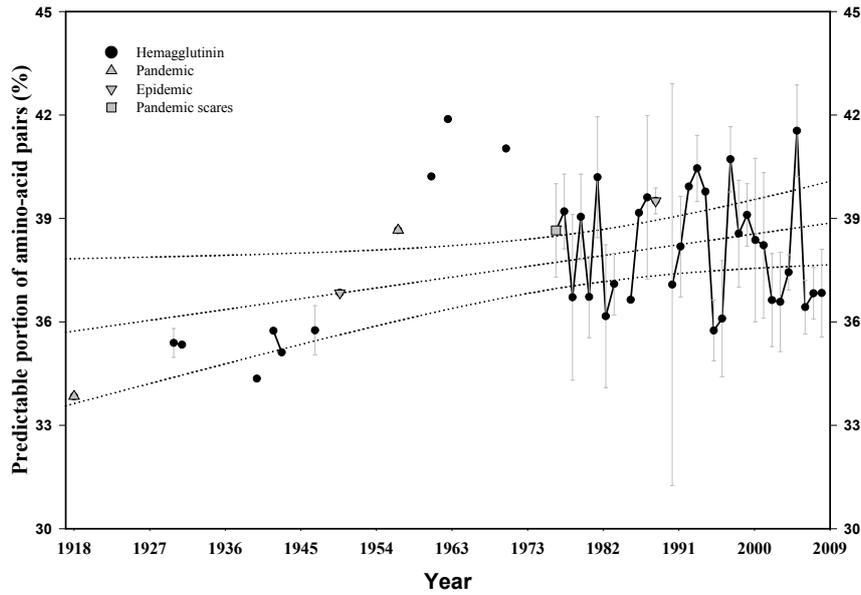


Figure 2. Evolution of 448 hemagglutinins of North America H1 influenza viruses. The data are presented as mean±SD. The dotted lines are regressed lines 95% confidence intervals.

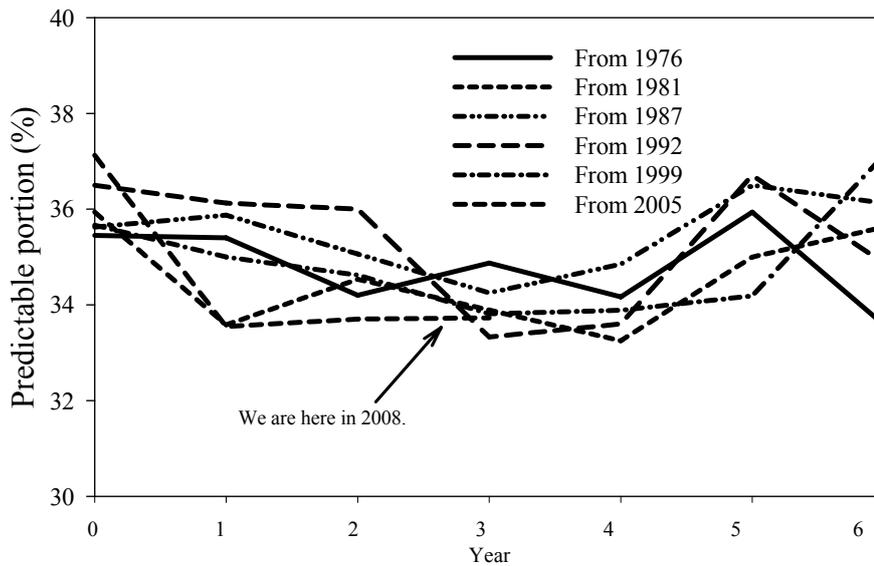


Figure 3. Stratification of hemagglutinin evolution after finding the periodicity using fast Fourier transform.

Furthermore, an attempt was made to time the mutation by stratifying the hemagglutinin evolution in **Figure 2** according to its periodicity, and **Figure 3** shows such an example, where the hemagglutinin evolution in **Figure 2** is stratified according to 6-year periodicity because it was the periodicity with biggest number of mutations. **Figure 3** shows that there would be a 2-year stable period before possible more mutations would occur in 2010.

At this stage of development, it is yet to verify the predictions made in this study. However, this is not uncommon phenomenon in science, because the first step is to find a way to transfer the measurements into the domain, where a mathematical model can be applied, the

second step is build a model, and the third step is to make the predictions. These three steps are more related to theoretical work. Thereafter the last step would be the verification experimentally, which is certainly beyond the scope of this paper. On the other hand, the science advances so much, it is impossible to verify each hypothesis and prediction, for example, the humans cannot create another earth without global warming to compare the effects on subjects of interests. With respect to the predictions in this study, the verifications can be done by using the same method in man-made mutations in industrial enzymes, where each mutation can be recorded and compared with prediction.

The frequency of mutations is not identical along a hemagglutinin sequence, namely, the different position has different chance of mutations. In fact, the prediction made in this study is consistent with this observation as seen in **Figure 1**, where the predicted mutation probability is not identical along the ABX58602 hemagglutinin.

To the best of knowledge, there are several models conducted at different levels for the prediction of possible pandemic/epidemic of influenza. At epidemiological level, the predictions were made using early indicators [40], time series analysis [41,42], etc. At clinical level, the prediction was made using medical visit [43], outbreak signatures [44], etc. At social level, the prediction was made using sales of computer printers, elections, and the Federal Reserve's decisions about interest rates [45]. At seroarcheological level, the prediction was made using accumulation of mutations or true recombinational events [46]. At protein level, the prediction was made with epitope [47-49], conformation [50]. However, no similar prediction was made with respect to the approached used in this study. The difference includes: (1) the quantification of protein sequences in this study was based on the random principle, (2) the occurrence and non-occurrence of mutation was quantified as yes-no event, (3) the cause-mutation relationship was defined using neural network, (4) the would-be-mutated amino acid was determined using the amino-acid mutating probability, and (5) the time of mutation was determined using the fast Fourier transform to stratify the time interval between outbreak of influenza.

This study paved a possible way for accurate, precise and reliable prediction of mutation in proteins from influenza A virus, because the model in prediction was the cause-mutation model, which was helpful for understanding of underlined mutation mechanism.

ACKNOWLEDGEMENTS

This study was supported in part by National Natural Science Foundation No. 20666002 (Guangxi Assignment No. 0728001).

REFERENCES

- [1] G. Wu & S. Yan. (2008) Lecture Notes on Computational Mutation. Nova Science Publishers, New York.
- [2] E. Ghedin, N. A. Sengamalay, M. Shumway, J. Zaborosky, T. Feldblyum, V. Subbu, D.J. Spiro, J. Sitz, H. Koo, P. Bolotov, D. Dernovoy, T. Tatusova, Y. Bao, K. St George, J. Taylor, D.J. Lipman, C.M. Fraser, J.K. Taubenberger & S.L. Salzberg. (2005) Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. *Nature*, 437, 1162-1166.
- [3] J. C. Obenauer, J. Denson, P. K. Mehta, X. Su, S. Mukatira, D. B. Finkelstein, X. Xu, J. Wang, J. Ma, Y. Fan, K.M. Rakestraw, R.G. Webster, E. Hoffmann, S. Krauss, J. Zheng, Z. Zhang & C.W. Naeye. (2006) Large-scale sequence analysis of avian influenza isolates. *Science*, 311, 1576-1580.
- [4] K. C. Chou. (2004) Structure bioinformatics and its impact to biomedical science. *Curr. Med. Chem*, 11, 2105-2134.
- [5] K. C. Chou. (2004) Modelling extracellular domains of GABA-A receptors: subtypes 1, 2, 3, and 5. *Biochem. Biophys. Res. Commun*, 316, 636-642.
- [6] K. C. Chou. Insights from modelling the 3D structure of extracellular domain of alpha 7 nicotinic acetylcholine receptor. *Biochem. Biophys. Res. Commun*, 319, 433-438.
- [7] K. C. Chou. (2004) Coupling interaction between thromboxane A2 receptor and alpha-13 subunit of guanine nucleotide-binding protein. *J. Proteome Res.* 2005, 4, 1681-1686.
- [8] X. Xiao, S. Shao, Y. Ding, Z. Huang, X. Chen & K.C. Chou. (2005) An application of gene comparative image for predicting the effect on replication ratio by HBV virus gene missense mutation. *J. Theo. Biol.* 235, 555-565.
- [9] X. Xiao, S. H. Shao & K. C. Chou. (2006) A probability cellular automation model for hepatitis B viral infections. *Biochem. Biophys. Res. Commun*, 342, 605-610.
- [10] S. Yan & G. Wu. (2008) Quantitative relationship between mutated amino-acid sequence of human copper-transporting ATPases and their related diseases. *Mol. Divers*, 12, 119-129.
- [11] Q. S. Du, S.Q. Wang & K. C. Chou. (2007) Analogue inhibitors by modifying oseltamivir based on the crystal neuraminidase structure for treating drug-resistant H5N1 virus. *Biochem. Biophys. Res. Commun*, 363, 525-531.
- [12] J. R. Schnell & J. J. Chou. (2008) Structure and mechanism of the M2 proton channel of influenza A virus. *Nature*, 451, 591-595.
- [13] S. Q. Wang, Q. S. Du & K. C. Chou. (2007) Study of drug resistance of chicken influenza A virus (H5N1) from homology-modeled 3D structure of neuraminidases. *Biochem. Biophys. Res. Commun*, 354, 634-640.
- [14] D.Q. Wei, Q.S. Du, H. Sun & K.C. Chou. (2006) Insights from modeling the 3D structure of H5N1 influenza virus neuraminidase and its binding interactions with ligands. *Biochem. Biophys. Res. Commun*, 344, 1048-1055.
- [15] D. C. Wiley & J. J. Skehel. (1987) The structure and function of the hemagglutinin membrane glycoprotein of influenza virus. *Annu. Rev. Biochem*, 56, 365-394.
- [16] Y. Kanegae, S. Sugita, K. F. Shortridge, Y. Yoshioka & K. Nerome. (1994) Origin and evolutionary pathways of the H1 hemagglutinin gene of avian, swine and human influenza viruses: cocirculation of two distinct lineages of swine virus. *Arch. Virol*, 134: 17-28.
- [17] A.H. Reid, T.G. Fanning, J.V. Hultin & J.K. Taubenberger. (1999) Origin and evolution of the 1918 "Spanish" influenza virus hemagglutinin gene. *Proc. Natl. Acad. Sci. USA*, 96, 1651-1656.
- [18] J. K. Taubenberger, A. H. Reid, A. E. Krafft, K. E. Bijwaard & T. G. Fanning. (1997) Initial genetic characterization of the 1918 "Spanish" influenza virus. *Science*, 275, 1793-1796.
- [19] Influenza virus resources. (2008) <http://www.ncbi.nlm.nih.gov/genomes/FLU/Database/multiple.cgi>.
- [20] G. Wu & S. Yan. Randomness in the primary structure of protein: methods and implications. *Mol. Biol. Today* 2002, 3: 55-69.
- [21] G. Wu & S. Yan. (2006) Mutation trend of hemagglutinin of influenza A virus: a review from computational mutation viewpoint. *Acta Pharmacol. Sin*, 27: 513-526.
- [22] Amino-acid pair predictability. (2008) <http://www.dreamscitech.com/Service/rationale.htm>.
- [23] G. Wu & S. Yan. (2000) Prediction of distributions of amino acids and amino acid pairs in human haemoglobin α -chain and its seven variants causing α -thalassemia from their occurrences according to the random mechanism. *Comp. Haematol. Int*, 10, 80-84.
- [24] G. Wu & S. Yan. (2001) Analysis of distributions of amino acids, amino acid pairs and triplets in human insulin precursor and four variants from their occurrences according to the random mechanism. *J. Biochem. Mol. Biol. Biophys*, 5, 293-300.
- [25] G. Wu & S. Yan. (2001) Analysis of distributions of amino acids and amino acid pairs in human tumor necrosis factor precursor and its eight variants according to random mechanism. *J. Mol. Model*, 7, 318-323.
- [26] G. Wu & S. Yan. (2002) Random analysis of presence and absence of two-and three-amino-acid sequences and distributions of amino acids, two-and three-amino-acid sequences in bovine p53 protein. *Mol. Biol. Today*, 3: 31-37.

- [27] G. Wu & S. Yan. (2002) Analysis of distributions of amino acids in the primary structure of apoptosis regulator Bcl-2 family according to the random mechanism. *J. Biochem. Mol. Biol. Biophys.*, 6, 407-414.
- [28] G. Wu & S. Yan. (2002) Analysis of distributions of amino acids in the primary structure of tumor suppressor p53 family according to the random mechanism. *J. Mol. Model.*, 8, 191-198.
- [29] G. Wu & S. Yan. (2004) Determination of sensitive positions to mutations in human p53 protein. *Biochem. Biophys. Res. Commun.*, 321, 313-319.
- [30] G. Wu & S. Yan. (2005) Searching of main cause leading to severe influenza A virus mutations and consequently to influenza pandemics/epidemics. *Am. J. Infect. Dis.*, 1, 116-123.
- [31] G. Wu & 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, 475-482.
- [32] W. Feller. (1968) *An Introduction to Probability Theory and Its Applications*. 3rd ed, Vol. I. Wiley, New York, p. 34-40.
- [33] Amino-acid distribution probability. (2008) <http://www.dreamscitech.com/Service/timing.htm>.
- [34] G. Wu & S. Yan. (2005) Determination of mutation trend in proteins by means of translation probability between RNA codes and mutated amino acids. *Biochem. Biophys. Res. Commun.*, 337, 692-700.
- [35] G. Wu & S. Yan. (2006) Determination of mutation trend in hemagglutinins by means of translation probability between RNA codons and mutated amino acids. *Protein Pept. Lett.*, 13, 601-609.
- [36] G. Wu & S. Yan. (2007) Translation probability between RNA codons and translated amino acids, and its applications to protein mutations. In: *Leading-Edge Messenger RNA Research Communications*. ed. Ostrovskiy M. H. Nova Science Publishers, New York, Chapter 3, 47-65.
- [37] Amino-acid mutating probability. (2008). <http://www.dreamscitech.com/Service/lag.htm>.
- [38] MathWorks Inc. (2001) *MatLab-The Language of Technical Computing* (version 6.1.0.450, release 12.1), 1984-2001.
- [39] Systat Software Inc. *Systat for Windows*, version 11.00.01. 2004.
- [40] E. Andersson, S. Kühlmann-Berenzon, A. Linde, L. Schiöler, S. Rubinova & M. Frisé. (2008) Predictions by early indicators of the time and height of the peaks of yearly influenza outbreaks in Sweden. *Scand. J. Public Health*, 36, 475-482.
- [41] Y. T. Li, H. W. Zhang, H. Ren, J. Chen & Y. Wang. (2007) Application of time series analysis in the prediction of incidence trend of influenza-like illness in Shanghai. *Zhonghua Yu Fang Yi Xue Za Zhi*, 41, 496-498.
- [42] J. Saltyte Benth & D. Hofoss. (2008) Modelling and prediction of weekly incidence of influenza A specimens in England and Wales. *Epidemiol. Infect.*, 136, 1658-1666.
- [43] R. Sebastian, D. M. Skowronski, M. Chong, J. Dhaliwal & J. S. Brownstein. (2008) Age-related trends in the timeliness and prediction of medical visits, hospitalizations and deaths due to pneumonia and influenza, British Columbia, Canada, 1998-2004. *Vaccine*, 4, 1397-1403.
- [44] P. F. Craigmile, N. Kim, S. A. Fernandez & B. K. Bonsu. (2007) Modeling and detection of respiratory-related outbreak signatures. *BMC Med. Inform. Decis. Mak.*, 7: 28.
- [45] P. M. Polgreen, F. D. Nelson & G. R. Neumann. (2007) Use of prediction markets to forecast infectious disease activity. *Clin. Infect. Dis.*, 44: 272-279.
- [46] R. G. Webster. (1997) Predictions for future human influenza pandemics. *J. Infect. Dis.*, 176 (Suppl 1): S14-S19.
- [47] A. Suhrbier, C. Schmidt & A. Fernan. (1993) Prediction of an HLA B8-restricted influenza epitope by motif. *Immunology*, 79, 171-173.
- [48] P. Somvanshi, V. Singh & P. K. Seth. (2008) Prediction of epitopes in hemagglutinin and neuraminidase proteins of influenza A virus H5N1 strain: a clue for diagnostic and vaccine development. *OMICS*, 12, 61-69.
- [49] P. Gogolak, A. Simon, A. Horváth, B. Réthi, I. Simon, K. Berkics, E. Rajnavölgyi & G.K. Tóth. (2000) Mapping of a protective helper T cell epitope of human influenza A virus hemagglutinin. *Biochem. Biophys. Res. Commun.*, 270: 190-198.
- [50] M. Young, K. Kirshenbaum, K. A. Dill & S. Highsmith. (1999) Predicting conformational switches in proteins. *Protein Sci.*, 8, 1752-1764.